Spatial Correlation for Flood Risk Assessment in Yom River Basin

Cholticha ARSSIRI[1]
Department of Water Resources Engineering
Faculty of Engineering, Chulalongkorn University
Bangkok, Thailand
Cholticha.arssiri@gmail.com

Piyatida RUANGRASSAMEE[2,*]
Department of Water Resources Engineering
Faculty of Engineering, Chulalongkorn University
Bangkok, Thailand
Piyatida.H@chula.ac.th

***Abstract*—In consideration of appropriate measures for flood mitigation, annual maximum flow data are one of the important data to be considered. Yom River Basin in Thailand has experienced floods frequently due to its topography of mountainous area in the upper basin and low-lying area in the lower basin as well as lack of largereservoirs in the basin. In flood risk assessment using flood frequency analysis, distribution function of annual maximum flow and estimated parameters of the distribution need to be determined. Additionally, a stochastic simulation of annual peak flow from multisite streamflow stations could provide valuable information for flood risk assessment. This study aims to develop a spatial correlation of peak flows in sub-basins ofYom River Basin as a step towards stochastic simulation. The methods are consisted of four parts: (1) standardization of peak flow data, (2) parameter estimation using method of L-moments, (3) goodness-of-fit test using L-moment ratio diagram, and (4) estimation of spatial correlation using semi-variogram models. The semi-variogram is fitted to three models: exponential, Gaussian, and spherical models using an ordinary least squares method. It was found that peak flow data from all four streamflow stations follow the Gumbel distribution. For the semi-variogram, the exponentialmodel is best fit. Furthermore, a stochastic simulation of annual maximum flow among stations can be developed using the obtained semi-variogram to estimate a covariance function. The stochastic simulation is a useful tool for quantifying probability of flooding in order to compare risk reduction of proposed flood mitigation measures.**

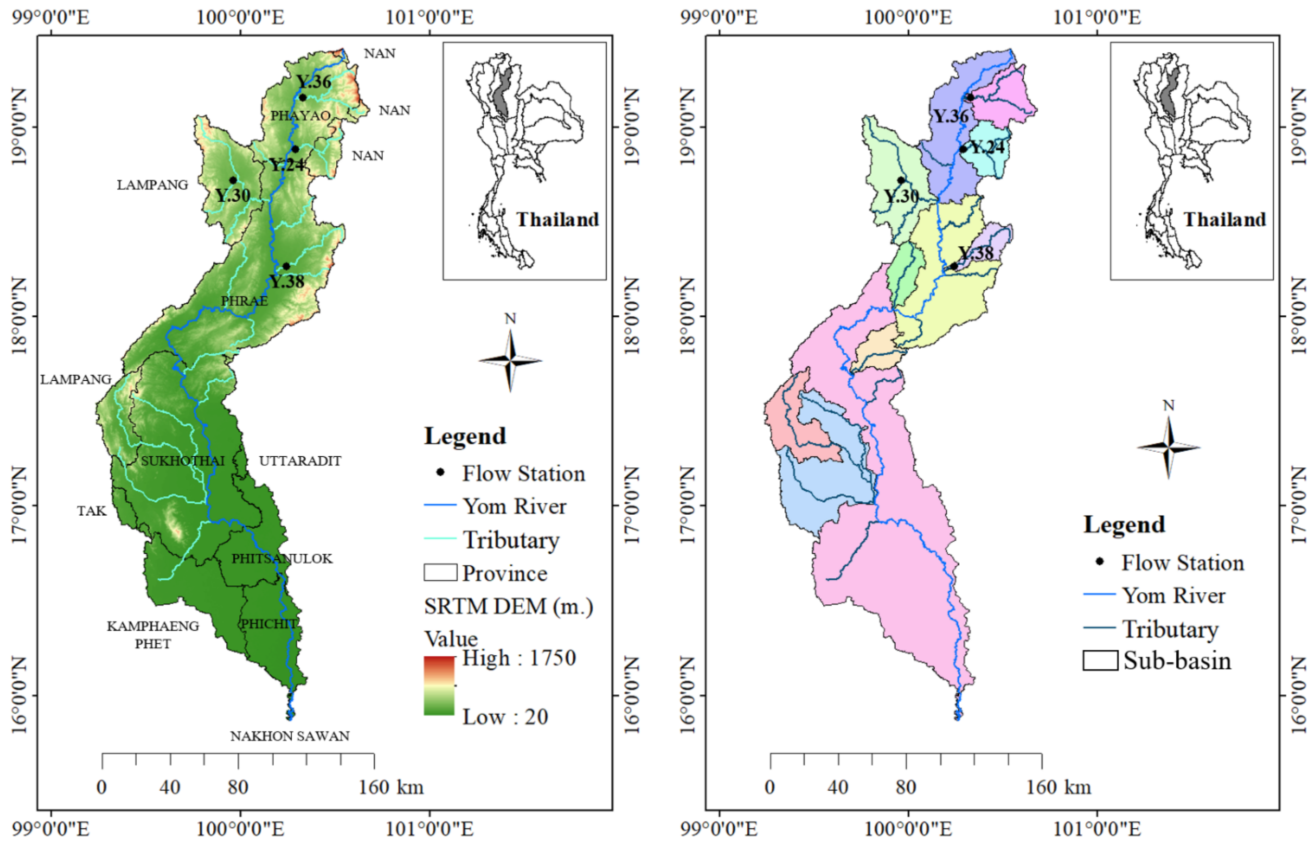
Keywords—flood, Yom River Basin, Gumbel distribution, L-momets method, semi-variogram

I. INTRODUCTION

Flow data are one of the most essential data for flood mitigation and flood risk management. Yom River Basin has experienced floods frequently due to its topography of mountainous area in the upper basin and low-lying area in the lower basin as shown in the topographic map in Fig 1. Yom River is the main river of the basin. The capacity of Yom River varies from 2,000 m^3/s in the upstream section to 300 m^3/s in the downstream section around Sukhothai province.For this reason, Sukhothai province is often flooded due to overflow.Furthermore, Yom River Basin lacks oflarge reservoirs for flood mitigation and water supply.Average annual flow in the basin is 4,720.7 million cubic meters but the existing reservoirs in the basin are medium and small sizes with the total capacity of only 406 million cubic meters and their main use is for irrigation (Royal Irrigation Department). The largest reservoir of the basin Mae Mok reservoir has a capacity of 96 million cubic meters and is located in Lampang province.There have been several development plans to increase water supply and mitigate floods in the basin including large, medium, and small reservoirs. However, there is still a lack of coordinated planning for the entire basin due to separate planning and management of each province in the basin.Considering spatial correlation of peak flows of multisite would provide useful information for flood risk analysis [1,2,3]. This study aims to develop a spatial correlation of peak flows insub-basins of Yom River Basin.

II. STUDY AREA AND DATA

Yom River Basin is one of the major upstreams of the Chao Phraya River. The average annual flows from the Yom River Basin contributes about 20 percent of the average annual flows of the Chao Phraya River (Royal Irrigation Department) at C.2 station in Nakhon Sawan. It is located in the northern part of Thailand. Heavy rainfall causesflash flood in the Upper Yom River Basin. In the Lower Yom River Basin where there is low lying area as well as narrowing river cross section in urban areas, floods frequently occurs and results in high economic and social loss [3,4]. The area of the basin is 23,616 $km^2$and it consists of eleven sub-basins. The basin boundary includes partial areas of eleven provinces, i.e. Nan, Phayao, Lampang, Phrae, Uttaradit, Sukhothai, Tak, Phitsanulok, KampaengPhet, Phichit, and Nakhon Sawan provinces.There are twenty-five flow stations in Yom River BasinbyRoyal Irrigation Department.Four flow stations were selected in this study area as shown in Fig 1. These four stations Y.36, Y.24, Y.30, and Y.38are located in the four sub-basins in the Upper Yom River basin. These stations are not affected by regulating infrastructure and have long historical records.In this study, we focus on the analysis of the variation in peak flows at different locations of the basin. Annual maximum flow data of four stations were collected from Royal Irrigation Department within the period of 1999-2017 as shown in Table I. Annual maximum flow usually occurs in September duringmonsoon season (May-October). The four flow stations are located on the separated tributaries. The highest annual maximum flow and the highest annual flow among the four stations arefrom station Y.36. The lowestannual maximum



(a) (b)

Fig. 1. Map of Yom River Basin and Locations of Selected Flow Stations. (a) Topography of the basin and (b) Sub-basins of the Yom RiverBasin

flow and the lowest annual flow among the four stations are from station Y.30 even though the drainage area of station Y.30 is largest. Station Y.30 is located on the right side of the main Yom River while the other three stations are located on the left side where tropical storms pass through from the South China Sea.

According to the land use data from Land Development Department in 2015-2016, the forest area and agriculture areaare major part of the basin. The upper part of the basin is mountain terrain and the lower part is flood plain. There is nolarge reservoir to reduce peak flow and store water for dry season use.

III. Methodology

This study is consisted of four parts:(A) standardization of peak flow data, (B) parameter estimation using method of L-moments, (C) goodness-of-fit test usingL-moment ratio diagram, and (D) estimation of spatial correlation using semi-variogram models.

A. Standardization of Peak Flow Data

The peak flow data from 1999 – 2017 at each station were first standardized to study the spatial correlation of the variation of the peak flows among different stations in different sub-basins. The goodness-of-test was carried out to find appropriate distribution for developing a stochastic simulation using a semi-variogram.

B. Parameter Estimation using Method of L-moments

The Extreme Value Type I or Gumbel distribution was developed in 1941. It played an important role in flood frequency analysis by Natural Environment Research Council (NERC) in 1975.In this study, the Gumbel distribution is

TABLE I. List of Annual Peak Flow Stations

No.	Station	River	Province	Sub-Basin	Drainage Area (km^2)	Average Annual Flow (MCM)	Maximum Annual Peak Flow (m^3/s)	Year
1.	Y.36	Khuan	Phayao	Mae Nam Khuan	870.74	397.03	525.92	2011
2.	Y.24	Pi	Phayao	Nam Pi	656.66	146.06	312.40	2008
3.	Y.30	Huai Pong	Lampang	Mae Nam Ngao	1,753.86	41.24	78.27	2011

4.	Y.38	Mae Kham Mi	Phrae	Nam Mae Kham Mi	456.18	174.46	500.00	2004

selected based on several studies of peak flow in Thailand that was found to fit with this distribution. There are several methods of parameter estimation. One of the simplest and general approaches for estimating parameter is the method of moments [5]. For small sample size, the method of maximum likelihood sometimes gives more accurate than the method of moments. The alternative approach that was developed to reduce bias in parameter estimation is the method of L-moments. This method is more robust to outliers in the sample data and occasionally give more accurate than the method of maximum likelihood in small sample size [6].

For this reason, two parameters of the Gumbel distribution were estimated using L-moments method. The probability density function for the Gumbel distribution is given as follows [7]

$$f(x) = \frac{1}{\alpha}\exp\left[-\frac{x-u}{\alpha} - \exp\left(-\frac{x-u}{\alpha}\right)\right], -\infty < x < \infty \qquad (1)$$

where u and α are the location and scale parameters respectively.

The probability weighted moments was developed by J. A. Greenwood et al. in 1979. Then L-moments were defined by J. R. Hosking in 1990 and they are linear combinations of probability weighted moments. They are similar to the ordinary moments that characterize measures of location, spread, and shape of the sample data and probability distributions [6,8].

L-moments for probability distributions (λ_r) are defined in terms of probability weighted moments (β_r, r = 0, 1, 2, 3,…).

$$\beta_r = \int_0^1 x(F)\{F(x)\}^r \, dF(x) \qquad (2)$$

where $F(x)$ and $x(F)$ are respectively cumulative distribution function and quantile function.

$$\lambda_1 = \beta_0 \qquad (3)$$

$$\lambda_2 = 2\beta_1 - \beta_0 \qquad (4)$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \qquad (5)$$

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \qquad (6)$$

$$\tau_r = \lambda_r / \lambda_2 \qquad (7)$$

where τ_r is L-moment ratios (r = 3, 4,…). L-skewness (τ_3) and L-kurtosis (τ_4) have range of values between -1 to 1.

L-moments for the random sample (X) are similar to L-moments for probability distributions. Let $X_1, X_2, X_3, ..., X_n$ be the sample data that have sample of size n and arrange value in ascending order [6]. The sample probability weighted moments (b_r, r = 1, 2, 3,…) are defined as [8]

$$b_0 = \frac{1}{n}\sum_{j=1}^{n} X_j \qquad (8)$$

$$b_r = \frac{1}{n}\sum_{j=r+1}^{n} \frac{(j-1)(j-2)...(j-r)}{(n-1)(n-2)...(n-r)} X_j \qquad (9)$$

The sample L-moments (l_r) and sample L-moment ratios (t_r, r = 3, 4,…) are given by

$$l_1 = b_0 \qquad (10)$$

$$l_2 = 2b_1 - b_0 \qquad (11)$$

$$l_3 = 6b_2 - 6b_1 + b_0 \qquad (12)$$

$$l_4 = 20b_3 - 30b_2 + 12b_1 - b_0 \qquad (13)$$

$$t_r = l_r / l_2 \qquad (14)$$

where t_3 and t_4 are respectively the sample L-skewness and sample L-kurtosis. L-moment ratios are dimensionless and have range of values between -1 to 1.

Parameter estimators of the Gumbel distribution using the method of L-moments can be expressed by [6]

$$\hat{\alpha} = \frac{l_2}{\ln 2} \qquad (15)$$

$$\hat{u} = l_1 - \gamma\hat{\alpha} \qquad (16)$$

where γ is Euler's constant with approximate value 0.5772.

C. Goodness-of-Fit Test using L-moment Ratio Diagram

The goodness-of-fit test is used to decide whether sample data resembles a particular kind of population [9]. L-skewness and L-kurtosis are characteristic of each distribution. They are less biased than the ordinary skewness and kurtosis [10]. The Gumbel distribution has constant values of L-skewness (τ_3 = 0.1699) and L-kurtosis (τ_4 = 0.1504).

The relationship between L-skewness and L-kurtosis is performed in L-moment ratio diagram which can be used to identify an appropriate distribution. The null hypothesis of this method of goodness-of-fit test is the sample data follows the Gumbel distribution if and only if L-skewness and L-kurtosis are inside the acceptance region of the Gumbel distribution. The acceptance region with 95% confident intervals of L-moment ratio diagram for the Gumbel distribution that depend on the sample size n was established [11].

D. Estimation of Spatial Correlation using Semi-Variogram Models

Let $Z(x_i)$ be a random variable located at $x_i, i=1,2,3,...,n$ and $\{Z(x_i), x_i \in D\}$ be random function in spatial domain (D). Under the second-order stationarity assumption is defined as

$$E[Z(x_i)] = \mu, \forall x_i \in D \tag{17}$$

$$\text{Var}[Z(x_i)] = \sigma^2, \forall x_i \in D \tag{18}$$

$$\text{Cov}(Z(x_i), Z(x_i)) = \text{Cov}\left(\left|x_i - x_j\right|\right) = C\left(\left|x_i - x_j\right|\right), \forall x_i, x_j \in D \tag{19}$$

where Cov(□) and C(□) are covariance function of random variables. Spatial correlation structure does not depend on special location but it depends only on distance vector between random variables. For isotropic, covariance function is a function only of magnitude between two random variables, $\left\|x_i - x_j\right\|$ [12]. The spatial correlation structure is often characterized as a semi-variogram, $\gamma\left(\left|x_i - x_j\right|\right)$. The semi-variogram describes about dissimilarity of the data that are separated by distance $\left(h = \left|x_i - x_j\right|\right)$ but covariance function describes about similarity.

$$\gamma(h) = \gamma\left(\left|x_i - x_j\right|\right) = \frac{1}{2}\text{Var}[Z(x_i) - Z(x_j)] = \frac{1}{2}E[Z(x_i) - Z(x_j)]^2, \forall x_i, x_j \in D \tag{20}$$

The conditions of the basic semi-variogram are used under the standardization of the data, second-order stationary, and isotropic. The basic semi-variogram models that commonly used are exponential, Gaussian, and spherical model are used in this study. The three models can be expressed by [13]

$$\text{Exponential model:} \quad \gamma(h) = 1 - \exp\left(-\frac{3h}{a}\right) \tag{21}$$

$$\text{Gaussian model:} \quad \gamma(h) = 1 - \exp\left(-\frac{3h^2}{a^2}\right) \tag{22}$$

$$\text{Spherical model:} \quad \gamma(h) = \begin{cases} 1.5\left(\frac{h}{a}\right) - 0.5\left(\frac{h}{a}\right)^3, & h \le a \\ 1, & h > a \end{cases} \tag{23}$$

where h and a are distance between two random variables and range respectively.

IV. Results and Discussion

The annual maximum flow of four flow stations in Yom River Basin was first standardized to study the spatial correlation of the variation of peak flow. The goodness-of-fit test usingL-moment ratio diagram was carried out in each station.All flow stations fit to the Gumbel distribution with 95% confidence interval as shown in Fig 2. However, L-moment ratios of two flow stationsY.30 and Y.38 are around the edge of the 95% acceptance region. Station Y.30 is located in the city of Lampang province and station Y.38 is located in Phrae province.

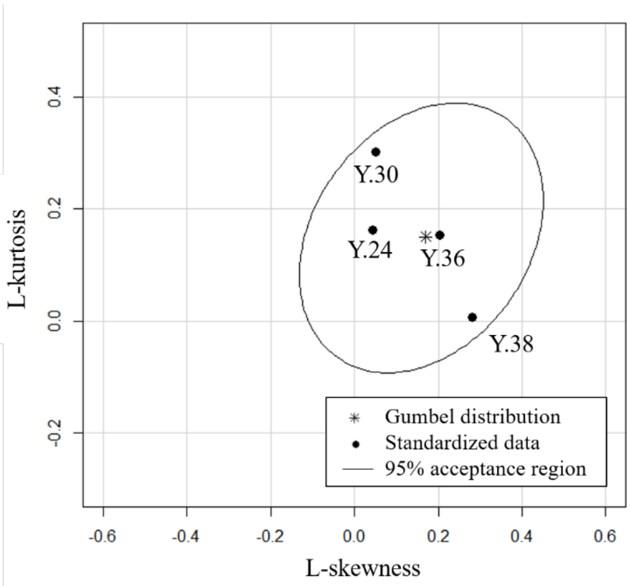


Fig. 2. L-moment ratio diagram with 95% confident intervals for the Gumbel distribution corresponds to the sample size ($n = 19$) at four flow stations in Yom River Basin

The two parameters of the Gumbel distribution of standardized annual maximum flow are estimated by the method of L-moments as shown in Table II.

TABLE II. Estimated Parameters of the Gumbel Distribution with the Method of L-Moments

Station	Parameter estimators	
	α	u
Y.36	0.81	-0.47
Y.24	0.83	-0.48
Y.30	0.80	-0.46
Y.38	0.80	-0.46

Under the second-order stationary and isotropic, the spatial correlation was constructed using a semi-variogram. The experimental semi-variogram was fitted to three theoretical models, i.e. exponential, Gaussian, spherical models. The ordinary least squares method was applied to determine the optimal model for spatial correlation as shown in Fig 3. The minimum sum of square errors between the experimental and theoretical semi-variogram models are criteria for the best fit. The sum of square errors from fitting to exponential, Gaussian, and spherical models are 0.070, 0.208, and 0.106 respectively. The result showed that the exponential model is the best fit. The parameter estimation of three theoretical models using the ordinary least squares method can be represented by the following equations:

Exponential model: $$\gamma(h)=1-\exp\left(-\frac{3h}{214.29}\right) \tag{21}$$

Gaussian model: $$\gamma(h)=1-\exp\left(-\frac{3h^2}{122.47^2}\right) \tag{22}$$

Spherical model:

$$\gamma(h)=\begin{cases}1.5\left(\dfrac{h}{155.48}\right)-0.5\left(\dfrac{h}{155.48}\right)^3, & h\le 155.48\\ 1 & ,h>155.48\end{cases} \tag{23}$$

The range of exponential model is 214.29 km indicating independency of peak flows between stations beyond this range. Station Y.36 and Y.24 are both located in Phayao province and 30 km apart. They have low semi-variance indicating strong spatial correlation. The sub-basins of station Y.36 and Y.24 contribute significantly to the flow in the main Yom River on average 397 and 146 million cubic meters annually. The maximum annual peak flow at station Y.36 is 525.92 m^3/s and at station Y.24 is 312.40 m^3/s. The semi-variance of stations Y.36 and Y.38 is 0.71 indicating weaker spatial correlation. Stations Y.36 and Y.38 are 100 km apart. Station Y.38 also contributes a large amount of annual flow of 174 million cubic meters on average to the main Yom River. The maximum annual peak flow of station Y.38 is 500 m^3/s. The obtained semi-variogram can be used to estimate a covariance function for developing a stochastic simulation to be used for risk assessment of flood mitigation measures.

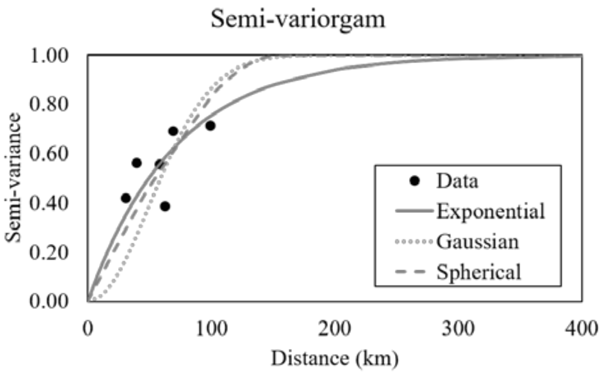


Fig. 3. The experimental semi-variogram model is fitted to the theorectical semi-variogram models (exponential, Gaussian, and spherical models)

V. CONCLUSIONS

Yom River Basin has experienced floods frequently due to its topography of mountainous area in the upper basin and low-lying area in the lower basin. Yom River Basin lacks oflarge reservoirs for flood mitigation and water supply. This study aims to develop a spatial correlation of peak flows in sub-basins of Yom River Basin to be used for building a stochastic simulation for risk assessment. Annual maximum flow at the selected four flow stations, namely Y.36, Y.24, Y.30, and Y.38 located in different sub-basins in the Upper Yom River Basin was fitted to the Gumbel distribution.The two-parameter Gumbel distribution is estimated by the method of L-moments. The goodness-of-test was carried out using L-moment ratio diagram. The annual maximum flow at the four stations fit with the Gumbel distribution with 95% confidence interval.

Spatial correlation structure for annual maximum flow in Yom River Basin is constructed using the semi-variogram under the second-order stationary and isotropic assumption. The sample semi-variogram of the annual maximum flows was best fitted to the exponential model with the range of 214.29 km. It was found that stations Y.36 and Y.24 has strong spatial correlation while stations Y.36 and Y.38 has weaker spatial correlation. The obtained semi-variogram can be used to estimate a covariance function for developing a stochastic simulation to be used for risk assessment of flood mitigation measures. It could be very challenging to construct a large reservoir in Yom River Basin nowadays but options of multiple medium sized reservoirs might be feasible. Understanding the spatial as well as temporal correlation of flows from each tributary would be very useful in quantifying risk reduction of each flood mitigation option or measure.

ACKNOWLEDGMENT

The authors gratefully acknowledge the data support provided by Royal Irrigation Department, Department of Water Resource, Department of Provincial Administration, and Land Development Department. The authors would also like to thank Prof. Ke-Sheng Cheng from National Taiwan University for his advice. The first author would like to acknowledge a partial funding from Department of Water Resources Engineering, Faculty of Engineering, Chulalongkorn University. The authors also would like to thank the reviewers for their constructive comments.

REFERENCES

[1] H. Hsin-I, S. Ming-Daw and C. Ke-Sheng, “Multisite spatiotemporal streamflow simulation with an application to irrigation water shortage risk assessment”. Terrestrial, Atmospheric & Oceanic Sciences, vol. 2, pp. 25, 2014.

[2] H. Hsin-I, S. Ming-Daw and C. Ke-Sheng, “Water shortage risk assessment using spatiotemporal flow simulation” Geoscience Letters, vol. 3, 2016.

[3] Committee on Agriculture and Cooperatives. “Mornitoring the development of the Yom River Basin”, 2012

[4] Department of Water Resources. “Exploration and implementation of river monitoring telemetry system in Yom and Nan Basins project”, 2012.

[5] S. Allan and S. Kannan. “Investigation and comparison of sampling properties of L-moments and conventional moments”. Journal of Hydrology, vol. 1-2, pp. 13-34, 1999.

[6] H. Jonathan. “L-moments: analysis and estimation of distributions using linear combinations of order statistics”. Journal of the royal statistical society. Series B (Methodological), pp. 105-124, 1990.

[7] V.T Chow, M. David R and M. Larry. “Applied hydrology”. New York : McGraw-Hill, 1988.

[8] G. Arthur, L. Maciunas, N. Matalas, J.R Wallis. “Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form”. Water resources research, vol. 15, pp. 1049-1054, 1979.

[9] D. David and S. Lori. "Applied statistics in business and economics". United States of America : McGraw-Hill Education, 2016.

[10] H. Jonathan and J. R. Wallis. "Regional frequency analysis: an approach based on L-moments". Cambridge University Press, 2005.

[11] L. Jun-Jih, W. Yii-Chen and C. Ke-Sheng. "Establishing acceptance regions for L-moments based goodness-of-fit tests by stochastic simulation". Journal of hydrology, vol. 1-4, pp. 49-62, 2008.

[12] C. Noel A. C. "Statistics for spatial data, revised edition". New York : Wiley, pp. 52-62, 1993.

[13] I. Edward H. and S. Mohan. "An introduction to applied geostatistics". New York : Oxford University Press, 1989.