Parameter estimation of Gumbel distribution for flood peak data

2102531 Term Project Report

Jitin Khemwong
Tiwat Boonyawiwat
Tanakorn Kriengkomol
Jitkomut Songsiri
Piyatida Hoisungwan

Department of Electrical Engineering
Chulalongkorn University

December 9, 2015

1 Introduction

Thailand usually suffers from flood condition during rainy season. The most recent severe flood occurred in 2011 due to unorganized water management and unpredictable nature of rain. It is estimated by the World Bank that it caused more than 45.7 billion USD [1]. Since flood significantly affects economy and life of people in Thailand, the prediction of Chao Praya Rivers water level which consists of Ping, Wang, Yom and Nan rivers would warn the control station to handle the water situation carefully.

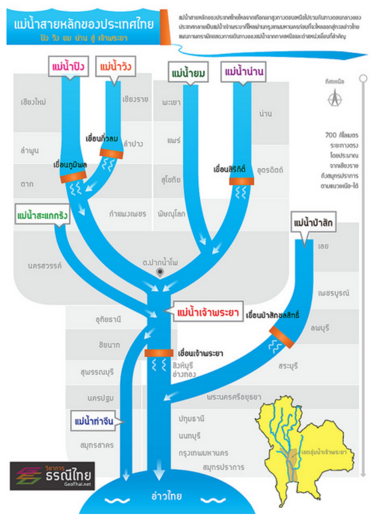


Figure 1: Thailand river map. Courtesy of `http://geothai.net`

Since the water level of each river is related to its water flow rate, then it is possible to analyse the water level by considering the flow rate instead. A plot of flow rate (in m^3/s) in each year is called a hydrograph. The maximum point of flow rate in hydrograph is called the flood peak.

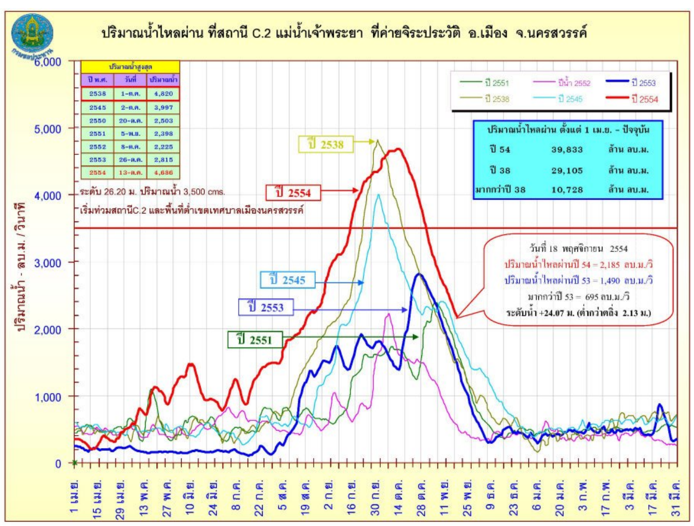


Figure 2: Hydrograph of Chao Phraya river. Courtesy of Thai royal irrigation department `http://www.rid.go.th`

Figure 2 shows the hydrograph of Chao Phraya river which is measured at C.2 (control station 2) in Nakhon Sawan province in 1995, 2002, 2008, 2010, and 2011. The y-axis shows daily flow rate of Chao Phraya river in m^3/s unit. The x-axis is the reference time which starts from the April that is the start of water calender year. Then for Thailand, the flood peaks of each hydrograph that is the maximum flow over time in each year are always in the middle zone of the graph.

Before analysing flood peak data, we need to choose a distribution that fits well with the data. From [2], a class of 2-parameter models that means the class of models which each model has two parameters are considered before others because they are easier to fit the data with model. There are several models that were generally used to model the problems in hydrology such as Normal, Log normal (2-parameter), Exponential, log Pearson type III, and generalized logistic.

Gumbel distribution which is a type of extreme value distribution, has been chosen to be a model in this problem since it is easy to compute the parameters needed in the model and also well fitted with the sample data which are flood peak value of a river. The CDF of Gumbel distribution is

$$F(x) = e^{(-\exp(-\frac{(x-\mu)}{\alpha}))} \tag{1}$$

for x greater than zero and reduced variable of Gumbel distribution is defined as

$$y = \frac{(x-\mu)}{\alpha} = -\log(-\log(F(x)))$$

where μ and α, which are two parameters of Gumbel distribution, are positive value. From the equation (1), the value of $F(x)$ lies between 0 and 1, then possible value of $\log(F(x))$ is less than 0. Therefore, value of y could be all possible value in the real line. If x is less than μ, then y is positive. In the same way, y is negative for the opposite case.

The plot of flood peak and reduced variable for each river is fitted well with Gumbel distribution.

Since correlation can determine the dependency between two rivers and the value is easy to compute, the correlation is examined to show basic relationship between two rivers. The correlation is calculated from the flood peaks of years 1972-1974 and 1976-2010 (38 samples). We denote $p_N, p_1, p_2, p_3, andp_4$ are annual flood peaks of Chao Phraya, Ping, Wang, Yom and Nan river respectively.

Figure 3 shows the correlation coefficient between two flood peaks of two rivers that consists of ten combinations. It is found that the correlation between Ping River and Wang River is the highest value of 0.725, while the correlation of Wang-Nan river and the correlation of Ping-Nan river are two the lowest values of 0.368 and 0.362, respectively. When the influence of the flood peak of each river to the flood peak of Chao Phraya river are being considered, the flood peak of Wang river and Nan river have the strongest relation with flood peak of Chao Phraya river.

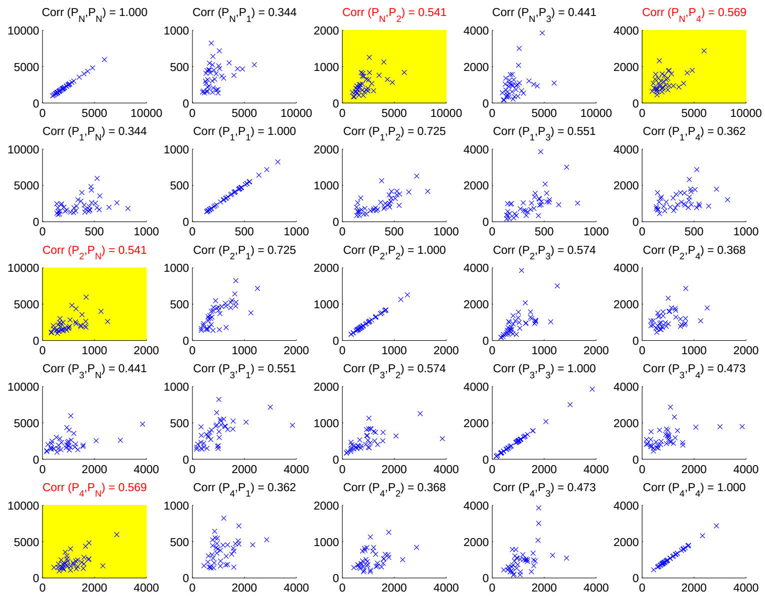


Figure 3: Scatter plot of flood peak data.

2 Problem Descriptions

This paper focuses on fitting Gumbel distribution to flood peak data of Chao Phraya river which depends on four rivers Ping, Wang, Yom, and Nan. First, the marginal probabilities of the river is obtained by fitting the old flood peak data with the Gumbel distribution where the parameters are estimated by using Maximum likelihood and Method of moments technique. Second, the relationship between the Chao Phraya river and others is investigated by considering the return period of bivariate Gumbel distribution. Third, multivariate Gumble distribution is being considered since it can describe the joint probability density function of flood peaks of all five rivers. Although the multivariate Gumbel distribution is expected to provide better information about the rivers, the formulation is too complicated (see Appendix 5.4). So this paper will focus solely on univariate and bivariate Gumbel distribution.

3 Parameter estimation of marginal Gumbel pdf

3.1 Method of moment estimation

In this subsection, method of moment estimator (MM) is used to estimate parameters μ and α of Gumbel probability density function. (see the detail in the appendix 5.1) First, deriving first and second (variance) moment of Gumbel distribution. Let Y be a Gumbel random variable with cdf and pdf as follows:

$$F(y) = \exp(-\exp(-y)),$$

$$f(y) = \exp(-y - \exp(-y)).$$

The moment generating function of Y is

$$m(t) = \mathrm{E}[e^{tY}] = \int_{-\infty}^{\infty} e^{ty} \exp(-e^{-y}) e^{-y} dy.$$

The substitution $x = e^{-y}, dx = -e^{-y}$give

$$m(t) = \int_0^{\infty} x^{-t} e^{-x} dx = \Gamma(1-t), \quad t \in (-\infty, 1)$$

where $\Gamma(1-t)$ is Gamma function defined as $\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$. And recall that $\gamma = -\Gamma'(1)$ is defined as

$$\gamma = \lim_{n\to\infty} (-\log(n) + \sum_{k=1}^{n} \frac{1}{k}) \approx 0.57722$$

(see [5]) and γ iscalled the Euler's constant. It is obtained from moment generating function that

$$\mathrm{E}[Y] = -\Gamma'(1) = \gamma$$

and

$$\mathrm{E}[Y^2] = \Gamma''(1) = \gamma^2 + \frac{\pi^2}{6}.$$

(see [6])
Therefore,

$$\mathrm{Var}(Y) = \mathrm{E}[Y^2] - \mathrm{E}[Y]^2 = \frac{\pi^2}{6}$$

If we define $X = \alpha Y + \mu$, then

$$\mathrm{E}[X] = \alpha\gamma \ + \mu$$

$$\mathrm{Var}(X) = \frac{\pi^2 \alpha^2}{6}$$

Then the parameters which are estimated with method of moments are

$$\hat{\mu} = \bar{x} - \alpha\gamma \tag{2}$$

$$\hat{\alpha} = \frac{\sqrt{6}S}{\pi} \tag{3}$$

where $\bar{x}$ is the sample mean and S is the sample standard deviation. Then the parameters μ and α for each river are computed by solving two linear equations. The parameters obtained are shown in Table 1.

3.2 Maximum likelihood estimation

In this subsection, We use maximum likelihood estimator(ML) to estimate parameters μ and α of Gumbel probability density function. From probability density function of Gumbel distribution,

$$f(x|\mu,\alpha) = \frac{1}{\alpha}e^{\left(-\frac{(x-\mu)}{\alpha}\right)}e^{\left(-\exp\left(-\frac{(x-\mu)}{\alpha}\right)\right)}$$

If $x_1, x_2, \ldots, x_N$ are iid Gumbel, then the likelihood function for given μ and α is

$$f(x_1, x_2, \ldots, x_N|\mu,\alpha) = \frac{1}{\alpha^N}\prod_{i=1}^{N} e^{\left(-\frac{(x_i-\mu)}{\alpha}\right)}e^{\left(-\exp\left(-\frac{(x_i-\mu)}{\alpha}\right)\right)}.$$

To maximize f, it is convenient to consider the log-likelihood function

$$\mathcal{L}(\mu,\alpha) = \log(f(x_1, x_2, \ldots, x_N|\mu,\alpha)) = -N\log(\alpha) - \sum_{i=1}^{N}\frac{x_i-\mu}{\alpha} - \sum_{i=1}^{N}\exp\left(-\frac{(x_i-\mu)}{\alpha}\right).$$

To find estimated parameters which is the maximizer of the log-likelihood function, MATLAB command 'fminunc' is used to solve an unconstrained nonlinear optimization that is

$$(\hat{\mu}, \hat{\alpha}) = \text{argmax } \mathcal{L}(\mu,\alpha).$$

Since the estimated parameters must maximize the log-likelihood function, function

$$f = N\log(\alpha) + \sum_{i=1}^{N}\frac{x_i-\mu}{\alpha} + \sum_{i=1}^{N}\exp\left(-\frac{(x_i-\mu)}{\alpha}\right).$$

is used as input function for 'fminunc'. After trying different starting points including the result that was calculated from MM method, the estimated parameters μ and α are obtained and are shown in Table 1.

3.3 Estimation results

The parameters μ and α which obtained from MM and ML estimation are shown in Table 1.(see Listing 1)

River	ML method		MM method	
	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\mu}$	$\hat{\alpha}$
Ping (p_1)	319.93	125.21	323.94	110.37
Wang (p_2)	421.63	222	420.65	236.77
Yom (p_3)	682.12	445.64	642.26	553.38
Nan (p_4)	998.31	444.1	1004.6	442.18
Chao Phraya (p_N)	1996.5	829.71	1989.3	899.71

Table 1: Estimated Gumbel distribution parameters from MM and ML method.

Method	Ping (p_1)	Wang (p_2)	Yom (p_3)	Nan (p_4)	Chao Phraya (p_N)
MM	-594.40	-332.12	-457.38	-1,030.49	-1,420.87
ML	-592.49	-325.39	-433.55	-964.89	-1,293.98

Table 2: Log-likelihood value evaluated using parameters from MM and ML estimation.

The parameters derived from method of moments and maximum likelihood are fitted using method in [4](p.92) and compared with scatter plot of data using Gringorten plotting position.

Gringorten plotting position (using non-exceedance Probability which is the probability that x is less than some value, $P(X \leq x)$) is computed by k^{th} ranked sorting for N data samples from smallest to largest. Then non-exceedance probability is

$$\mathrm{P} = \frac{k - 0.44}{N + 0.12}, \quad k = 1, 2, \ldots, N.$$

Reduced variable are acquired by $y = -\log(-\log(\mathrm{P}))$.

Figure 4 displays plots of reduced variable compare to Gumbel distribution with parameters from MM and ML, the Gumbel plot of parameters that derived from method of moments is closer to the linear least-squares model (polynomial of order 1) of data than Gumbel plot of parameters from maximum likelihood estimation.

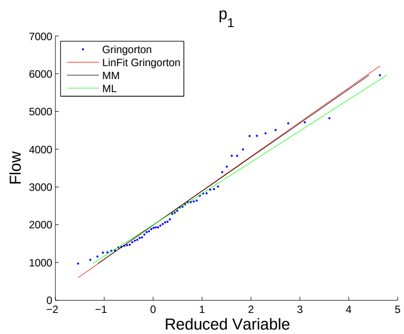
4 Parameters estimation of bivariate Gumbel pdf

4.1 Maximum likelihood estimation

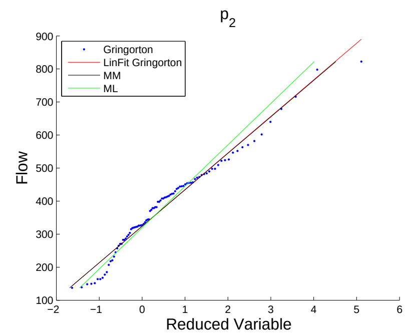
In this section, bivariate Gumbel distribution will be considered and its parameters will be estimated by using ML estimation. If $X = (X_1, X_2, ..., X_N)$ and $Y = (Y_1, Y_2, ..., Y_3)$ are two vectors of i.i.d. Gumbel, then the pdf of bivariate gumbel distribution (see Appendix 5.3) is

$$f(X, Y|\mu_x, \mu_y, \alpha_x, \alpha_y, \theta) = \frac{1}{\alpha_x^N \alpha_y^N} \prod_{i=1}^{N} F(x_i, y_i) e^{-c_i} \times \left\{ 1 - \theta \frac{e^{\frac{2(x_i - \mu_x)}{\alpha_x}} + e^{\frac{2(y_i - \mu_y)}{\alpha_y}}}{d_i^2} + 2\theta \frac{e^{2c_i}}{d_i^3} + \theta^2 \frac{e^{2c_i}}{d_i^4} \right\} \tag{4}$$

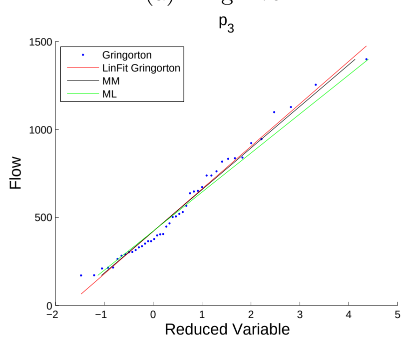
where θ is the parameter that explains the relation between X and Y that have the value between 0 and 1. Then the log-likelihood function is



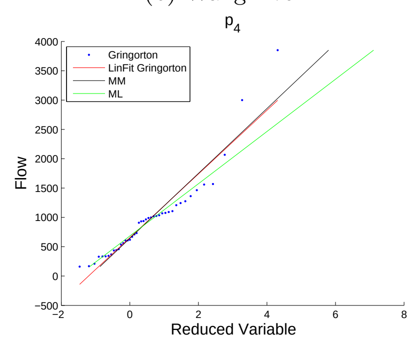
(a) Ping river



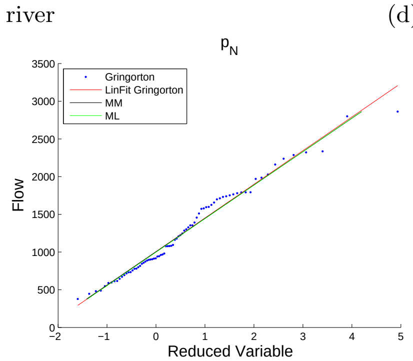
(b) Wang river



(c) Yom river



(d) Nan river



(e) Chao Phraya river

Figure 4: Plot of reduced variable compared to Gumbel distribution with parameters from MM and ML estimation.(see Listing 2)

$$
\begin{aligned}
\mathcal{L}(\mu_x,\mu_y,\alpha_x,\alpha_y,\theta) &= \log f(X,Y|\mu_x,\mu_y,\alpha_x,\alpha_y,\theta) \\
&= -N\log\alpha_x\alpha_y + \sum_{i=1}^{N}\log F(x_i,y_i) - \sum_{i=1}^{N} c_i \\
&\quad + \sum_{i=1}^{N}\log\left\{1-\theta\frac{e^{\frac{2(x_i-\mu_x)}{\alpha_x}}+e^{\frac{2(y_i-\mu_y)}{\alpha_y}}}{d_i^2}+2\theta\frac{e^{2c_i}}{d_i^3}+\theta^2\frac{e^{2c_i}}{d_i^4}\right\}.
\end{aligned}
$$

where

$$
\begin{aligned}
\log F(x_i,y_i) &= \log F(x_i)+\log F(y_i)-\theta\frac{\log F(x_i)\log F(y_i)}{\log F(x_i)+\log F(y_i)} \\
&= \exp\left(-\frac{x_i-\mu_x}{\alpha_x}\right)+\exp\left(-\frac{y_i-\mu_y}{\alpha_y}\right)+\theta\frac{\exp\left(-\frac{x_i-\mu_x}{\alpha_x}\right)\exp\left(-\frac{y_i-\mu_y}{\alpha_y}\right)}{\exp\left(-\frac{x_i-\mu_x}{\alpha_x}\right)+\exp\left(-\frac{y_i-\mu_y}{\alpha_y}\right)}.
\end{aligned}
$$

The ML problem is

$$
\max_{\mu_x,\mu_y,\alpha_x,\alpha_y,\theta}\mathcal{L}(\mu_x,\mu_y,\alpha_x,\alpha_y,\theta)
$$

with the constraints $0 < \theta < 1$,then parameters which maximize the log-likelihood function is computed by using MATLAB command 'fmincon'. The results are shown in Table 3.

4.2 Computation of parameter θ from population product-moment correlation coefficient

For another method to find parameter θ, we can compute parameter θ of bivariate Gumbel distribution from the relationship between θ and population product-moment correlation coefficient ρ (see [4]) :

$$
\theta = 2\left[1-\cos\left(\pi\sqrt{\frac{\rho}{6}}\right)\right] \quad \text{for} \quad 0\le\rho\le\frac{2}{3}. \tag{5}
$$

The parameters $\hat{\theta}$ that are estimated by using (5) are shown in Table 4.

4.3 Estimation results

Only the Gumbel parameter θ is estimated

From the Table 3, notice that the parameter θ of bivariate Gumbel distribution between Ping river and Wang river is -, since its correlation coefficient is greater than $\frac{2}{3}$, so the formula (5) is not valid to use.

River	Chao Phraya (p_N)	Ping (p_1)	Wang (p_2)	Yom (p_3)	Nan (p_4)
Chao Phraya (p_N)	-	0.4852 \ 0.5393	1 \ 0.8255	0.3566 \ 0.6825	0.8564 \ 0.8653
Ping (p_1)	0.4852 \ 0.5393	-	1.0000 \ -	0.8561 \ 0.8393	0.5611 \ 0.5667
Wang (p_2)	1.0000 \ 0.8255	1.0000 \ -	-	0.9099 \ 0.8726	0.6049 \ 0.5758
Yom (p_3)	0.3566 \ 0.6825	0.8561 \ 0.8393	0.9099 \ 0.8726	-	0.6367 \ 0.7292
Nan (p_4)	0.8564 \ 0.8653	0.5611 \ 0.5667	0.6049 \ 0.5758	0.6367 \ 0.7292	-

Table 3: Bivariate Gumbel distribution parameters θ which are computed (see Listing 3) from MLE and formula (5). (MLE\Formula)

Non-exceedance probability (value of joint CDF) of observed data can be calculated by Gringorten plotting position similar to the one using in univariate with the observed data are arranged in ascending order for two data sets. (see [4])

$$F(x_i, y_i) = P(X \leq x_i, Y \leq y_j) = \frac{\sum_{m=1}^{i} \sum_{l=1}^{j} n_{ml} - 0.44}{N + 0.12} \tag{6}$$

where N_{ml} is the number of occurences of the combinations of x_i and y_j and N is total numbers of samples.

We compute the non-exceedance proability $P(X \leq x_i, Y \leq y_i)$ by the following three methods: i) Gringorten plotting position as in equation (6), ii) $F(x_i, y_i)$ when parameters $\mu_x, \mu_y, \alpha_x, \alpha_y$ are estimated from marginal MLE and parameter θ is computes by the fomular (5) iii) $F(x_i, y_i)$ when parameter θ is estimated from bivariate MLE of one parameter (other parameters $\mu_x, \mu_y, \alpha_x, \alpha_y$ are replaced by the estimated value from marginal MLE. To compare non-exceedance probability of the observed data with non-exceedance probability from the theory with parameter θ from formulation given in (5) and MLE. 2-Norm error has been chosen to be a comparator of the parameter θ from the method ii) and the method iii). It was found that 2-norm error for each combination of two rivers between two method are slightly different.

River	Chao Phraya (p_N)	Ping (p_1)	Wang (p_2)	Yom (p_3)	Nan (p_4)
Chao Phraya (p_N)	-	0.8689 \ 0.8583	1.2025 \ 1.2377	0.8928 \ 0.8217	1.1125 \ 1.1105
Ping (p_1)	0.8689 \ 0.8583	-	- \ 1.1385	0.8214 \ 0.8252	0.7062 \ 0.7050
Wang (p_2)	1.2025 \ 1.2377	- \ 1.1385	-	0.9905 \ 0.9989	0.8232 \ 0.8299
Yom (p_3)	0.8928 \ 0.8217	0.8214 \ 0.8252	0.9905 \ 0.9989	-	0.7577 \ 0.7345
Nan (p_4)	1.1125 \ 1.1105	0.7062 \ 0.7050	0.8232 \ 0.8299	0.7577 \ 0.7345	-

Table 4: 2-Norm error of non-exceedance probability of observed and theory (θ calculated (see Listing 4) from formula). (MLE\Formula)

All Gumbel parameters are estimated by ML

Parameters was estimated by maximum the sum of log-likelihood function for Chao Phraya river and other rivers. The results of ML estimation are shown in Table 5. It is interesting that the parameters of Chao Phraya river are not constant when estimating across different rivers that are clear from the values of μ and α are around 1800 and 700 which compare with ML estimation of marginal probability, which μ and α are about 2000 and 800 respectively. The parameters of bivariate model are lesser than univariate model. Besides, the parameters of Chao Phraya river are not same as in univariate model. Parameters of other rivers are slightly different and θ which are estimated from ML (all parameters were estimated simultaneously) compare with θ from formula are also different.

River	μ_{River}	α_{River}	$\mu_{\text{ChaoPhraya}}$	$\alpha_{\text{ChaoPhraya}}$	θ
Ping (p_1)	289.07	141.65	1,791.44	698.49	0.2773
Wang (p_2)	403.59	193.53	1,807.88	733.72	0.8776
Yom (p_3)	716.58	498.84	1,748.89	618.80	0.4277
Nan (p_4)	939.91	334.89	1,794.46	701.58	0.6661

Table 5: ML estimated of Gumbel mixed models.(see Listing 5)

4.4 Return period analysis

ML method

When considering the return period of Chao Phraya river given others flood peak, Figure 7b shows that when the given flood peak of Wang river is varied, the return period of each given flood peak tends to be the most diverge compared to other rivers which shows that the given flood peak of Wang river has the biggest impact on Chao Phraya river. Figure 7d shows the return period of Chao Phraya river given Nan's flood peak. The graph is similar to figure 7b but the return period of each given flood peak is less diverge compared to figure 7d. Figure 7a shows that when the given flood peak of Ping river is varied. Even though at its extreme (1 percent exceedance probability), the return period of Chao Phraya river does not differ much from other given flood peaks. Since the return period of Chao Phraya river given Ping river's data does not differ much when the given data are changed, it is considered that Chao Phraya's flood peak is less dependent of Ping's flood peak. For figure 7c, it can be observed that the graph differs somewhat between each flood peak hence slightly more impact than Ping river

However, when comparing Figure 5 to Figure 7a,7b,7c,7d, it is discovered that all conditional return periods (Figure 7a,7b,7c,7d) have higher return periods than the marginal return period of Chao Phraya river (Figure 5) at every values of Chao Phraya river's flood peak. When given a river flood peak, the conditional probability is less than the marginal probability of Chao Phraya river. Notice that when high magnitude flood peak (*i.e.* at 1 percent exceedance probability or 100 years return period) occurs in either Ping, Wang, Yom, or Nan, probability of Chao Phraya river flooding should be greater than when no prior knowledge of previously mentioned river is known.

In order to investigate the problem of results above, we consider marginal

CDF of Gumbel distribution

$$\mathrm{F}(x) = e^{(-\exp(-\frac{(x-\mu)}{\alpha}))}$$

and conditional CDF of Gumbel distribution

$$\mathrm{F}(x \mid Y = y) = F(x)e^{-\theta(\frac{1}{\log F(x)} + \frac{1}{\log F(y)})^{-1}}.$$

To campare the value of the marginal CDF and the conditional CDF when all parameters of the function are identical, the term $e^{-\theta(\frac{1}{\log F(x)} + \frac{1}{\log F(y)})^{-1}}$ is to be considered. Since $\frac{1}{\log F(x)} + \frac{1}{\log F(y)}$ is negative, $-\theta(\frac{1}{\log F(x)} + \frac{1}{\log F(y)})^{-1}$ will be positive. This leads to the value of $e^{-\theta(\frac{1}{\log F(x)} + \frac{1}{\log F(y)})^{-1}}$ to be greater than 1. Finally, it can be concluded that a conditional CDF of bivariate Gumbel distribution is greater than the marginal CDF of univarite Gumbel distrition provided that the parameters $\alpha_x, \alpha_y, \mu_x, \mu_y$ used in the univariate CDF and in the bivariate CDF are the same values.

At this point, we can not utilize the bivariate Gumbel model to predict the relationship between Chao Phraya river and the others.

MM method

Figure 8 and Figure 10 show the marginal return period of Chao Phraya river and the conditional return period of Chao Phraya river given others flood peak computed using the parameters which is derived from method of moment respectively. Figure 10a, 10b, 10c, and 10d suggest that the conditional return period of Chao Phraya river depends on the peak flow of the other rivers, which is in contrast to the result obtained in 4.4 (in 4.4 the return period of Chao Phraya river seems to independent of the peak flow of Ping and Yom). As expected, the conditional return period of Chao Phraya river is higher than the marginal return period.

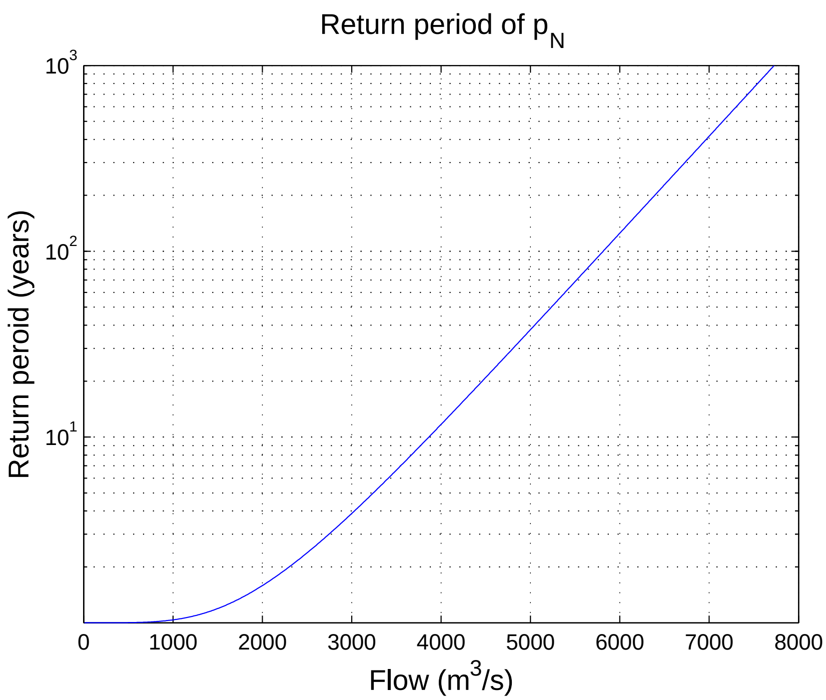


Figure 5: Marginal return period of Chao Phraya river by using ML estimate. (see Listing 6)

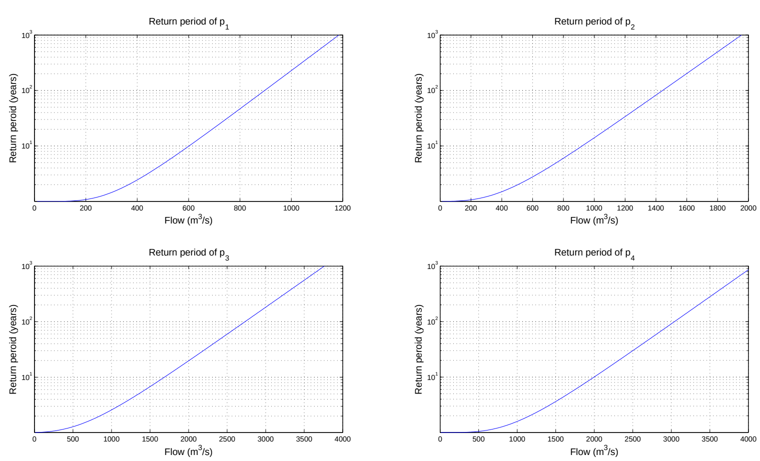
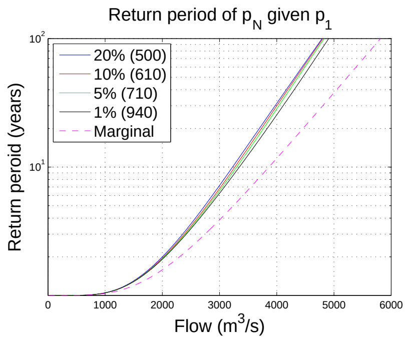
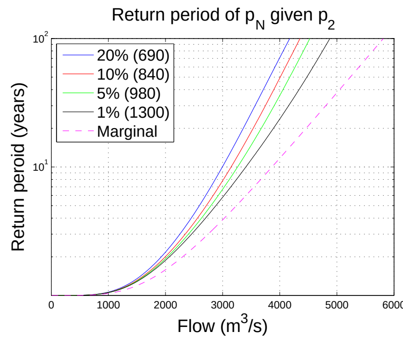


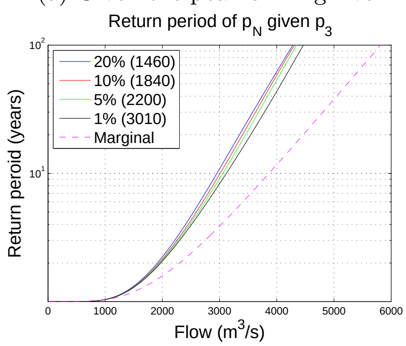
Figure 6: Marginal return period of Ping, Wang, Yom, Nan river by using ML estimate. (see Listing 6)



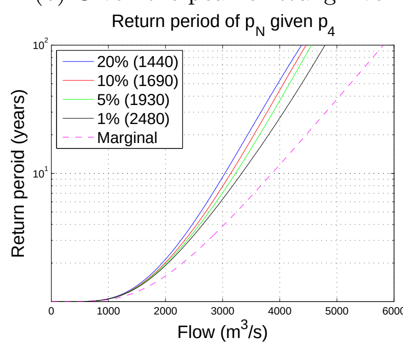
(a) Given the peak of Ping river



(b) Given the peak of Wang river



(c) Given the peak of Yom river



(d) Given the peak of Nan river

Figure 7: Conditional return period by using ML estimate. (see Listing 6)

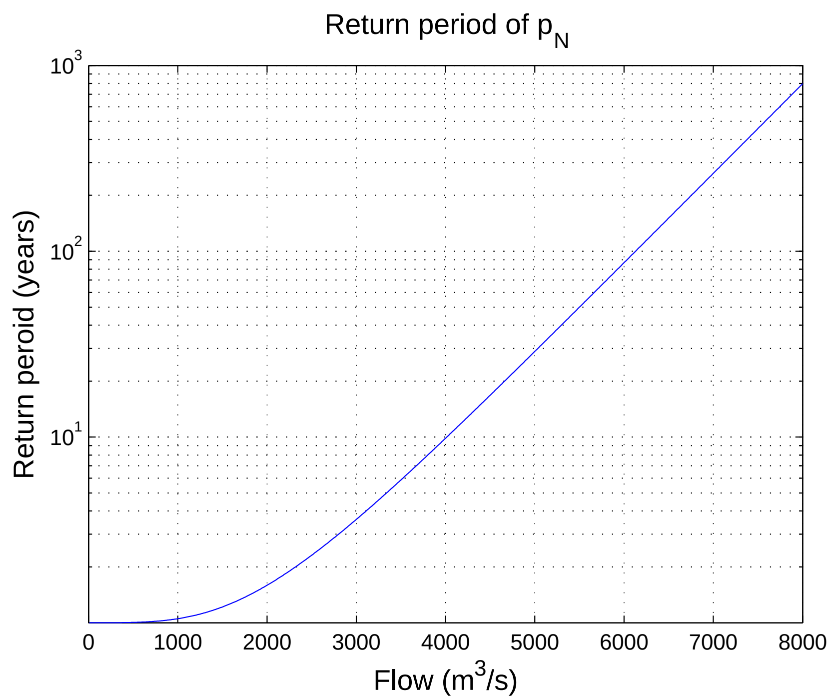


Figure 8: Marginal return period of Chao Phraya river by using MM estimate. (see Listing 6)

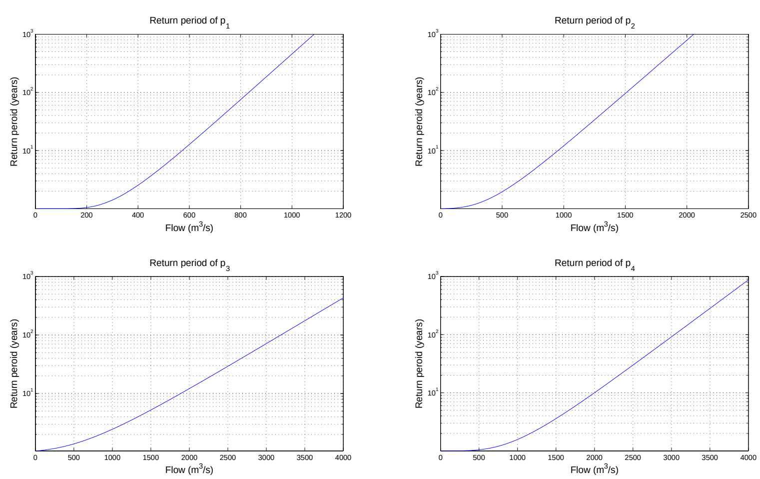
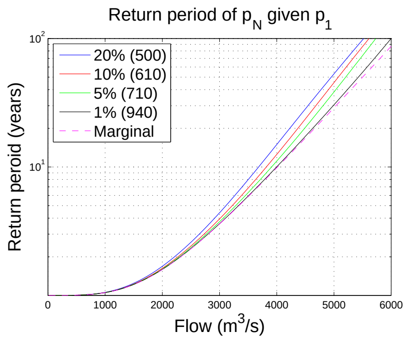
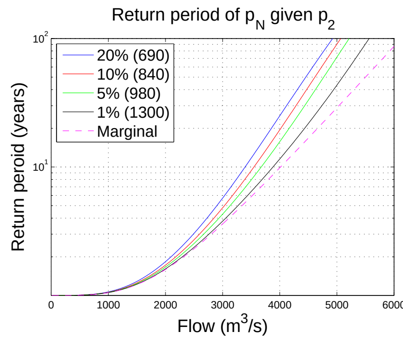


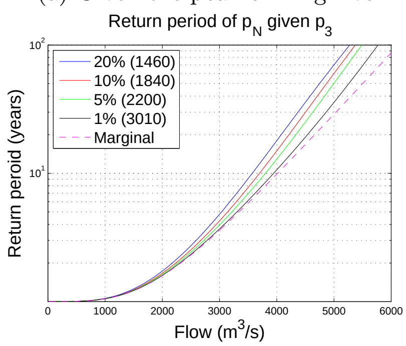
Figure 9: Marginal return period of Ping, Wang, Yom, Nan river by using MM estimate. (see Listing 6)



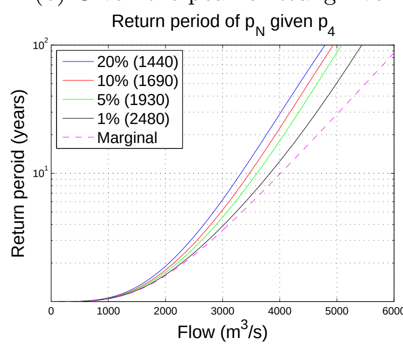
(a) Given the peak of Ping river



(b) Given the peak of Wang river



(c) Given the peak of Yom river



(d) Given the peak of Nan river

Figure 10: Conditional return period by using the estimated parameters computed from the formula 5 and MM method. (see Listing 6)

5 Appendix

5.1 Principle of method of moments

(see [7]) Method of moments is a simple method of parameters estimation by using population parameters. Let X be a random variable following some distribution. Then the kth moment of the distribution is defined as,

$$\mu_k = \mathrm{E}[X^k].$$

For example, $\mu_1 = \mathrm{E}[X]$ and $\mu_2 = \mathrm{Var}(X) + (\mathrm{E}[x])^2$.

The sample moments of observations $X_1, X_2, ..., X_n$ independent and identically distributed from some distribution are defined as,

$$\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k.$$

For example, $\hat{\mu}_1 = \bar{X}$ is the familiar sample mean and $\hat{\mu}_2 = \hat{\sigma}^2 + \bar{X}^2$ where $\hat{\sigma}$ is the standard deviation of the sample.

The method of moments estimator simply equates the moments of the distribution with the sample moments ($\mu_k = \hat{\mu}_k$) and solves for the unknown parameters. Note that this implies the distribution must have finite moments.

For example, if $X_1, X_2, ..., X_n$ are i.i.d. a Poisson distribution with proability mass function,

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, x = 0, 1, 2, ...$$

where λ is an unknown parameter. Check that $\mathrm{E}[X] = \lambda$. So, $\mu_1 = \mathrm{E}[X] = \lambda = \bar{X} = \hat{\mu}_1$. Hence, the method of moments estimator of λ is the sample mean.

5.2 Principle of maximum likelihood estimation

Suppose $x_1, x_2, ..., x_N$ are i.i.d. observation with joint probability density function

$$f(x_1, x_2, ..., x_N|\theta) = f(x_1|\theta) \times f(x_2|\theta) \times \cdots \times f(x_N|\theta)$$

called the likelihood function, where θ is a vector of unknown parameters $\theta_1, \theta_2, ..., \theta_N$ for pdf of a family of x_i. It is often more convenient to work with the logarithm of likelihood function, called the log-likelihood function:

$$\mathcal{L}(\theta) = \log(f(x_1, x_2, ..., x_N|\theta) = \sum_{i=1}^{N} \log(f(x_i|\theta))$$

The method of maximum likelihood estimates $\hat{\theta}$ by finding a value of θ that maximizes the log-likelihood function $\mathcal{L}(\theta)$

5.3 Bivariate Gumbel distribution

Bivariate Gumbel distribution called Gumbel mixed model which its marginal pdf is Gumbel distribution was used by [4] to create joint pdf of Flood peaks-Volume peak and Flood duration-Volume. The general form of cdf is:

$$F(x, y) = F(x)F(y)\exp\left\{-\theta\left[\frac{1}{\log F(x)} + \frac{1}{\log F(y)}\right]^{-1}\right\}, \quad (0 \le \theta \le 1).$$

θ is the parameter that describes the relation between two random variables. The formulation to calculate θ is given in [4] which is

$$\theta = 2\left[1 - \cos\left(\pi\sqrt{\frac{\rho}{6}}\right)\right] \quad \text{for} \quad 0 \le \rho \le \frac{2}{3}$$

where ρ is correlation coefficient of the two random variables.
probability density function of bivariate Gumbel distribution is

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{1}{\alpha_x \alpha_y} F(x,y) e^{-c} \left\{ 1 - \theta \frac{e^{\frac{2(x-\mu_x)}{\alpha_x}} + e^{\frac{2(y-\mu_y)}{\alpha_y}}}{d^2} + 2\theta \frac{e^{2c}}{d^3} + \theta^2 \frac{e^{2c}}{d^4} \right\} \tag{7}$$

where $c = \frac{x-\mu_x}{\alpha_x} + \frac{y-\mu_y}{\alpha_y}$, $d = e^{\frac{x-\mu_x}{\alpha_x}} + e^{\frac{y-\mu_y}{\alpha_y}}$.

5.4 Multivariate Gumbel distribution

Multivariate Gumbel distribution is obtain from transforming multivariate log-gamma distribution. [3] Multivariate log-gamma distribution

$$f_Y(y_1,\ldots,y_p) \approx \delta^v \sum_{n=0}^{\infty} \frac{\Gamma(v+n)}{\Gamma(v)n!} (1-\delta)^n \prod_{j=1}^{p} \frac{1}{\Gamma(v+n)} \left(\frac{y_j}{\lambda_j \delta}\right)^{v+n} \times \exp\left(-\frac{y_j}{\lambda_j \delta}\right), \quad y_j > 0$$

Using the transformation,

$$Z_i = \frac{\ln(Y_i/\delta)}{\mu_i}$$

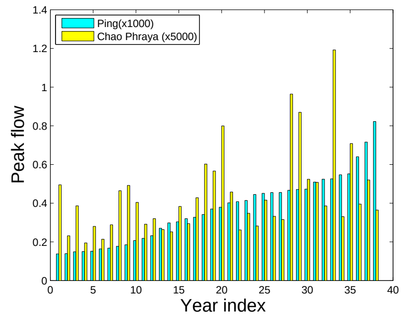
Gives

$$f(z_1,\ldots,z_p) = \delta^v \sum_{n=0}^{\infty} \frac{(1-\delta)^n \prod_{i=1}^{p} \mu_i \lambda_i^{-v-n}}{[\Gamma(v+n)]^{p-1}\Gamma(v)n!} \times \exp\left\{ (v+n) \sum_{i=1}^{p} \mu_i z_i - \sum_{i=1}^{p} \frac{1}{\lambda_i} \exp(\mu_i z_i) \right\},$$

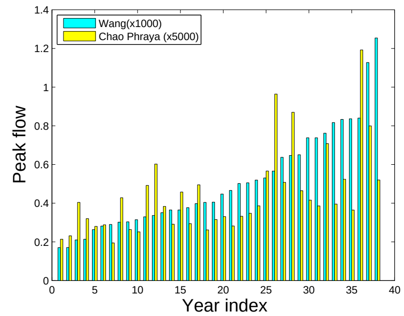
where $z_i \in \mathbb{R}$

5.5 Detail of Ping river and Chao Phraya river

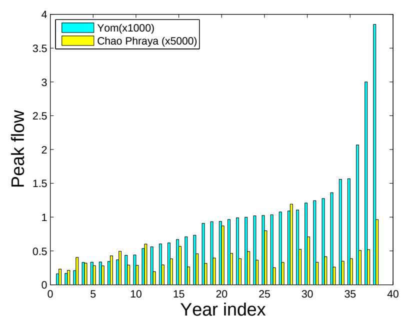
In this section, we plot the flow peak of Ping, Wang, Yom, Nan rivers. For each river, we sort the peak values ascendingly and plot the peak of Chao Phraya river from the corresponding year. This is to see the result of cause and effect from each river to Chao Phraya similar to the scatter plots to see the correlations. Figure 11 suggest that there are no clear distinct effect from the peak magnitude of the four rivers to Chao Phraya.



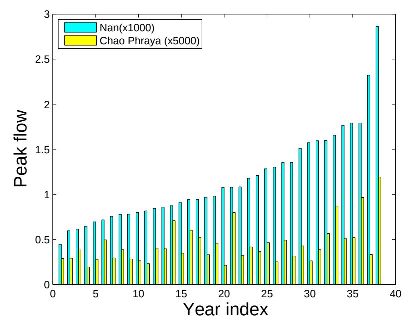
(a) Ping



(b) Wang



(c) Yom



(d) Nan

Figure 11: Bar plot of scaled flow peak values of each of Ping, Wang, Yom, Nan rivers and Chao Phraya river. (see Listing 7)

5.6 Code

`Univar_est.m` is a code to compute the value of parameters for univariate distribution by MM and MLin Section 3.3. The sample correlation of all rivers have been computed in this code and represent the correlation as shown in Figure 3.

```
load('Peak_data.mat')

% Create necessary data for MM and Ml estimation
% mean
avg = [mean(pn);mean(p1);mean(p2);mean(p3);mean(p4)];
%std deviation
s   = [std(pn);std(p1);std(p2);std(p3);std(p4)];
% alpha from MM
a_mm = sqrt(6)*s/pi;
% mu from MM
u_mm = avg - 0.5772*a_mm;
% number of sample data
n = [length(pn),length(p1),length(p2),length(p3),length(p4)];
par_mm = [u_mm';a_mm'];

% Reformat sample data for convenience in coding
P = {pn,p1,p2,p3,p4};
par_ml = [];

% ML estimation of parameters using fminunc to find local minimum
for i=1:5
```

```
    fun1 = @(x) n(i)*log(x(2)) + sum((P{i}-x(1))/x(2)) + sum(exp(-(P{i}-x(1))/x(2)));
    par1 = fminunc(fun1,par_mm(:,i));
    par_ml = [par_ml,par1];
end

% Display parameters from MM and ML
par_mm = [u_mm';a_mm']
par_ml = par_ml

% Calculate log likelihood values
likelihood_mm = [];
likelihood_ml = [];
for i=1:5
    likelihood_mm=[likelihood_mm;n(i)*log(par_mm(2)) + sum((P{i}-par_mm(1))/par_mm(2)) + sum(exp(-(P{i}-par_mm(1))/par_mm(2)))];
    likelihood_ml=[likelihood_ml;n(i)*log(par_ml(2)) + sum((P{i}-par_ml(1))/par_ml(2)) + sum(exp(-(P{i}-par_ml(1))/par_ml(2)))];
end
% Display log likelihood values
likelihood_mm = likelihood_mm'
likelihood_ml = likelihood_ml'

% Calculate sample correlation of all combination
correlation = corr(p_mat)
% Display scatter plot of each pairs
figure
k=1;
for i= 1:5
    for j = 1:5
        subplot(5,5,k)
        scatter(p_mat(:,i),p_mat(:,j),'x')
        str1 = num2str(i-1);
        str2 = num2str(j-1);
        if i == 1 && j==1
            str1 = 'N';
            str2 = 'N';
        elseif j==1
            str2 = 'N';
        elseif i==1
            str1 = 'N';
        end
        str = sprintf('Corr (P_%s,P_%s) = %1.3f',str1,str2,correlation(i,j));
        title(str)
        k = k+1;
    end
end

clear i j k fun1 par1 f hes1 avg s a_m u_m n str str1 str2 a_mm u_mm
save('Peak_Par_data.mat')
```

Listing 1: Univar_est.m

Figure 4 in Section 3.3 was generated by `Univar_reduced.m`.

```
load('Peak_Par_data.mat');

% Create array that keep the number of samples for each rivers
n = [length(pn),length(p1),length(p2),length(p3),length(p4)];
% Sort data in ascend order
p1 = sort(p1,'ascend');
```

```
p2 = sort(p2,'ascend');
p3 = sort(p3,'ascend');
p4 = sort(p4,'ascend');
pn = sort(pn,'ascend');
% Create cell for convenience of coding
P={pn,p1,p2,p3,p4};
% Create new variables
Y_g = cell(5,1);
Y_mm = cell(5,1);
Y_ml = cell(5,1);
X_ls = cell(5,1);
par_ls_gringorten =zeros(2,5);

% Calaculate reduced variable value
for i = 1:5
    % Observed data using Gringorten method
    y = -log(-log((((1:n(i))-0.44)/(n(i)+0.12))));
    % Reduced var using parameters form MM
    y_mm = (P{i}-par_mm(1,i))/par_mm(2,i);
    % Reduced var using parameters form ML
    y_ml = (P{i}-par_ml(1,i))/par_ml(2,i);
    % Least square to find parameters from Gringorten method
    A = ones(n(i),2);
    A(:,2) = y;
    par_ls_gringorten(:,i) =  A\P{i};
    % x_ls is flow peak compute from par_ls_gringorten. computed for purpose
    % of plotting.
    x_ls = (y*par_ls_gringorten(2,i))+par_ls_gringorten(1,i);

    Y_g{i} = y';
    Y_mm{i} = y_mm;
    Y_ml{i} = y_ml;
    X_ls{i} = x_ls;
end

% Plot reduced variable vs flow peak
for i = 1:5
    figure;
    scatter(Y_g{i},P{i},'.');
    hold on
    plot(Y_g{i},X_ls{i},'r');
    hold on
    plot(Y_mm{i},P{i},'k');
    hold on
    plot(Y_ml{i},P{i},'g');
    legend('Gringorton','LinFit Gringorton','MM','ML','Location','northwest');
    xlabel('Reduced Variable','FontSize',15);
    ylabel('Flow (m^3/s)','FontSize',15);
    str = sprintf('p_%i',i);
    if i == 5
        str = 'p_N';
    end
    title(str,'FontSize',15);
end
clear str y y_mm y_ml A x_ls i
```

Listing 2: Univar_reduced.m

The parameters of Gumbel mixed model in Section 4.3 was obtained from `Bivar_est.m`. The concern of this code is only to provide θ from formula provided in [4] and from ML estimation which only θ has been determined not all parameters.

```
load('Peak_Par_data.mat')

% Create new variables
theta_formulation = zeros(5);
theta_ml = zeros(5);
Fx = zeros(38,5);
lnFx = zeros(38,5);
c = cell(5);
d = cell(5);
std = zeros(38,5);
exp_std = zeros(38,5);

for i=1:5
    % std = (x-mu)/alpha
    std(:,i) = (p_mat(:,i)-par_ml(1,i))/par_ml(2,i);
    % exp_std = exp(std)
    exp_std(:,i) = exp(std(:,i));
    % lnFx is log of non-exceedance prob. lnFx = -exp(-std)
    lnFx(:,i) = -exp(-std(:,i));
    % Fx is non-exceedance prob. Fx = exp(-exp(-))
    Fx(:,i) = exp(lnFx(:,i));
    for j=1:i
        % Sample correlation must less than or equal to 2/3 to make the
        % formulation to be valid.
        if correlation(i,j) <= 2/3
            theta_formulation(i,j) = 2*(1-cos(pi*sqrt(correlation(i,j)/6)));
            theta_formulation(j,i) = theta_formulation(i,j);
        end
        % Calculate c, d for convenience of coding
        % c = (x-mu_x)/alpha_x + (y-mu_y)/alpha_y
        c{i,j}= std(:,i)+std(:,j);
        % d = exp{(x-mu_x)/alpha_x} + exp{(y-mu_y)/alpha_y}
        d{i,j}= exp_std(:,i)+exp_std(:,j);

        % Find theta from ML
        if i~=j
            % A,B,C are constant terms, pre-calculated for convenience of
            % coding
            A= (exp(2*std(:,i))+exp(2*std(:,j)))./(d{i,j}.^2);
            B= exp(2*c{i,j})./(d{i,j}.^3);
            C= exp(2*c{i,j})./(d{i,j}.^4);
            % Minus of Log-likelihood function
            func = @(x) -(sum(-x*(1./((1./lnFx(:,i))+(1./lnFx(:,j))))+log(1- x*A + 2*x*B + (x^2)*C)));
            % Find the minimum point. Choose starting point from theta
            % calculated from formula.
            t = fmincon(func,theta_formulation(i,j),[],[],[],[],0,1);
            theta_ml(i,j) = t;
            theta_ml(j,i) = t;
        end
    end
end
clear i j func A B C c d t std exp_std
save('bivar.mat')
```

Listing 3: Bivar_est.m

Joint non-exceedance probability of observed data calculated from Gringorten method and theoretical non-exceedance probability of Gumbel mixed model using parameters (μ, α) from ML estimation and θ from both formula and ML estimation in Section 4.3.

```
% N is number of samples for each rivers
N=length(p_mat);
% Create new variables
% Occurence is matrix of the number of occurences of X<=x, Y<=y
Occurrence=cell(5);
% Non-exceedance prob from Gringorten, Formulation, ML
jnt_nonexc_Gri=cell(5);
jnt_nonexc_For=cell(5);
jnt_nonexc_ML=cell(5);

for i=1:5
    for j=1:i
        Occurrence{i,j}=zeros(N,1);
        % Count the number of occurences
        for k=1:N
            x=(p_mat(:,i) <= p_mat(k,i));
            y=(p_mat(:,j) <= p_mat(k,j));
            Occurrence{i,j}(k) = sum(x&y);
        end
        % Calculate non-exceedance prob
        jnt_nonexc_Gri{i,j}= (Occurrence{i,j}-0.44)./(N+0.12);
        jnt_nonexc_For{i,j}=Fx(:,i).*Fx(:,j).*exp(theta_formulation(i,j)*(1./((1./lnFx(:,i))+(1./lnFx(:,j)))));
        jnt_nonexc_ML{i,j} =Fx(:,i).*Fx(:,j).*exp(theta_ml(i,j)*(1./((1./lnFx(:,i))+(1./lnFx(:,j)))));
    end
end

% Calculate error compare to observed data by 2-norm and percentage
Obse_Form_percent=zeros(5);
Obse_Form_norm=zeros(5);
Obse_ML_percent=zeros(5);
Obse_ML_norm=zeros(5);
for i=1:5
  for j=1:i
   Obse_Form_percent(i,j)=mean((jnt_nonexc_Gri{i,j}-jnt_nonexc_For{i,j})./jnt_nonexc_Gri{i,j});
   Obse_Form_norm(i,j) = norm(jnt_nonexc_Gri{i,j}-jnt_nonexc_For{i,j})/N;
   Obse_ML_percent(i,j)= mean((jnt_nonexc_Gri{i,j}-jnt_nonexc_ML{i,j})./jnt_nonexc_Gri{i,j});
   Obse_ML_norm(i,j)    = norm(jnt_nonexc_Gri{i,j}-jnt_nonexc_ML{i,j})/N;
  end
end
clear i j k x y
```

Listing 4: jnt_prob.m

ML estimation of all parameters in Gumbel mixed model was calculated by `ML_mixed.m` in Section 4.3.

```
load('bivar.mat')
% Create cell to record parameter from estimation
parameter = cell(5);
for i=1:5
    for j=1:i
        if i~=j
```

```
            % Starting point of optimization choose from marginal ml and
            % theta from formulation
            start = [par_ml(:,i);par_ml(:,j);theta_formulation(i,j)];
            % Minus of log likelihood function
            fun1 = @(x) -(sum(-log(x(2)*x(4))...
                               -exp(-(p_mat(:,i)-x(1))/x(2))...
                               -exp(-(p_mat(:,j)-x(3))/x(4))...
                               -x(5)*(1./((1./(-exp(-(p_mat(:,i)-x(1))/x(2))))+(1./(-exp(-(p_mat(:,j)-x(3))/x(4))))))...
                               -((p_mat(:,i)-x(1))/x(2)) -((p_mat(:,j)-x(3))/x(4)) ...
                               +log(1 - x(5)*((exp(2*(p_mat(:,i)-x(1))./x(2))+exp(2*(p_mat(:,j)-x(3))./x(4)))...
                               ./((exp((p_mat(:,i)-x(1))./x(2))+exp((p_mat(:,j)-x(3))./x(4))).^2))...
                               +2*x(5)*((exp(2*((p_mat(:,i)-x(1))./x(2) + (p_mat(:,j)-x(3))./x(4))))...
                               ./((exp((p_mat(:,i)-x(1))./x(2))+exp((p_mat(:,j)-x(3))./x(4))).^3))...
                               +(x(5)^2)*((exp(2*((p_mat(:,i)-x(1))./x(2) + (p_mat(:,j)-x(3))./x(4))))...
                               ./((exp((p_mat(:,i)-x(1))./x(2))+exp((p_mat(:,j)-x(3))./x(4))).^4)))));
            temp = fmincon(fun1,start,[],[],[],[],[0,0,0,0,0],[inf,inf,inf,inf,1]);
            parameter{i,j} = temp;
        end
    end
end
clear i j fun1 temp start
save('bivar_ml.mat')
```

Listing 5: ML_mixed.m

Figure 5, 6, 7, 8, 9, and 10 in Section 11 were generated by `Return_period_bivar.m`.

```
load('bivar_ml.mat')

F=cell(10,1);
cF=cell(4);
% Create matrix of peak flow
qn = (10:10:8000)';
% Calculate non-exceedance prob
F{1} = exp(-exp(-(qn-par_ml(1,1))/par_ml(2,1)));
% Calculate marginal return period
T_chao = 1./(1-F{1});
figure
% Plot semilog graph of flow vs return period
semilogy(qn,T_chao)
ylim([0,1000])
xlim([0,8000])
grid on
title('Return period of p_N','FontSize',13)
xlabel('Flow (m^3/s)','FontSize',13)
ylabel('Return peroid (years)','FontSize',13)

% Create matrix of peak flow
q = (10:10:4000)';
figure
for i=1:4
    % Calculate non-exceedance prob
```

```
    F{i+1} = exp(-exp(-(q-par_ml(1,i+1))/par_ml(2,i+1)));
    % Calculate marginal return period
    T = 1./(1-F{i+1});
    % Plot semilog graph of flow vs return period
    subplot(2,2,i)
    semilogy(q,T)
    ylim([0,1000])
    str = sprintf('Return period of p_%i',i);
    title(str,'FontSize',13)
    xlabel('Flow (m^3/s)','FontSize',13)
    ylabel('Return peroid (years)','FontSize',13)
    grid
    ax=gca;
    set(ax,'LineWidth',.05,'XGrid','on','YGrid','on');
end

C={'b','r','g','k'};
Fx_bivar=cell(4,1);
Fy_bivar=cell(4,1);
lnFx_bivar=cell(4,1);
lnFy_bivar=cell(4,1);
% Each row of Confidence refer to index of flow peak of each river (Ping,
% Wang, Yom, Nan) which have exceedance prob 20%, 10%, 5%, 1% respectively
Confidence=[50,61,71,94;69,84,98,130;146,184,220,301;144,169,193,248];
for i=1:4
    figure
    % Calculate marginal non-exceedance prob & log of non-exceedance prob
    % using parameters from ML
    Fx_bivar{i}=exp(-exp(-(qn-parameter{i+1,1}(3))/parameter{i+1,1}(4)));
    Fy_bivar{i}=exp(-exp(-(qn-parameter{i+1,1}(1))/parameter{i+1,1}(2)));
    lnFx_bivar{i}=log(Fx_bivar{i});
    lnFy_bivar{i}=log(Fy_bivar{i});
    for j=1:4
        % Calculate conditional non-exceedance prob & conditional return
        % period
        cF{j,i} = Fx_bivar{i}(1:600).*exp(-parameter{i+1,1}(5)*(1./((1./lnFx_bivar{i}(1:600))+(1/lnFy_bivar{i}(Confidence(i,j))))));
        cT = 1./(1-cF{j,i}(1:600));
        semilogy(qn(1:600),cT,'color',C{j})
        hold on
    end
    % Plot marginal of Chao Phraya river
    semilogy(qn(1:600),T_chao(1:600),'--m')
    str = sprintf('Return period of p_N given p_%i',i);
    ylim([0,100])
    title(str,'FontSize',17)
    leg1=sprintf('20%% (%i0)',Confidence(i,1));
    leg2=sprintf('10%% (%i0)',Confidence(i,2));
    leg3=sprintf('5%% (%i0)',Confidence(i,3));
    leg4=sprintf('1%% (%i0)',Confidence(i,4));
    legend(leg1,leg2,leg3,leg4,'Marginal','Location','northwest')
    ch = get(gcf,'children')
    set(ch(1), 'FontSize',15)
    xlabel('Flow (m^3/s)','FontSize',17)
```

```
    ylabel('Return peroid (years)','FontSize',17)
    grid
    ax=gca;
    set(ax,'LineWidth',.01,'XGrid','on','YGrid','on');
end

% Calculate marginal return period using parameter from MM
F{6} = exp(-exp(-(qn-par_mm(1,1))/par_mm(2,1)));
T_chao = 1./(1-F{6});
figure
% Plot semilog graph of flow vs return period
semilogy(qn,T_chao)
ylim([0,1000])
xlim([0,8000])
grid on
title('Return period of p_N','FontSize',13)
xlabel('Flow (m^3/s)','FontSize',13)
ylabel('Return peroid (years)','FontSize',13)

figure
for i=1:4
    % Calculate non-exceedance prob
    F{i+6} = exp(-exp(-(q-par_mm(1,i+1))/par_mm(2,i+1)));
    % Calculate marginal return period
    T = 1./(1-F{i+6});
    % Plot semilog graph of flow vs return period
    subplot(2,2,i)
    semilogy(q,T)
    ylim([0,1000])
    str = sprintf('Return period of p_%i',i);
    title(str,'FontSize',13)
    xlabel('Flow (m^3/s)','FontSize',13)
    ylabel('Return peroid (years)','FontSize',13)
    grid
    ax=gca;
    set(ax,'LineWidth',.05,'XGrid','on','YGrid','on');
end
% Reusing old variable since we only consider the plot
for i=1:4
    figure
    % Calculate marginal non-exceedance prob & log of non-exceedance prob
    % using parameters from MM and formula
    Fx_bivar{i}=exp(-exp(-(qn-par_mm(1,1))/par_mm(2,1)));
    Fy_bivar{i}=exp(-exp(-(qn-par_mm(1,i+1))/par_mm(2,i+1)));
    lnFx_bivar{i}=log(Fx_bivar{i});
    lnFy_bivar{i}=log(Fy_bivar{i});
    for j=1:4
        % Calculate conditional non-exceedance prob & conditional return
        % period
        cF{j,i} = Fx_bivar{i}(1:600).*exp(-theta_formulation(i+1,1)*(1./((1./lnFx_bivar{i}(1:600))+(1/lnFy_bivar{i}(Confidence(i,j))))));
        cT = 1./(1-cF{j,i}(1:600));
        semilogy(qn(1:600),cT,'color',C{j})
        hold on
    end
    % Plot marginal of Chao Phraya river
    semilogy(qn(1:600),T_chao(1:600),'--m')
    str = sprintf('Return period of p_N given p_%i',i);
    ylim([0,100])
```

```
    title(str,'FontSize',17)
    leg1=sprintf('20%% (%i0)',Confidence(i,1));
    leg2=sprintf('10%% (%i0)',Confidence(i,2));
    leg3=sprintf('5%% (%i0)',Confidence(i,3));
    leg4=sprintf('1%% (%i0)',Confidence(i,4));
    legend(leg1,leg2,leg3,leg4,'Marginal','Location','northwest')
    ch = get(gcf,'children')
    set(ch(1), 'FontSize',15)
    xlabel('Flow (m^3/s)','FontSize',17)
    ylabel('Return peroid (years)','FontSize',17)
    grid
    ax=gca;
    set(ax,'LineWidth',.01,'XGrid','on','YGrid','on');
end
clear leg1 leg2 leg3 leg4 ax str C i j
```

Listing 6: Return_period_bivar.m

Figure ??, ??, ??, and ?? in Section 4.3 were generated by

```
% Scaling factor is set to be 5000 for Chao Phraya and 1000 for others
scaling = [5000,1000];
p_mat_scale = p_mat;
% Scale data of Chao Phraya river
p_mat_scale(:,1) = p_mat(:,1)./scaling(1);
str = {'Ping','Wang','Yom','Nan'};
for i=1:4
   % Sorting p_i data
   p_mat_scale = sortrows(p_mat_scale,i+1);
   % Scale data of p_i river
   p_mat_scale(:,i+1) = p_mat_scale(:,i+1)./scaling(2);
   temp = [p_mat_scale(:,i+1),p_mat_scale(:,1)];
   figure
   b=bar(temp);
   set(b(1),'FaceColor','c')
   set(b(2),'FaceColor','y')
   set(b,'BarWidth',1)
   legend(strcat(str{i},'(x1000)'),'Chao Phraya (x5000)','Location','northwest');
   ylabel('Peak flow','FontSize',17)
   xlabel('Year index','FontSize',17)
end
```

Listing 7: bar_plot.m

References

[1] "The World Bank Supports Thailand's Post-Floods Recovery Effort". World Bank. 13 December 2011. Retrieved 25 January 2012.

[2] Kidson, R., & Richards, K. S. (2005). *Flood frequency analysis: assumptions and alternatives*. Progress in Physical Geography, 29(3), 392-410.

[3] Demirhan, H., & Hamurkaroglu, C. (2011). *On a multivariate log-gamma distribution and the use of the distribution in the Bayesian analysis*. Journal of Statistical Planning and Inference, 141(3), 1141-1152.

[4] Yue, S., Ouarda, T. B. M. J., Bobee, B., Legendre, P., & Bruneau, P. (1999). *The Gumbel mixed model for flood frequency analysis*. Journal of hydrology, 226(1), 88-100.

[5] Weisstein, Eric W. *Euler-Mascheroni Constant*. From MathWorld–A Wolfram Web Resource. `http://mathworld.wolfram.com/Euler-MascheroniConstant.html`

[6] Weisstein, Eric W. *Euler-Mascheroni Integrals*. From MathWorld–A Wolfram Web Resource. `http://mathworld.wolfram.com/Euler-MascheroniIntegrals.html`

[7] Bowman, K. O., & Shenton, L. R. *Estimation: Method of Moments*. Encyclopedia of statistical sciences, Wiley (1998): 2092-2098.