Development of an improved flood frequency curve applying Bulletin 17B guidelines

M.S. Alam[1]

[1]*Department of Water, PO Box K822, Perth, WA 6842, Australia*
Email: Muhammad.alam@water.wa.gov.au

Abstract: Flood frequency curves for gauging stations are developed by fitting the logarithms of annual peak flows to a Pearson Type 3 distribution. Such a curve is sensitive to the mean, standard deviation and skew coefficient computed from the logarithms of the annual peak flows. If a gauging station is identified as a short record site then adjustments of the mean, standard deviation and skewness are required as suggested in Bulletin 17B, developed by the Hydrology Subcommittee, Office of Water Data Coordination, U.S. Geological Survey. The mean and standard deviation can be improved by undertaking a regression analysis for the short record site with a nearby long record site. The estimate of the station's skew coefficient can be improved if it is weighted by the regional skew coefficient. In this study, annual peak flow data of 80 gauging stations in Western Australia are analysed to estimate the regional skew coefficient for the area of interest. The regional skew coefficient estimation technique applied in this study is rudimentary which does not follow the complete instruction of the Bulletin 17B. The improvement of the mean and standard deviation is done by following the complete instruction provided in Bulletin 17B. A software tool is developed using MATLAB to carry out the current study. This tool is GUI-enabled which takes flood data from the users and generates flood frequency curves based on the two station method mentioned in the Bulletin 17B. A flood frequency curve is developed for the Kentish Farm gauging station (614005) using this software tool. The Mundlimup gauging station (614073) is used as a long record site for the adjustments of the mean and standard deviation to improve the flood frequency estimates.

Keywords: *Generalized skew coefficient, flood frequency curve, LP3 distribution, Bulletin 17B, Western Australia*

1. INTRODUCTION

Flood modelling is necessary when a catchment is targeted for developments. The post development scenarios are modelled using hydraulic modelling software tools to identify flood hazards, develop a flood plain map and, design water conveyance and transportation structures in or near streams. One of the main inputs to the hydraulic model is design flood hydrographs. Design flood hydrographs at gauging stations are generated based on the estimated flood frequency curves. Constructing a flood frequency curve requires information of the magnitude and frequency of floods at gauged sites. The magnitude of certain flood frequencies can be estimated using the annual peak flows of a gauged site. In this study, flood magnitudes of certain frequencies are developed following the guidelines mentioned in Bulletin 17B of the Hydrology Subcommittee of the Interagency Advisory Committee on Water Data (1982). As suggested in Bulletin 17B, the log Pearson Type 3 (LP3) probability distribution is used to estimate flood frequencies at the gauged site. It should be mentioned that so far the LP3 distribution is the most common statistical distribution used in Australian practice. The three key parameters required to fit this type of distribution to the observed annual peaks are the mean, standard deviation and skew coefficient. Accuracy of fitting a LP3 distribution to the recorded peak flows depends on reliable estimates of these parameters. Bulletin 17B recommends that the at-site skew coefficient calculated from the station's data has to be weighted with the regional skew coefficient. It is because the skew coefficient is very sensitive to the extreme events. If the site has a short record then it is difficult to have an accurate estimation of the skew coefficient from the small samples. The regional skew coefficient is determined from the pooled data at nearby long record sites located within around a 160 km radius of the gauged site for which the flood frequency curve is to be developed. All long record sites should have at least 25 years of record. The actual procedure of regional skew analysis includes three methods: a) development of skew isoline map; b) skew prediction equation; and c) the mean of the station skew values. Normally all the three methods are applied in obtaining the generalized skew coefficient. The MSE (mean square error) from the isoline map is compared with the MSE for the prediction equation. The smaller MSE is then compared with the variance of the data. If the MSE is significantly smaller than the variance, then that MSE is used for the equation of calculating MSE of the generalized skew coefficient of the study area. If the MSE is not significantly smaller than the variance, neither the isoline map nor the prediction equation method provides more accurate estimation of skew coefficient than the mean value method. The regional skew analysis of this study was carried out as a partial fulfilment of the flood study project for the study area. The allocated time to develop a flood frequency curve was very short as the main focus of the project is to develop a hydraulic flood model for the study area to estimate flows for design events. This is why no time was allocated in the project to apply isoline map to skew prediction equation methods in this study. For simplicity and time constraints, it is assumed that MSE from isoline map and prediction equation would not be significantly smaller than the variance of the data of the study area. Based on the assumption, only the mean value method is exercised to estimate the regional skew coefficient. The estimated regional skew coefficient is then used to calculate the weighted skew coefficient of the gauged site for which the flood frequency curve is to be developed. Once the weighted station skew coefficient is determined with the help of the regional skew coefficient, the mean and standard deviation of the station's peak flow data are required for the log Pearson 3 type distribution. If the gauging station has a short record, the improvement of these parameters is recommended by Bulletin 17B. This is done using a nearby long record site's discharge data. The logarithmic mean and standard deviation of the short record site are adjusted on the basis of a regression analysis as mentioned in the Bulletin 17B with a nearby long record site.

The main objective of this study is twofold. First, the flood frequency curve for the Kentish Farm gauging station is improved. The improvement involves adjusting the mean, standard deviation and skew coefficient implementing the guidelines written in Bulletin 17B. Second, a GUI-enabled software tool to improve and develop a flood frequency curve is presented which is developed as a partial fulfilment of this study. This tool is mainly developed based on the two station method described in Bulletin 17B. This tool can also be used by any hydrologist for the improvement of the mean and standard deviation of a short record site and the development of a flood frequency curve.

2. STATION SKEW COEFFICIENT

The discharge-frequency relation for a gauging station is established by fitting the logarithms of annual peaks to the log Pearson Type 3 distribution. The guidelines mentioned in Bulletin 17B are followed. It is assumed that the sample is the representation of the population. The recorded annual peaks are samples whereas all recorded and unrecorded annual peaks are the population. Fitting the distribution requires the skew coefficient estimation. The equation for computing station skew is mentioned below:

$$G_s = \frac{N \sum (X - \bar{X})^3}{(N-1)(N-2)S^3} \tag{1}$$

in which:

X=logarithms of annual peak flow; N=number of years of peak discharge record for the station; $\overline{X}$= average of annual peak flows; S= standard deviation of logarithms; and, G_s=station's skew coefficient.

There could be some low and high outliers in the peak flow dataset. Outliers need to be detected and removed from the dataset based on the guidelines mentioned in Bulletin 17B. The detection of high outliers is omitted in this study as useful historic information is not available to adjust for high outliers. If useful historic information is not available then high outliers (if there are any) should be retained as part of the systematic record (page 17, Bulletin 17B). All low outliers are detected based on the equations mentioned on page 18 in Bulletin 17B. They are then removed from the peak flow record. Once the dataset is free from outliers, the skew coefficient is estimated for the gauging stations. The station skew coefficient calculated using equation (1) is not unbiased. A bias-correction factor which is a function of record length in years, proposed by Tasker and Stedinger (1986) is applied to the calculated skew coefficient. The correction factor is as follows:

$$C_b = \left(1 + \frac{6}{N}\right),\tag{2}$$

where:

C_b= bias correction factor; and, N= station's record length in terms of number of years.

The skew coefficient for a gauging station is multiplied by the above mentioned correction factor. It is yet to be weighted with the regional skew coefficient as recommended by Bulletin 17B. The weighted skew coefficient of the station is obtained by weighting both the station and generalized skew coefficient in inverse proportion to their individual MSE's (mean square error). The equation of the weighted skew coefficient is as follows:

$$G_w = \frac{MSE_{\overline{G}} \times G_s + MSE_{G_S} \times \overline{G}}{MSE_{\overline{G}} + MSE_{G_S}}\tag{3}$$

in which:

G_w=weighted station's skew coefficient; G_s= station's skew coefficient; $\overline{G}$= station's unbiased generalized skew coefficient (see section 3); MSE_{G_S}= mean square error of station skew coefficient; and, $MSE_{\overline{G}}$= mean square error of the unbiased generalized skew coefficient.

3. GENERALIZED SKEW COEFFICIENT

There are three methods recommended by Bulletin 17B available for estimating generalized skew coefficient. They are: a) regression; b) mapping; and c) averaging methods. In this study, for simplicity and project time constraints, it is assumed that the MSE from mapping or regression would not be smaller than the variance of the station data. This implies that the averaging method would be the most accurate method of calculating generalized skew coefficient of the study region. The study region is defined according to the recommendation made by Bulletin 17B. All gauging stations within 160 km radius of the gauging station for which the flood frequency curve is to be estimated form the study region. According to the mean method, the regional skew coefficient ($\overline{G}$) is obtained by averaging station skew of all gauging stations of the region. As mentioned earlier in the previous section, a bias-correction factor is applied to all station skew values (see equation (2)) before considering them for averaging. The estimation of weighted skew coefficient requires the MSE of the regional skew. The MSE of the regional skew is estimated by averaging the sum of squared differences between the gauging station and the regional skew values. The equation is mentioned below:

$$MSE_{\overline{G}} = \sum_{i=1}^{n} \frac{(G_i - \overline{G})^2}{n}\tag{4}$$

where:

G_i stands for skew coefficient for the n gauging stations in the region,

$\overline{G}$ stands for the unbiased generalized skew coefficient.

4. TWO-STATION METHOD

The objective of implementing the two station method described in Bulletin 17B is to improve the logarithmic mean and standard deviation of a short term gauging station using the peak flow data of a nearby long term gauging station. This is done by regression with peak flows of the long term gauging station over a concurrent period. The regression equation, long term mean and standard deviation is used to improve the short term statistics. If the correlation between the two stations' flow datasets meets certain criteria (eqn. 7-4 and eqn. 7-9, Bulletin 17B), the improvement based on the longer record may reduce the possibility of bias due to the unrepresentative peak flow data of the short term gauging station. Readers are referred to Bulletin 17B for a

detailed description of the two station method used in this study. The key component of the method is the comparison of the variances of the short-term statistics (logarithmic mean and standard deviation) with the variances of the adjusted statistics. The smaller variances are the ones which are considered to be used in the flood frequency curve. At first, the slope of the regression line (b) and the correlation coefficient (r) are estimated from the logarithmic peak flow data for the concurrent period of record for short-term and long-term gauging stations. Then, r is compared with r_{min} (eqn. 7-4, Bulletin 17B). If $r \leq r_{min}$, the variance of the adjusted mean will be greater than the variance of the short-term mean. It implies the unadjusted short-term mean will be used in the flood frequency curve. If $> r_{min}$, the adjustment may be worthwhile. The variance of the adjusted and unadjusted mean are estimated employing equations mentioned in Bulletin 17B. The mean with smaller variance will then be considered as the estimated mean in the flood frequency curve. The next step which is followed in this study is to compare r with r_{min} (eqn. 7-9, Bulletin 17B). If $r \leq r_{min}$ the variance of the standard deviation is greater than the variance of the unadjusted short term standard deviation and the smaller one is used in the flood frequency curve calculations. On the other hand, if $r > r_{min}$, it indicates that an adjustment to the short-term standard deviation estimation may be worthwhile. The variance of the adjusted variance and the variance of the unadjusted variance for the short-term gauging station for the entire record length are estimated. The variance with the smaller variance will be used for the estimation of the standard deviation in the flood frequency curve calculations.

5. DEVELOPMENT OF A SOFTWARE TOOL TO ESTIMATE FLOOD FREQUENCY CURVES

A software tool is developed in this study to estimate flood frequency curves for gauging stations. The main idea of developing such a software tool is to provide a user friendly calculation platform with a graphical user interface (GUI) for other hydrologists who are required to estimate flood frequency curves based on the two station method. The software tool is developed using the programming language MATLAB. The tool is capable of producing improved flood frequency curves employing the guidelines for estimating the weighted skew coefficient, improved mean and improved standard deviation mentioned in Bulletin 17B. The main inputs of this software tool are the logarithmic flows of the long record and short record gauging stations, the unbiased generalized skew coefficient and MSE of the generali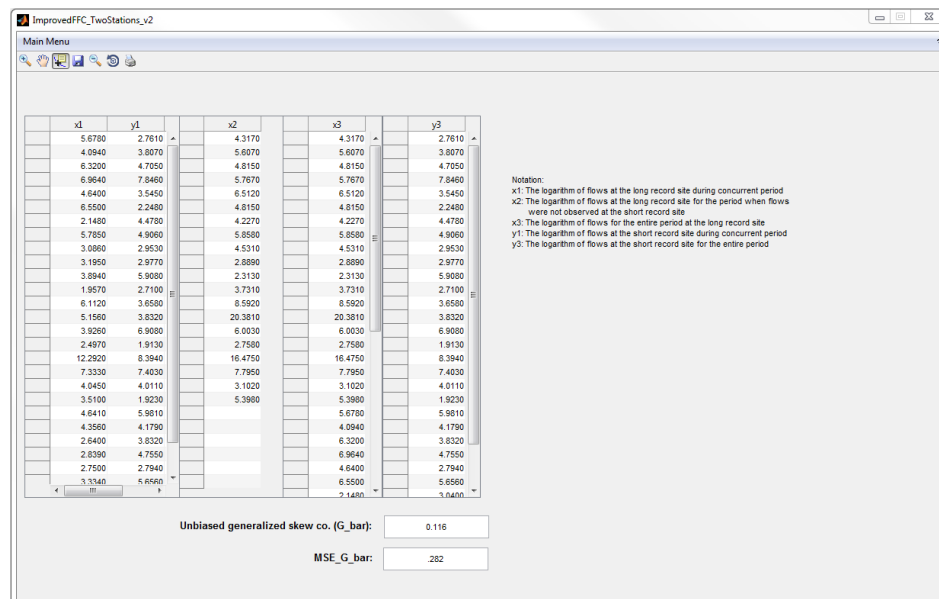zed skew coefficient. The software tool has two screens. The first screen (which is the default one) is named as *Input Page* (See Figure 1) where all inputs are required to be supplied. The second screen is called *Calculation Page* where all calculations in regard to the estimation of the flood frequency curve can be carried out using some button options. Users can go from one screen to the other using the Main Manu navigation option on the top. The details on the functionalities of the software tool

Figure 1. The Input Page screen of the developed Flood Frequency Curve program in MATLAB.

will be provided in the results section as it is easier to introduce the tool functionalities with examples.

6. RESULTS

The objective of the current study is to develop a flood frequency curve for the Kentish Farm gauging station which is located around 55 km south-east of Perth in the Darling Scarp. The blue dot with an unbiased skew coefficient of 0.17 located next to the blue polygon (flood study area) in Figure 2 represents this gauging station. It has 31 years of flow record (1971−2001). This record is not long enough and probably does not represent the full record of floods of the site. Bulletin 17B guidelines are applied to the logarithmic peak flows of the station to improve the skew, mean and standard deviation.

At first an attempt is made to improve the skew coefficient of the station (Kentish Farm). An estimation of the regional skew coefficient is required for that. According to the guidelines mentioned in Bulletin 17B, the regional skew coefficient is to be estimated based on all gauging stations within 160 km radius of the Kentish Farm gauging station. The total of 321 gauging stations that fall within the 160 km radius of this site are selected initially for the current study. Bulletin 17B recommends to remove the gauging stations from the analysis which are affected by anthropogenic effects such as regulation or urbanization, or which have less than 25 years of flow record. A total of 80 out of the 321 gauging stations are found eligible for the regional skew analysis as the rest of them have less than 25 years of flow record or anthropogenic effects are observed. The skew coefficients of all eligible gauging stations are calculated using the equation (1) and a correction factor is applied on each of them using the equation (2). Figure 2 shows the eligible gauging stations and their unbiased skew coefficients. The average of the unbiased skew coefficients is 0.116. This is considered as the unbiased generalized skew coefficient for the present study. The weighted skew coefficient of the Kentish Farm gauging station is
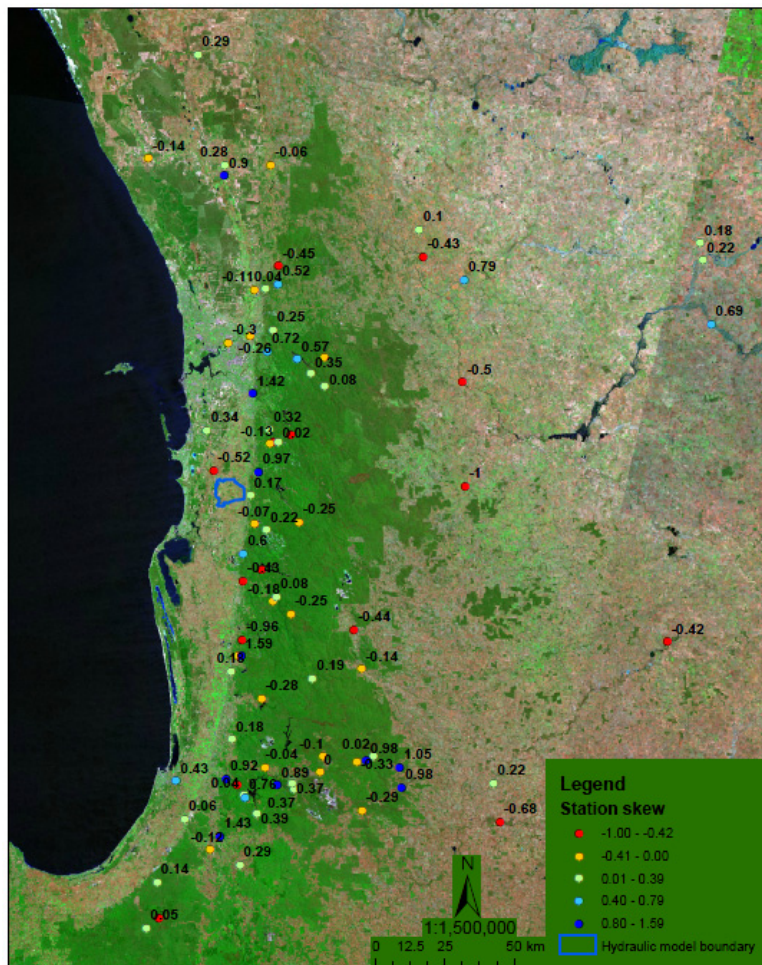

Figure 2. Station skew coefficients (unbiased).

estimated using equation (3). The skew coefficient of the station is 0.1405. The MSE of the unbiased generalized skew coefficient estimated using equation (4) is 0.282. The MSE of the station skew calculated using equation 6 mentioned in Bulletin 17B is 0.178. The estimated weighted skew coefficient of the Kentish Farm gauging station based on the parameters mentioned in the equation (3) is 0.131. This value will be used in estimating the flood frequency curve of the station. A 31 year flow record exists for the Kentish Farm gauging station. This could be treated as a short term gauging station as the peak flow data of this station may not represent the full range of floods at the site. The mean and standard deviation of the station could be improved by using the peak flows of a long term gauging station nearby. The Mundlimup gauging station (614073) which is around 9 km away from the Kentish Farm gauging station possesses 48 years flow record. This could be considered as a long term gauging station assuming that it captured the full range of floods of the catchment. In this study, the two station comparison method mentioned in Bulletin 17B is applied on the Kentish Farm and the Mundlimup gauging station considering them as a short term and a long term gauging station, respectively. There exists a good correlation (r=0.593) of peak flows for the Kenstish Farm gauging station with those for the Mundlimup gauging station.

Once the weighted skew coefficient of the Kentish Farm gauging station is estimated and the correlation between this and the Mundlimup gauging station is found good, the initiatives are taken to improve the mean and standard deviation of the Kentish Farm gauging station. The software tool developed in this study to estimate the improved flood frequency curve is applied to both gauging stations to get a better estimate of the flood frequency curve for the Kentish Farm gauging station. Figure 1 shows the Input Page screen of the software tool. There are three input tables in the screen. The first table takes the logarithmic peak flows of the short and long record gauging stations during the concurrent period. The data is inserted into the table by right clicking on the table and selecting "paste x1 or paste y1" option popped up. The data has to be copied from the source using Ctrl+C option before. Then the logarithms of peak flows of the long record site which are not concurrent are pasted into the column (x2) of the second table. The third table is filled using the entire peak flow records of the long and short record sites. The entire long record and short record peak flows are copied into the
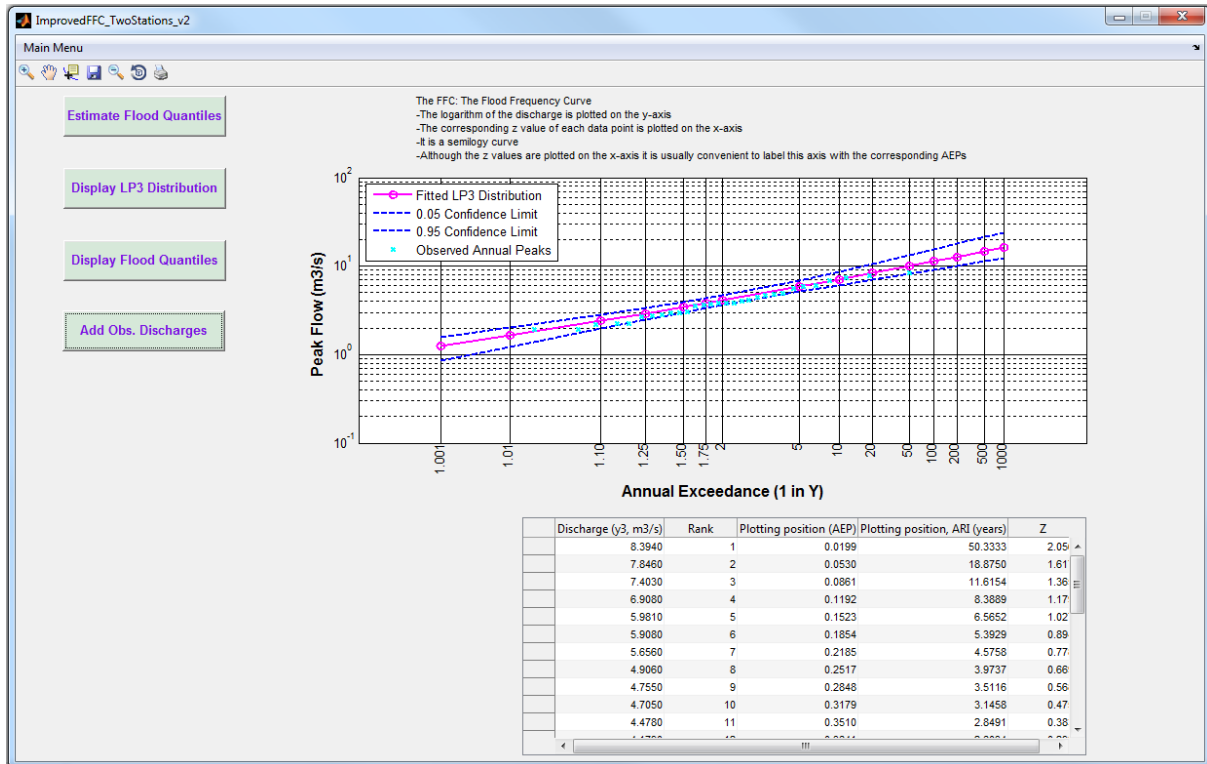
Figure 3. The Calculation Page screen for the estimation of flood frequency curve.

x3 and y3 column, respectively. The unbiased generalized skew coefficient and MSE of the generalized skew coefficient estimated following the procedure mentioned in the previous paragraph are given as inputs to the screen. The estimated unbiased generalized skew coefficient of this study area is 0.116 and the estimated MSE of the generalized skew coefficient is 0.282. Figure 3 shows the Calculation Page screen of the two station software tool. This screen is designed to analyse logarithmic peak flow data, display statistical results in new screens and plot the flood frequency curve for the study area. The Estimate Flood Quantiles button to the left is pressed to estimate the adjusted mean and standard deviation of the short record gauging stations using the two station method. Then the flood quantiles for 1.001, 1.01, 1.10, 1.25, 1.50, 1.75, 2, 5, 10, 20, 50, 100, 200, 500 and 100 year ARI events are estimated. A summary of calculations is provided into a new screen once the Estimate Flood Quantiles button is hit. Figure 4 shows the summary for the calculations done for Kentish Farm and Mundlimup gauging stations. It shows that the adjusted logarithmic mean of the short term gauging station

Figure 4. The calculation summary of the two station method for the Kentish Farm and Mundlimup gauging stations.

(Kentish Farm) is improved from 0.584 (Y3_bar) to 0.617 (Y_bar_final). It also shows that the standard deviation of the short term gauging station is not required to be adjusted as the variance of the variance of the short term station is smaller than the variance of the adjusted variance ($Var(Sy^2)<Var(Sy3^2)$). This screen also provides the estimated weighted skew coefficient of the Kentish Farm gauging station (0.131) for which the flood frequency curve is to be calculated. The *Display LP3 Distribution* button is to be executed to display a detailed summary on flood quantile information in terms of ARI (Average Recurrence Interval), AEP (Annual Exceedance Probability), Z (standard normal deviate) and peak flows. Pressing this button also plots flood quantiles with 5% and 95% confidence intervals. It means the flood frequency curve is displayed as this button is pressed. Figure 5 shows the flood quantile summary on the Kentish Farm Gauging station. It shows that the peak flows for 5, 10, 20, 50, 100, 200 and 500 year ARI design events for Kentish Farm gauging station are 5.9, 7.1, 8.3, 10.0, 11.4, 12.8 and 14.7 m^3/s respectively. *Display Flood Quantiles* button is needed to be pressed to display the annual flood series with their ranks, calculated plotting positions in AEPs and ARIs. The observed peak flows of the gauging station

are not plotted in the graph yet. This can be done once the Z column in the table of the screen is filled and then *Add Obs. Discharges* button to the left side of the page is pressed. Z values against all AEPs are required to be calculated using the standard normal distribution chart available in literature. The Z values are needed to be inserted into the Z column by right clicking on the table and using the "paste z-scores" option appeared. Pressing the *Add Obs. Discharges* button displays all observed peak flows in the flood frequency curve. Figure 6 shows the flood frequency curve for the Kentish Farm gauging station developed by the software. The 5% and 95% confidence intervals with observed annual peak flow data are also displayed in the figure.

7. CONCLUSIONS

The flood frequency curve for the Kentish Farm gauging station is estimated. The skew coefficient, mean and standard deviation of the station are adjusted following the guidelines mentioned in Bulletin 17B. The regional skew coefficient is estimated using peak flows of 80 out of 321 gauging stations within 160 km radius of the Kentish Farm gauging station. The rest of the stations are rejected as they do not comply with the criteria set by Bulletin 17B. The weighted skew coefficient of the Kentish Farm gauging station is calculated using the regional and the station's skew coefficient and their MSEs. The mean and standard deviation of the station is improved using the two station method described in Bulletin 17B. A GUI-enabled MATLAB software tool is developed to estimate the flood frequency curve employing the two station method. The functionalities of the software tool are described with examples and figures. The Mundlimup gauging station located 9 km north-east from the Kentish Farm gauging station is considered as the long term gauging station to improve the mean and standard deviation. The flood frequency curve is finally developed using the software tool. It should be mentioned that to keep the calculation procedure simple an assumption is made in estimating the regional skew coefficient.
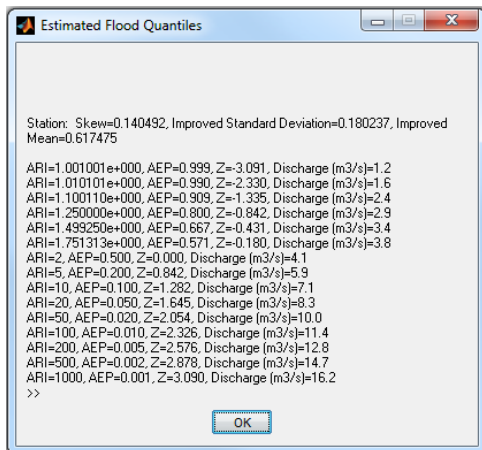

Figure 5. The estimated flood quantiles for the Kentish Farm gauging station.

In this study, it is assumed that MSE from the isoline map and the prediction equation would not be significantly smaller than the variance of the data of the study area. Therefore (avoiding the isoline map and prediction equation methods) only the mean value method is exercised to estimate the regional skew coefficient.
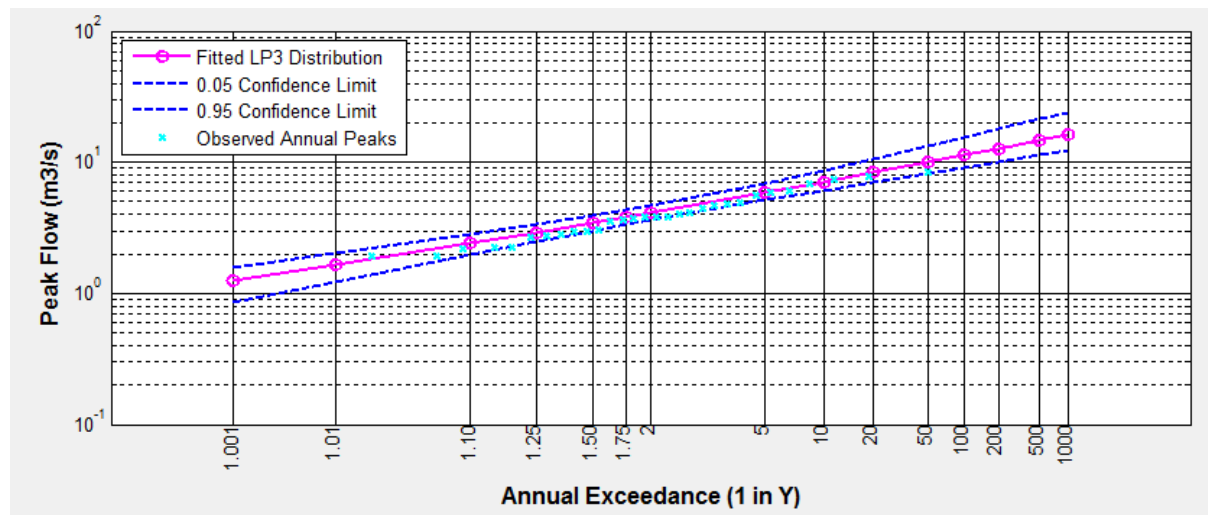

Figure 6. The plotted flood frequency curve using the developed software tool for the Kentish Farm gauging station.

REFERENCES

Interagency Advisory Committee on Water Data (1982), Guidelines for determining flood flow frequency-Bulletin 17B of the Hydrology Subcommittee, *U.S. Geological Survey, Office of Water-Data Coordination*.

Tasker, G.D., and J.R. Stedinger (1986). Regional skew with weighted LS regression, *Journal of Water Resources*, 112(2), 225–237.