

12-1-2019

## Evaluation the Best Random Component in Modified Thomas-Fiering Model in Generating Rainfall Data for Akre station

Fawaz Kh. Aswad

*Department of Water Resources Engineering, University of Duhok, Duhok, Kurdistan Region, Iraq,*  
Fawaz\_khl@uod.ac

Ali Abduljabar Yousif

*Department of Water Resources Engineering, ali.yousif@uod.ac*

Sayran A. Ibrahim

*University of Duhok, Duhok, Kurdistan Region, Iraq, sayran.ahmed@uod.ac*Follow this and additional works at: <https://polytechnic-journal.epu.edu.iq/home>Part of the [Life Sciences Commons](#)

### How to Cite This Article

Aswad, Fawaz Kh.; Yousif, Ali Abduljabar; and Ibrahim, Sayran A. (2019) "Evaluation the Best Random Component in Modified Thomas-Fiering Model in Generating Rainfall Data for Akre station," *Polytechnic Journal*: Vol. 9: Iss. 2, Article 30.

DOI: <https://doi.org/10.25156/ptj.v9n2y2019.pp186-192>

This Research Article is brought to you for free and open access by Polytechnic Journal. It has been accepted for inclusion in Polytechnic Journal by an authorized editor of Polytechnic Journal. For more information, please contact [karwan.qadir@epu.edu.iq](mailto:karwan.qadir@epu.edu.iq).

---

## Evaluation the Best Random Component in Modified Thomas-Fiering Model in Generating Rainfall Data for Akre station

### Abstract

In this research, the effect of random component in the modified Thomas-Fiering model to generate daily rainfall data was studied, and Akre station considered a case study. A random component with special distributions: Normal random numbers, Wilson-Hilferty (W-H) transformation, truncated W-H, and Kirby modification to W-H transformation were used. The model applied to the daily rainfall data for Akre station for available years 2000–2006 and the model used to generate the rainfall data for the years 2006 and 2007. The results showed that the correlation coefficients between the observed and generated data were 0.82 for normal random numbers, 0.77 for W-H transformation, 0.89 for truncated W-H, and 0.87 for KM to W-H transformation. The tests of Chi-square test, Kolmogorov–Smirnov test, root mean squared error (RMSE) test, and mean absolute error (MAE) test were used to compare between observed and generated data. All the results have passed the Chi-square test and Kolmogorov–Smirnov, where the calculated values were less than the tabulated value at 5% significance. For the test RMSE and MAE, the truncated W-H transform was the values of at least two. Therefore, W-H transform is the best for generating the rainfall data at Akre station

### Keywords

Thomas-Fiering, Rainfall, forecasting, Random component, Stochastic models

## RESEARCH ARTICLE

# Evaluation the Best Random Component in Modified Thomas-Fiering Model in Generating Rainfall Data for Akre station

Fawaz Kh. Aswad\*, Ali A. Yousif, Sayran A. Ibrahim

*Department of Water Resources Engineering, University of Duhok, Duhok, Kurdistan Region, Iraq*

### \*Corresponding author:

Fawaz Kh. Aswad,  
Department of Water  
Resources Engineering,  
University of Duhok, Duhok,  
Kurdistan Region, Iraq.  
E-mail: Fawaz\_khll@uod.ac

**Received:** 21 October 2019

**Accepted:** 01 December 2019

**Published:** 10 December 2019

### DOI

10.25156/pj.v9n2y2019.pp186-192

## ABSTRACT

In this research, the effect of random component in the modified Thomas-Fiering model to generate daily rainfall data was studied, and Akre station considered a case study. A random component with special distributions: Normal random numbers, Wilson-Hilferty (W-H) transformation, truncated W-H, and Kirby modification to W-H transformation were used. The model applied to the daily rainfall data for Akre station for available years 2000–2006 and the model used to generate the rainfall data for the years 2006 and 2007. The results showed that the correlation coefficients between the observed and generated data were 0.82 for normal random numbers, 0.77 for W-H transformation, 0.89 for truncated –W –H, and 0.87 for KM to W-H transformation. The tests of Chi-square test, Kolmogorov–Smirnov test, root mean squared error (RMSE) test, and mean absolute error (MAE) test were used to compare between observed and generated data. All the results have passed the Chi-square test and Kolmogorov–Smirnov, where the calculated values were less than the tabulated value at 5% significance. For the test RMSE and MAE, the truncated W-H transform was the values of at least two. Therefore, W-H transform is the best for generating the rainfall data at Akre station

**Keywords:** Thomas-Fiering; Rainfall; forecasting; Random component; Stochastic models; Generation data

## INTRODUCTION

Water is the key element to sustain life and most of the civilizations have risen in the vicinity of water sources to acquire water for drinking and agricultural purposes. In addition, water resources provide fertile land during flood events. The development of human civilization has increased the need for water in various fields. The availability of water resources has become a rare resource which requires quantity assessment, projects planning, and managing it through scientific means.

The evaluation and planning of water projects demand and forecasting water resources through utilizing the hydrological available data of sufficient period (time series). However, the researchers in the hydrology field suffer from the lack of hydrological series monitored for a sufficient period, where the hydrological monitoring stations are few and in most cases, there is a lack of recorded data. Therefore, it is of prime importance in such cases to depend on historical facts that describe the phenomena and natural disasters such as floods, heavy rainfall, or long-term

drought to add these data to the existing data in the design of water resources projects.

For many years, the hydrologic time series forecasting used in the planning, design and management of water resources systems (Kim et al., 2004). The forecasting methods have been widely used in the prediction of stream flows (Stedinger and Taylor, 1982) and the prediction of rainfall (Yousif et al., 2016). The predicted time series from the stochastic models is as important as the real available data in the planning, design, and operating of water resources systems (Wijayarathne and Chan, 1987).

Thomas and Fiering (1962) were first explored the role of using stochastic methods in the water resources systems planning and operating through the generation of synthetic times series of stream flows through Monte Carlo simulation. A forecasting model developed to generate the streamflow time series data with similar correlation behavior of the original data, and this was an example of Markovian type models.

The Thomas-Fiering (T.F.) model can be considered as a typical forecasting model in hydrology (Harms and

Campbell, 1967; Joshi and Gupta, 2009; McMahon and Miller, 1971; and Thomas and Fiering, 1962).

In 1967, Harms and Campbell have modified T.F. (MTF) model to preserve the normal distributions of annual flow, the log-normal distributions of monthly flows, and the autocorrelation of both annual and monthly flows. Over the last decade, the TF model used in forecasting stream flows (Kurunç et al., 2005), reservoir operations (Joshi and Gupta, 2009), and rainfall predictions (Yousif et al., 2016).

This article deals with the selection of the best random component of the Modified – T.F. model for forecasting the daily rainfall time series for Akre station to predict the rainfall data for the purposes of agricultural and water resources projects use.

## MATERIALS AND METHODS

### Study Area and Data

The climate of the Akre city is hot and dry in summer and other seasons are usually rainy. Akre City is located on the Kurdistan region of Iraq at about 115 km from east of Duhok city and 85 km north of Mosul city at coordinate (430 53' 07" E, 360 44' 13" N), as shown in [Figure 1]. In July, temperature degree is about 46 Co as a maximum, and in January, temperature degree is about -1.9

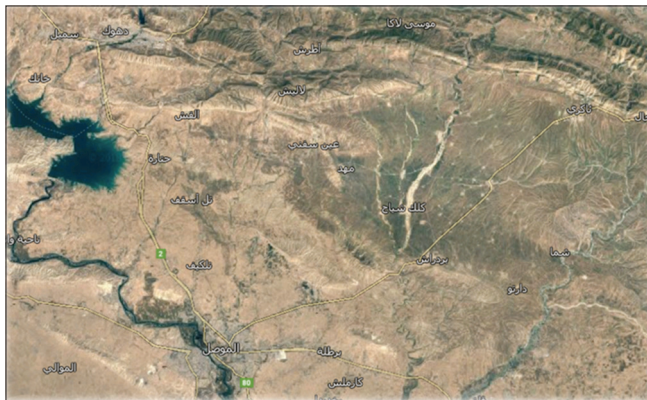


Figure 1: Location of Akre City (Google Earth)

Co as a minimum. The rate of relative humidity ranged between 46.06% and 66.9%.

In general, the meteorological data at the selected station in this research showed seasonal rainfall from the beginning of winter to the end of spring, where the average annual rainfall in the Akre city is 587.3 mm. The highest rainfall in the Akre city is recorded in December and January.

In the past 10 years, the Kurdistan Region has been affected by drought due to the global warming. Table 1 shows the values of the highest, lowest, and average rainfall for the period 1941–1982, and Table 2 shows the highest, lowest, and the average rainfall for the period 1993–2007. These data were directly collected from the general directorate of agriculture in Akre.

### MTF Model

It is sometimes necessary to generate discharge series for missing data (rainfall or discharge) during the dry season. In 1984, the researcher Clark has MTF model to generate hydrological data to the basin, which has a dry period of some months during the year, such as rainfall or seasonal rivers.

The following procedures used to generate the rainfall data using MTF model (Clark, 1984):

1. Find the percentage of record, which is equal to the number of month, which has non-zero data ( $n_j$ ) to the total number of year ( $N$ ). For each month  $j$  (except the month of dry) this can be estimated by:

$$P_j = \frac{n_j}{N} \quad (1)$$

2. For each month, calculate the mean ( $\bar{X}_j$ ) and variance ( $S_j$ ) for the rainfall data, and then find the regression coefficient ( $B_j$ ) and correlation coefficient ( $R_j$ ) between two consecutive months as follows:-

$$\bar{X}_j = \frac{\sum_{i=1}^N X_{i,j}}{N} \quad (2)$$

Table 1: Highest, lowest, and average monthly rainfall for the period 1941–1982

Months	January	February	March	April	May	June	July	August	September	October	November	December
Highest	336.7	466.4	469.7	274.8	231.8	6.7	0	0	4.4	71.1	271.9	472.4
Average	161.5	160.8	150.4	110.5	37.2	0.35	0	0	0.10	16.0	85.0	124
Lowest	14.1	23.4	11.0	0.1	0.1	2.3	0	0	0.0	0.1	3.6	12.4

Table 2: Highest, lowest, and average monthly rainfall for the period 1993–2007

Months	January	February	March	April	May	June	July	August	September	October	November	December
Highest	257.10	254.5	186.50	168.00	186.70	1.50	0.00	0.00	0.00	126.1	259.70	259.00
Average	142.92	137.17	101.14	76.73	29.85	0.10	0.00	0.00	0.00	31.08	89.24	112.33
Lowest	39.20	36.20	11.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.00	7.50

Where:

$\bar{X}_J$ : Average of rainfall or discharge for the month J.

$$S_J = \sqrt{\frac{\sum_{I=1}^N (X_{I,J} - \bar{X}_J)^2}{N-1}} \quad (3)$$

Where:

$S_J$ : Variance or Standard deviation of rainfall or discharge for the month J.

$$B_J = \frac{\sum_{i=1}^n (X_{I,J} - \bar{X}_J)(X_{I,J+1} - \bar{X}_{J+1})}{\sum_{i=1}^n (X_{I,J} - \bar{X}_J)^2} \quad (4)$$

$B_J$ : Regression coefficient of rainfall or discharge between two consecutive months (J and J+1).

$$R_J = \frac{\sum_{i=1}^n (X_{I,J} - \bar{X}_J)(X_{I,J+1} - \bar{X}_{J+1})}{\sqrt{\sum_{i=1}^n (X_{I,J} - \bar{X}_J)^2 * \sum_{i=1}^n (X_{I,J+1} - \bar{X}_{J+1})^2}} \quad (5)$$

$R_J$ : correlation coefficient of rainfall or discharge between two consecutive months (J and J+1).

Steps to generate synthetic series of monthly rainfall:

- Choose a Sudo – random number, rectangular distribution over (zero, 1); for month j, if this number is less than  $P_j$  (but  $>0$ ), then the rainfall is to occur in month j, otherwise, no rainfall is to occur.
- If no rainfall is to occur in month J, repeat step 3, 1 for month +1.
- For the 1<sup>st</sup> month in the year (I), and if rainfall is to occur in this month according to the procedure of step a, then the selected pseudo-random numbers (Y) deviate for a distribution with mean and variance equal to the mean monthly rainfall ( $\mu$ ) and variance of rainfall ( $\sigma$ ) for month J using the relation (6) and the value of rainfall for month J will be equal to the value of random number:

$$Y = \mu + \sigma * z \quad (6)$$

The value of z is normally and independently random number with zero mean and unit variance.

- If the rainfall is to occur in month J, and rainfall occurs in month J – 1, use the regression equation of the T.F. (eq. 7) to obtain the rainfall for month J.

$$X_{I,J} = \bar{X}_J - B_J (X_{I,J-1} - \bar{X}_{J-1}) + YY * S_{J-1} * \sqrt{1-R_{J-1}} \quad (7)$$

$X_{I,J}$  = the generated rainfall value for month J and year I.  
 $YY$  = Random number with special distribution.

### The Random Number Component

In any stochastic model, the process is naturally random. The random numbers are essential parts of the process of generating rainfall data. Computers have a sub-function to generate random numbers distributed regularly (uniformity distribution). The generation of each element in a series of random numbers with uniform distribution is random in the chain compared with the rest of the time series (Clark, 1984).

Random numbers used in the models incidental form random numbers follow the normal distribution. In this research, four different formulas used to generate random numbers with different distributions. The following section summarizes these formats:

### Normally – Distributed Pseudo-Random Number

#### Box – Muller's method

In 1958, the researchers Box and Muller's developed a method for generating the random number with normal distribution from the varieties  $v_1$  and  $v_2$  rectangular – distributed over the interval (0, 1).

$$z_1 = (-2 \log_e x_1)^{1/2} * \cos(2\pi v_1) \quad (8)$$

$$z_2 = (-2 \log_e x_2)^{1/2} * \cos(2\pi v_2) \quad (9)$$

Where:

$v_1$  and  $v_2$ : the random numbers with uniform distribution.

#### Wilson-Hilferty (W-H) transformation

The transformation of W-H for random component is to generate random component with special mean and variance. This transformation is as follow (Srikanthan, 1978):

$$\epsilon_J = \frac{2}{SC_\epsilon} \left( 1 + \frac{SC_\epsilon * v}{6} - \frac{SC_\epsilon^2}{36} \right)^3 - \frac{2}{SC_\epsilon} \quad (10)$$

Where:

$\epsilon_J$  = random component generated by W-H Transformation  
 $SC_\epsilon$ : Skewness Coefficient of  $\epsilon_J$  and can calculate from the following equation:

$$SC_\epsilon = \frac{1 - r_k^3}{(1 - r_k^2)^{3/2}} * SC_J \quad (11)$$

Where:

$SC_J$  = Skewness Coefficient of the observed series.

$r_k$  = autocorrelation coefficient at lag = 1, for the observed values, can be found using the following equation:

$$r_k = \frac{\sum_{t=1}^{N-k} (X_t - \bar{X})(X_{t+1} - \bar{X})}{\sum_{t=1}^n (X_t - \bar{X})} \quad (12)$$

Where:

$$SC_J = \frac{N * \sum_{i=1}^n (X_{I,J} - \bar{X}_J)^3}{(N-1)(N-2) * S_J^3} \quad (13)$$

#### Truncated W-H

In 1969, Payne et al. found that at higher values of  $r_k$ , the obtained values for  $\varepsilon_j$  can be lower than the theoretical values, so to correct this Payne et al. (1969) transformed W-H transformation to truncated W-H. This transform can be expressed as follows:

$$\varepsilon_j^T = \max\left(\frac{-2}{r_k}, \varepsilon_j\right) \quad (14)$$

Where:  $\varepsilon_j^T$  = Truncated W-H random component.

#### Kirby modification for W-H transformation

Kirby (1972) modified W-H transformation so that they could remain satisfied for all data and the modification is as follows:

$$\varepsilon_j^M = A \left[ \max \left\{ H, 1 - \left( \frac{C}{6} \right)^2 + \frac{C}{6} * Z \right\}^3 - B \right] \quad (15)$$

Where:  $\varepsilon_j^M$  = Kirby modification for W-H transformation and can find the value of H as follows:

$$H = \left[ B - \left\{ \frac{2/CS_J}{A} \right\}^{\frac{1}{3}} \right] \quad (16)$$

Where: H, A, B, C = constants dependent on the value of Skewness coefficient  $CS_J$ . To find, the values of these constants in the equation (16) use the following formulas:

$$A = \max \left( \frac{2}{CS_J}, 0.4 \right) \quad (17)$$

$$B = 1 + 0.0144 * \max(0, CS_J - 2.25)^2 \quad (18)$$

$$C = CS_J - 0.063 * \max(0, CS_J - 1)^{1.85} \quad (19)$$

#### Statistical Tests

##### Chi-square test ( $\chi^2$ )

The Chi-square test ( $\chi^2$ ) can be determined as following (Barnes, 1994):

$$\chi_{ca}^2 = \frac{\sum_{i=1}^n (O_i - G_i)^2}{G_i} \quad (20)$$

$O_i$  = the frequency of observed values within the range of i.  
 $G_i$  = the frequency of Generated values within the range of i.

To pass the Chi-square test, the tabular value ( $\chi_o^2$ ) given at the level of significance ( $\alpha$ ) is greater than the calculated value of ( $\chi_{ca}^2$ ) from the abovementioned equation. Further, the performance of the model is acceptable if ( $\chi_o^2 > \chi_{ca}^2$ ), otherwise, the test is rejected (Barnes, 1994).

#### Kolmogorov-Smirnov test ( $\Delta$ )

The general equation for this test is:

$$\Delta_c = \max \left| \frac{O_i - G_i}{N} \right| \quad (21)$$

If the critical  $\Delta_c$  calculated from the above formula is less than the tabular value  $\Delta_o$  at the level of significance ( $\alpha$ ), the application of the model is accepted.

#### Mean absolute error (MAE)

The MAE test defined as following (Wang, 2006):

$$MAE = \frac{1}{N} \sum_{i=1}^N |X_i - \bar{X}_i| \quad (22)$$

#### Applying the MTF Model

The model of MTF is applied to generate daily data for Akre station, in which a series of daily rainfall for a period of 2000–2006 is considered. The model applied for four cases of the random numbers as described above to find the best random component suitable for generating rainfall data at the studied station.

The model is implemented to generate daily rainfall data of the year 2006–2007, following the steps mentioned in section 3. Then, the generated data compared with the observed data of rainfall for the same period years. Figures 2–5 show that there is a convergence between the observed data and generated data.

Figures 6–9 show that there is a convergence between the values of average rainfall in month for observed data and values of generated data.

## RESULTS DISCUSSION AND CONCLUSIONS

To identify the best random component model and in order to evaluate the efficiency of the model in generating data, the three aforementioned statistical tests presented in



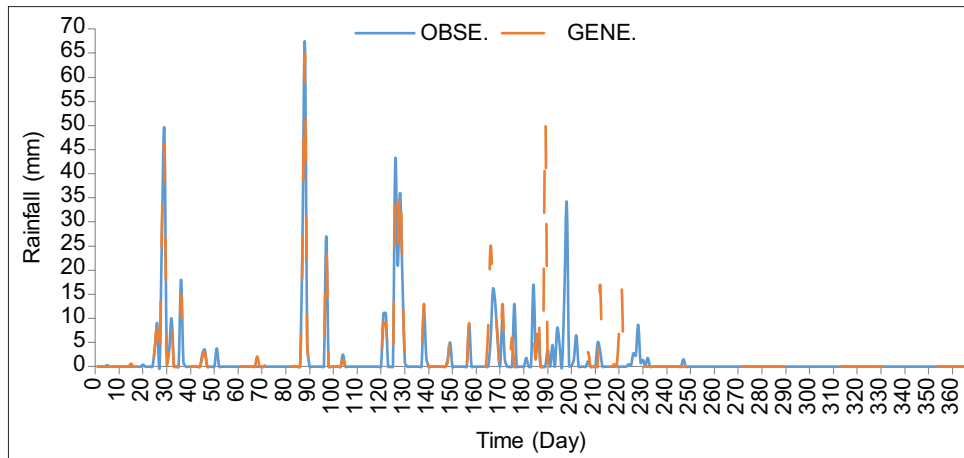


Figure 2: Observed and generated data with normal random component

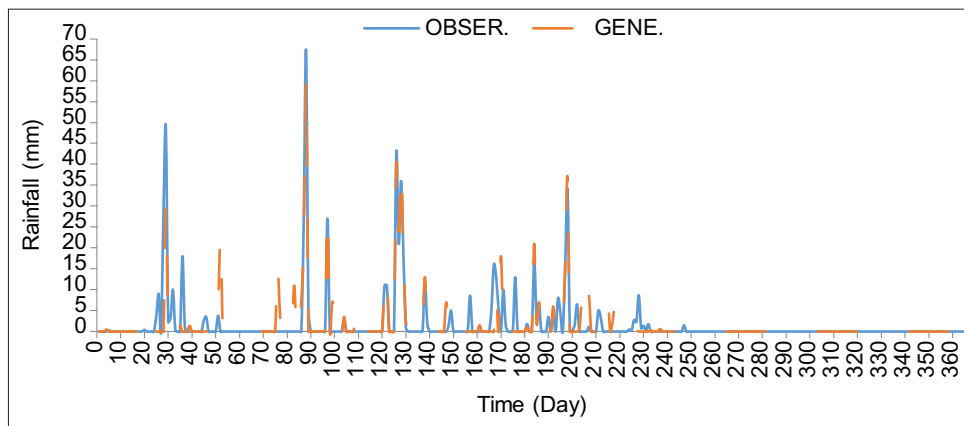


Figure 3: Observed and generated data with Wilson-Halfirtly transformation random component

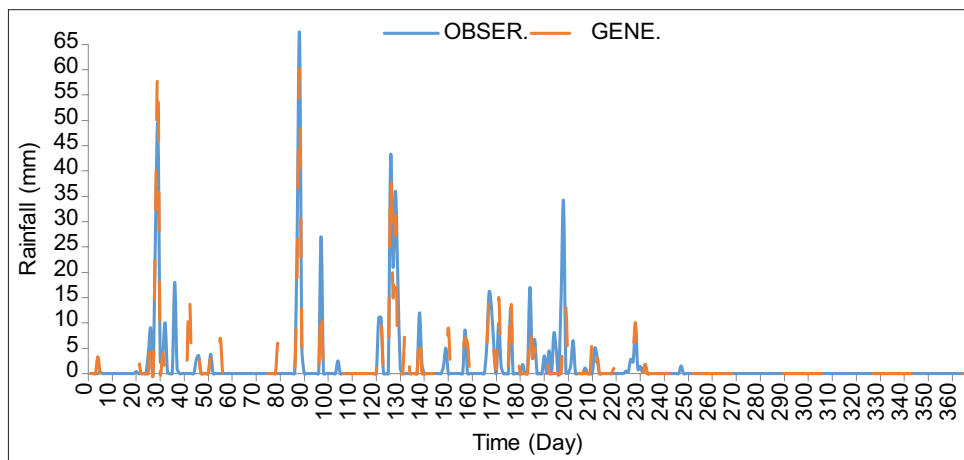


Figure 4: Observed and generated data with truncated Wilson-Halfirtly transformation random component

section 5 are applied to generate the desired data. Table 3 illustrates the results of these tests.

In Table 3, it can be seen clearly that the random component models used in this study have passed the Chi square – test where the results of this test are less than the tabulate value

at level of significant ( $(\alpha) 5\%$ ) ( $X^2_c = 2.73$ ). The table also shows that  $\Delta$  – test for all random component models has passed the test at level of signification 5%, and for the MAE the model with truncated W-H random component represent the smallest value from the other, according to that we can say that is the best model.

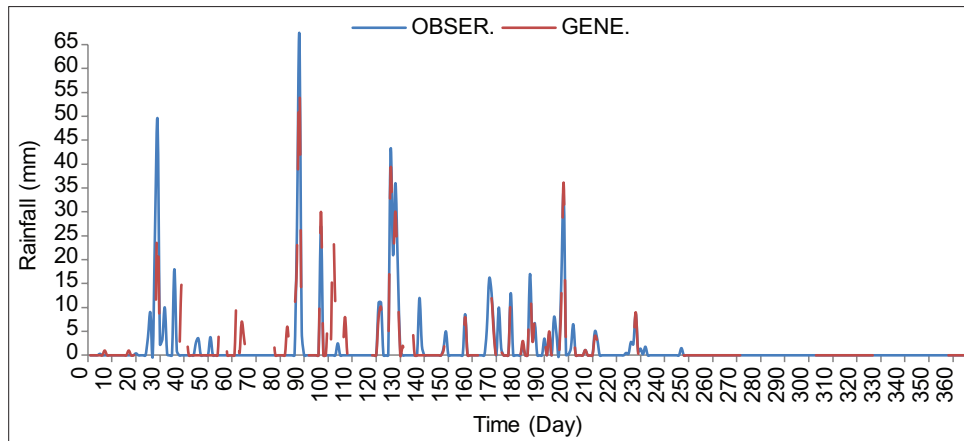


Figure 5: Observed and generated data with Kirby modification to Wilson-Halferty transformation random component

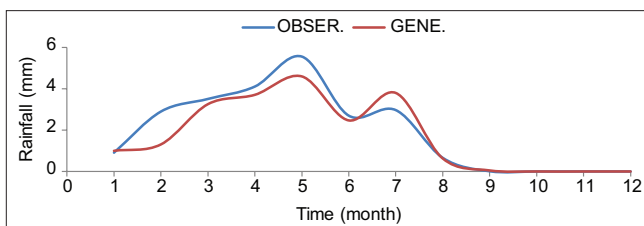


Figure 6: The scheme of average values for observed data and generated data using normal random component

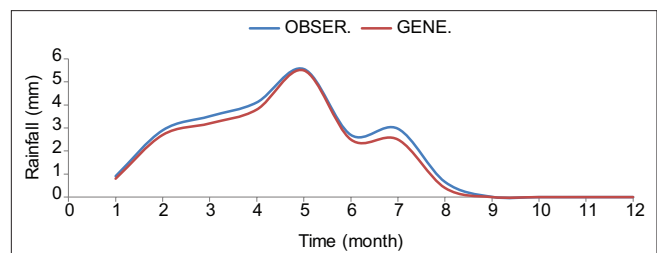


Figure 9: The scheme of average values of observed data and generated data using Kirby modification to Wilson-Halferty transformation as a random component

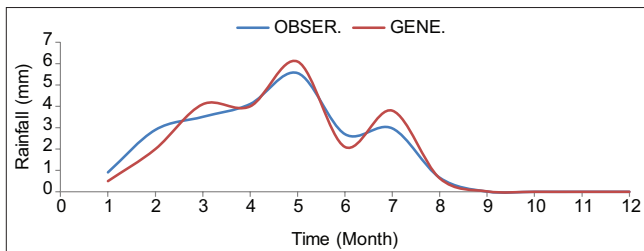


Figure 7: The scheme of average values for observed data and generated data using Wilson-Halferty transformation as a random component

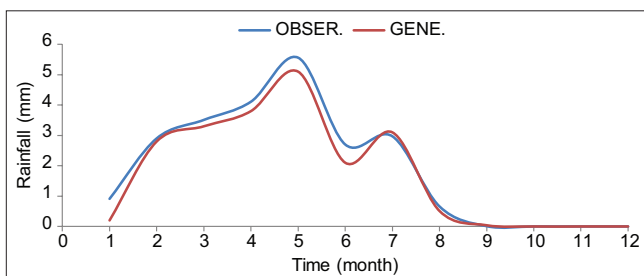


Figure 8: The scheme of average values for observed data and generated data using truncated Wilson-Halferty transformation as a random component

Table 3: The results of statistical tests

R. C. Tests	MAE	$\chi^2$	$\Delta$	R
Normal R. C.	0.91	2.77	0.42	0.81
Wilson-Halferty R.C.	1.32	2.51	0.61	0.77
Truncated Wilson-Halferty	0.90	2.3	0.41	0.89
Kirby modification to Wilson-Halferty	0.97	2.6	0.56	0.87

1. Significant differences in the results of each model can be observed, which mean that the choice of the random component model considerably affects the performance of the stochastic model.
2. The results of generated rainfall data of Akre station, using all random components models, showed that no negative values are appeared in the generated data, and this demonstrates the validity of the models.
3. It was found that the correlation coefficient between generated data and observed data ranges between 0.770 and 0.89 for all forms of random components.
4. The results of statistical tests used in this research show that the random truncated W-H method could generate more accurate results compared with the other methods applied in this study. This leads to conclude that this method can be recognized as the best and most suitable method to represent the data of daily station Akre.

## CONCLUSIONS

Based on the results predicted by the random component models and the statistical tests used in the current work, one can draw the following conclusions:



## REFERENCES

- Barnes, J. W. 1994. Statistical Analysis for Engineer and Scientists. McGraw-Hill, Inc., USA.
- Clark, R. T. 1984. Mathematical Models in Hydrology. FAO Consultant, Institute of Hydrology, Wallingford, UK.
- Harms, A. A. and T. H. Campbell. 1967. An extension to the Thomas-Fiering model for the sequential generation of streamflow. *Water Resour. Res.* 3: 653-661.
- Joshi, G. S. and K. Gupta. 2009. A simulation model for the operation of multipurpose multireservoir system for River Narmada. *India. J. Hydro Environ. Res.* 3: 96-108.
- Kim, B. S., H. S. Kim and B. H. Seoh. 2004. Streamflow simulation and skewness preservation based on the bootstrapped stochastic models. *Stoch. Environ. Res. Risk Asmt.* 18(6): 386-400.
- Kirby, W. 1972. Computer-oriented Wilson-Halferty transform that preserves the first three moments are that lower bound of the Pearson Type 3 distribution. *Water Resour. Res.* 8(5): 1254-1254.
- Kurunç, A., K. Yurekli and O. Çevik. 2005. Performance of two stochastic approaches for forecasting water quality and stream flow data from Yesilirmak River. *Turkey Environ. Mod. Softw.* 20: 1195-1200.
- McMahon, T. A. and A. J. Miller. 1971. Application of the Thomas-Fiering model to skewed hydrologic data. *Water Resour. Res.* 7:133e-134.
- Payne, K., W. Human and D. Kern. 1969. Stream flow simulation. *J. Hydrol. Dev. ASCE.* 95: 1163-1179.
- Srikanthan, R. 1978. Sequential generation of monthly stream flows. *J. Hydrol.* 38: 71-80.
- Stedinger, J. R. and M. R. Taylor. 1982. Synthetic stream flow generation, 1. Model verification and validation. *Water Resour. Res.*, 18(4): 909-918.
- Thomas, H. A. and M. B. Fiering. 1962. Mathematical synthesis of stream flow sequences for the analysis of river basins by simulation. In: *Design of Water Resources Systems*. Harvard University Press, Cambridge, Massachusetts. p459-493.
- Wang, W. 2006. Stochasticity, Non Linearity and Forecasting of Stream Flow Processes, PhD Thesis. Technical University of Delft, Netherlands.
- Wijayaratne, L. H. and P. C. Chan. 1987. Synthetic flow generation with stochastic models. In: Singh, V. P. editors. *Flood Hydrology*. Springer, Dordrecht. p175-185.
- Yousif, A. A., F. K. Aswad and S. A. Ibrahim. 2016. Performance of ARIMA model and modified Thomas-Fiering model for predicting the monthly rainfall data for Tallafar station. *J. Poly Tech.* 6(1): 17