Probable Maximum Flood (PMF)Estimation Using a Bayesian Model(Case Study: Latian Dam)

Ahmad Agheshlouei[1],Mojtaba Shourian[1]✉

1. Faculty of Civil, Water and Environmental Engineering, ShahidBeheshti University
✉Email: m_shourian@sbu.ac.ir

Abstract
Probable Maximum Flood (PMF) is one of the important factors indam designing.In addition, it is the base for sizing the spillway of a dam to prevent overtopping during a flood. PMF is the response of a catchment to the Probable Maximum Precipitation (PMP). Among the different methods, the data-driven models try to estimate the unobserved events only based on a series of observed flows.
Bayesian models are used effectively topredict flood discharge. They help hydrologists and other responsible people to makemore appropriate decisions to save lives and properties in downstream of watershed. Although, the frequency analysis is popular yet, Bayesian model uses knowledge of an expert together with observation data to produce a more precise estimation.
In this study, the hydrometric station located near Latian dam is selected. Its historical report is available for more than 70 years.Two scenarios namely, Annual maximum flows and Peak floods are modeled using WinBUGS software, which estimates the posterior distribution function, using MCMC method, Metropolis-Hastings algorithm with Gibs sampling. An upper bounded LN4 is chosen as prior function.
Keywords: Probable Maximum Flood (PMF),Bayesian Analysis, Latian Dam.

1. INTRODUCTION

During recent decade, due to climate change, the frequency and severity of floods have been increasing dramatically.It is essential for a water resource planner to know the magnitude of the maximum flood. Knowledge of PMF is required for dams design, because it is a critical input parameter for designing the reservoir of a Dam, and the capacity of the Spillway. Furthermore, it is used for design of a flood control system around cities or industrial facilities (e.g. Oil and Gas plants), since industrial plantsare usually located at the foot of a mountain.

PMF maybe the most used criterion in the UnitedStates and other countries to design major hydraulic flood defensestructures, such aslarge dam spillways(ICOLD 1987; FEMA 2004; USNRC 1977). [1,2,3]

WMO (1986) defined PMF as the response of a catchment to PMP, probable maximum precipitation that is meteorologically possible tooccur over that catchment, during a given season of the year [4].Usually, an appropriate rainfall-runoff simulation model performs the PMP–PMF transformation [5]. Dawdy and Lettenmaier (1987) stated, since the PMF is astatistic,as other statistic must be estimated, regarded thePMF as ‘‘quasi-deterministic’’. They declared that its exceedance probabilities is small but in fact, it is not zero [6]. Although engineers accepted this ‘‘quasi-deterministic’’ PMF-like floods as a safety standard for high-hazard dams, the probability of extreme floods should be estimatetoincorporate it into quantitative risk assessment studies. This study is needed for both new and existing structures (Stedinger et al. 1989;Dawdy and Lettenmaier 1987; Dubler and Grigg 1996;USBR 2002).[7,6,8,9]

Nathan and Weinmann(2001)have categorizes floods [10], based on their return periods, Figure 1. Floods with return period of more than 2000 years (Pr$\leq 2 \times 10^{-3}$) are considered outside theextrapolation range because of very large amount of uncertainty leve. However, PMF or some percent of PMF is used for flood control system, as well as for the detention and retention damsto check the Dam failure for both new and existing ones, andalsothe spillway’s sizing.Therefore,instead of frequency analysis, Bayesian model is used as an alternative solution.

Fernandez (2010) studied American river basin [5], near the Folsomreservoir, in California, USA. He combinedconventionalanalysis withthe logical context in a Bayesian model to reach a more reliable estimation ofPMF. He used two types of non-systematic data: (a) the historical floods so-called paleofloods which has been recorded by human, (b) the remaining physical, geological, and/or botanical evidences that show the risen water level in some pre-historical time.

Mauro Naghettini [11] explaineda comprehensive discussion of this method. Anyhow, in this case the non-systematic data are not available in this study.

Bayesian models are established based on three parts:first priories including the background knowledge of researcher,second, likelihood functionwhich is obtained usingobserved floods and third, marginal function. A LN4 upper bond function is usedas prior function. The estimation of its parameters is explained later.

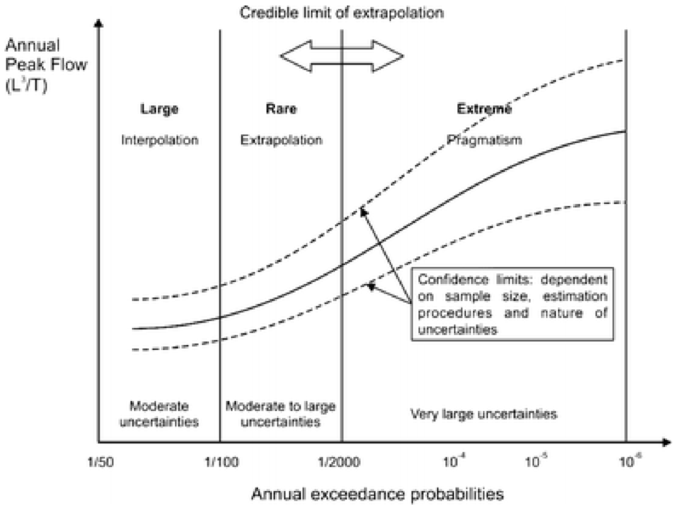


Figure 1. Flood categories (adapted from Nathan and Weinmann 2001)

2. METHODOLOGY

Some important assumptions had to be made; (a) despite the fact that flood is non-stationary, the series assumed as stationary homogeneity; (b) non-systematic data is not available for this study; (c) the upper bond LN4 is the best fitted one, considering the available data; (d) the short data (only 70 years) is used to model an event with T>1000 years. These assumptions caused an uncertainty to the results. Some of them can be resolved to some extent in further studies but some is inevitable because of the nature of random process.

Flood estimation is not deterministic and cannot produce any firm result. Only the statistical analysis, either frequency analysis or Bayesian, can be used to estimate flood confidential or credential interval. The result is not unique and it is strongly dependent on availability and accuracy of data. Any bias in input data will result in more uncertainty in output estimation.

In a brief and short form, the Bayesian Model can be explained as Equation 1.

$$\pi(\theta|x,\Omega) = \frac{f(x|\theta,\Omega)\pi(\theta|\Omega)}{f(x|\Omega)} \text{Or} Posterior\ PDF = \frac{Likeliy\ hood\ \ Function\ \ \times Prior\ \ PDF}{Marginal\ \ Probablity} \quad (1)$$

Where, θ, x and Ω are the parameter,observation, and prior data about the parameter θ respectively. Further information can be found in chapter 11 of Fundamentals of Statistical Hydrology book, by Ma Mauro Naghettini[11]. In fact, this equation for hydrological event and in particular for flood cannot be solved analytically in most cases. So, numerical estimation is needed. One of the common solutions is Markov chain Monte Carlo (MCMC). To solving the MCMC of Bayesian model, using Metropolis-Hastings algorithm within Gibbs sampling from probability function is recommendedas an effective and reliable method [12].

The input data then are used to develop a Bayesian model in WinBUGS(Bayesian inference Using Gibbs Sampling). It is developed in the MRC Biostatistics Unit, Cambridge University asan open source software.WinBUGSis used for the Bayesian analysis of complex statistical models using Markov chain Monte Carlo (MCMC) methods [13].

The upper bound Log Normal Distribution with four parameters (LN4) is used for prior distribution. The transformation of Equation2is used to normalize X to Y. Where X, and NEX are the possible PMF in the catchment area and the number of “Annual flood-peaks with exact values” respectively.

$$\text{Transformation:} Y = \ln\left(\frac{X-\varepsilon}{\alpha-X}\right), Y\sim N(\mu_Y,\sigma_Y) \quad (2)$$

The Probable Distribution Function (PDF) and Cumulative Distribution Function (CDF) of LN4 are formulated in Equations 3 and 4 respectively:

$$\text{PFD: } f_x(x|\Theta) = \frac{\alpha}{x(\alpha-x)\sigma_Y\sqrt{2\pi}} e^{-\frac{1}{2\sigma_Y^2}[\ln\left(\frac{x}{\alpha-X}\right)-\mu_Y]^2},\ 0\leq x\leq\alpha \quad (3)$$

CDF: $F_x(x|\Theta) = \phi[\frac{1}{\sigma_Y}\ln\left(\frac{x}{\alpha-x}\right) - \frac{\mu_Y}{\sigma_Y}]$ (4)

Where, Φ isthe notion for CDF of Normal Distribution. In addition, $\alpha \in \mathcal{R}_+$ is the upper bond of LN4 and analogouslyε $\in \mathcal{R}_+$ is the lower bond of LN4.

There are some hyper-parameters: $\rho_\alpha, \beta_\alpha, \rho_\sigma, \beta_\sigma, \mu_\mu, \sigma_\mu$. Takara and Tosa (1999) reported that there would be someeffects of a zero lower bound Ɛ=0 on the LN4 goodness-of-fit and upper tail behavior[14].

$$Pr(\alpha \leq PMF_{local} \,|\rho_\alpha, \hat{\beta}_\alpha) = 0.5 \quad (5)$$

To assess the accuracy of the posterior estimations, two criteria isused. The first criterion is the Coefficient of Variation (CV) which can be calculated by dividing variance by mean. The second one is MC_error, which means that the Monte Carlo error for each parameter that is an estimation of the difference between the mean of the sampled values (posterior mean) and the true posterior mean using the metropolis algorithm. The poorer the convergence, the higher the Monte Carlo error [15].

3. CASE STUDY

Latian Dam is located 25 km away from Tehran, the capital of Iran, constructed in 1968. It plays a very important role for providing potable water for the capital city. The main water resource for the dam is JajroodRiver. This river springsfrom AlborzMountain and flows to the Namak Lake in the center of Iran. In this catchment, there are 16 meteorological stations.Latian Dam has two hydropower plants;the capacity of each one isabout 22.5MW.Watershed code in “wrs.wrm.ir” [16] system is 41-36 and its area is 987 Km^2. Annual input rainfall is about 334MM^3, and output is about 307 MM^3, from which 171 $MM^3$of it is consumed for potable water of Tehran.The reservoir capacity is 95MM^3, and the height of the dam crest from foundation is 107m[17].

The hydrometric station (Code 41-119) near the dam, is located at E=51-41-00, N=35-47-00, Z=1560m. The measured discharge meanis about 2.87 m3/s, with the mean of annual maximum flows of 2.71. This is shown in Table 1, which clearly indicates a bias in the input data.

Figure 2.Site Location of Latian Dam

The data used in this study nelongs to Latian Dam; located on Jajrood river basin. Due to 70 years long record, it forms a suitable likelihood function. “Iran Water Resource Management Co” has provided this record [16], from 1946 to 2016, at Latian Hydrometric station coded 41-119, as shown in Figure 3.

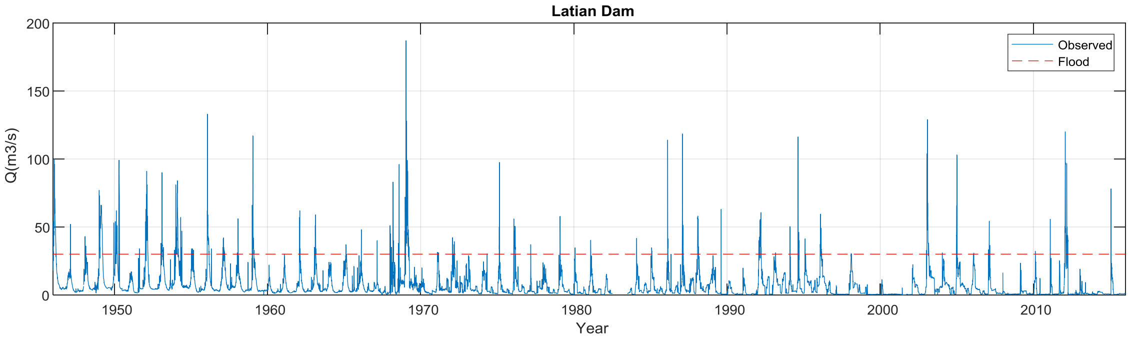


Figure 3.The record of discharge at hydrometric Station

The classical definition of flood is that when the discharge fills the whole stream and starts to flow on the floodplain. However, in this study, for practical engineering reason, the return period of 25 years is selected as threshold. First, all discharge history record is given to a Matlab code to sort the time history based on the magnitude. The frequency analysis is done by assigning the relevant rank as Weibullrankings to each discharge Figure4 and 5. Considering25-year return period (T) threshold, or Exceedance probabilities Pr=0.04, resulted in flood with 30 m^3/s flow rate as shown in Figure 3.

During 70 years,1006 days have been recorded with flood (i.e. $Q>30m^3/s$), amongst those 692 days with return period larger than 25 years and less than 80 years. Meanwhile, there is only 314observationswith T>80 years or Pr<0.0125. Due to rare observations, the flood should be estimated using extreme method of statistical interferences.

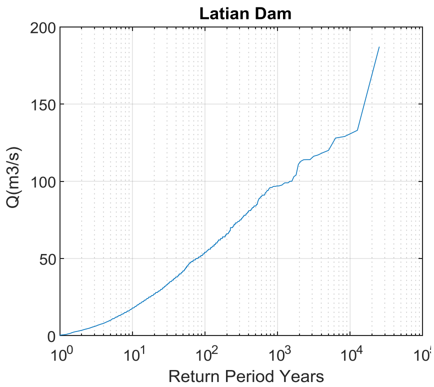


Figure 4. Frequency Analysis of the Discharge

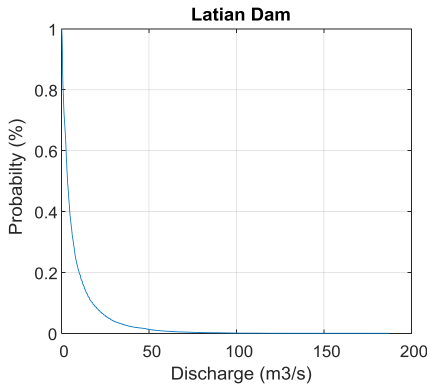


Figure 5. Flow-Duration-Curve (FDC)

Accordingly, the Flow-Duration-Curve (FDC) presented in Fig 5, the daily peak flow of each flood is extracted and saved as “Peak Latian Dam” as the first scenario for the next step; including 139 items, see Fig 6.

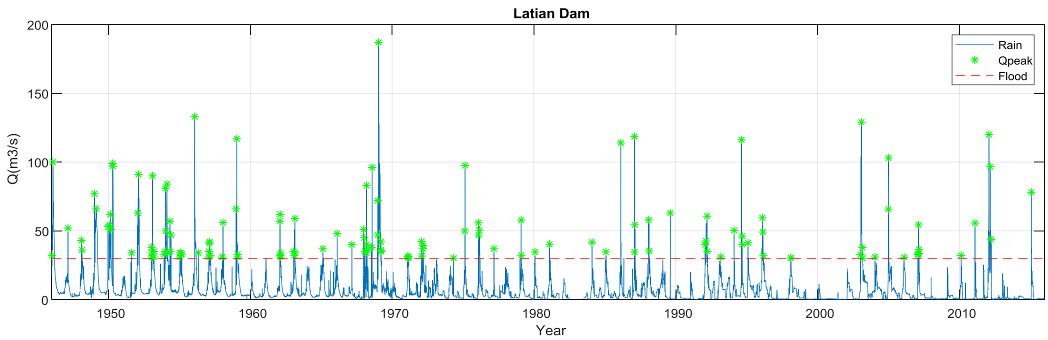


Figure 6.Site Location of Latian Dam

The second scenario includes the set annual maximum discharge in hydrometric station, called “Annual Latian Dam”. It comprises of 70 item,some of them are less than the threshold as illustrated in Figure 7.

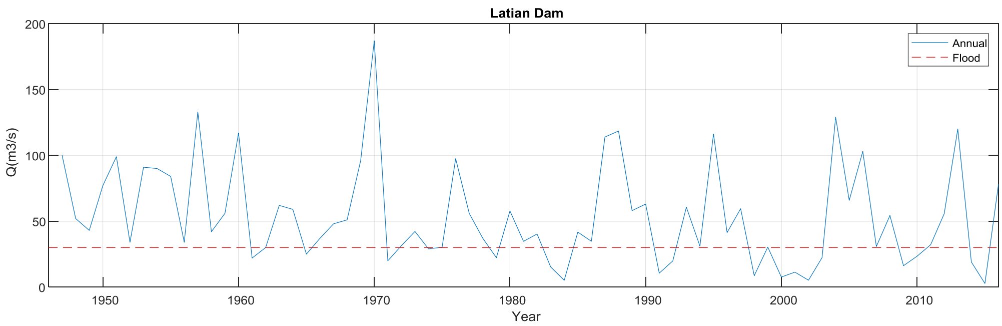


Figure 7.Annual maximum Discharge

4. Bayesian Model

Two selected scenarios (Peak and Annal) of Latian Dam are compared to the American river basin at the Folsom dam site done by Fernandes (2010) did [5]. The input data of Folsom Dam is reported by Ma Mauro Naghettini, [11].Refer to Table 1; it is obvious that CV and variance of Peak series are smaller than Annual’s on, but the skew of Peak is greater than Annuals’.

Table 1- Input Data (X)Comparison

Dam	No of Observation	Max Q	Mean Q	Min Q	Variance	CV= Var./Mean	Skew
Folsom [5]	52	8,438	2,259.423	280	4,386,216	1,941	1.431
Latian (Annual)	70	187	53.887	2.71	1,433	26	1.078
Latian (Peak)	139	187	51.241	30.2	733	14	2.059

The transformation of Equation 2normalize X input data to Y, Table 2. This transformation helps to reduce the CV and variance of data.

Table 2- Transformed Data (Y) Comparison

Dam	α (PMF)	No	Max	Mean	Min	Variance	CV= Var./Mean	Skew
Folsom [3]	25,655	52	-0.713	-2.725	-4.507	0.984	-0.361	0.263
Latian (Annual)	3,000	70	-2.711	-4.288	-7.009	0.737	-0.172	-0.833
Latian (Peak)	3,000	139	-2.711	-4.153	-4.588	0.183	-0.044	1.134

The available knowledge and prior data about the scenarios are summarized in the Table 3. It is entered in Bayesian model.

Table 3- LN4 model as prior distribution for Bayesian Analysis

parameter	Distribution of parameter	Description	Selected priors	Remarks
α	$\alpha \sim Gamma(\rho_\alpha, \beta_\alpha)$	Upper bound of the LN4 distribution	Table 4	Marginal Distribution
ε	-	Lower bound of the LN4 distribution	Zero	[6]
σ_Y	$\sigma_Y \sim Gamma(\rho_\sigma, \beta_\sigma)$	Scale parameter of the LN4 distribution	$\sigma_Y \sim Gamma(1.0, \frac{1}{10^7})$	non-informative prior distributions
μ_Y	$\mu_Y \sim Normal(\mu_\mu, \frac{1}{\sigma_\mu})$	Location parameter of the LN4 distribution	$\mu_Y \sim Normal(1.0, \frac{1}{10^5})$	non-informative prior distributions

Where,ρ_α and $\beta_\alpha$are the parameters of Gamma PDF. The method of momentsyieldsto $\hat{\rho}_\alpha = CV_\alpha^{-2}$. In this study,$CV_\alpha$ is discrete with values of 0.1, 0.2, 0.3, 0.4 and 0.5;ρ_α and $\beta_\alpha$are calculated accordingly. The existing spillways forthe Dam are sized to the capacity of **1930** m3/s [18]. So, the initial value for PMF is considered to be**3'000** m^3/s (about 50% more than the existing PMF). Table 4demonstrates results for discrete value of CV_α using Equation 5.

Table 4- The Hyper Parameter for α

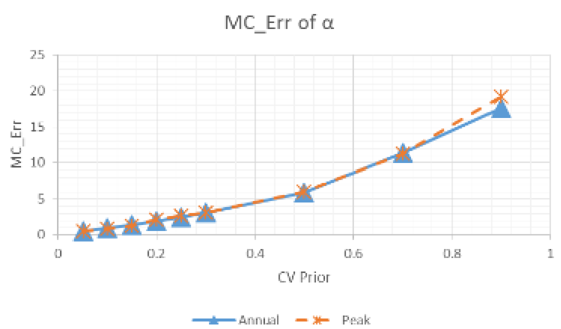
CV_α	Pr. γ	ρ_α	β_α	$1/\beta_\alpha$	Mean	Median
0.05	0.50	400	7.506	1.332E-01	3,003	3,000
0.10	0.50	100.00	30.100	3.322E-02	3,010	3,000
0.15	0.50	44.444	68.009	1.47039E-02	3,023	3,000
0.20	0.50	25.00	121.618	8.222E-03	3,040	3,000
0.25	0.50	16	191.474	5.22264E-03	3,023	3,000
0.30	0.50	11.11	278.303	3.593E-03	3,092	3,000
0.50	0.50	4.00	816.980	1.224E-03	3,268	3,000
0.70	0.50	2.04	1,745.307	5.730E-04	3,562	2,998
0.90	0.50	1.23	3,254.522	3.073E-04	4,018	2,989

Where, for the gamma distribution, mean is$\rho_\alpha \times \beta_\alpha$ and median is about Mean × ($3\rho_\alpha$-0.8) / ($3\rho_\alpha$+0.2).

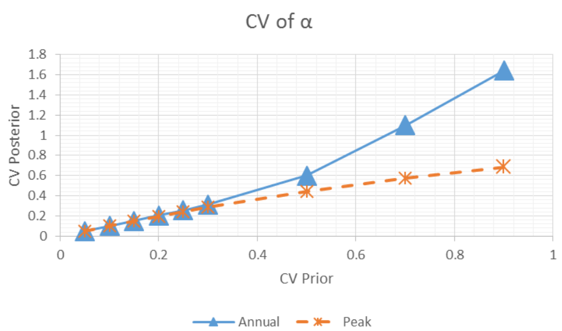
5. Results

Two scenarios (Annual and Peak) have been studied. Based on the results, it is proven that Latian Peak scenario is not sensitive to the changes of σ_Y, μ_Y: the set of (10^{-8},10^{-6}) or (10^{-7},10^{-6}) or (10^{-7},10^{-5}). However,in the second scenario, Latian Annual shows a little improvement in this order: of (10^{-8},10^{-6}) or (10^{-7},10^{-6}) or (10^{-7},10^{-5}). Therefore, the distribution of σ_Y, $\mu_Y$is selected as(10^{-7},10^{-5})to run the final models.

Both scenarios are shownad compared in Figure 8 (a,b). The MC_Error for both scenarios are similar and the lower prior CV gives the lower Error. Nonetheless, the Peak Flow scenario is more accurate than the Annual. Analogously,in both scenarios, smaller prior CV results in smaller posterior CV.



(a) MC_Error Criterion

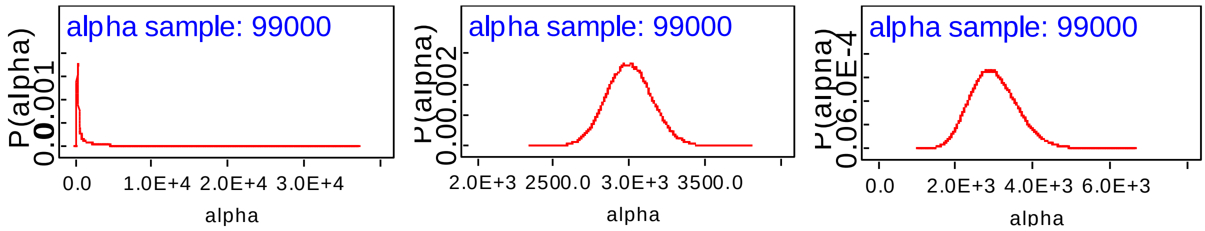


(b) Posterior CV Criterion

Figure 8.Comparison of the accuracy of posterior α

Running the models leads to the posterior joint distribution of the LN4 parameters: $\alpha, \sigma_Y$and μ_Y. Lunn et al. (2000) stated that WinBUGSor OpenBugsmakes practical MCMC methods andMetropolis–Hastings algorithm generate the samples from the posterior joint distributionof the LN4 parameters [19]. In this study, WinBUGS sampled 1,000,000 times from the posterior joint distribution of parameters, lagged by 10 samples to avoid correlation. First 10,000 samples as 'burn in' were discarded and the 99,000 samples left were used to build the posterior data set.

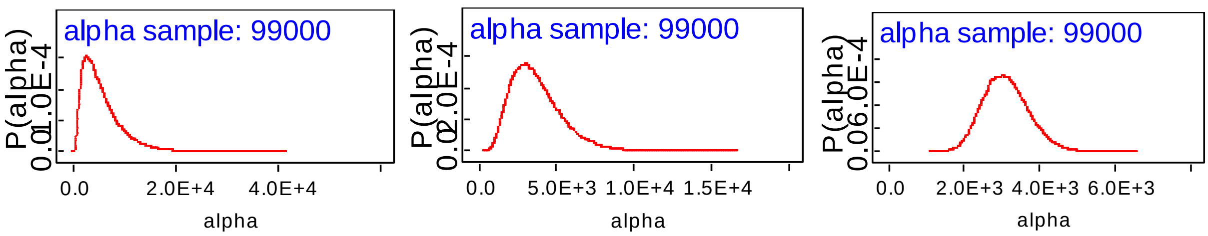
The shape of posterior probability distribution function varies according to different prior CVs. As shown in Figure 9, the shape of Annual scenario changes from sharp Gamma type to Normal, and to shifted bell shape.



(a) CV=0.7, Annual **(b) CV=0.5, Annual** **(c) CV=0.2, Annual**

Figure 9.Posterior Distribution of α, Annual Scenario

Similarly, the shape of Peak Flow scenario is gradually changing from sharp Gamma to bell type, but it has a smoother distribution as shown in Figure 10.

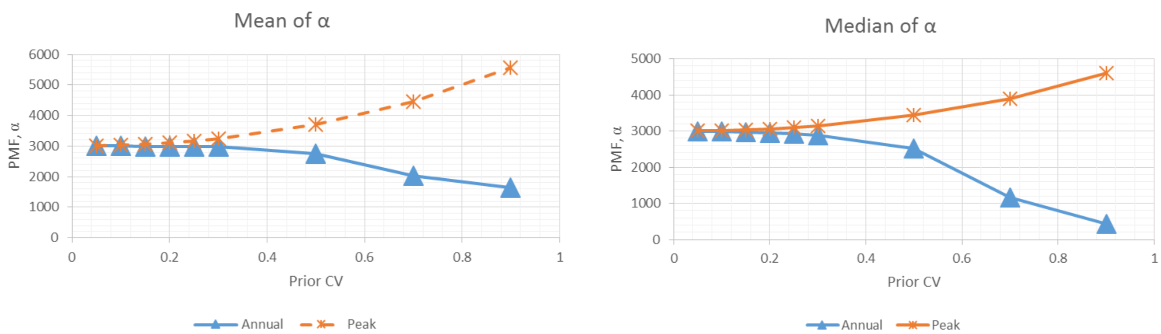


(a) CV=0.7, Peak **(b) CV=0.5, Peak** **(c) CV=0.2, Peak**

Figure 10.Posterior Distribution of α, Peak Scenario

According to Figure 11, having a suitable estimation of α and a lower amount of CV, the results gotfrom two scenarios are very close. However, forhigher CV values, peak flow scenario ends up with a larger amount of flow. Conversely, the annual maximum scenario tends to lower amount of flow.

It can be concludedthat the annual maximum series underestimatesthe PMF, yet the peak flow overestimates the magnitude of PMF, Table 5.



(a) Mean of Estimated α (or PMF) **(b) Median of Estimated α (or PMF)**

Figure 11.Comparison of estimated posterior α

Furthermore, without non-systematic data, the estimation of Bayesian Model will lead to the same amount of PMF in both scenarios. In the future, with non-systematic data, Bayesian Model can be run and the outcomes can be compared. According to research conducted byFernandes [5], this can improve the results.

Table 5- The result of estimation α, prior CV=0.1

	scenario	mean	sd	MC_error	val5.0pc	median	val95.0pc	start	sample
α	Peak	3025	299.3	0.9041	2550	3015	3535	10000	99000
	Annual	2999	301.7	0.9569	2519	2989	3513	10000	99000
μ_Y	Peak	-4.156	0.1073	0.0003234	-4.33	-4.157	-3.977	10000	99000
	Annual	-4.282	0.147	0.0004661	-4.522	-4.283	-4.039	10000	99000
σ_Y	Peak	0.4312	0.0263	0.00009196	0.3905	0.4298	0.4765	10000	99000
	Annual	0.8747	0.07661	0.0002189	0.7584	0.8693	1.009	10000	99000

CONCLUSIONS

This study shows that the estimated PMF by annual maximum flow underestimates the PMF, although the Peak flow overestimates the PMF. Bayesian Model can reduce the difference between the results of the two scenarios by deploying an effective Likelihood function. The upper bond parameter of LN4 distribution function estimates the PMF properly with a CV of less than 0.1. OpenBUGS software generates the sampling by Gibs method and estimates the posterior function by Metropolis–Hastings algorithm.

To improve the estimation of PMF, the nonsystematic floodsshould be added. The nonsystematic floods can be extracted from the responsible sources as the needed input into the Bayesian models.

REFERENCES

1. ICOLD—International Commission of Large Dams (1987) Bulletin 59—dam safety guidelines, Paris, France
2. FEMA—Federal Emergency Management Agency (2004) Federalguidelines for dam safety: selecting and accommodating inflowdesign floods for dams. The Interagency Committee of the U. S.Department of Homeland Security, Washington, DC
3. USNRC—U.S. Nuclear Regulatory Commission (1977) Design BasisFloods for Nuclear Power Plants. Regulatory Guide 1.59,Washington, DC
4. WMO—World Meteorological Organization (1986) Manual for estimation of probable maximum precipitation, 269 pp. Operational Hydrology Report 1, WMO 332, Geneva, Switzerland
5. Fernandes, Wilson &Naghettini, Mauro &Loschi, Rosângela. (2010), "A Bayesian approach for estimating extreme flood probabilities with upper-bounded distribution functions", Stochastic Environmental Research and Risk Assessment. 24. 1127-1143. 10.1007/s00477-010-0365-4.
6. Dawdy DR, Lettenmaier DP (1987) Initiative for risk-based flood design. J HydrolEng 113(8):1041–1051
7. Stedinger JR, Heath DC, Thompson K (1989) Risk analysis for damsafety evaluation. Report TCN 87-653 prep. for the U. S. ArmyCorps of Engineers
8. Dubler JR, Grigg NS (1996) Dam safety policy for spillway designfloods. J Prof Issues EngEducPract 122(4):163–169
9. USBR—United States Bureau of Reclamation (2002) Flood hazardanalysis, Folsom Dam Central Valley Project. Flood HydrologyGroup, Denver, CO
10. Nathan RJ, Weinmann PE (2001), "Estimation of large and extreme floods for medium and large catchments: book VI. Australian rainfall and runoff—a guide to flood estimation", 4th ed. Engineers Australia, National Committee for Water Engineering, Australia.
11. Mauro Naghettini (2017), "Fundamentals of Statistical Hydrology", Cham, Switzerland : Springer.
12. Geweke, John &Tanizaki, Hisashi. (2001). Bayesian estimation of state space models using Metropolis-Hastings algorithm within Gibbs sampling. Computational Statistics & Data Analysis. 37. 151-170. 10.1016/S0167-9473(01)00009-3.
13. WinBUGS official site: https://www.mrc-bsu.cam.ac.uk/software/bugs/
14. Takara K, Tosa K (1999), "Storm and flood frequency analysis using PMP/PMF estimates. Proc. International symposium on floods and droughts, Nanjing, China", pp 7–17, IHP-V Technical Documents in Hydrology No. 4, UNESCO Jakarta Office, Hohai University Press.
15. The tutorial, manuals of software: OpenBUGS,http://www.openbugs.net/Manuals/Tutorial.html
16. Iran Water Resources Management Company-Office of Basic Water Resources Studies, (in Persian), http://wrbs.wrm.ir/.
17. Iran Water Statistical Yearbook, (2014-2015), http://isn.moe.gov.ir/getattachment/a5d92d49-dcce-466e-ad35-818090c3f98b/Iran-Water-Statistical-yearbook-2014-2015(1393-94)
18. Regional Water Company of Tehran, Dam under operation, LatianDam, Construction Specifications,http://www.thrw.ir/SC.php?type=static&id=66
19. Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000) WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. Stat Comput 10:325–337