Estimating Exceedance Probability of Extreme Water Levels of the Akosombo dam

Eric Ocran[1] , Kwabena Doku-Amponsah[1] and Ezekiel N. N. Nortey[1*]

Department of Statistics, University of Ghana, Legon, Ghana

***Corresponding author:** ennnortey@ug.edu.gh

ABSTRACT

The Akosombo dam is a major source of electric energy in Ghana. Considering the current increase in the demand for electricity in the country, where such an increase in demand implies more pressure on the dam, it is of key interest to study the tail behaviour of the water levels of the dam. Such a study is important because the level of water in the dam determines the amount of electricity generated. The study employed the Univariate Extreme Value Theory to model the monthly maximum and minimum water levels of the dam. The Generalized Extreme Value Distribution was fitted to the data and the Maximum likelihood estimation method was employed to estimate the model parameters. The study indicated that, the water levels cannot fall below 226.00ft which is the critical water level of the Akosombo dam. It further showed that, the lowest ever level of water the dam can attain is 226.69ft and the highest 279.07ft. The study also found that, though the water cannot fall below the critical level, there was evidence of its falling below the minimum operation head.

Keywords: Extreme value theory; Generalized extreme value distribution; Maximum likelihood estimation; Exceedance probabilities; Tail behaviour

Introduction

Since the late 1980s, Ghana has been facing an energy crisis. A major source of energy in the country is electrical energy and this has been insufficient due to generation problems. According to the Ghana Energy Commission's 2016 energy supply and demand outlook, currently the country's sources of electrical energy are hydro power, thermal power and renewables (solar). Hydro power alone contributes 49.80% (1,580MW) of the country's electrical energy. Of this quantity, the Akosombo hydroelectric plant contributes 32.14% (1,020MW) (Ghana Energy Commission, 2016). The water levels of the dam are a key determinant of the amount of power generated. On the one hand, water levels below the required level for operation result in low power output since not all turbines can function at such a minimal level. On the other hand, very high water levels above the maximum operation level can result in the collapsing of the dam, leading to flooding. In the event of low or very high water levels, lives, businesses, social activities and the economy of the country can be affected. For these reasons, the tail behaviour of the Akosombo dam, which is the largest dam and a major contributor to Ghana's energy, needs to be investigated.

Extreme Value Theory (EVT) is a field of statistics which deals with the statistical techniques for modeling and the estimation of rare (infrequent) events. By definition, extreme values are infrequent, meaning that estimates are often required for levels of a process that are much greater (or lower) than those which have already been observed, which implies extrapolation from observed events to unobserved events (Coles, 2001). Extreme value theory also provides a class of models to enable such extrapolation. EVT is an exclusive statistical discipline, since it develops techniques and models for describing the unusual rather than the usual, and because it offers a framework in which an estimate of expected forces could be made using past data (Coles, 2001).

Extreme value distributions arises as limiting distributions of maximum or minimum values in random samples, as the samples become sufficiently large (Katz *et al.*, 2002). There are two main families of extreme value distributions, namely the Generalized Extreme Value (GEV) distribution introduced by Jenkinson (1955) and the Generalized Pareto (GP) distribution introduced by Pickands (1975). This paper investigates the tail behaviour of the Akosombo dam using the GEV distribution. Minkah (2016) studied the tail distribution of the Akosombo dam using the GP distribution. The objective of this paper is to complement Minkah (2016) by studying the tail distribution of the Akosombo dam using the GEV distribution to estimate the lower and upper bounds of the water levels of the dam and to determine the exceedance probabilities for very low water levels and very high water levels of the dam which have not yet been observed.

The rest of the paper is organized as follows: section 2 details the GEV methodology and parameter estimation of some extreme events. Section 3 provides a brief background of the Akosombo dam and a brief description of the data employed. Section 4 examines the fitness of the model and an estimation of some extreme events, that is, the bounds and the exceedance probabilities. Section 5 concludes the discussions of the results.

Method

Extreme value theory enables us to study the tail behaviour of stochastic processes. A finite-sample approximation to asymptotic results is the basis of extreme value analyses (Watts *et al.*, 2006).

This paper employs the definitions of Coles (2001). Let $X_{n,n} = \max\left(X_1, X_2, \ldots, X_n\right)$ denote the sample maximum, where $X_1, X_2, \ldots, X_n$ is a sequence of independent and identically distributed random variables with distribution function . Suppose there exists a suitable sequence $\{a_n > 0, n \geq 1\}$ and $\{b_n, n \geq 1\}$ such that for a sufficiently large sample and at all continuity points x of F, with x taken from the set of real numbers:

$$\lim_{n\to\infty} P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) = \lim_{n\to\infty}\left\{F\left(a_n x + b_n\right)^n\right\} \to H\left(x\right) \qquad (1)$$

where F belongs to the domain of attraction of H and H is a non-degenerate distribution function, then $H\left(x\right)$ is the Generalized Extreme Value distribution given by:

$$H(x) = \begin{cases} \exp\left\{-\left[1+\gamma\left(\dfrac{x-\mu}{\sigma}\right)\right]^{-1/\gamma}\right\}, \gamma \neq 0 \\ \exp\left\{-\exp\left[-\left(\dfrac{x-\mu}{\sigma}\right)\right]\right\}, \quad \gamma = 0 \end{cases} \qquad (2)$$

$-\infty < \mu < \infty, \sigma > 0$ *and* $-\infty < \gamma < \infty$ with $-\infty < \mu < \infty, \sigma > 0$ and $-\infty < \gamma < \infty$. This distribution is a three parameter distribution: a location parameter μ, a scale parameter σ and a shape parameter γ. The GEV distribution is a combination of three families of extreme value distributions, namely the Gumbel, the Fréchet and the

Weibull distributions. If $\gamma = 0$, then the GEV corresponds to the Gumbel distribution; $\gamma > 0$ corresponds to the Fréchet distribution with $\alpha = 1/\gamma$; and $\gamma < 0$ corresponds to the Weibull distribution with $\alpha = -1/\gamma$.

The GEV distribution provides a model for the distribution of block maxima (or minima). The block maxima (or minima) method consists of partitioning the sample into r disjoint and independent blocks. The blocks are usually of equal length and the block sizes are usually selected naturally, for example, yearly. The maximum (or minimum) of each block is chosen as r independent sample maxima (or minima) from the observed data and then fitted to the GEV distribution. The choice of block size is a sensitive issue since it is a trade-off between bias and variance (Coles, 2001). Blocks of small size are likely to yield poor approximation by the GEV distribution and blocks of larger size produce few block maxima (or minima), which might lead to large variance. In this paper, the block size will be monthly (see Katz *et al.*, 2002; Reiss and Thomas, 2007; Mendez and Menendez, 2006). The monthly maxima and minima water levels of the Akosombo dam will be used, instead of the ussal annual extremes. The block sizes in this case may differ since the months do not have the same number of days. We however do not expect this to affect the result significantly. We used monthly extremes in order to consider more than one datum per year. Also, according to Mendez and Menendez (2006), monthly random extreme variables are expected to be more homogeneous than yearly ones and therefore expected to improve the asymptotic approximation..

The GEV distribution is for modelling only block maxima; hence to be able to model the block minima, we employ the duality between the distribution for maximum and minimum. Thus, given that $x_1, x_2, \ldots, x_n$ are realizations from the GEV distribution for minima with parameter (μ_*, σ, γ), then fitting the GEV distribution for maxima to the data $-x_1, -x_2, \ldots, -x_n$ yields the estimates for the minima with just a correction of $\mu_* = -\mu$ (Coles, 2001).

The parameters of the GEV distribution can be estimated using the Maximum Likelihood (ML), the Probability Weighted Moment (PWM), the Generalized Probability Weighted Moment (GPWM) or the L-moment method (Ribereau *et al.*, 2008; Katz *et al.*, 2002; Bezak *et al.*, 2014; Hawks *et al.*, 2008). However, this paper employed the ML estimation method due to its "off-the-shelf" large sample inference properties (Coles, 2001).

Considering the assumption that the $X_1, X_2, \ldots, X_n$ are independent and identically distributed, the log-likelihood function of the GEV is given as:

If $\gamma \neq 0$, we have:

$$l(\mu,\sigma,\gamma) = -n\log\sigma - (1+1/\gamma)\sum_{i=1}^{n}\log\left[1+\gamma\left(\frac{x_i-\mu}{\sigma}\right)\right] - \sum_{i=1}^{n}\left[1+\gamma\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-1/\gamma} \quad (3a)$$

Provided that $.1+\gamma\left(\frac{x_i-\mu}{\sigma}\right) > 0$ for $i = 1, 2, \ldots, n$

If $\gamma = 0$, the log-likelihood reduces to:

$$l(\mu,\sigma,\gamma) = -n\log\sigma - \sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right) - \sum_{i=1}^{n}\exp\left\{-\left(\frac{x_i-\mu}{\sigma}\right)\right\} \quad (3b)$$

The derivatives of (3a) and (3b) with respect to the parameters $\theta=(\gamma, \mu, \sigma)$ yield the likelihood system of equations which has no explicit solution; therefore the system of equations is solved iteratively. For computational details see Prescott and Walden (1980).

When the estimates μ, σ and γ of the GEV distribution have been obtained, the extreme quantile, the exceedance probabilities and the endpoints (bounds) of the GEV can be estimated.

By inverting the GEV distribution function, we obtain the quantiles for the original X data as:

$$q_{X,p} = \begin{cases} \hat{\mu} - \dfrac{\hat{\sigma}}{\hat{\gamma}}\left[1 - \left\{-\log\left(1-p\right)\right\}^{-\hat{\gamma}}\right], \hat{\gamma} \neq 0 \\ \hat{\mu} - \hat{\sigma}\log\left[-\log\left(1-p\right)\right], \qquad \hat{\gamma} = 0 \end{cases} \tag{4}$$

Where $H\left(q_{X,p}\right) = 1 - p$ and $0 \le p \le 1$, p is the exceedance probability.

Hence $P(X > q_{X,p}) = 1 - H\left(q_{X,p}\right)$

Equation (4) is known as the extreme quantile. The parameters $\hat{\mu}, \hat{\sigma}$ and $\hat{\gamma}$ are corresponding ML estimates of μ, σ and γ respectively.

For $\gamma < 0$, an inference can be made on the upper endpoint of the distribution and the ML estimate of the upper endpoint is $q_{X,0} = \hat{\mu} - \hat{\sigma}/\hat{\gamma}$, with probability zero (Coles, 2001; Beirlant *et al.*, 2004).

Data and brief background of the Akosombo dam

The data used for the study is the daily water level readings of the Akosombo dam obtained from the office of the Volta River Authority (VRA) at the Akuse station. The data consists of 17,525 data point, covering the period January 1, 1966 to December 31, 2013.

The Akosombo dam was commissioned into operation in January 1965. The dam is located on the Volta river in South-eastern Ghana in the Akosombo gorge. The construction of the Akosombo dam resulted in the creation of the Lake Volta, the world's largest man-made lake, covering 3.6% of Ghana's land area. The Akosombo hydroelectric plant has six 170MW Francis turbines. Each turbine is supplied with water via a penstock which is 112-116 meters long and 7.2 meters in diameter. The dam operates, under normal conditions between a minimum water level of 240ft and a maximum water level of 278ft. Increased power demand, coupled with unanticipated environmental conditions, has resulted in rolling blackouts and major power outages over the years. The dam has over the years experienced some trends of water levels below the minimum requirement for operation. The critical minimum level of the dam is 226ft and at this level all the turbines must be shut down completely. The dam has also had a fair share of very high water levels, very close to the maximum requirement for the dam's operation. For instance, in November 2010, the water peaked at 277.54ft, and to safeguard the integrity of the dam, the spill gate was opened. This resulted in the flooding of some houses and farmlands around the dam (VRA, 2010).

Results and discussion

This paper applied EVT to the water levels of the Akosombo dam. All the extreme value analyses in this paper are limited to the water levels of the dam recorded over the period 1966 to 2013. The methodology described in the previous section is applied to the data and discussed in this section.

Table 1 presents some descriptive statistics of the Akosombo water levels. The lowest ever water level for the period under study is 234.00 ft., and this was recorded on June 28, 1966. The electricity demand during that time was at a minimal level, so the electricity generated by such low water levels was enough to go round. However, with the nation's current increasing demand for electricity, any water level below 240ft may pose a problem which may lead to nationwide load shedding. The highest ever water level recorded is 277.54ft, and this occurred on 8th and 9th November 2010. The Skewness value indicates the distribution is asymmetric and therefore has a tail much lighter than that of the normal distribution.

Table 1: Descriptive Statistics of the Water levels

Minimum	1st Quartile	3rd Quartile	Mean	Skewness	Maximum
234.00	248.06	265.73	256.67	-0.124	277.54

The partitoning of the data into monthly blocks resulted in 576 data points in each case (i.e., for maxima blocks and minima blocks). In modelling the minimum, we employed the duality between maximum and minimum, that is, we converted the minimum into a maximum problem by using the negated forms of the minimum values. According to Coles (2001), the ML estimates of the parameters of this distribution correspond exactly to those of the required GEV distribution for minima except the sign correction $\mu_* = -\mu$.

Figure 1 shows the time series plots of the sample maxima and minima water levels for the monthly blocks. From the figure, it can be seen that the maximum and minimum plots are similar in terms of variations and trends. The Extreme Value Index (EVI), that is, the shape parameter, is the primary problem in extreme value analyses.

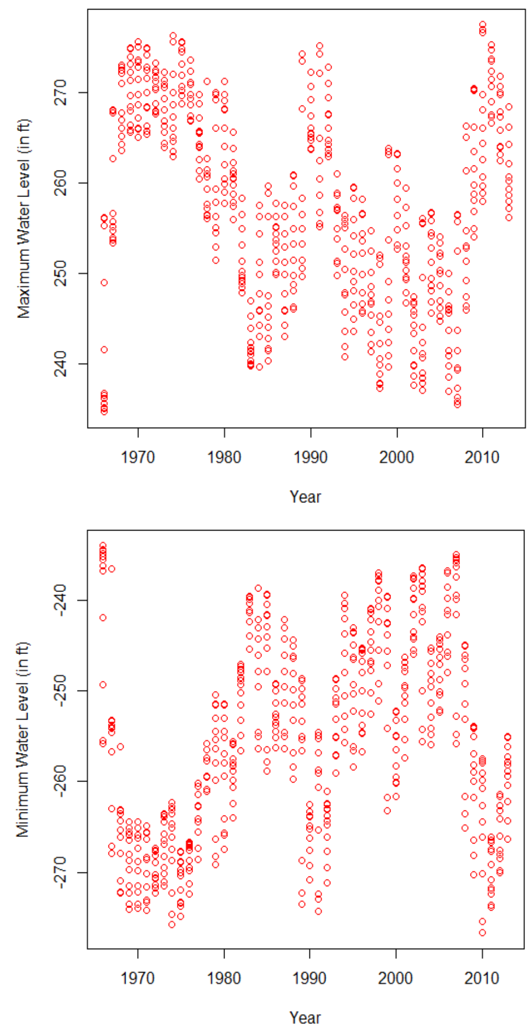

Fig. 1: Monthly Maximum Water Levels, top panel; Monthly Minimum Water levels, bottom panel

Table 2: Parameter estimates of the GEV distribution

Estimates	Right tail	Left til
Location ($\hat{\mu}$)	254.90 (0.5287)	259.10 (0.52)
Scale ($\hat{\sigma}$)	11.42 (0.4198)	10.69 (0.40)
Shape ($\hat{\gamma}$)	-0.47 (0.0338)	-0.33 (0.041)
95% conf. Interval ($\hat{\gamma}$)	(-0.54, -0.41)	(-0.4, -0.25)

* Standard errors are in parentheses

Table 2 presents the parameter estimates for the GEV distribution for the right and left tails of the Akosombo water levels. The EVI for the right tail of the Akosombo water levels is $\hat{\gamma} = -0.47$ with an associated standard error of 0.034, and that of the left tail is $\hat{\gamma} = -0.33$ with associated standard error of 0.04. In both cases, $\gamma > -0.5$ indicates that, regularity conditions are satisfied and the ML estimators have the usual asymptotic properties (Kotz *et al.*, 2000; Coles, 2001). The negativity of EVI's in both cases indicates that, the Akosombo dam water levels belong to the Weibull family of distribution.

We further used diagnostic plots to assess the fitted GEV models for the monthly maximum and minimum water levels (figure 2a and 2b). The plots are the QQ-plot (a), PP-plot (b) and a histogram overlay with the density curves of the fitted GEV (c). For the QQ-plot and the PP-plot, a good fit should yield a straight one-to-one line of points. Generally, the QQ-plot is preferred to the probability plot. The approximate linearity of the QQ-plot and the PP-plot indicates that the model is valid. From the histogram, the density also appears to be consistent with the data points. We therefore conclude that the diagnostic plots are in favour of the fitted model and that the GEV model is a good fit for the Akosombo data.

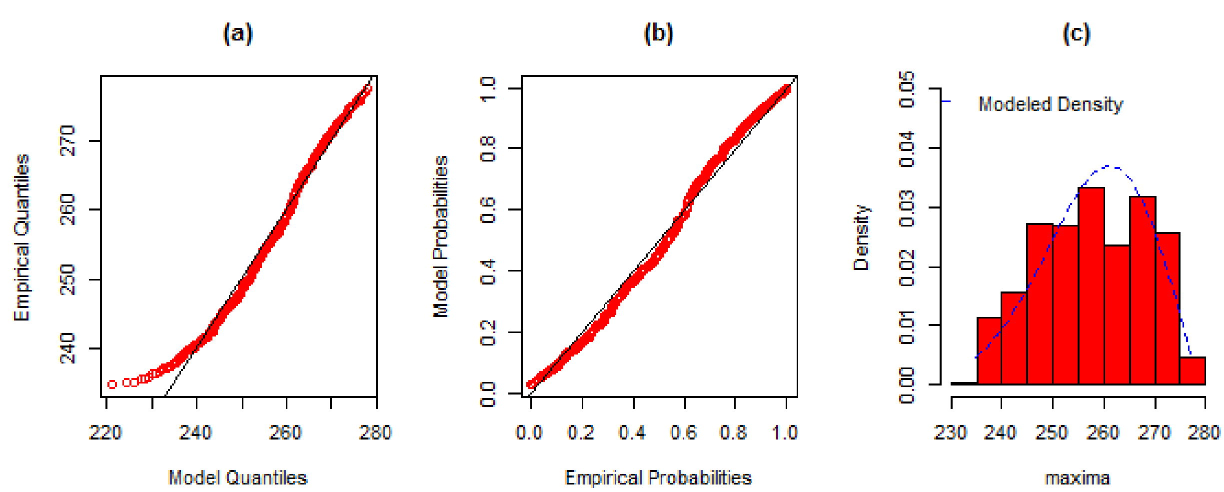

Fig. 2a: Diagnostic plots for the fitted GEV model for the monthly maximum water levels

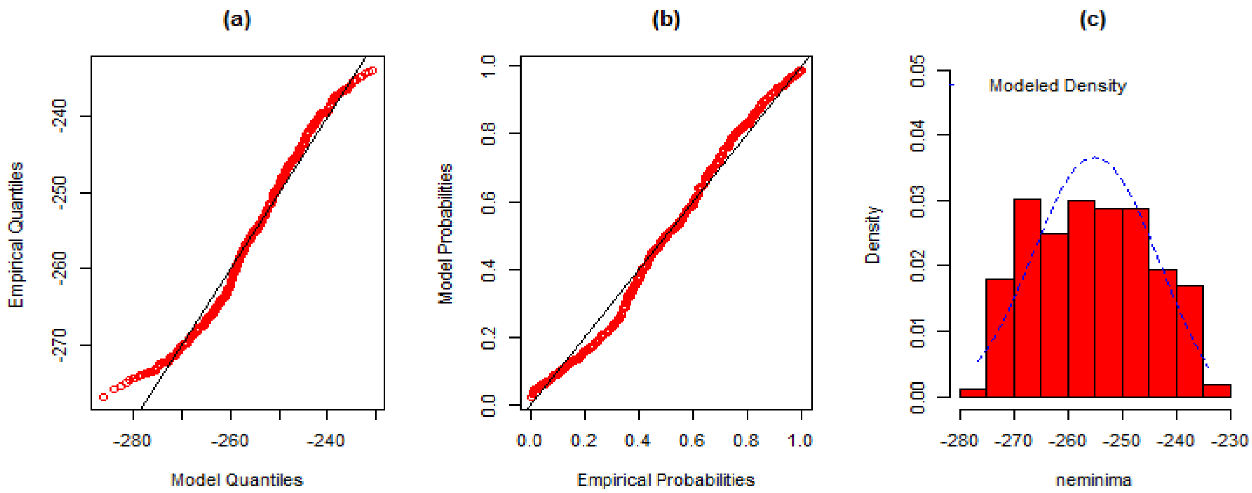

Fig. 2b: Diagnostic plots for the fitted GEV model for the monthly minimum water levels

Now that there is no doubt that the GEV model is valid for modeling the Akosombo dam data, we can estimate the upper endpoint of the distribution. The upper bound for the left tail is 226.69ft, which is much below the lowest water level ever recorded for the period under study but very close to the critical level of the dam. The upper bound for the right tail is 279.07ft, which is higher than the maximum water capacity (278ft) of the dam. Deducing from the endpoint estimates of the left tail and right tail, the water levels of the dam are bounded below by 226.69ft and above by 279.07ft. This implies that, the water level can neither fall below 226.67ft nor go above 279.07ft.

Table 3: Exceedance Probabilities of water levels of the dam

Tail	Water level (ft.)	Probability
Right	277.6	$1.872*10^{-4}$
	277.90	0.0000
Left	234.00	$1.09*10^{-2}$
	232.00	$4.154*10^{-3}$
	228.00	$6.04*10^{-5}$
	226.00	0.0000

Table 3 presents the probability of exceedances for some selected water levels of the dam. We observed that, very high water levels and very low water levels are associated with small probabilities. The exceedance probability for the right tail is interpreted as, the probability that the water level of the dam will exceed the specified level. Additionally, any level greater than the maximum capacity of the dam has an exceedance probability of zero, which implies that it is impossible to observe a level higher than what the dam can actually contain. Since there is a ceiling (278ft.) on the amount of water in the dam, any level close to it leads to spillage of water from the dam to save it from collapsing; therefore it is impossible for the dam to attain a level higher than 278ft. The exceedance probabilities for the left tail are interpreted as, the probability that the water level of the dam will fall below the specified level. Very low probabilities implies that the chances of observing such water levels are rare, but not impossible, and the zero probabilities imply that it is impossible to observe such water levels.

Conclusion

This study set out to investigate the tail behaviour of the water levels of the Akosombo dam using EVT, particularly the GEV distribution. The paper had two objectives: to estimate the bounds of the water levels of the dam and to estimate the exceedance probabilities of some selected (very high and low) water levels.

The ML estimation method was employed to estimate the EVIs and the other parameters. The EVI determines the tail behaviour and the domain of attraction of the underlying distribution (Coles 2001; Kotz and Nadarajah, 2000; Beirlant *et al*., 2004). The tail distribution of the Akosombo water levels follows the Weibull distribution; this confirms Minkah (2016).

The diagnostic plots indicate that the block maxima approach, that is, the GEV distribution, is very good for the description of the Akosombo water levels.

Considering only the water levels of the dam and all other things being equal, a complete shutdown of the dam is impossible, but a shutdown of a couple of turbines is inevitable. Therefore, the dependence on the Akosombo hydroelectric power plant for electricity supply should be reduced so that even when some turbines are shut down, "communities" depending on the dam for electricity supply would have enough power to carry out their activities without any power interruptions.

It can be observed from the study that water levels very close to the maximum operation head which would necessitate water spillage from dam are possible. Therefore, the dam's reservoir should be extended to retain the excess water to prevent flooding downstream and reduce the rate at which the water levels fall below the minimum operation head.

We conclude by stating that multivariate extreme value analysis which would include other factors likely to influence the water levels of the dam should be employed to investigate the problem further.

Acknowledgement

This work was supported by the Carnegie Corporation of New York under the University of Ghana-Carnegie Next Generation of Academics in Africa project through an MPhil thesis research grant awarded to EO.

References

Beirlant, J., Goegebeur, Y., & Teugels, J. (2004). *Statistics of Extremes Theory and Applications*. West Sussex, England: John Wiley & Sons, Ltd.

Bezak, N., Brilly, M., & Sraj, M. (2014). Comparison between the Peaks-Over-Threshold Method and annual maximum Method for Flood Frequency Analysis. *Hydrological Sciences Journal, 59*(5), 959 - 977. doi: 10.1080/02626667.2013.831174.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Great Britain: Springer - Verlag London.

Energy Commission, Ghana. (2016). Energy Supply and Demand Outlook for Ghana.

Hawkes, P. J., Ganzalez - Marco, D., Sanchez - Arcilla, A., & Prinos, P. (2008). Best Practice for the Estimation of Extremes: A Review. *Journal of Hydraulic Research, 46*(S2), 324 - 332. doi: 10.1080/00221686.2008.9521965.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society, 81*(348), 158-171.

Katz W. Richard , Parlange B. Marc , & Naveau P. (2002). Statistics of Extremes in Hydrology. *Advances in Water Resources, 25*, 1287 - 1304.

Kotz, S. & Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications*: World Scientific.

Mendez, F. J. & Menendez, M. (2006). Analyzing Monthly Extreme Sea levels with a Time-Dependent GEV Model. *Journal of Atmospheric and Oceanic Technology, 24*, 894 - 910. doi: 10.1175/JTECH2009.1.

Minkah, R. (2016). An application of extreme value theory to the management of a hydroelectric dam. *Springerplus, 5*, 96. doi: 10.1186/s40064-016-1719-2.

Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics, 3*(1), 119 - 131.

Prescott, P. & Walden A. T. (1980). Maximum Likelihood Estimation of the Parameters of the Generalized Extreme-Value Distribution. *Biometrika, 67*(3), 723 - 724.

Reiss, R-D. & Thomas, M. (2007). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields* (Third ed.). Germany: Birkhauser Basel. Boston. Berlin.

Ribereau, P. A. G. & Naveau, P. (2008). Estimating Return levels from Maxima of Non-stationary Random Sequences using the Generalized PWM Method. *Nonlinear Processes in Geophysics, 15*, 1033 - 1039.

VRA. (2010). Forty Ninth Annual Report and Accounts 2010: Volta River Authority (VRA), Ghana.

Watts, K. A., Dupuis, D. J. & Jones, B. L. (2009). An Extreme Value Analysis of Advance Ade Mortality Data. *North American Actuarial Journal, 10*(4), 162 - 178. doi: 10.1080/10920277.2006.1059741.