Bivariate Flood Frequency Analysis using Copula with Parametric and Nonparametric Marginals

Subhankar Karmakar

Slobodan P. Simonovic

The University of Western Ontario

The Institute for Catastrophic Loss Reduction

Western

Presentation outline

- Introduction
- Objectives of the study
- Study area
  - Data and flood characteristics
- Marginal distributions of flood characteristics
  - Parametric and nonparametric estimation
- Joint and conditional distributions using copula
- Conclusions

Introduction

- Flood management (design, planning, operations) requires knowledge of flood event characteristics

Flood characteristics

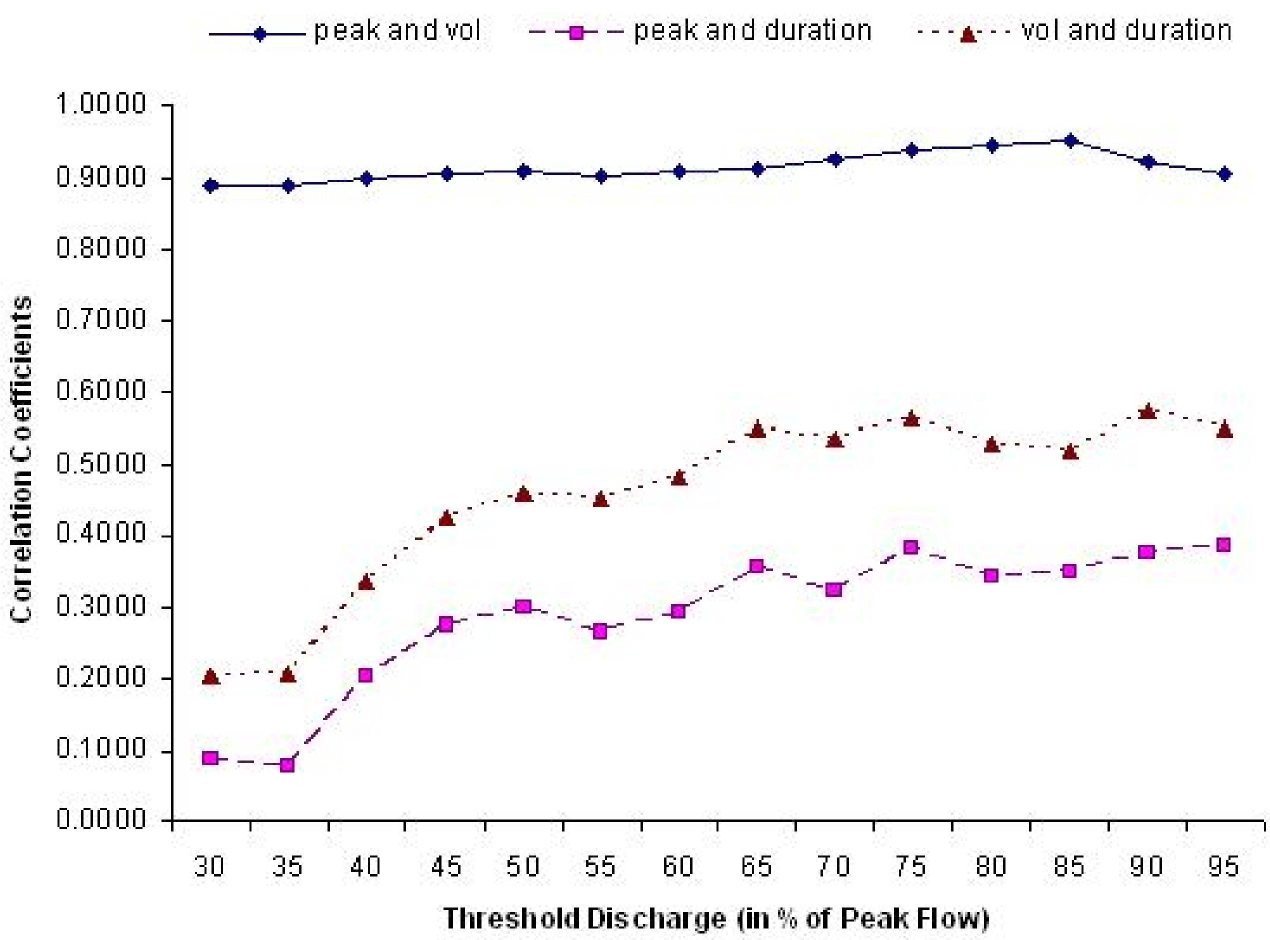
Peak flow

Volume

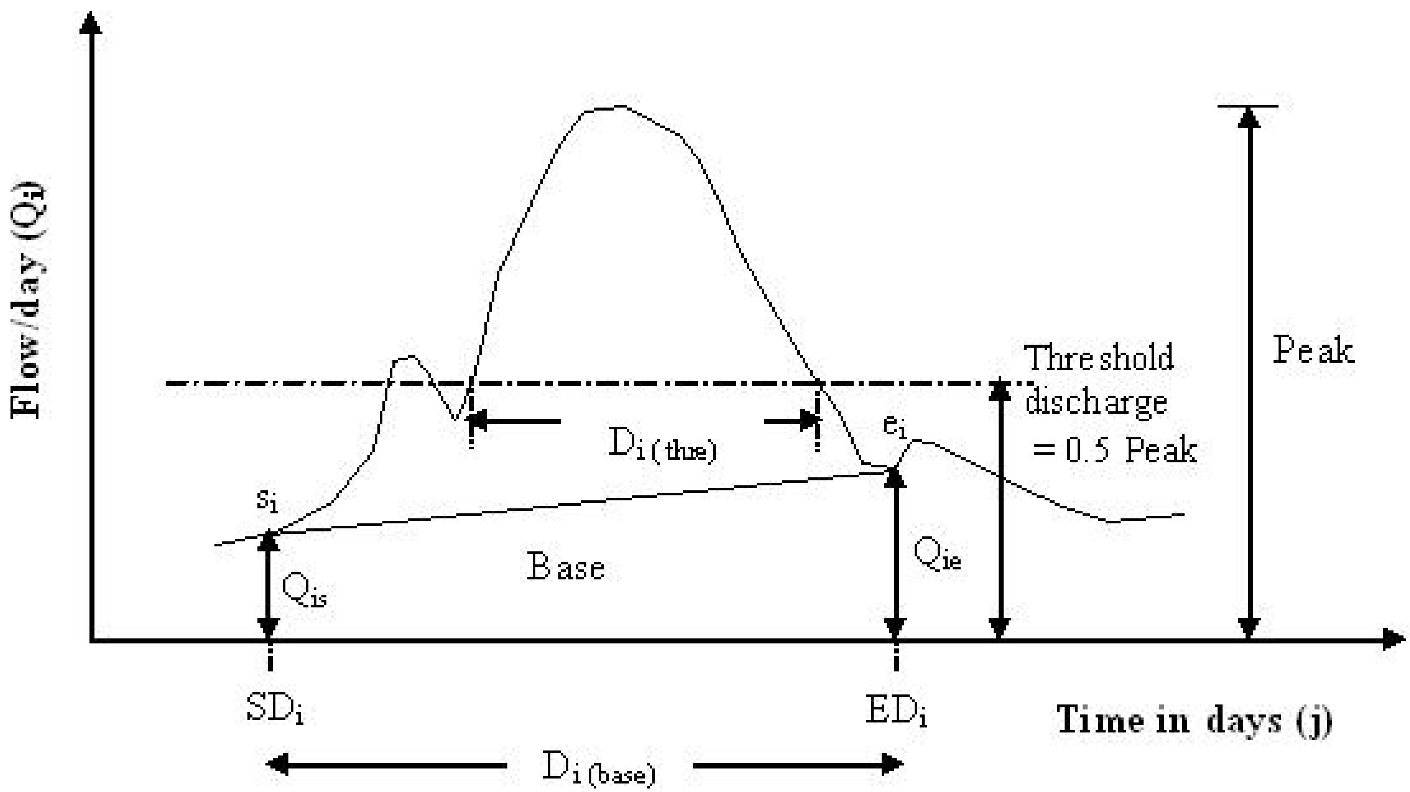
Duration

Western

Introduction



Introduction

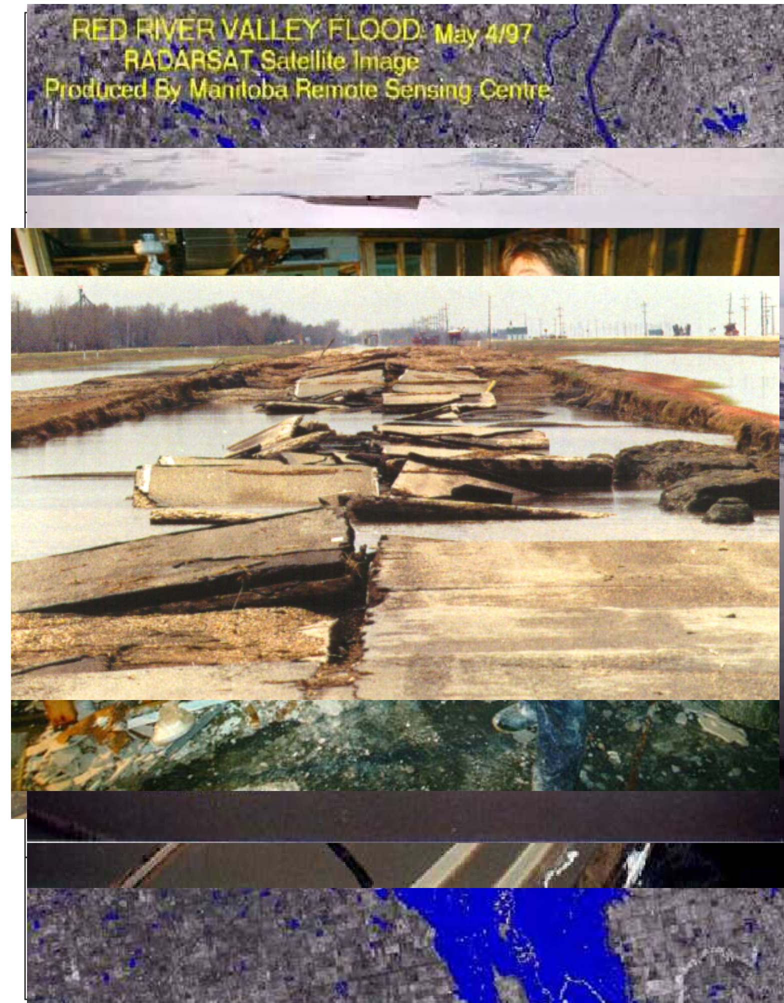


Study objectives

- To determine appropriate marginal distributions for peak flow, volume and duration using parametric and nonparametric approaches
  - To define marginal distribution using orthonormal series method
- To apply the concept of copulas by selecting marginals from different families of probability density functions
- To establish joint and conditional distributions of different combinations of flood characteristics and corresponding return periods

Study area

- Red River Basin
- 116,500 km^2 (89% in USA 11% in CDN)
- Flooding in the basin is natural phenomena
- Historical floods: 1826; 1950; 1997
- Size of the basin and flow direction
- No single solution to the flood mitigation challenge



Study area - data

- Daily streamflow data for 70 years (1936-2005)
- Gauging station (05082500) - Grand Forks, North Dakota, US
  - Location - latitude 47°55'37"N and longitude 97°01'44"W
  - Drainage area - 30,100 square miles
  - Contributing area - 26,300 square miles
- http://waterdata.usgs.gov

Study area – flood characteristics

- Dependence between P, V and D

Flood Characteristics	Pearson's Linear Correlation Coefficient	Kendall's Coefficient of Correlation	Spearman's rho Correlation Coefficient
Peak Flow-Volume	0.9359	0.7892	0.9150
Volume-Duration	0.6934	0.5756	0.7313
Peak Flow-Duration	0.5306	0.4033	0.5182

- P and V highly correlated
- All the correlations positive

Marginals - parametric

PDF		Parameters: Peak Flow (P)	Parameters: Volume (V)	Parameters: Duration (D)
Exponential: $f_X(x) = \frac{1}{\eta} e^{-x/\eta}; x > 0$	η	0.0159	0.0013	0.0245
Gamma: $f_X(x) = \frac{\lambda^\beta x^{\beta-1} e^{-\lambda x}}{\Gamma(\beta)}; x \geq 0, \lambda > 0, \beta > 0$	λ	39.956	885.81	5.6490
and $\Gamma(\beta) = \int_0^\infty u^{\beta-1} e^{-u} du$	β	1.5787	0.8921	7.2251
Gumbel or EV1: $f_X(x) = \frac{1}{\alpha} \exp[-\frac{x-\varepsilon}{\alpha} - \exp(-\frac{x-\varepsilon}{\alpha})]$	ε	40.485	413.69	33.981
	α	39.144	652.34	11.839
Lognormal: $f(x) = \frac{1}{x\sqrt{2\pi}\sigma_y} \exp\left(-\frac{(y-\mu_y)^2}{2\sigma_y^2}\right)$	μ_y	3.8562	6.1470	3.6416
$y = \log x, x > 0, -\infty < \mu_y < \infty, \sigma_y > 0$	σ_y	0.7971	1.0873	0.3697

Western

Marginals - nonparametric

- Nonparametric kernel estimation of flood frequency

$$\hat{f}(x) = (nh)^{-1} \sum_{l=1}^{n} K\{(x - x_l)/h\}$$

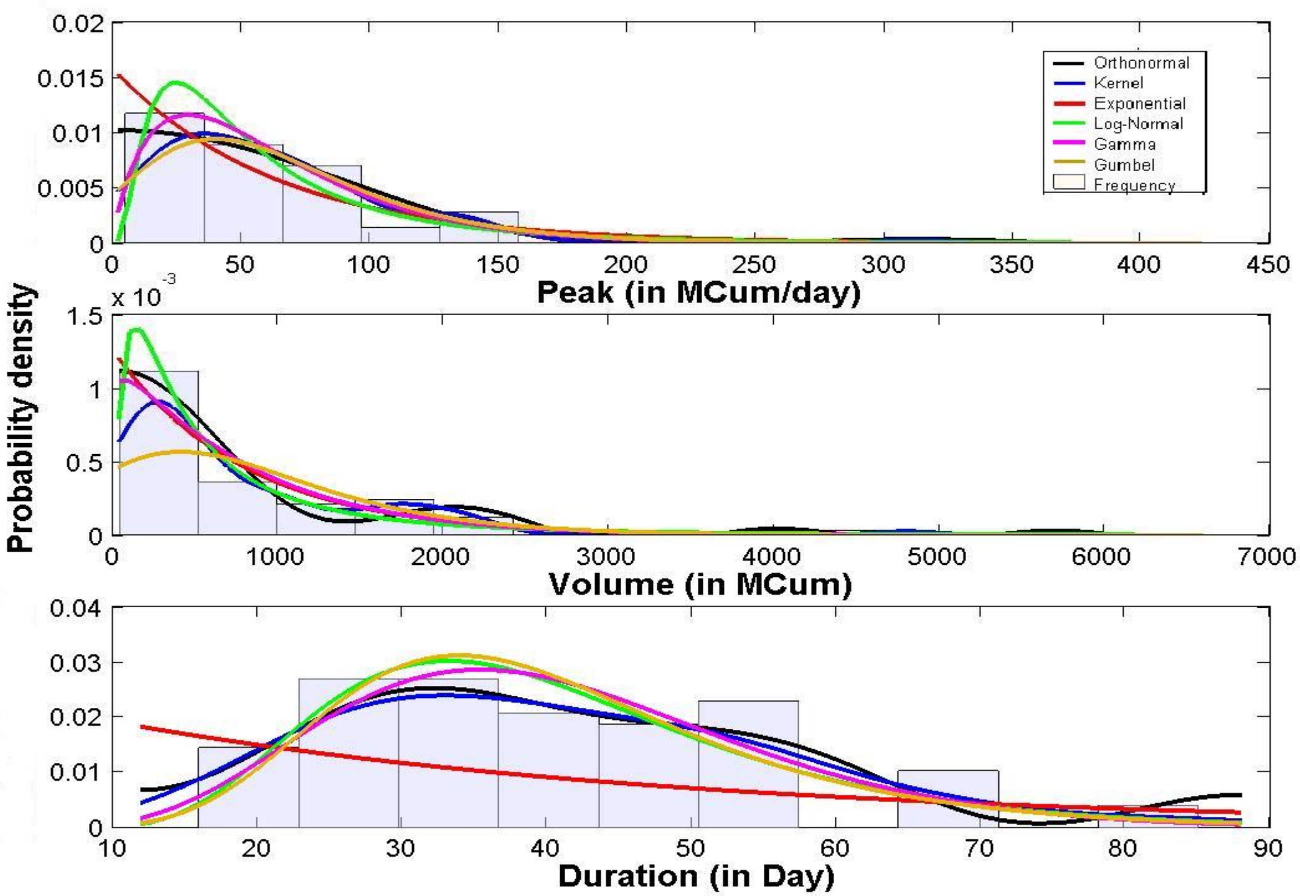
- Orthonormal series method

$$\int \Phi_s(x)\Phi_j(x)dx = 0 \quad \forall s \neq j \qquad \int \{\Phi_j(x)\}^2\, dx = 1 \quad \forall j$$

$$\Phi_0(x) = 1 \qquad\qquad \Phi_j(x) = \sqrt{2}\cos(\pi jx)$$

Western

Results


Probability density
Orthonormal
Kernel
Exponential
Log-Normal
Gamma
Gumbel
Frequency
Peak (in MCum/day)
x 10^-3
Volume (in MCum)
Duration (in Day)

Results

Distribution Function	RMSE			AIC			BIC		
	P	V	D	P	V	D	P	V	D
Kernel	0.054	0.103	0.026	-583.03	-454.18	-730.34	-583.03	-454.18	-730.34
Orthonormal	0.020	0.021	0.019	-781.24	-773.36	-788.02	-781.24	-773.36	-788.02
Exponential	0.047	0.045	0.273	-610.31	-618.72	-257.97	-607.70	-616.12	-255.37
Gamma	0.017	0.039	0.023	-813.25	-647.89	-746.61	-808.04	-642.68	-741.40
Gumbel	0.066	0.171	0.027	-540.47	-349.01	-719.57	-535.26	-343.80	-714.36
Lognormal	0.021	0.025	0.023	-771.86	-730.35	-746.27	-766.64	-725.14	-741.06

- P follows gamma distribution (parametric) and
- V and D follow distribution function obtained from orthonormal series method (nonparametric)
- Mixed marginals

Results

- Second test

Sl. No.	Marginal Distribution	χ^2 - Value	Significance Level, α	Cutoff obtained from Chi-Square Probability Table, $\chi^2_{(\alpha,k-c)}$	Conclusion
1	Peak Flow Fitted by Gamma Dist. (Parameter = 2)	0.2835	99.5%	0.989	Accepted
2	Volume Fitted by Orthonormal Series Function (Parameter = 0)	1.3790	99.5%	1.735	Accepted
3	Duration Fitted by Orthonormal Series Function (Parameter = 0)	0.0986	99.5%	1.735	Accepted

Copula

- An alternative way of modeling the correlation structure between random variables.
- They dissociate the correlation structure from the marginal distributions of the individual variables.
- n - dimensional distribution function can be written:

$$F(x_1,......,x_n) = C(F_1(x_1),......,F_n(x_n))$$

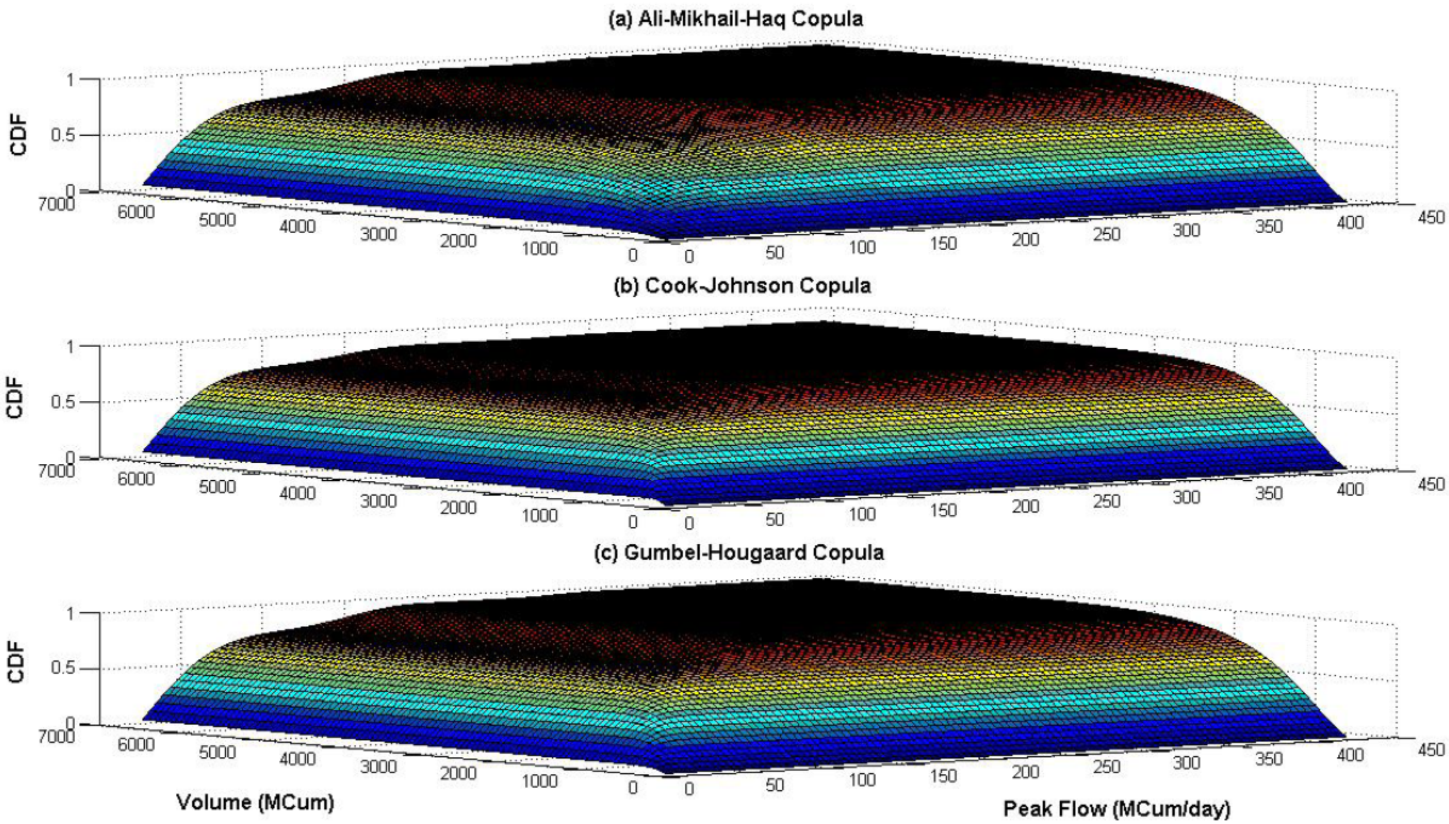
$$C(u_1,......,u_n) = F(F_1^{-1}(u_1),......,F_n^{-1}(u_n)), \quad 0 \le u_1,.....,u_n \le 1$$

Copula

Copula $[C_\theta(u_1,u_2)]$	$\theta \in$	Generating Function $\phi(t), t = u_1$ or u_2	$\tau = 1 + 4\int_0^1 \frac{\phi(t)}{\phi'(t)} dt$
Ali-Mikhail-Haq Family: $\frac{u_1 u_2}{[1-\theta(1-u_1)(1-u_2)]}$	[-1,1)	$\ln\{\frac{[1-\theta(1-t)]}{t}\}$	$[\frac{(3\theta-2)}{\theta}] - [\frac{2}{3}(1-\theta^{-1})^2 \ln(1-\theta)]$
Cook-Johnson Family: $\{\max[(u_1)^{-\theta} + (u_2)^{-\theta} - 1, 0]\}^{-1/\theta}; \theta \geq 0$	$[-1,\infty)\backslash\{0\}$	$\frac{[(t)^{-\theta} - 1]}{\theta}$	$\frac{\theta}{(\theta+2)}$
Gumbel-Hougaard Family: $\exp\{-[(-\ln u_1)^\theta + (-\ln u_2)^\theta]^{1/\theta}\}$	$[1,\infty)$	$(-\ln t)^\theta$	$(1-\theta^{-1})$

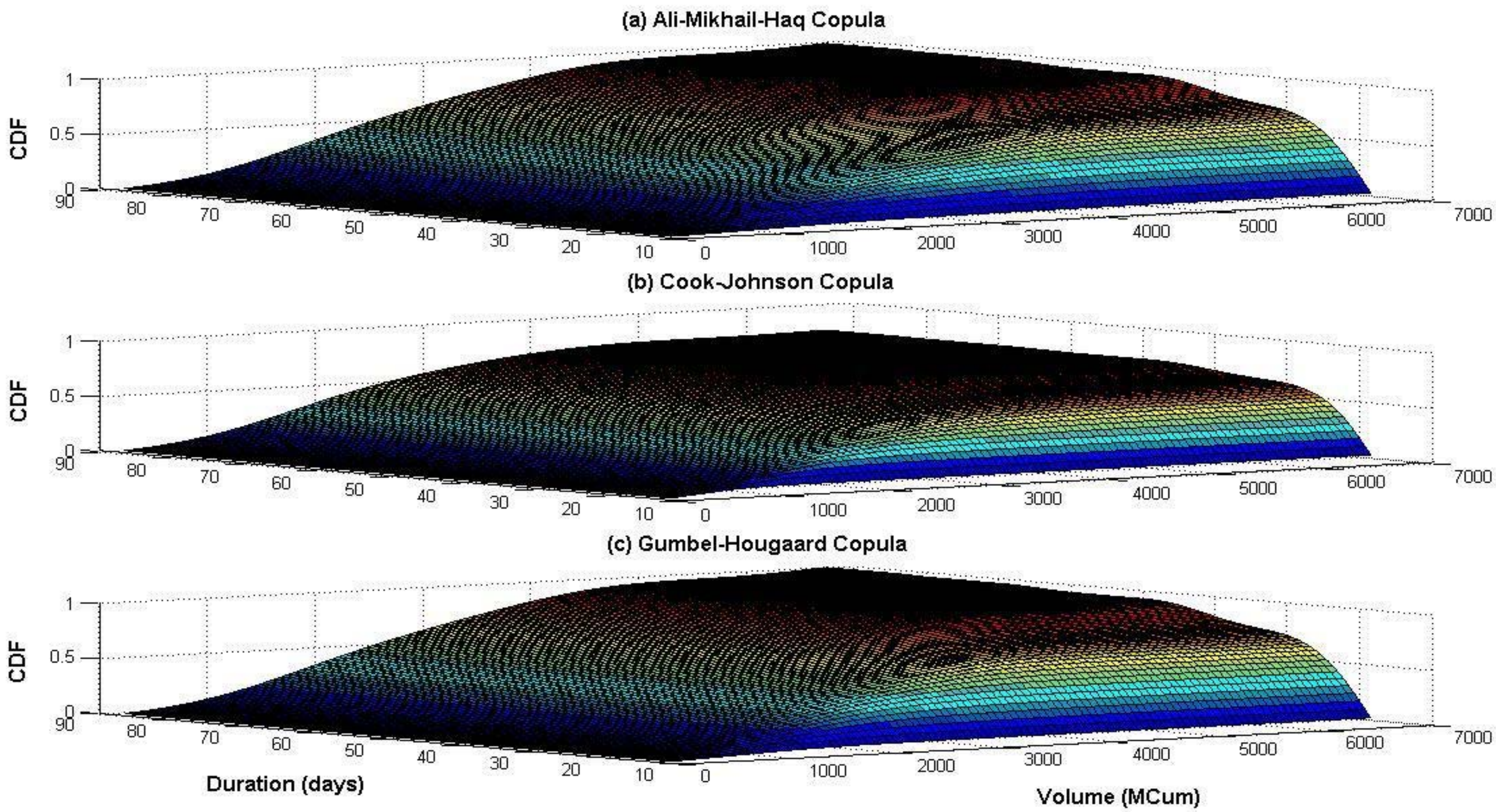
Copula	RMSE			AIC			BIC		
	P-V	V-D	P-D	P-V	V-D	P-D	P-V	V-D	P-D
Ali-Mikhail-Haq	0.141	0.090	0.056	-68.43	-84.64	-101.76	-67.54	-83.75	-100.87
Cook-Johnson	0.031	0.058	0.055	-122.80	-100.50	-102.31	-121.91	-99.61	-101.42
Gumbel-Hougaard	0.025	0.027	0.020	-130.53	-128.74	-138.33	-129.64	-127.84	-137.44

Results – joint distributions



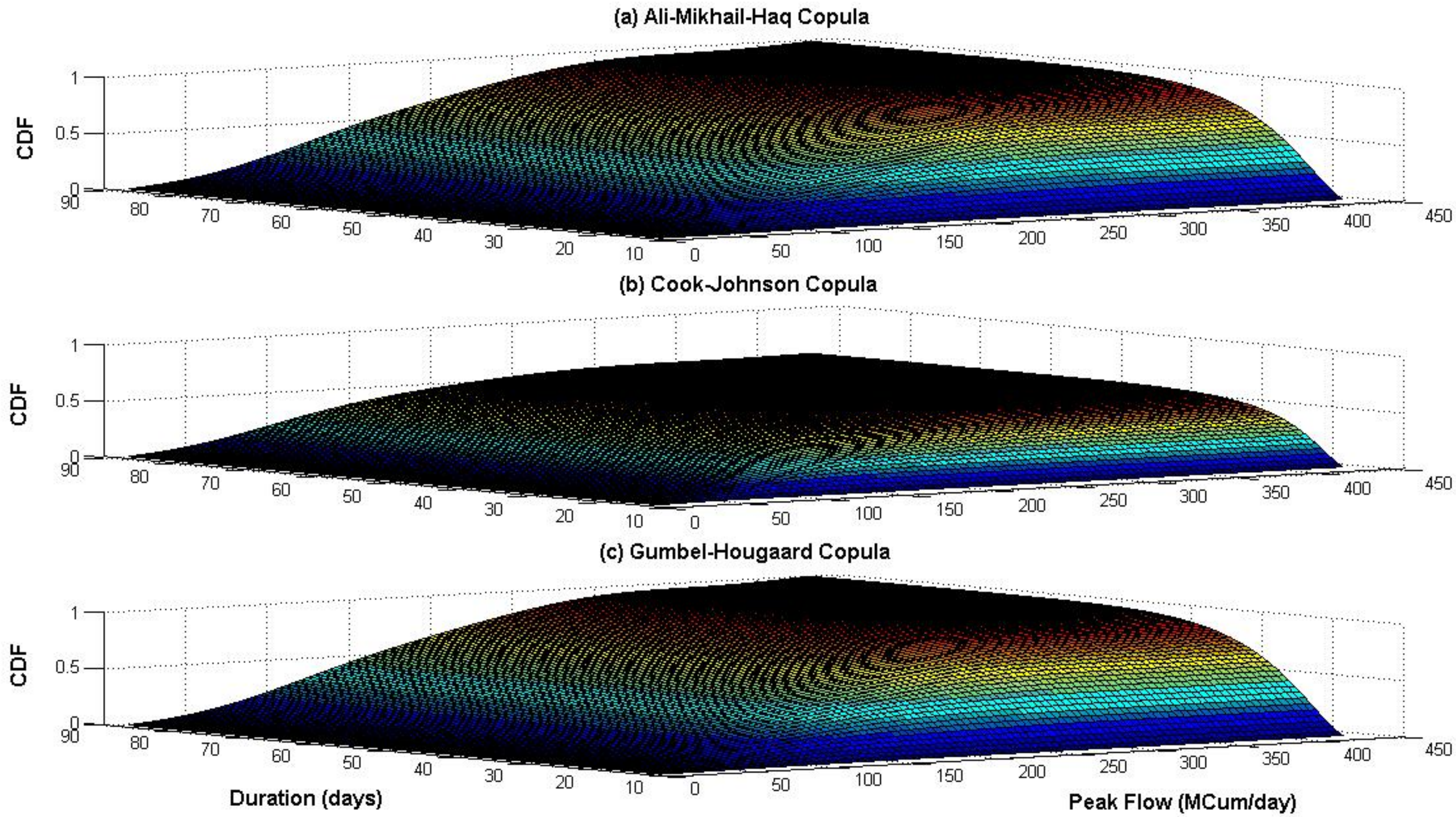
Peak flow - Volume

Results – joint distributions



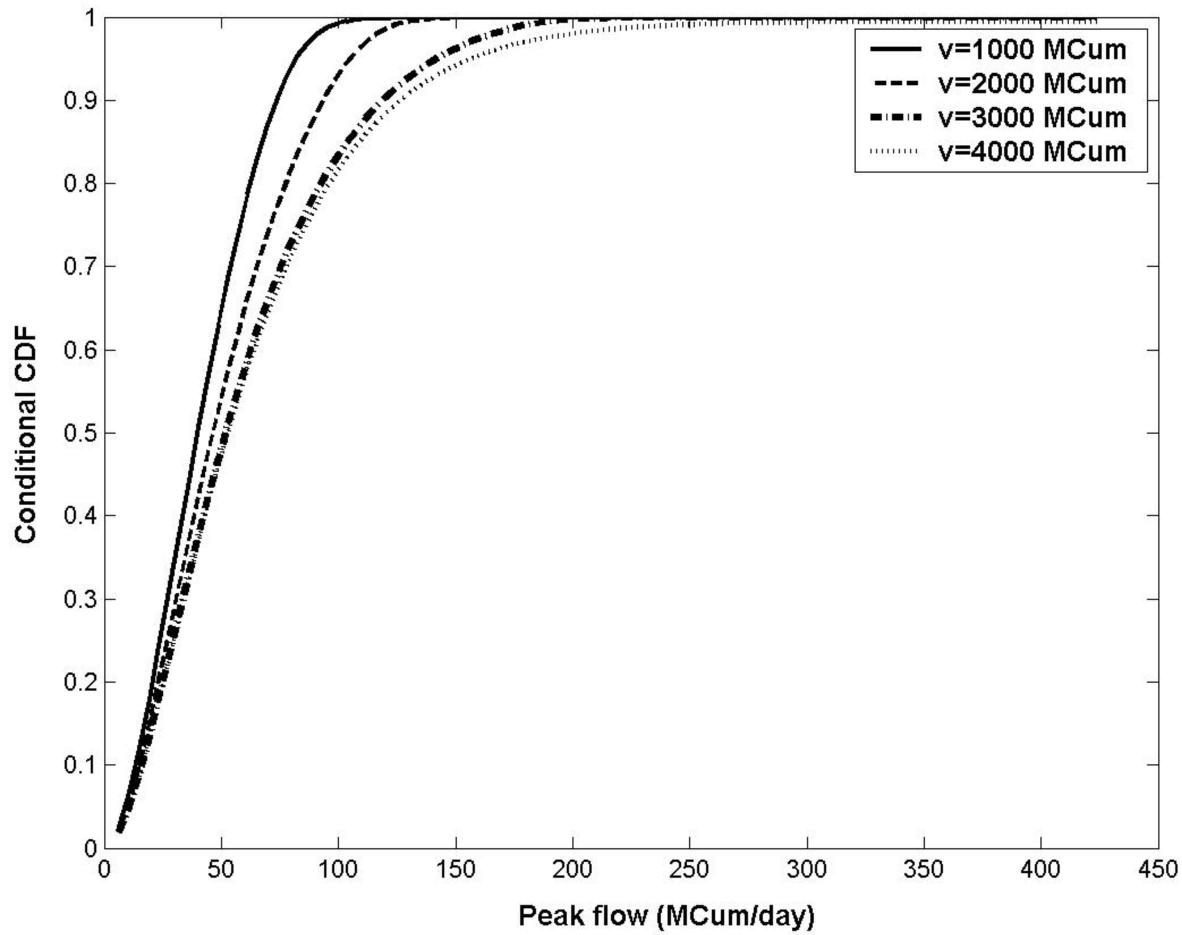
Volume - Duration

Results – joint distributions

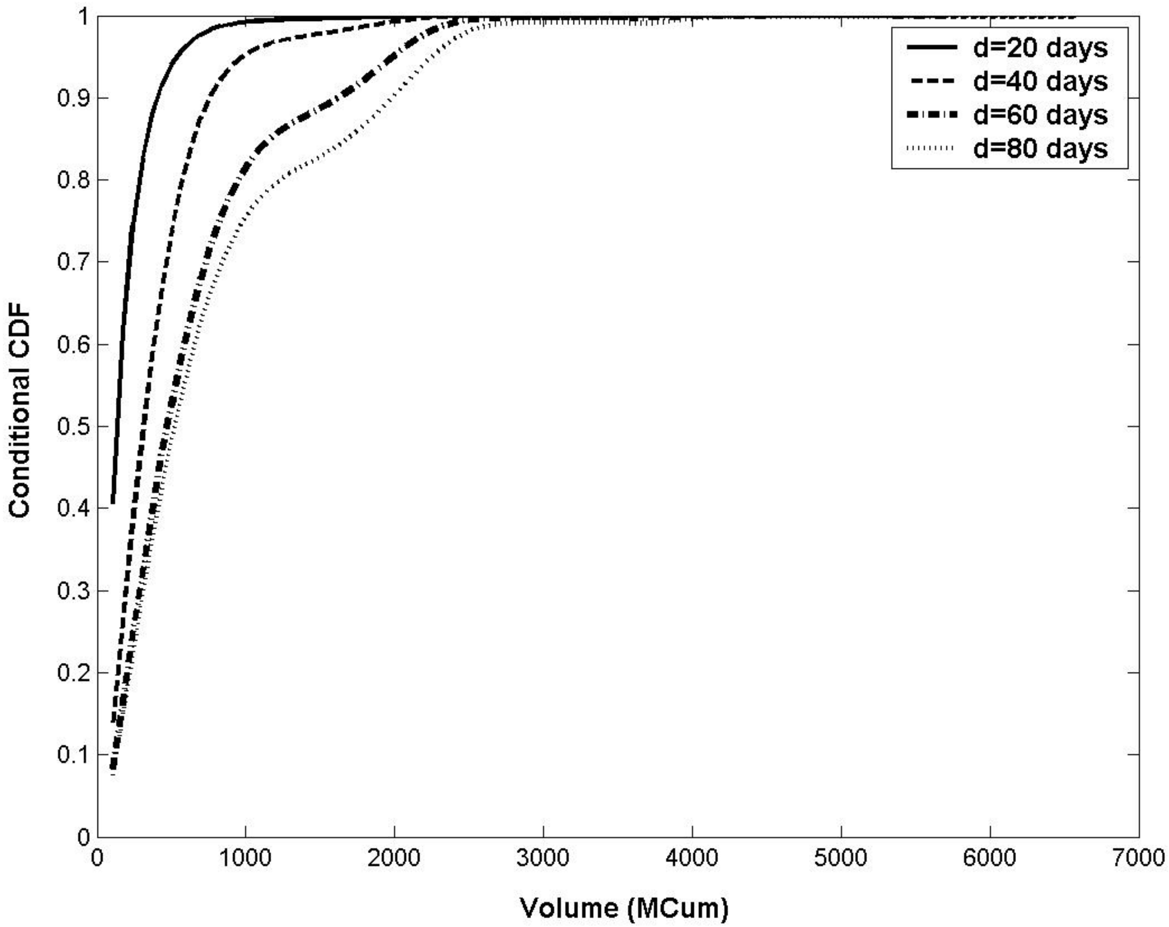


Peak flow - Duration

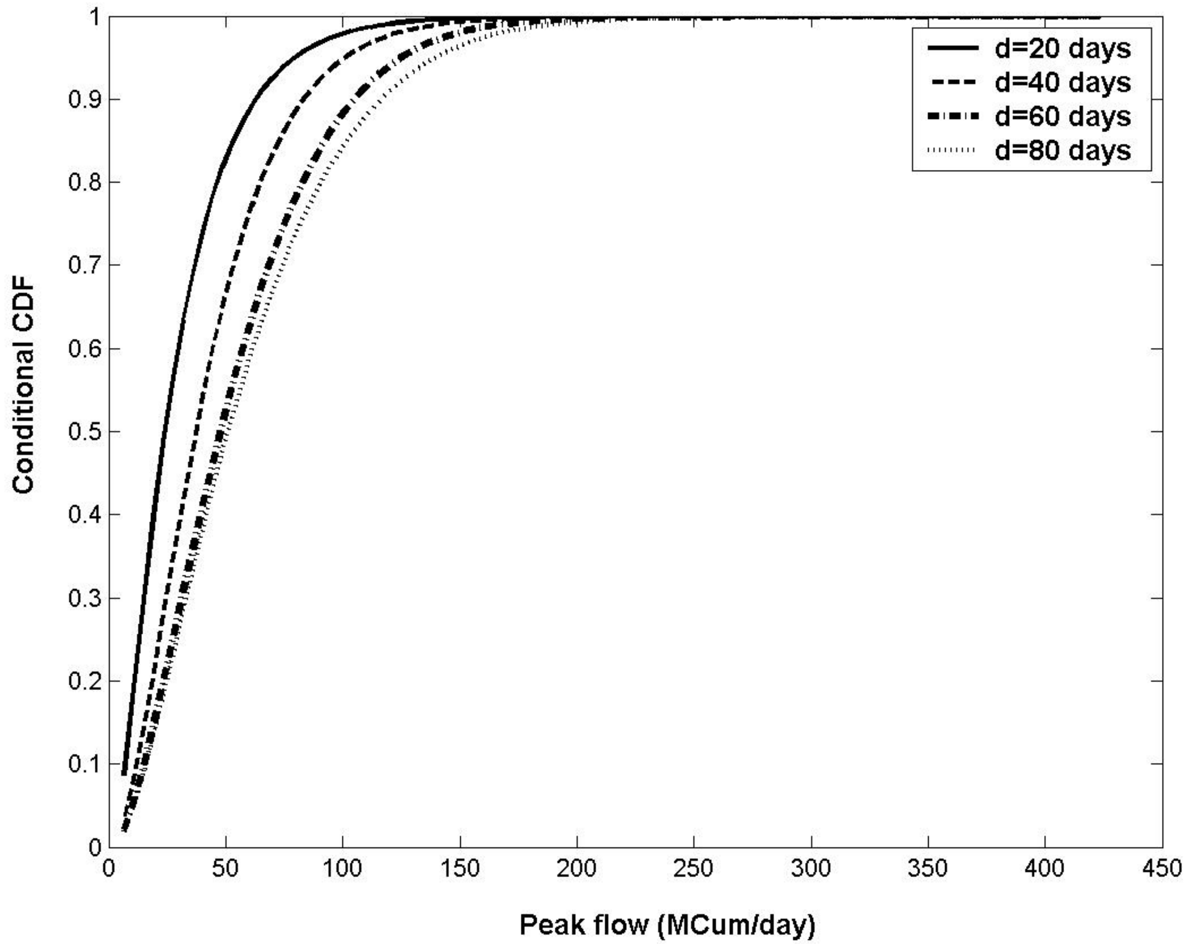
Results – conditional distributions



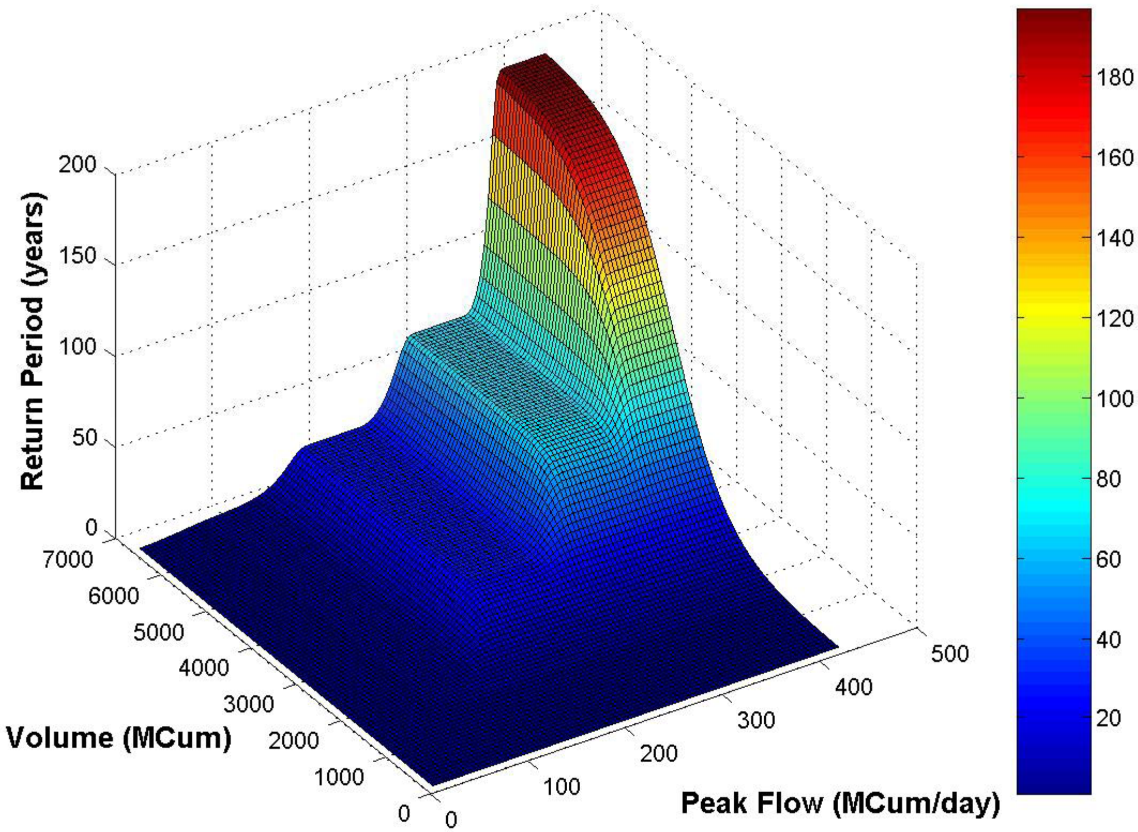
Results – conditional distributions



Results – conditional distributions



Results – return period



Western

Conclusions

- Concept of copula is used for evaluating joint distribution function with mixed marginal distributions - eliminates the restriction of selecting marginals for flood variables from the same family of probability density functions.
- Nonparametric methods (kernel density estimation and orthonormal series) are used to determine the distribution functions for peak flow, volume and duration.
- Nonparametric method based on orthonormal series is more appropriate than kernel estimation.