IJNES International Journal of Natural and Engineering Sciences 4 (3): 01-06, 2010
ISSN: 1307-1149, www.nobel.gen.tr

A Study on the Determining the Best Probability Distribution for the Annual Average Amount of Water Entering The Porsuk Dam

Arzu ALTIN YAVUZ* Barış ERGUL
Eskişehir Osmangazi University, Art and Science Faculty, Statistics Department, Eskisehir, TURKEY

***Corresponding Author**
e-mail: aaltin@ogu.edu.tr

Received : March 20, 2010
Accepted : May 28, 2010

Abstract

Available water resources should be assessed in an efficient manner for the effective use of water resources. Dams have an important place in the solution of water problems in general. We can make use of water resources plan according to the amount of water coming to the dams. That's why, prediction of future values of random variables such as precipitation, water flow, water coming to the dam has great importance. In this study, some probability distributions have been used for modeling the average annual amount of water coming in to Eskisehir Porsuk dam. For this purpose, the best fit of 33-years of data to the Normal, Lognormal, Logistic, Gamma and Weibull distributions have been tested with the Kolmogorov-Smirnov (KS) test. Test results showed that the annual average amount of water coming in to Porsuk dam best fits to the two-parameter Weibull distribution. Then, the shape and scale parameters of Weibull distribution, were determined by using the least squares, maximum likelihood and the method of moments. According to the criteria of mean square error, the most effective method was found to be the method of moments for the best estimates of the parameters.

Key Words: Kolmogorov-Smirnov Test, Porsuk dam, Prediction, Probability distribution, Water

INTRODUCTION

Climatic change, or global warming as it is more commonly termed, appears to be the chief issue that is increasingly raising alarm across the world. In view of the continually changing climate, modeling natural phenomena has gained as much importance as the issue itself. Despite the fact that our country is known to be rich in natural water resources, it has been suffering from water problems due to recent irregular precipitation and geographical conditions. Drought can be defined as a hydrological phenomenon that results in adverse impacts upon natural resources when the amount of the water entering the dam falls below average. One of the efficient ways of lessening the effect of drought is using the available water economically. However, using water efficiently becomes possible only if available water resources can be taken advantage of in a cost-effective way, apart from the necessity of monitoring water consumption [1]. It is at this crucial point that dams gain importance in coping with water-related problems. Dam lakes are generally built for purposes of energy production, irrigation, provision of drinkable water, and prevention of floods. Evaluation and analysis of hydrological data, that is, prediction of the amount of the water entering in dams, water-flow during droughts and floods, are of great importance on the subject of efficient water use.

Determining the best way to making most of available water resources requires that characteristics of water should be ascertained in consideration of such diverse issues as storing water without damaging nature, irrigation, and provision of water. However, the most difficult aspect of hydrological studies is measuring and evaluating the amount of water entering the dam and water flow. By analyzing the data obtained from the measurements undertaken, it is possible to determine the potential of a resource of water, in addition to determining values and probability of drought and overflowing. Since hydrological occurrences may be attributed to a large number of variables, they tend to have a trait that cannot be determined beforehand, that is, they are known for their unpredictability. For this reason, it is necessary to employ some probability and statistical methods in an effort to determine the values of such hydrological phenomena as precipitation, evaporation and leakage [1].

The present study aims to determine probability distribution and relevant parameters of the average annual amount of water that enter The Porsuk Dam from 1976 through 2008, a vital dam for the city of Eskişehir because of its huge capacity and its status as providing most of the available water for the city.

MATERIALS and METHODS

The Porsuk Dam has been built on the Porsuk River, which is located 40 km off the Southwest of Eskişehir. The location of the Porsuk Dam has been illustrated in Figure-1.

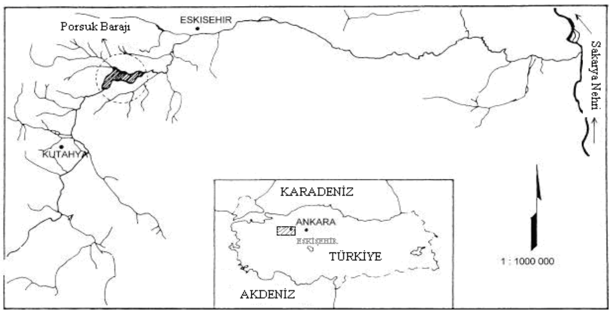

Figure-1 The Porsuk Dam

The Porsuk River had already brimmed over several times before it was extended with a view to increasing its volume of body of water in 1973 and thus caused extensive damage to Eskişehir. The maximum amount of water that this dam can take in is 525.000.000 m^3. However, considering the fact that the water level may rise at any time due to a sudden increase in the amount of water that may enter the dam on account of various factors, care is taken to always keep the level of the water in the dam far below the limit. Thus, floods and uncontrolled water release that would otherwise cause extensive damage to the available agricultural areas are intended to be kept under control.

The Porsuk River flows through the city of Eskişehir. The maximum amount of the water that is allowed to run through the city-centre is 40m^3 per second, because people living in this city face big problems in the case of a rise above this limit. The Porsuk Dam is what meets almost all the need for water in Eskişehir.

During the periods in which there is a shortage of water, the need for water cannot be met in the available agricultural areas, for which reason the optimum levels of crop cannot be achieved. Thus, determining the probability model of the amount of the water that runs into the Porsuk Dam per year is of great importance as far as an efficient way of using water is concerned. The average amount of the water entering this dam from 1976 through 2008 was obtained from the 3rd region designated by The State Hydraulic Works (SHW). The relevant data has been presented in Figure-2.

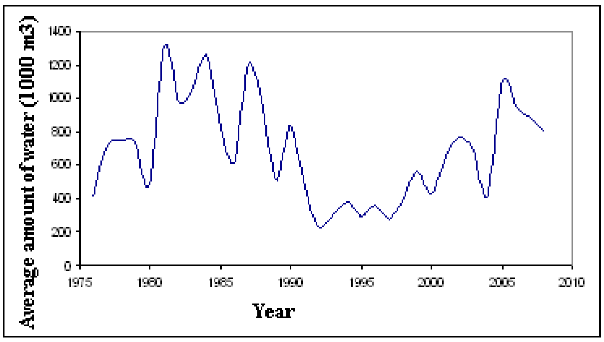

Figure 2. The average amount of annual water entering the Porsuk Dam

It is possible to make predictions for the upcoming years by looking into the fitting of various probabilities concerning series that exhibit random behavior. Among the most common models of probability distribution used for hydrological data are Normal, Lognormal, Weibull, Logistic, Gamma, Log-lojistic, Pearson, Gumbel and Student t. It is also possible to determine fitting of such models for the data obtained by using the Kolmogorov-Smirnov (KS) test at the desired level of probability [2, 3].

The Kolmogorov-Smirnov goodness of fit test was proposed by Kolmogorov for a single sample method in 1933 [4], which is based upon analyzing double cumulative distribution function [5]. The first of these functions is the cumulative distribution function as determined by null hypothesis, the second one being the observed cumulative distribution function obtained from the sample itself. The following lines show the way hypotheses are proposed by the Kolmogorov-Smirnov single-sampled test [4];

H_0 : O_i – E_i (The observed frequencies are in fitting with the expected ones)

H_0 : $O_i \neq E_i$ (The observed frequencies are not in fitting with the expected ones. The difference is of significance).

The statistical test to be used in putting the proposed hypotheses to test is showed by D, which is the largest of the absolute differences between cumulative relative frequencies of the observed and expected values [4].

$D = \max | F_o - F_e |$

F_o = The observed cumulative relative frequency

F_e = The expected cumulative relative frequency

The value of D is determined by comparing it with the critical value to be obtained from the table of Kolmogorov-Smirnov critic values. If the statistical value of this D islarger than that of the critical table, the hypothesis of H_0 cannot be rejected, and thus the decision is made in such a way that the observed frequencies are not in fitting with the expected ones, known as α *significance level*. Or else, the H_0 hypothesis is accepted and the decision is made in favor of the assumption that the observed frequencies are in fitting with those of the expected ones [5].

The present study aims to determine the fitting of Normal, Lognormal, Logistic, Gamma and Weibull distributions with that of the amount of the annual water that enters the Porsuk Dam. The chief reason for the choice of these distributions is that random variables assume only positive values in all the distributions apart from the Normal one. It is more logical to analyze the random variables that assume only positive values for the purpose of modeling the amount of the water entering the dam. The fitting analysis of the aforementioned distributions was tested by the Kolmogorov-Smirnov Test, the results of which have been presented in Table 1.

Table 1. Kolmogorov-Smirnov Testing Result

	Gamma (3 Parameter)	Lojistik	Lognormal	Normal	Weibull
DPLUS	0.078322	0.112609	0.087799	0.118100	0.104116
DMINUS	0.112655	0.076701	0.113183	0.066264	0.052366
DN	0.112655	0.112609	0.113183	0.118100	0.104116
p-value	0.796406	0.796814	0.791723	0.746774	0.866787

By looking at Table 1, we can see that the *p* values provided for the fitting of the 5 aforementioned distributions were all larger than the significance level of α=0,05. Thus, it is possible to take advantage of these 5 models in modeling of the average annual amount of the water having been entering the Porsuk dam over the years. However, as the high level of *p* reflects the acceptance probability of the hypothesis of H_0, it may also serve as an indicator of the fitting of the model in the problem. For this reason, distribution parameters have been predicted by trying out different methods, on the basis of the Weibull distribution which has the highest value of *p*.

The Weibull Distribution was developed by a Swedish physicist, Waloddi Weibull, in 1939. Among the situations to which this distribution has so far been applied to are the expectance model, life tables, duration of epidemic diseases, traveling periods, the amount of rainfall per m^2, and velocity of the wind, in all of which cases positive random variables may occur [6-8]. Furthermore, the Weibull Distribution may take shapes like a symmetric, positive or negative asymmetric distribution, based upon the value of a shape parameter. Because of this flexibility of the Weibull Distribution, it is commonly preferred in modeling phenomena that can be defined by means of positive random variables. It has been commonly used particularly in modeling annual flows [9-11] and predicting velocity of the wind [12, 13].

The general definition of the probability function of the Weibull Distribution with two parameters, which was used for the data on the annual amount of the water entering the dam, has been presented as follows;

$$f_w(v)=\left(\frac{k}{c}\right)\left(\frac{v}{c}\right)^{k-1}e^{-\left(\frac{v}{c}\right)^k} \tag{1}$$

Where *k* is the shape parameter of the Weibull Distribution, while *c* is the scale parameter. The cumulative distribution function of the Weibull Distribution has been presented in the Equation 2.

$$F_w(v)=1-e^{-\left(\frac{v}{c}\right)^k} \tag{2}$$

There are known to be various methods that have been developed to determine parameters of the Weibull distribution in the literature. We mostly employed the Maximum Likelihood Method, Method of Moments and Least Square Methods for the present study. Using these methods, we sought to predict the shape and scale parameters of the Weibull Distribution, which was found to fit the data on the annual amount of the water entering the Porsuk dam.

Method of Moments (MOM) is among the widely-applied methods in predicting parameters of the Weibull Distribution. The moment predictors of the shape and scale parameters of the Weibull Distribution have been represented in Equations 3 and 4.

$$k=\left(\frac{\hat{\sigma}}{v_{ort}}\right)^{-1.089} \tag{3}$$

$$c=\frac{v_{ort}}{\Gamma\left(1+1/k\right)} \tag{4}$$

V_{ort} is mean levels of the water entering the Porsuk Dam annually, while $\hat{\sigma}$ is the standard deviation value and *n* is the number of observations in a given period.

$\Gamma(\ )$, is gamma function whose value is obtained via 5 equation for any α value [14].

$$\Gamma(\alpha)=\int_0^\infty e^{-x}x^{\alpha-1}dx \tag{5}$$

Least Square Methods (LSM) is another widely-applied method in predicting parameters of the Weibull Distribution. First of all, in order to be able to benefit from the linear relationship between two parameters, it is necessary that logarithm of both sides of Equation (2) should be taken. Thus, the following equation will be achieved;

$$\ln\left[-\ln\left(\frac{1}{1-F(v)}\right)\right] = k\ln v - k\ln c$$

However, if we are to achieve the conversions

$Y = \ln[-\ln(1-F(v))]$, $X = \ln(v)$,
$a = -k\ln(c)$, $b = k$

we will arrive at the following linear equation [15-17].

$$Y = a + bX \tag{6}$$

In this conversion, the least square predictors of *a* and *b* parameters are determined by replacing them with X and Y. Following are LS predictions of the shape and scale of the Weibull Distribution, which are determined by Equation 7.

$$\hat{c}_{EKK} = e^{-\frac{\hat{a}}{\hat{b}}} \quad \text{ve} \quad \hat{k}_{EKK} = \hat{b} \tag{7}$$

y_i, v_i and $F(v_i)$ are ordered values that are used in solving equations. Previous studies have stipulated that $F(v_i)$ values should be determined independent of parameters in order to obtain LS predictions of the Weibull distribution $F(v_i)$. Different equations are available for determination of $F(v_i)$ values, the most commonly of which is Equation 8 that is derived by using the median values of y_i [16,18].

$$F(v_i) \approx \frac{i-0.3}{n+0.4} \tag{8}$$

The present study took advantage of this equation with a view to determining prediction values of $F(v_i)$.

Maximum Likelihood Method (MLM) enables prediction of the unknown *k* shape and *c* scale parameters by means of the following equations [19, 20].

$$\sum_{i=1}^{n}\ln v_i - n\left(\frac{\sum_{i=1}^{n}\ln v_i (v_i)^{\hat{k}}}{\sum_{i=1}^{n}(v_i)^{\hat{k}}}\right) + \frac{n}{\hat{k}} = 0 \tag{9}$$

$$\hat{c} = \left(\frac{\sum_{i=1}^{n}(v_i)^{\hat{k}}}{n}\right)^{1/\hat{k}} \tag{10}$$

Equation 9 may be solved by any of the available numerical root finding methods. Afterwards, the prediction value of the scale parameter can be directly determined via $\hat{c}$ in Equation 10. As to the solution of Equations 3, the Newton-Raphson Method tends to be preferred due to its speed of convergence regarding the root [20,21]. The general cycle equation for Newton-Raphson repetitions is as follows;

$$\hat{k}_{t+1} = \hat{k}_t + \frac{A + \left(1/\hat{k}_t\right) - C_t/B_t}{1/\hat{k}_t^2 + \left(B_t H_t - C_t^2\right)/B_t^2}$$

Herewith, the following repetitions

$$A = \sum_{i=1}^{n}\frac{\ln v_i}{n},\quad B = \sum_{i=1}^{n} v_i^{\hat{k}_t},\quad C_t = \sum_{i=1}^{n} v_i^{\hat{k}_t}\ln v_i,$$

$$H_t = \sum_{i=1}^{n} v_i^{\hat{k}_t}\left(\ln v_i\right)^2$$ ’ dir.

were determined using the starting point appearing below as

$$k_0 = \left[\frac{6}{\left(\pi^2(n-1)\right)}\left(\sum_{i=1}^{n}(\ln v_i)^2 - \frac{\left(\sum_{i=1}^{n}\ln v_i\right)^2}{n}\right)\right]^{-1/2}$$

[20, 21].

A numerical comparison of the three different methods employed in predicting parameters of the Weibull Distribution that fits the data on the annual amount of the water entering the Porsuk dam will benefit from Mean Square Error (MSE), which is a criterion used for the accuracy of estimators. It can be determined as follows;

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{F}(v_i) - F(v_i)\right\}^2 \tag{11}$$

Where,

$$\hat{F}(v_i) = 1 - \exp\left(-v_i/\hat{c}\right)^{\hat{k}},\quad F(v_i) = \frac{i-0.3}{n+0.4}$$

and *n* is the observation number [22].

RESULTS

Table 1 illustrates the results of the Kolmogorov-Smirnov Test that was undertaken in effort to determine the most fitting distribution among the probability distributions chosen for the data on the average annual amount of water entering the Porsuk dam. The Weibull distribution was determined to be the most fitting one for the data with a significance level of α=0.05. It is necessary that the shape and scale parameters of the Weibull probability density function should be determined before looking forward predictions can be made. The present study used 3 different predictions methods that have so far been commonly used in the literature in order to determine parameters of the Weibull distribution, the prediction values of which have been illustrated in Table 2.

Table 2. Prediction values of the shape and scale parameters of the Weibull Distribution

Method	**LSM**	**MLM**	**MOM**
Shape (k)	2.4676	2.4905	2.4212
Scale (c)	774.9310	777.0160	774.6970

Table 2 shows prediction values of parameters obtained by means of LSM and MOM to be very quite similar, while the predictions obtained by MLE were less similar to those of LSM and MOM. It is at this crucial point to determine which parameter values should be used for the prediction of the parameters of the Weibull Distribution. The present study used the MSE criterion presented in Equation 11 in determining which predictions values to use as parameters of the average annual amount of the water entering the Porsuk Dam. The MSE values obtained by 3 different methods have been presented in Table 3.

Table 3. **MSE values of** 3 different methods

Method	**MSE**
LSM	0.139684
MLM	0.140981
MOM	0.135995

According to the MSE value, which is an indicator of the difference between real values and predicted ones, the best predictions for the shape and scale parameters of the Weibull Distribution appear to be the ones made by the MOM. The next less fitting one was LSM, while the least fitting one was MLE. For this reason, we suggest that prediction values obtained by MOM should be used as the parameter predictions of the distribution of the annual amount of water entering the Porsuk Dam. The probability density function of the average annual amount of water entering the Porsuk Dam has been presented in Equation 12.

$$f_w(v) = \left(\frac{2.4212}{774.6970}\right)\left(\frac{v}{774.6970}\right)^{1.4212} e^{-\left(\frac{v}{774.6970}\right)^{2.4212}} \quad (12)$$

DISCUSSION

There has been a rapid depletion in the amount of available water because of the increasing population of the world and industrialization. Furthermore, difficulty in predicting precipitation due to global warming contributes to difficulty in predicting such natural disasters as drought and flood. This being the case, it is of great importance to use available water in an efficient way. However, the water resources be tended efficiently should be stored in such a way that balance of nature is not disturbed in any way, so that the water in these reservoirs can be consumed for diverse purposes. It is at this crucial step that dams gain so much importance in solving water problems. Therefore, it is vital that the actual amount of available water in dams should be estimated accurately in consideration of designing dams, managing water in these dams and preventing floods. If the amount of the water in dams is overestimated, it could result in significant increases in costs, while loss of life or material damage could occur if underestimated. It is for this reason that several probability distributions are employed in analysing hydrological phenomena.

As it cannot be known for sure whether or not a certain distribution is fitting for given data, the chief task should be to determine the closest of these distributions in relation to the given data. Determining only the shape of the distribution is not sufficient for the estimations to be made, though how accurate the estimations that have been made are should be evaluated in the meantime.

Büyükkaracığan and Kahya (2009), who determined the probability distribution fitting best for the annual peak currents of the 13 streams occurring in the vicinity of Konya, obtained estimators for the parameters of the distributions as far as various methods of estimating parameters are concerned. Anlı (2006) reported that they used some probability distributions for modelling maximum current frequencies belonging to the base of Giresun Aksu, and that they determined the Weibull Distribution with 3 parameters to be fitting for these currents.

The present study aims to determine the best probability distribution in determining the possible amount of water that might enter the Porsuk Dam in future years by taking advantage of the observations concerning the average annual amount of the water entering this dam, which meets most of the need for water in Eskişehir. The conclusion was that distributions of Normal, Lognormal, Logistic, Gamma and Weibull could be fitting for the data regarding the average annual amount of the water entering this dam. Whether these distributions are fitting for these data was investigated by the Kolmogorov-Smirnov fitting test, the results of which Weibull appears to be the bets one. Afterwards, in an effort to obtain predictability of the shape and scale parameters of the Weibull Distribution, the methods of LSM, MLM and MOM were employed. The criterion of MSE was used in determination of the most fitting estimation values of parameters obtained for 3 different methods. According the MSE, the parameter estimations values achieved by means of the MOM method were those that best reflected the real distribution of the data obtained. We suggest that the probability distribution given in Equation 12 could be used in determination of the possible amount of the water that might enter the Porsuk dam in following years. The Porsuk Dam is the most important dam that is used for generating power, as well as providing water to be used in homes, meeting the need for water to be used in irrigation and industrialization. With the help of the distribution method emphasized in our study, it could be possible to predict the amount of water to enter the dam. Also, it might be possible to make decisions with probability in planning how to distribute the available water, as well as being able to determine the risk of flood and uncontrolled water release. Damage to agricultural areas could be lessened by taking precautions against drought periods. Finally, it is of great importance for our country, which is not rich enough in water and is increasingly feeling the need for more water because of its growing population, to take advantage of its water resources in a scientific way.

REFERENCES

[1] Ülke A, Baran T, 2005. Köprüçay yıllık akım verilerine uygun olasılık dağılım fonksiyonu ve kuraklık analizi. Antalya Yöresi İnşaat Mühendisliği Sorunları Kongresi, 20-24 Eylül 2005, Antalya.

[2] Tülücü, K., 1996. Uygulamalı Hidroloji. Ç.Ü. Ziraat Fakültesi Yayınları No: 138, Adana.

[3] Tonkazı T, Çetin M, Çullu MA, 2007. Diyarbakır meteoroloji istasyonu aylık ortalama sıcaklıklarının olasılık analizi. I. Türkiye İklim Değişikliği Kongresi – TİKDEK 2007, 11 - 13 Nisan, 2007, İTÜ, İstanbul.

[4] Kartal, M., 1998. Hipotez Testleri. Şafak Yayınevi, Erzurum.

[5] Gamgam, H., 1998. Parametrik Olmayan İstatistiksel Teknikler. Gazi Üniversitesi Yayınları, Ankara.

[6] Aytaç, M., 1991. Matematiksel İstatistik. Ezgi Kitabevi Yayınları, Bursa.

[7] Şenoğlu B, 2005. Robust 2(k) factorial design with Weibull error distributions. Journal of Applied Statistics, 32 (10): 1051-1066.

[8] Şenoğlu B, 2007. Estimating parameters in one-way analysis of covariance model with short-tailed symmetric error distributions. Journal of Computational and Applied Mathematics, 201(1): 275-283.

[9] Anlı, AS., 2006. Giresun aksu havzası maksimum akımlarının frekans analizi. Akdeniz Üniversitesi Ziraat Fakültesi Dergisi, 19(1), 99-106.

[10] Lee KS, Kim SU, 2007. Identification of uncertainty in low flow frequency analysis using Bayesian MCMC method, Hydrol. Process., 22 (12): 1949–1964.

[11] Wilks DS, 1988. Rainfall intensity, the Weibull distribution, and estimation of daily surface runoff. Journal of Applied Meteorology, 28: 52-58.

[12] Kantar YM, Şenoğlu B, 2008. A Comparative study for the location and scale parameters of the Weibull distribution with a given shape parameter. Computers&Geosciences, 34(12) :1900-1909.

[13] Kamio T, Iida M, ARAKAWA C, 2007. Influences of the estimating methods on weibull parameter for wind speed distribution in Japan. European Wind Energy Conference & Exhibition MIC - Milano Convention Centre, 7-10 May 2007, Milan, Italy.

[14] Gökçek M, Bayülken A, Bekdemir Ş, 2007. Investigation of wind characteristics and wind energy potential in Kirklareli, Turkey. Renewable Energy, 32, 1739-1752.

[15] Hossain AM, Zimmer WJ, 2003. Comparison of estimation methods for weibull parameters: Complete and censored samples. Journal of Statistical Computation and Simulation, 73(2), 145-153.

[16] Almalı MN, Okut H, 2009. İki Parametreli weibull dağılımında parametre tahmin yöntemlerinin karşılaştırılmasına ilişkin bir simülasyon. Anadolu University Journal of Science and Technology, 10(2): 455-465.

[17] Lei Y, 2008. Evaluation of three methods for estimating the Weibull distribution parameters of Chinese pine (*Pinus tabulaeformis*). J. For. Sci., 54: 566-571.

[18] Ross R, 1999. Comparing linear regression and maximum likelihood methods to estimate weibull distribution on limited data sets: systematic and random errors. IEEE Conference on Electrical Insulation and Dielectric Phenomena, Cilt (I): 170-173.

[19] Thoman DR, Bain LJ, Antle CE., 1969. Inferences on the parameters of the Weibull distribution. Technometrics, 11(3), 445-460.

[20] Danacı MA, Birgören B, Ersöz S, 2009. Weibull parametreleri ve yüzdelikleri için güven aralığı tahmin algoritmaları. Gazi Üniv. Müh. Mim. Fak. Der., 24 (1): 119-128.

[21] Law, AM, Kelton, WD. 2000. Simulation modeling and analysis. McGraw-Hill, New York.

[22] Al-Fawzan MA, 2000. Methods for estimating the parameters of the weibull distribution, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia.