Multivariate Statistical Analysis of Indiana Hydrologic Data

Shih-Chieh Kao
Ph.D. Candidate
School of Civil Engineering
Purdue University
Extreme Events

Feb. 5, 2008
Delphi, Indiana
Flooding of Tippecanoe River

Sept., 2007
George H. Sparks Reservoir
Lithia Springs, Georgia

(AP Photo/Journal & Courier, Michael Heinz)

(Barry Gillis, http://www.drought.unl.edu/gallery/2007/Georgia/Sparks1.htm)
Outline

• Background and motivation
  – Limitations in univariate approach

• Introduction to copulas

• Research objectives
  – Topic 1: Probabilistic structure of surface runoff
  – Topic 2: Extreme rainfall frequency analysis
  – Topic 3: Drought frequency analysis

• Summary and concluding remarks
Limitations in Univariate Approach

- Example: Selection of annual maximum precipitation events in constructing design rainfall estimates
  - *Durations* are not the actual durations of rainfall events
  - Long-term maximum may cover multiple events
  - Short-term maximum encompasses only part of the extreme event
**Bivariate Distribution Example**

Bivariate Gaussian distribution, $\rho = 0.8$

**Marginals**

$$f_x(x) = \int_{-\infty}^{\infty} h_{XY}(x, y) dy$$

$$f_y(y) = \int_{-\infty}^{\infty} h_{XY}(x, y) dx$$

**Joint density**

$$h_{XY}(x, y)$$

Gaussian marginals with Clayton Copulas

$$\rho = 0.8$$
Copulas

• Transformation of joint cumulative distribution
  – \( H_{XY}(x,y) = C_{UV}(u,v) \)
    marginals: \( u = F_X(x), \ v = F_Y(y) \)
  – Sklar (1959) proved that the transformation is \textit{unique} for continuous r.v.s

• Use copulas to construct joint distributions
  – Marginal distributions => selecting suitable PDFs
  – Dependence structure => selecting suitable copulas
  – Together they form the joint distribution
Use of Copulas in Hydrology

• Since 2003, over 20 papers has been published in water resources related journals
  – Topics include: rainfall and flood frequency analysis, groundwater parameters estimation, sea storms analysis, rainfall IDF curves, and etc.
  – Full potential of copulas is yet to be realized (Genest and Favre, 2007)

• For copulas in rainfall frequency analysis:
  – The definition of extreme events was not clear
  – Few stations were examined

• For copulas in drought frequency analysis:
  – Bivariate streamflow drought analysis
Data Sources & Study Area

• Precipitation
  – NCDC hourly precipitation dataset
    • 53 stations with record length greater than 50 years
  – NCDC daily precipitation dataset
    • 73 stations with record length greater than 80 years

• Streamflow
  – USGS unregulated daily mean flow
    • 36 stations with record length greater than 50 years
Topic 1
Probabilistic Structure of Surface Runoff (I)

- Classical problem in derived flood frequency analysis
  - For regular rainfall events, duration (D) and average intensity (I) are assumed to be exponentially distributed
  - Eagleson (1972) assumed independence between D & I
  - Córdova and Rodríguez-Iturbe (1985) assumed positive dependence between D & I
  - Copulas are found to be a more mathematical efficient approach in solving probabilistic feature of rainfall excess ($P_e$)
• Dependence between D & I cannot be neglected
Topic 2
Extreme Rainfall Frequency Analysis

• Definitions of Extreme Rainfall Events
  – Hydrologic designs are usually governed by depth (volume) or peak intensity
  – Annual maximum volume (AMV) events
    • Longer duration
  – Annual maximum peak intensity (AMI) events
    • Shorter duration
  – Annual maximum cumulative probability (AMP) events
    • The use of empirical copulas between volume and peak intensity
    • Wide range of durations
Estimate of depth for known duration

T-year depth $p_T$ given duration $d$

- The estimates of depth are similar for durations larger than 10 hours
- AMP definition seems to be an appropriate indicator for defining extreme events
Estimate of peak intensity for known duration

T-year peak intensity $i_T$ given duration $d$

- Conventional approach fails to capture the peak intensity
Rainfall Peak Attributes

- Given depth (P) and duration (D), compute the conditional expectation of peak intensity (I) and percentage time to peak ($T_p$)

![Graphs showing expected peak intensity and time to peak for different rainfall depths and durations.](image-url)
Temporal Accumulation Curves

- Given depth (P) and duration (D), compute the conditional expectation of percentage accumulations at each 10% temporal ordinates ($A_{10}$, $A_{20}$, ..., $A_{90}$)

Expectation of % accumulation given P & D
Drought Frequency Analysis

• Challenges in characterizing droughts
  – No clear (scientific) definition: deficit of water for prolonged time
  – Phenomenon dependent in time, space, and between various variables such as precipitation, streamflow, and soil moisture

• Classification of droughts
  – Meteorological drought: precipitation deficit
  – Hydrologic drought: streamflow deficit
  – Agricultural drought: soil moisture deficit

• Various drought indices
  – Palmer Drought Severity Index (PDSI), Crop Moisture Index (CMI), Surface Water Supply Index (SWSI), Vegetation Condition Index (VCI), CPC Soil Moisture, Standardized precipitation index (SPI)
US Drought Monitor

• Overall drought status (D0 ~ D4) determined based on various indices together (Svobada et al., 2002)
  – PDSI
  – CPC Soil moisture
  – USGS weekly
  – Percentage of normal
  – SPI
  – VCI

• Linear combination of selected indices (OBDI, objective blend of drought indicator) was adopted as the preliminary overall drought status

• The decision of final drought status relies on subjective judgment

http://drought.unl.edu/dm/monitor.html
Standardized Index Method

- Proposed by McKee *et al.* (1993)
- Generalizable to various types of observations
  - For precipitation: SPI
- For a given window size, the observed precipitation is transformed to a probability measure using Gamma distribution, then expressed in standard normal variable

<table>
<thead>
<tr>
<th>Probabilities of Occurrence (%)</th>
<th>SI Values</th>
<th>Drought Monitor Category</th>
<th>Drought Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 ~ 30</td>
<td>-0.84 ~ -0.52</td>
<td>D0</td>
<td>Abnormally dry</td>
</tr>
<tr>
<td>10 ~ 20</td>
<td>-1.28 ~ -0.84</td>
<td>D1</td>
<td>Drought - moderate</td>
</tr>
<tr>
<td>5 ~ 10</td>
<td>-1.64 ~ -1.28</td>
<td>D2</td>
<td>Drought - severe</td>
</tr>
<tr>
<td>2 ~ 5</td>
<td>-2.05 ~ -1.64</td>
<td>D3</td>
<td>Drought - extreme</td>
</tr>
<tr>
<td>&lt; 2</td>
<td>&lt; -2.05</td>
<td>D4</td>
<td>Drought - exceptional</td>
</tr>
</tbody>
</table>

- Though SIs for different windows are dependent, no representative window can be determined
**Modified SI**

- Limitations of the conventional SI approach
  - Significant auto-correlation exists in samples
  - Cannot account for seasonal variability
  - Gamma distribution may not be suitable

- Modified algorithm
  - Group samples by the “ending month”
  - KS test with 5% significant level

<table>
<thead>
<tr>
<th></th>
<th>Precipitation</th>
<th>Streamflow</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>G2</td>
<td>G2</td>
</tr>
<tr>
<td>SI</td>
<td>142 / 876</td>
<td>287 / 432</td>
</tr>
<tr>
<td>mod. SI</td>
<td>122 / 10512</td>
<td>190 / 5184</td>
</tr>
</tbody>
</table>
Dependence Structure

- Precipitation marginals \( \{u_1, u_2, \ldots, u_{12}\} \) and streamflow marginals \( \{v_1, v_2, \ldots, v_{12}\} \) are selected
  - Annual cycle accounts for the seasonal effect naturally
  - Avoid overlaying in samples
  - Allow for a month-by-month assessment for future conditions

<table>
<thead>
<tr>
<th>i</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>j</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.71</td>
<td>0.57</td>
<td>0.48</td>
<td>0.41</td>
<td>0.38</td>
<td>0.37</td>
<td>0.36</td>
<td>0.35</td>
<td>0.33</td>
<td>0.31</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.89</td>
<td>0.82</td>
<td>0.70</td>
<td>0.61</td>
<td>0.55</td>
<td>0.53</td>
<td>0.51</td>
<td>0.49</td>
<td>0.47</td>
<td>0.44</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.80</td>
<td>0.93</td>
<td>0.87</td>
<td>0.76</td>
<td>0.69</td>
<td>0.64</td>
<td>0.61</td>
<td>0.59</td>
<td>0.56</td>
<td>0.54</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.73</td>
<td>0.85</td>
<td>0.94</td>
<td>0.90</td>
<td>0.81</td>
<td>0.75</td>
<td>0.70</td>
<td>0.67</td>
<td>0.65</td>
<td>0.62</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.67</td>
<td>0.78</td>
<td>0.87</td>
<td>0.95</td>
<td>0.92</td>
<td>0.85</td>
<td>0.79</td>
<td>0.75</td>
<td>0.72</td>
<td>0.69</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.63</td>
<td>0.72</td>
<td>0.81</td>
<td>0.89</td>
<td>0.96</td>
<td>0.93</td>
<td>0.87</td>
<td>0.82</td>
<td>0.78</td>
<td>0.75</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.59</td>
<td>0.68</td>
<td>0.75</td>
<td>0.83</td>
<td>0.90</td>
<td>0.96</td>
<td>0.94</td>
<td>0.89</td>
<td>0.85</td>
<td>0.81</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.57</td>
<td>0.64</td>
<td>0.72</td>
<td>0.79</td>
<td>0.85</td>
<td>0.91</td>
<td>0.97</td>
<td>0.95</td>
<td>0.90</td>
<td>0.86</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.55</td>
<td>0.62</td>
<td>0.69</td>
<td>0.75</td>
<td>0.81</td>
<td>0.87</td>
<td>0.93</td>
<td>0.97</td>
<td>0.96</td>
<td>0.91</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.53</td>
<td>0.60</td>
<td>0.66</td>
<td>0.72</td>
<td>0.78</td>
<td>0.83</td>
<td>0.89</td>
<td>0.94</td>
<td>0.98</td>
<td>0.96</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.51</td>
<td>0.58</td>
<td>0.64</td>
<td>0.70</td>
<td>0.75</td>
<td>0.81</td>
<td>0.85</td>
<td>0.90</td>
<td>0.94</td>
<td>0.98</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.50</td>
<td>0.56</td>
<td>0.62</td>
<td>0.68</td>
<td>0.73</td>
<td>0.78</td>
<td>0.83</td>
<td>0.87</td>
<td>0.91</td>
<td>0.95</td>
<td>0.98</td>
<td></td>
</tr>
</tbody>
</table>

Spearman's \( r_{i,j} \) between \( u_i \) and \( u_j \)

Spearman's \( r_{i,j} \) between \( v_i \) and \( v_j \)
Higher Dimensional Copulas

- Limited choices because of high mathematical complexity
  - Gaussian copulas
    - Derived from the well-known multivariate normal distribution
    - Preserving all bivariate marginal dependencies through the correlation matrix $\Sigma$
  - Empirical copulas
    - Multi-dimensional rank-based probabilities
    - Treated as the observed probabilities when performing model verification
- Empirical copulas were adopted in this study.
Joint Deficit Index (I)

- Assumption: events with the same value of copulas (joint cumulative probability) cause similar joint drought impact
  - Copula values are treated as joint deficit status

- Distribution function of copulas $K_C(t)$
  - Give probability measure for events with $C(u_1, u_2, ..., u_{12}) \leq t$

- Joint deficit index (JDI)
  - $JDI = \Phi^{-1}(K_C)$
  - Share the same classification with SI
Joint Deficit Index (II)
Joint Deficit Index (III)

- Comparison between 1-Mn, 12-Mn, and joint SPI
  - 12-Mn SPI changes slowly, weak in reflecting emerging drought
  - 1-Mn SPI changes rapidly, weak in reflecting accumulative deficit
  - Joint SPI reflects joint deficit
Precipitation vs. Streamflow

\[ r = 0.73 \]
Potential of Future Droughts

- Required precipitation for reaching joint normal status ($K_C = 0.5$) in the future
- Probability of drought recovery
Conclusions for Drought Frequency Analysis

• Modified SI provides better statistical footing and helps alleviate the effect of seasonal variability

• JDI can offer an objective and probability-based overall drought description. It is capable of capturing both emerging and prolonged droughts in a timely manner.

• JDI has potential to be applied on different types of hydrologic variables, and can be used to derive an inter-variable drought index

• Potential of future droughts can be assessed by using JDI, where the required precipitation and its exceedance probability can be determined.
Summary and Concluding Remarks

• Copulas are found to be flexible for constructing joint distributions (no specific marginals are required).

• The dependence structure can be faithfully preserved

• Caution when using copulas
  – Need sufficient historic records
    • NWS Atlas 14 adopted 50-year minimum recording length for univariate at-site rainfall frequency analysis
  – Difficulties arise in higher dimensions
    • Mathematical complexity
    • Hard to preserve all lower level mutual dependencies
    • Compatibility problem
Acknowledgements

• Deepest gratitude to my advisor, Dr. Rao S. Govindaraju

• Thank all of my committee members
  – Dr. Dennis A. Lyn
  – Dr. Devdutta S. Niyogi
  – Dr. Venkatesh M. Merwade
  – Dr. Bo Tao

• Special thanks to
  – Dr. A. Ramachandra Rao
  – Dr. Jacques W. Delleur

• Thank my family and friends
Thank you
Questions?