ARTICLE

JOURNAL OF UNIVERSITY OF BABYLON

For ENGINEERING SCIENCES (JUBES)

Vol.30, No.3.| 2022

Using Statistics in Hydrology for Analyzing the outflow by Minitab Program

Zainab Ali Omran

College of Civil Engineering, University of Babylon, Babylon, Iraq

Email: zainabalialtaee@yahoo.com

Wisam Shamkhi Jaber

College of Civil Engineering, University of Babylon, Babylon, Iraq

Email: eng.wisam.jaber@uobabylon.edu.iq

Mohamed Hussein Al rawaf

The General Authority for Dams and Reservoirs, Kufa and Abbasiya System Project Management

Shaker hanidi naema

The General Authority for Dams and Reservoirs, Kufa and Abbasiya System Project Management

Received:	26/7/2022	Accepted:	9/10/2022	Published:	10/10/2022

Abstract

The purpose of this study is to examine using statistics in hydrology for analyzing outflow discharge in Hadith Dam. Total annual outflow data of thirty years for period from 1991to 2020 were used. Mathematical equation for the probability distribution functions were established for the data. Three tests were used, namely (Anderson-Darling, Ryan- Joiner Similar to Shapiro-Wilk and Kolmogorov- Smirnov) to check the normality of the data; the predicted values were subjected to goodness of fit tests such as kolmogorov-smirnov to determine the distribution which is suitable for the data. Depending on the kolomogrov-smirnov (k-s) index, the distributions were suitable for the data, therefore, all the fitted distributions were typical for the data that were used in this paper depending on (k-s) index.

Keywords: Outflow, Cumulative distribution function, Normality test, Minitab Program

مـجلـة جـامعة بـابل للعلـوم الهندسية مـجلـة جـامعة بـابل للعلـوم الهندسية مـجلـة جـامعة بـابل للعلـوم الهندسية

ISSN: 2616 - 9916 | www.journalofbabylon.com | Journal.eng@uobabylon.edu.iq | info@journalofbabylon.com | www.journalofbabylon.com

ISSN: 2616 - 9916 | www.journalofbabylon.com | Journal.eng@uobabylon.edu.iq | info@journalofbabylon.com

1- Introduction

Statistical distributions and models are frequently used in many fields of research, including the social, biological, and engineering sciences as well as economics. They are used to simulate a range of real-world problems. For these probability models to be used successfully, a thorough knowledge of the theory and familiarity with the situations where particular distributions can be suggested were required [1].

The design of water structures, watershed management, and the identification of significant hydrologic cycle components can be studied using statistical distributions. But it is critical to identify the distribution that best fits the examined data [2, and 3]. Statistical analysis of any data can be described on certain parameters. Parameters like mean, standard deviation, variance, skewness, and kurtosis [4]. Finding appropriate probability distribution models for the Hadith Dam discharge data is the goal of this endeavor. To define which probability distribution best fits the data and to ensure that the data are normal, three widely used probability distributions—two-parameter standard and normal, two-parameter log and normal, and two-parameter gamma—are utilized.

2. Data and Analysis

In this paper the total annual outflow data of thirty years for period from 1991 to 2020 for Haditha Dam were analyzed as shown in Figure below

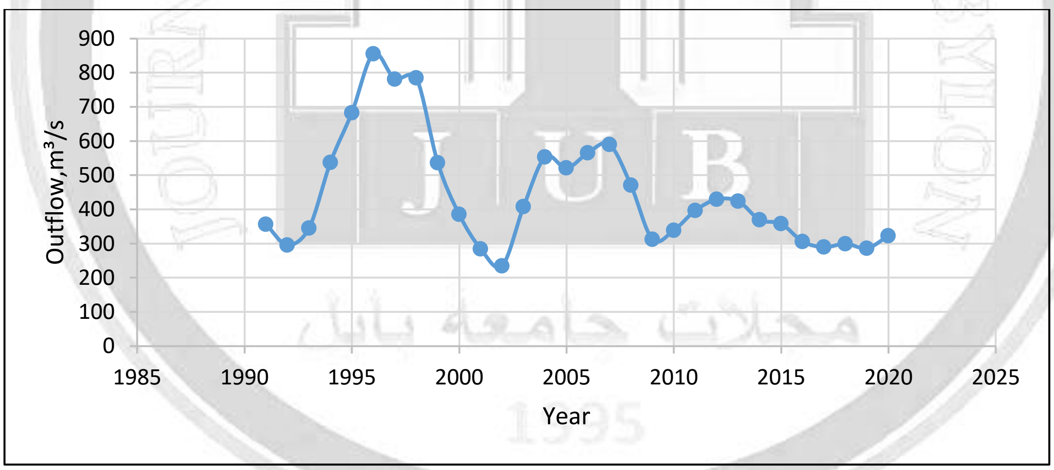

Fig. 1: Outflow data (m³/sec) for Hadith Dam.

The normal, log normal, and gamma distributions were the three probability distributions chosen. These distributions were chosen because they were effective in numerous earlier studies on rainfall modeling., [5] and [6].

Probability density models were tested statistically to identify the best-fit model, goodness of fit tests and Kolmogorov-Smirnov.

3. Probability Distributions

In statistical analysis, a variety of plausible probability distributions are used. Three probability distributions considered in this study are the Normal, Log Normal, and Gamma distributions.

3.1 Normal Distribution

This distribution, referred to as the Gaussian distribution, is one of the most significant illustrations of a continuous probability distribution. The distribution's density function is given by [7]:

$$f(x)= \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad -\infty \le x \le \infty \quad \ldots\ldots(1)$$

where μ and σ are the mean and standard deviation, respectively. The corresponding distribution function is given by Equation 2:

$$F(x)= P(X\le x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(x-\mu)^2/2\sigma^2} \, dx \quad \ldots\ldots(2)$$

3.2 Log Normal Distribution

The independent variable (x) is replaced by its logarithm in the Log Normal distribution; otherwise it is identical to the Normal distribution. The Log Normal distribution has a strong positive skew and is bounded on the left by zero, among other properties. If the logarithm of a random variable, often the natural logarithm, is normally distributed, the random variable is said to have a log normal distribution. The probability density function of such a variable $y=ln(x)$ [5].

$$f(x)=\frac{1}{\sqrt{2\pi}\,\sigma_y}\, exp\left[-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)\right] \quad \text{for} \quad 0\le x \le \infty \ldots\ldots(3)$$

3.3 Gamma Distribution

The two-parameter gamma distribution has one shape and one scale parameter. The random variable X follows gamma distribution with shape and scale parameters as $b > 0$ and, $F > 0$ respectively, it has the following probability density function (PDF); [8].

$$f(x) =\frac{1}{F^b\Gamma(b)} x^{b-1}e^{-x/F} \quad x > 0 \ldots\ldots (4)$$

Γ the gamma function is referenced. In contrast to the "normal" distribution, almost all distributions do not include the mean and/or standard deviation as parameters; the gamma distribution is typical in this regard. The mean of a gamma random variable is bF and the variance is bF^2 [9]. For fitting the distribution to the data the Cumulative Distribution Function (CFD) used for all the distributions; see Table 1 and Figure 2 to Figure 5 (These Figures shows the probability plot and cumulative probability distribution for normal and gamma distributions).

ISSN: 2616 - 9916

www.journalofbabylon.com | Journal.eng@uobabylon.edu.iq | info@journalofbabylon.com

مجلة جامعة بابل للعلوم الهندسية

مـجلــة جــامعة بــابـل للعلـــوم الهندسية مـجلــة جــامعة بــابـل للعلـــوم الهندسية مـجلــة جــامعة بــابـل للعلـــوم الهندسية

ISSN: 2616 - 9916 | www.journalofbabylon.com | Journal.eng@uobabylon.edu.iq | info@journalofbabylon.com

Table. 1: Calculations of probabilities for outflow data

Rank	Normal	Log normal	Gamma
	F(x) %	F(x) %	F(x) %
1	10.25	4.63	8.05
2	16.69	13.06	16.62
3	16.97	13.47	17.00
4	17.43	14.14	17.62
5	18.36	15.48	18.86
6	19.07	16.53	19.81
7	20.16	18.13	21.25
8	21.27	19.77	22.72
9	23.13	22.50	25.13
10	26.21	26.98	29.06
11	27.41	28.69	30.55
12	29.75	32.00	33.43
13	30.08	32.46	33.83
14	32.57	35.87	36.80
15	36.08	40.51	40.86
16	38.72	43.86	43.82
17	41.34	47.08	46.68
18	45.13	51.52	50.67
19	46.41	52.96	51.98
20	56.45	63.48	61.76
21	67.96	73.89	71.93
22	71.21	76.55	74.63
23	71.43	76.74	74.82
24	74.58	79.22	77.38
25	76.91	81.01	79.25
26	81.17	84.18	82.62
27	92.63	92.36	91.68
28	97.96	96.57	96.50
29	98.06	96.67	96.61
30	99.37	98.14	98.26

مـجلـة جـــامعة بــــابل للعلـــوم الهندسية مـجلـة جـــامعة بــــابل للعلـــوم الهندسية مـجلـة جـــامعة بــــابل للعلـــوم الهندسية

ISSN: 2616 - 9916 | www.journalofbabylon.com | Journal.eng@uobabylon.edu.iq | info@journalofbabylon.com

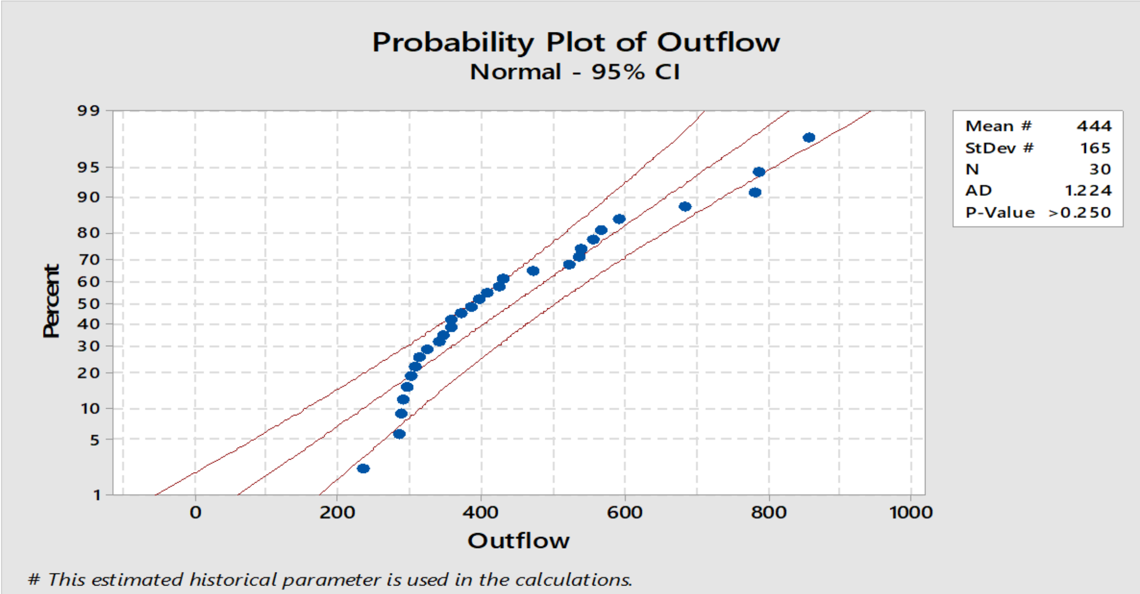

Fig.2: Probability plot for normal distribution

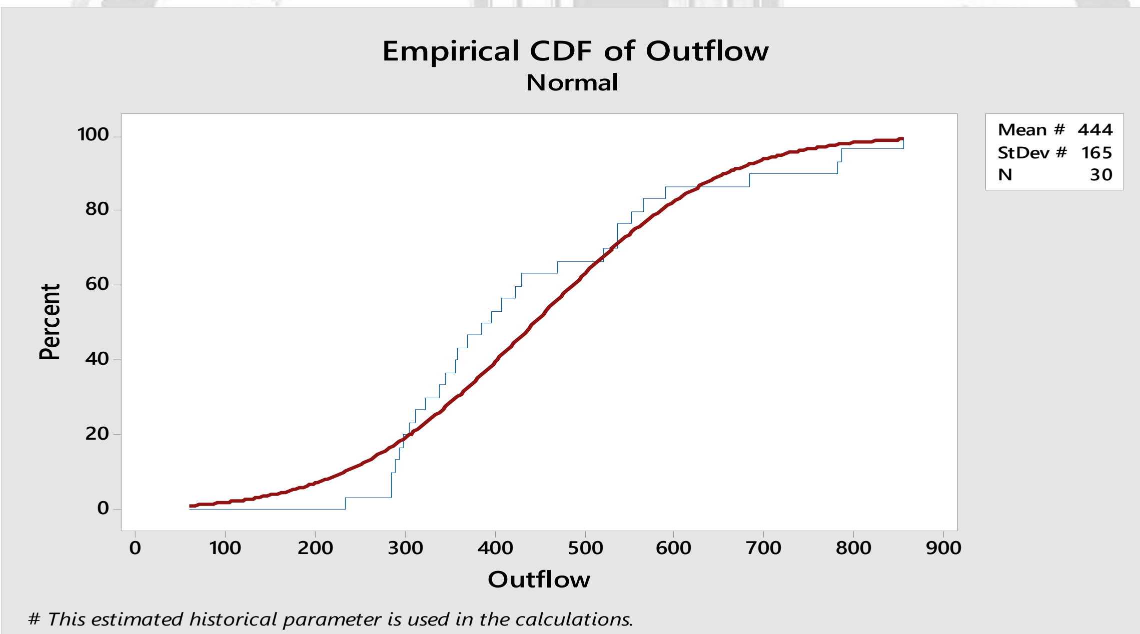

Fig.3: Cumulative probability distribution plot for normal distribution

مجلة جامعة بابل للعلوم الهندسية
ISSN: 2616 - 9916 | www.journalofbabylon.com | Journal.eng@uobabylon.edu.iq | info@journalofbabylon.com

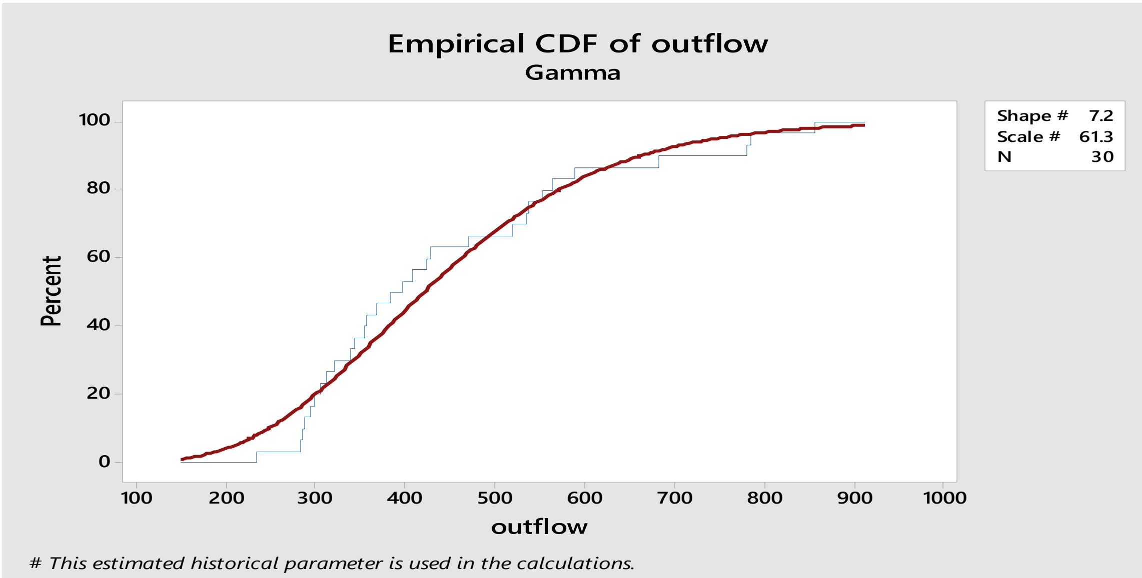

Fig.4: Cumulative probability distribution plot for Gamma distribution

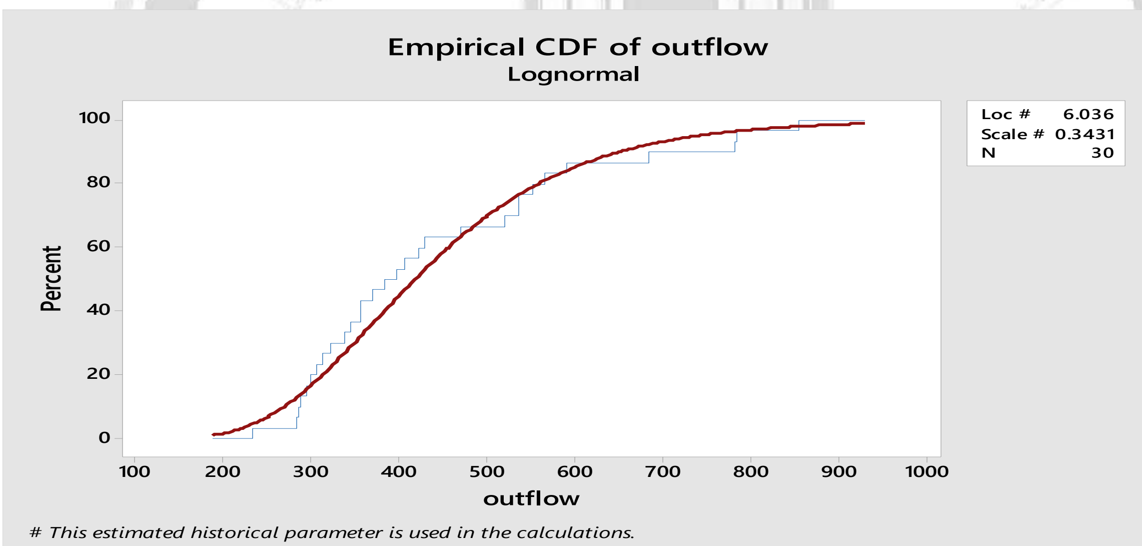

Fig.5: Cumulative probability distribution plot for log normal distribution

4. Goodness of Fit Tests

The goodness of fit problem has recently piqued the interest of the particle physics community. Modern particle research usually employs unbinned likelihood fits to the data. The experimenter must first measure the degree to which the t function fits the observed distribution. Many methods have been used in the past to solve this problem, and many more methods have recently been published in the physics literature [10].

مجلة جامعة بابل للعلوم الهندسية
ISSN: 2616 - 9916 | www.journalofbabylon.com | Journal.eng@uobabylon.edu.iq | info@journalofbabylon.com

Kolmogorov- Smirnov test

The Kolmogorov-Smirnov (K-S) goodness of fit test is based on a statistic that quantifies how far the cumulative histogram observed deviates from the cumulative distribution function expected.

- The maximum absolute value of $D_2 = S(x_i) - F(x_i)$ obtained from the above equation is compared with theoretical value shown in statistical tables. Reject hypothesis H if D_2 greater than theoretical value in tables (for critical values for the Kolmogorov-Smirnov test) Otherwise, accept H [11].

Table 2: Results of K-S test for the three distributions

i	Outflow	Normal	S=i/n	D
1	234.9	0.1025	0.0333	-0.0692
2	284.5	0.1669	0.0667	-0.1002
3	286.4	0.1697	0.1000	-0.0697
4	289.4	0.1743	0.1333	-0.0410
5	295.2	0.1836	0.1667	-0.0169
6	299.6	0.1907	0.2000	0.0093
7	306.1	0.2016	0.2333	0.0318
8	312.5	0.2127	0.2667	0.0539
9	322.8	0.2313	0.3000	0.0687
10	338.9	0.2621	0.3333	0.0713
11	344.9	0.2741	0.3667	0.0926
12	356.3	0.2975	0.4000	0.1025
13	357.8	0.3008	0.4333	0.1325
14	369.4	0.3257	0.4667	0.1410
15	385.2	0.3608	0.5000	0.1392
16	396.7	0.3872	0.5333	0.1461
17	407.9	0.4134	0.5667	0.1533
18	423.8	0.4513	0.6000	0.1487
19	429.1	0.4641	0.6333	0.1693
20	470.8	0.5645	0.6667	0.1022
21	521.0	0.6796	0.7000	0.0204
22	536.3	0.7121	0.7333	0.0213
23	537.4	0.7143	0.7667	0.0523
24	553.1	0.7458	0.8000	0.0542
25	565.4	0.7691	0.8333	0.0643
26	589.9	0.8117	0.8667	0.0549
27	683.1	0.9263	0.9000	-0.0263
28	781.4	0.9796	0.9333	-0.0462
29	785.0	0.9806	0.9667	-0.0140
30	855.2	0.9937	1.0000	0.0063

Max absolute 0.1693 (from Table 2 max absolute value of D)
From standard table D=0.24 (from standard Tables of K-S test)
Accepted for all distributions

5- Normality test

It is possible to quantify the likelihood that a random variable underlying a set of data is normally distributed using normality tests, which examine whether a set of data can be effectively described by a normal distribution. Since regularly distributed data is a fundamental assumption of all parametric testing, determining the normality of data is important for many statistical tests. In other words, a most of statistical methods must be applied in order for the data to behave in a Gaussian manner [12].

The time series are assumed to be normally distributed if the P value provided by the software is equal to or greater than 0.05 [12]. The Anderson-Darling, Ryan-Joiner (comparable to the Shapiro-Wilk), and Kolmogorov-Smirnov tests are chosen in this study to test for normalcy in Minitab.

6. Results

- Results of K-S test, all the fitted distributions were typical for the data that are used in this paper depends on (k-s) index as shown in Table 3 because the theoretical value of D from standard tables is greater than the observed value from calculations for all the distributions.

Table 3: Check the K-S test

Data	*n*	Theo. *D.*	Obs. *D*		
			Normal Dis.	Log normal Dis.	Gamma Dis.
Outflow discharge	30	0.24	0.1693	0.109	0.113

- Using the normality test for the three tests Anderson-Darling, Ryan- Joiner (Similar to Shapiro-Wilk) and Kolmogorov- Smirnov), The p- value is less than 0.05 for all the 3 tests which means 5%, the data are not normalized as shown in figure 6 to Figure 8.

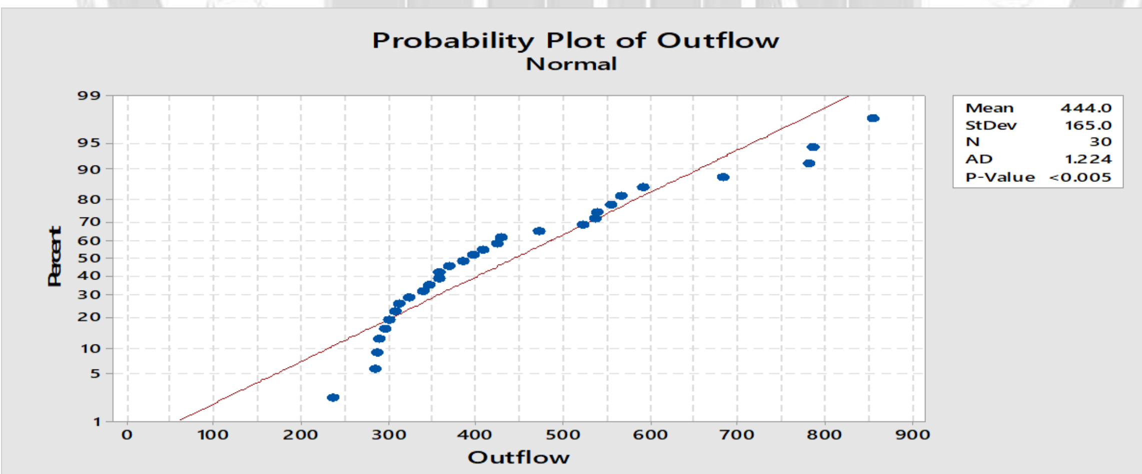

Fig. 6: Normality test using Anderson-Darling, p-value < 0.005

ISSN: 2616 - 9916 | www.journalofbabylon.com | Journal.eng@uobabylon.edu.iq | info@journalofbabylon.com
مجلة جامعة بابل للعلوم الهندسية

ISSN: 2616 - 9916 | www.journalofbabylon.com | Journal.eng@uobabylon.edu.iq | info@journalofbabylon.com
مجلة جامعة بابل للعلوم الهندسية

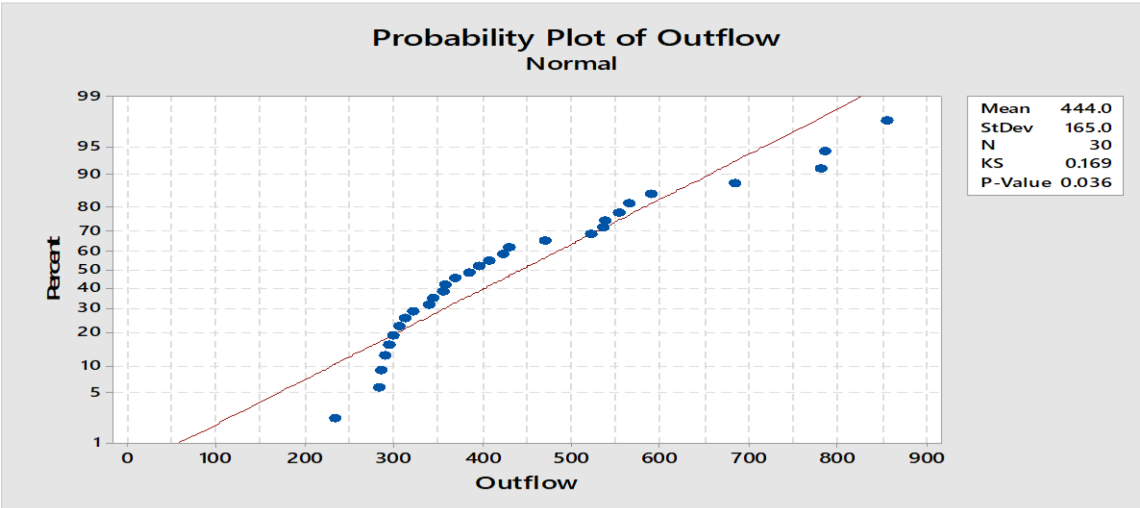

Fig .7: Normality test using Kolomogrov- Smirnov, p-value=0.036

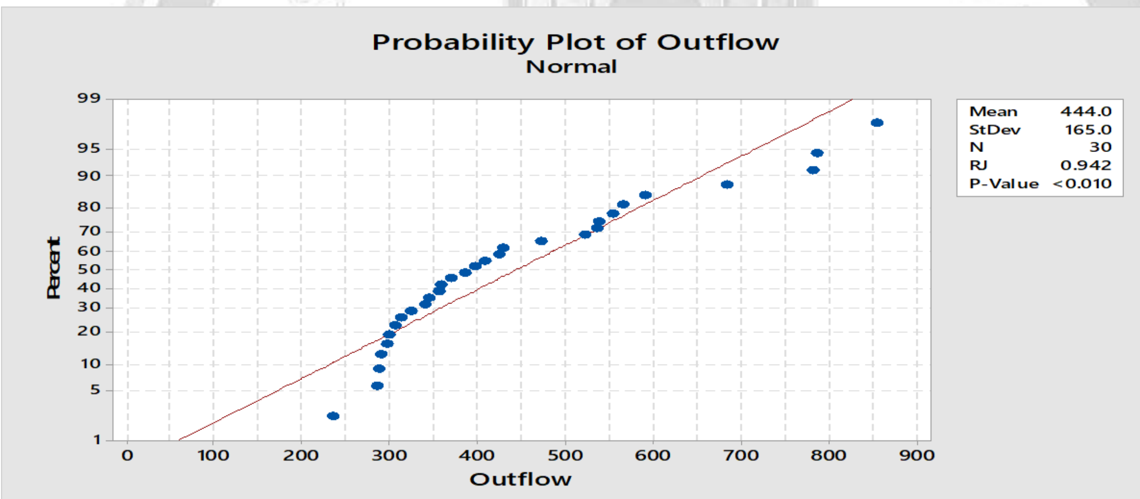

Fig .8: Normality test using Ryan- Joiner (Similar to Shapiro-Wilk), p-value < 0.01

7- Conclusion

In this paper total annual outflow data of thirty years for period from 1991 to 2020 were used, depending on K-S test, three probability distributions were suitable to the outflow data, and normality test also used to check the normality

8- References

[1] Catherine Forbes et. al, "Statistical Distributions" Fourth Edition, Published by John Wiley & Sons, Inc., Hoboken, New Jersey, 2011.

مجلة جامعة بابل للعلوم الهندسية مجلة جامعة بابل للعلوم الهندسية مجلة جامعة بابل للعلوم الهندسية مجلة جامعة بابل للعلوم الهندسية
ISSN: 2616 - 9916 | www.journalofbabylon.com | Journal.eng@uobabylon.edu.iq | info@journalofbabylon.com

[2] Omran Zainab Ali, "Determining Suitable Probability Distribution Models for Total Annual Rainfall Data (A Case Study of Najaf, Kerbala, and Diwaniya Stations) " The Iraqi Journal for Mechanical And Material Engineering, Special Volume Babylon First International Engineering Conference Issue (D), 2015.

[3] T. Jesan, et al, "Statistical analysis of rainfall events at Kalpakkam" Conference Paper, 2016.

[4] Kaamun et al, "Statistical Analysis of Rainfall in Haryana" IOP Conf. Series: Earth and Environmental Science 889, doi:10.1088/1755-1315/889/1/012003, 2021.

[5] Kwaku, Xeflide, Seth and Duke, Ophori, "Characterization and frequency analysis of one day annual maximum and two to five consecutive days' maximum rainfall of Accra, Ghana", ARPN Journal of Engineering and Applied Sciences, Vol. 2, No. 5, 2007.

[6] Olofintoye, O.O, Sule, B.F and Salami, A.W., "Best–fit probability distribution model for peak daily rainfall of selected cities in Nigeria", Department of Civil Engineering, University of Ilorin, P.M.B 1515, Ilorin, Nigeria, New York Science Journal, 2(3), ISSN 1554 0200, 2009. http://www.sciencepub.net/newyork, sciencepub@gmail.com.

[7] Chow, Ven Te, Maidment, David R., and Mays, Larry W., "Handbook of applied hydrology", McGraw-Hill series in water Resources and Environmental Engineering, New York, ISBN 0-07-010810-2, 1988.

[8] Pradhan, Biswabrata and Kundu, Debasis, "Bayes estimation and prediction of the two-parameter Gamma distribution", Department of Mathematics and Statistics, Indian Institute of Technology, Kanpur, Pin 208016, India, 2010.

[9] Ryan, Thomas P., "Modern engineering statistics", A John Wiley & Sons, INC., Publication, Acworth, Georgia, ISBN 978-0- 470-08187, 2007.

[10] I. Narsky, "Goodness of Fit: What Do We Really Want to Know?" Phystat, SLAC, Stanford, California, September 8-11, 2003, 2003.

[11] Frank J. Massey, Jr., " The Kolmogorov-Smirnov Test for Goodness of Fit" Journal of the American Statistical Association, Vol. 46, No. 253 (Mar., 1951), pp. 68- 78 Published by: American Statistical Association Stable URL: http://www.jstor.org/stable/2280095 Accessed: 14/06/2010 04:54, 2010.

[12] Mohammadbagher Amjadi, et al, " Statistical analysis of Rainfall Data (Case study: TRNC)" ENG Transactions, H & T Publication, 2021, 2, pp.1-8. ffhal-03229208, 2021.

ARTICLE
JOURNAL OF UNIVERSITY OF BABYLON
For ENGINEERING SCIENCES (JUBES)
Vol.30, No.3.| 2022

استخدام الإحصاء في الهيدرولوجيا لتحليل التدفق الخارج بواسطة برنامج Minitab

زينب علي عمران

كلية الهندسة المدنية، جامعة بابل، بابل، العراق

الايميل: zainabalialtaee@yahoo.com

وسام شمخي جابر

كلية الهندسة المدنية، جامعة بابل، بابل، العراق

الايميل: eng.wisam.jaber@uobabylon.edu.iq

محمد حسين الرواف

الهيئة العامة للسدود والخزانات، إدارة مشروع منظومة الكوفة والعباسية

شاكر هنيدي نعمه

الهيئة العامة للسدود والخزانات، إدارة مشروع منظومة الكوفة والعباسية

الخلاصة

الغرض من هذه الدراسة هو فحص استخدام الإحصاء في الهيدرولوجيا لتحليل تصريف التدفق الخارج في سد حديثة. تم استخدام إجمالي بيانات التدفق السنوي لمدة ثلاثين عامًا للفترة من 1991 إلى 2020. تم إنشاء معادلة رياضية لوظائف التوزيع الاحتمالي للبيانات. واستخدامت ثلاثة اختبارات هي Anderson–Darling و (Ryan– Joiner و Kolmogorov– Smirno) للتحقق من صحة البيانات، وقد تم إخضاع القيم المتوقعة لاختبار ملاءمة مثل kolmogorov–smirnov لتحديد التوزيع المناسب للبيانات. حسب اختبار K–s التوزيعات مناسبة للبيانات، لذا فإن جميع التوزيعات المجهزة كانت نموذجية للبيانات المستخدمة في هذه الدراسة بالاعتماد على اختبار K–S

الكلمات الدالة: التدفق الخارجي، دالة التوزيع التراكمي، اختبار الوضع الطبيعي، برنامج Minitab

ISSN: 2616 - 9916 | www.journalofbabylon.com | Journal.eng@uobabylon.edu.iq | info@journalofbabylon.com

مجلة جامعة بابل للعلوم الهندسية مجلة جامعة بابل للعلوم الهندسية مجلة جامعة بابل للعلوم الهندسية مجلة جامعة بابل للعلوم الهندسية