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
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
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 km² (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
Study area – flood characteristics

- Dependence between P, V and D

<table>
<thead>
<tr>
<th>Flood Characteristics</th>
<th>Pearson’s Linear Correlation Coefficient</th>
<th>Kendall’s Coefficient of Correlation</th>
<th>Spearman’s rho Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak Flow-Volume</td>
<td>0.9359</td>
<td>0.7892</td>
<td>0.9150</td>
</tr>
<tr>
<td>Volume-Duration</td>
<td>0.6934</td>
<td>0.5756</td>
<td>0.7313</td>
</tr>
<tr>
<td>Peak Flow-Duration</td>
<td>0.5306</td>
<td>0.4033</td>
<td>0.5182</td>
</tr>
</tbody>
</table>

- P and V highly correlated
- All the correlations positive
## Marginals - parametric

<table>
<thead>
<tr>
<th>PDF</th>
<th>Parameters</th>
<th>Peak Flow (P)</th>
<th>Volume (V)</th>
<th>Duration (D)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exponential:</strong> ( f_x(x) = \frac{1}{\eta} e^{-x/\eta}; x &gt; 0 )</td>
<td>( \eta )</td>
<td>0.0159</td>
<td>0.0013</td>
<td>0.0245</td>
</tr>
<tr>
<td><strong>Gamma:</strong> ( f_x(x) = \frac{\lambda^\beta x^{\beta-1} e^{-\lambda x}}{\Gamma(\beta)}; x \geq 0, \lambda &gt; 0, \beta &gt; 0 )</td>
<td>( \lambda )</td>
<td>39.956</td>
<td>885.81</td>
<td>5.6490</td>
</tr>
<tr>
<td>and ( \Gamma(\beta) = \int_0^\infty u^{\beta-1} e^{-u} du )</td>
<td>( \beta )</td>
<td>1.5787</td>
<td>0.8921</td>
<td>7.2251</td>
</tr>
<tr>
<td><strong>Gumbel or EV1:</strong> ( f_x(x) = \frac{1}{\alpha} \exp\left[-\frac{x-\varepsilon}{\alpha} - \exp\left(-\frac{x-\varepsilon}{\alpha}\right)\right] )</td>
<td>( \varepsilon )</td>
<td>40.485</td>
<td>413.69</td>
<td>33.981</td>
</tr>
<tr>
<td></td>
<td>( \alpha )</td>
<td>39.144</td>
<td>652.34</td>
<td>11.839</td>
</tr>
<tr>
<td><strong>Lognormal:</strong> ( f(x) = \frac{1}{x \sqrt{2\pi\sigma_y}} \exp\left(-\frac{(y-\mu_y)^2}{2\sigma_y^2}\right) )</td>
<td>( \mu_y )</td>
<td>3.8562</td>
<td>6.1470</td>
<td>3.6416</td>
</tr>
<tr>
<td>( y = \log x, x &gt; 0, -\infty &lt; \mu_y &lt; \infty, \sigma_y &gt; 0 )</td>
<td>( \sigma_y )</td>
<td>0.7971</td>
<td>1.0873</td>
<td>0.3697</td>
</tr>
</tbody>
</table>
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
  \[ \int \Phi_s(x) \Phi_j(x) \, dx = 0 \quad \forall s \neq j \quad \int \{ \Phi_j(x) \}^2 \, dx = 1 \quad \forall j \]

  \[ \Phi_0(x) = 1 \quad \Phi_j(x) = \sqrt{2} \cos(\pi j x) \]
Results
Results

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

<table>
<thead>
<tr>
<th>Distribution Function</th>
<th>RMSE</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P V D</td>
<td>P V D</td>
<td>P V D</td>
</tr>
<tr>
<td>Kernel</td>
<td>0.054</td>
<td>-583.03</td>
<td>-454.18</td>
</tr>
<tr>
<td>Orthonormal</td>
<td>0.020</td>
<td>-781.24</td>
<td>-773.36</td>
</tr>
<tr>
<td>Exponential</td>
<td>0.047</td>
<td>-610.31</td>
<td>-618.72</td>
</tr>
<tr>
<td>Gamma</td>
<td>0.017</td>
<td>-813.25</td>
<td>-647.89</td>
</tr>
<tr>
<td>Gumbel</td>
<td>0.066</td>
<td>-640.47</td>
<td>-349.01</td>
</tr>
<tr>
<td>Lognormal</td>
<td>0.021</td>
<td>-771.86</td>
<td>-730.35</td>
</tr>
</tbody>
</table>
## Results

### Second test

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

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n))$$

$$C(u_1, \ldots, u_n) = F(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)), \quad 0 \leq u_1, \ldots, u_n \leq 1$$
# Copula

<table>
<thead>
<tr>
<th>Copula [ C_\theta(u_1, u_2) ]</th>
<th>Generating Function</th>
<th>( \tau = 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} , dt )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ali-Mikhail-Haq Family:</strong> ( \frac{u_1 u_2}{[1 - \theta(1 - u_1)(1 - u_2)]} )</td>
<td>( \theta \in [-1, 1] ) ( \ln \left{ \frac{1 - \theta(1 - t)}{t} \right} ) ( \frac{(3\theta - 2)}{\theta} - \left[ \frac{2}{3} (1 - \theta^{-1})^2 \ln(1 - \theta) \right] )</td>
<td></td>
</tr>
<tr>
<td><strong>Cook-Johnson Family:</strong> ( { \max\left[ (u_1)^{-\theta} + (u_2)^{-\theta} - 1, 0 \right] }^{-1/\theta}, \theta \geq 0 ) ( \theta \in [-1, \infty) \setminus {0} ) ( \frac{\left[ (t)^{-\theta} - 1 \right]}{\theta} ) ( \frac{\theta}{(\theta + 2)} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Gumbel-Hougaard Family:</strong> ( \exp\left{ -\left[ (-\ln u_1)^\theta + (-\ln u_2)^\theta \right]^{1/\theta} \right} )</td>
<td>( \theta \in [1, \infty) ) ( (-\ln t)^\theta ) ( (1 - \theta^{-1}) )</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Copula</th>
<th>RMSE ( P-V )</th>
<th>RMSE ( V-D )</th>
<th>RMSE ( P-D )</th>
<th>AIC ( P-V )</th>
<th>AIC ( V-D )</th>
<th>AIC ( P-D )</th>
<th>BIC ( P-V )</th>
<th>BIC ( V-D )</th>
<th>BIC ( P-D )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ali-Mikhail-Haq</td>
<td>0.141</td>
<td>0.090</td>
<td>0.056</td>
<td>-68.43</td>
<td>-84.64</td>
<td>-101.76</td>
<td>-67.54</td>
<td>-83.75</td>
<td>-100.87</td>
</tr>
<tr>
<td>Cook-Johnson</td>
<td>0.031</td>
<td>0.058</td>
<td>0.055</td>
<td>-122.80</td>
<td>-100.50</td>
<td>-102.31</td>
<td>-121.91</td>
<td>-99.61</td>
<td>-101.42</td>
</tr>
<tr>
<td>Gumbel-Hougaard</td>
<td>0.025</td>
<td>0.027</td>
<td>0.020</td>
<td>-130.53</td>
<td>-128.74</td>
<td>-138.33</td>
<td>-129.64</td>
<td>-127.84</td>
<td>-137.44</td>
</tr>
</tbody>
</table>
Results – joint distributions

Peak flow - Volume
Results – joint distributions

(a) Ali-Mikhail-Haq Copula

(b) Cook-Johnson Copula

(c) Gumbel-Hougaard Copula

Volume - Duration
Results – joint distributions

(a) Ali-Mikhail-Haq Copula
(b) Cook-Johnson Copula
(c) Gumbel-Hougaard Copula

Peak flow - Duration
Results – conditional distributions
Results – conditional distributions

![Graph showing conditional CDF vs. Volume (MCum)]

Key:
- **d=20 days**
- **d=40 days**
- **d=60 days**
- **d=80 days**
Results – conditional distributions
Results – return period
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.