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

(AP Photo/Journal & Courier, Michael Heinz)

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

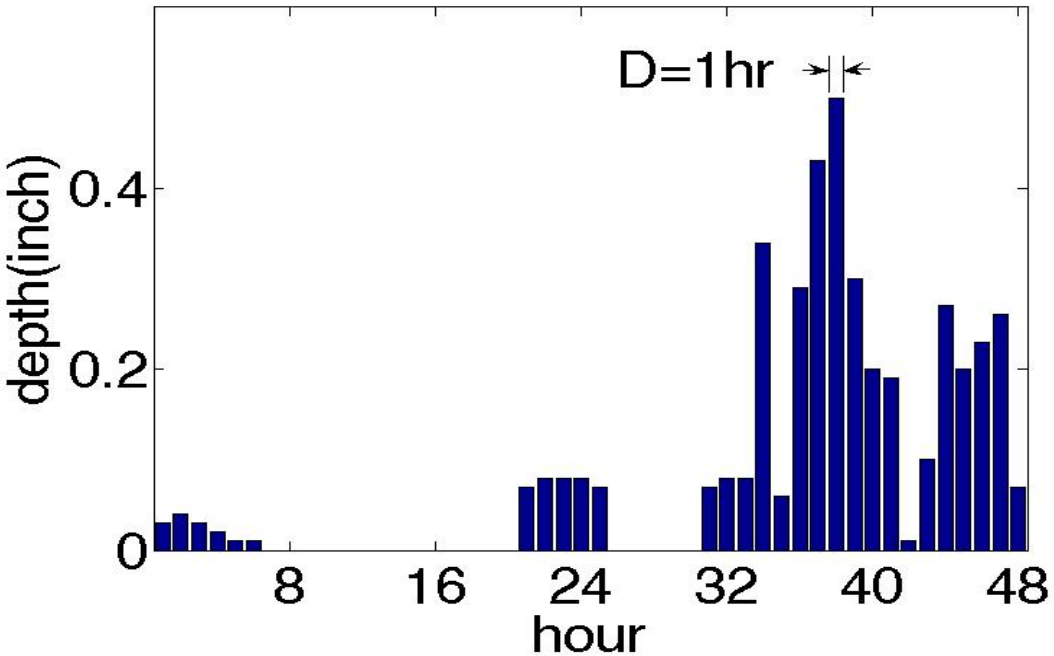
(Barry Gillis, http://www.drought.unl.edu/ gallery/ 2007/Georgia/Sparks1.htm)

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

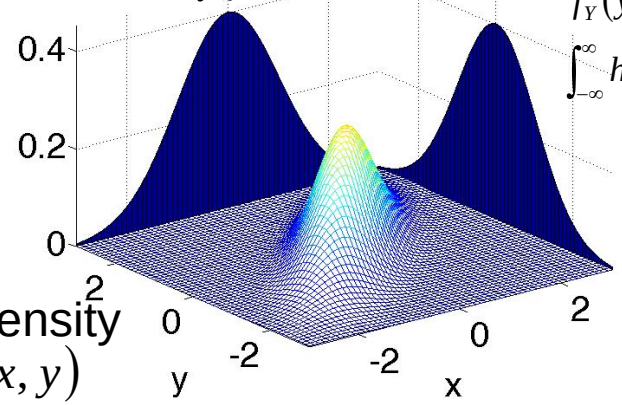


Bivariate Distribution Example

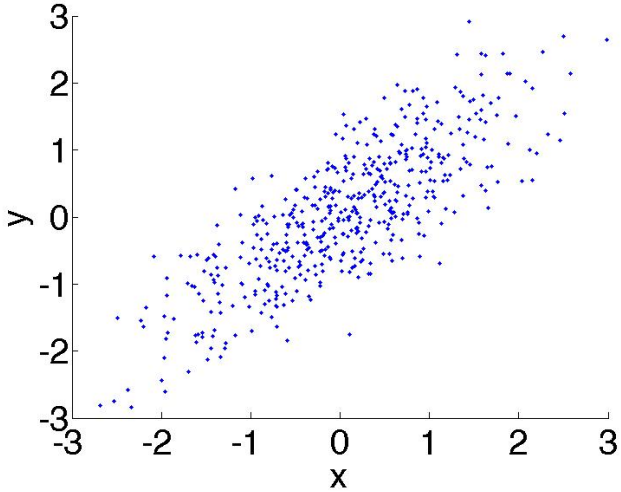
Bivariate Gaussian distribution, ρ = 0.8

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

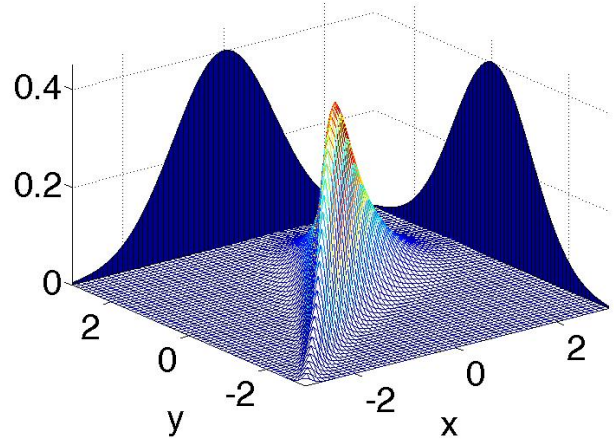
$f_Y(y) = \int_{-\infty}^{\infty} h_{XY}(x, y)dx$

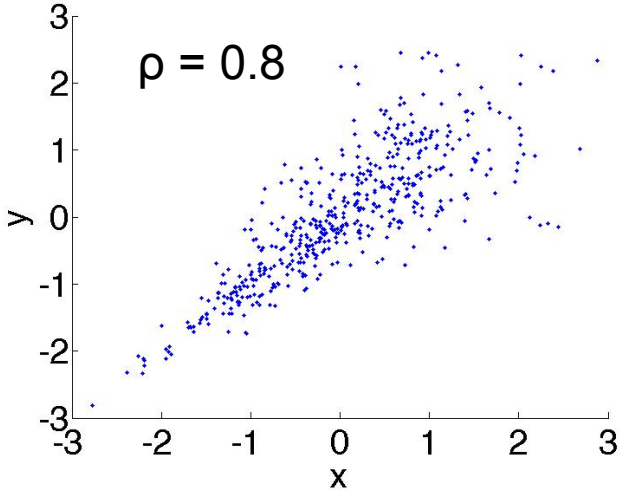


Joint density $h_{XY}(x, y)$



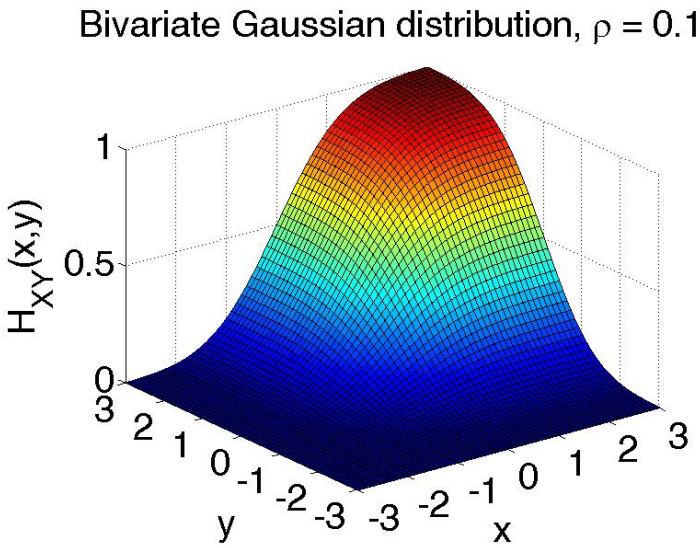
Gaussian marginals with Clayton Copulas

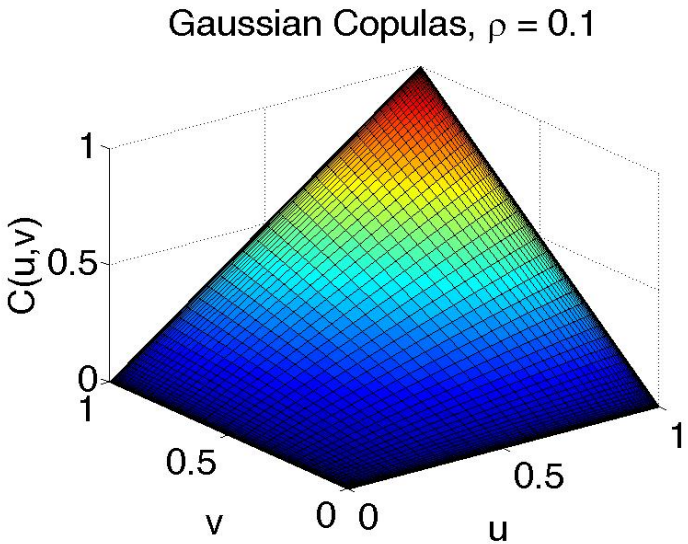




Copulas

- 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 *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

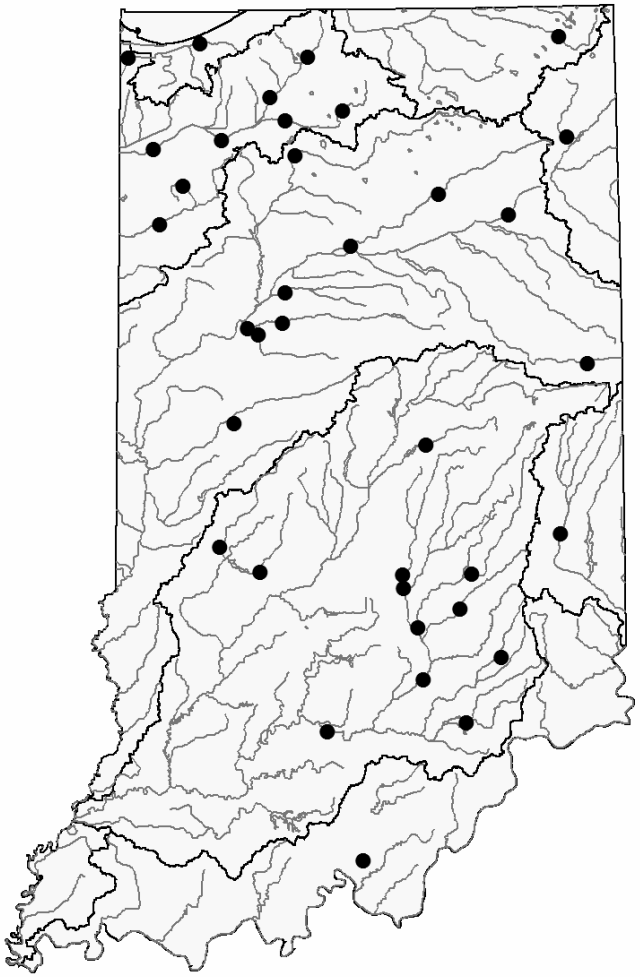
- Since 2003, over 20 papers has been published in water resources related journals
  - Topics include: rainfall and flood frequency analysis, groundwater parameters estimation, sea storms analysis, rainfall IDF curves, and *etc*.
  - Full potential of copulas is yet to be realized (Genest and Favre,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

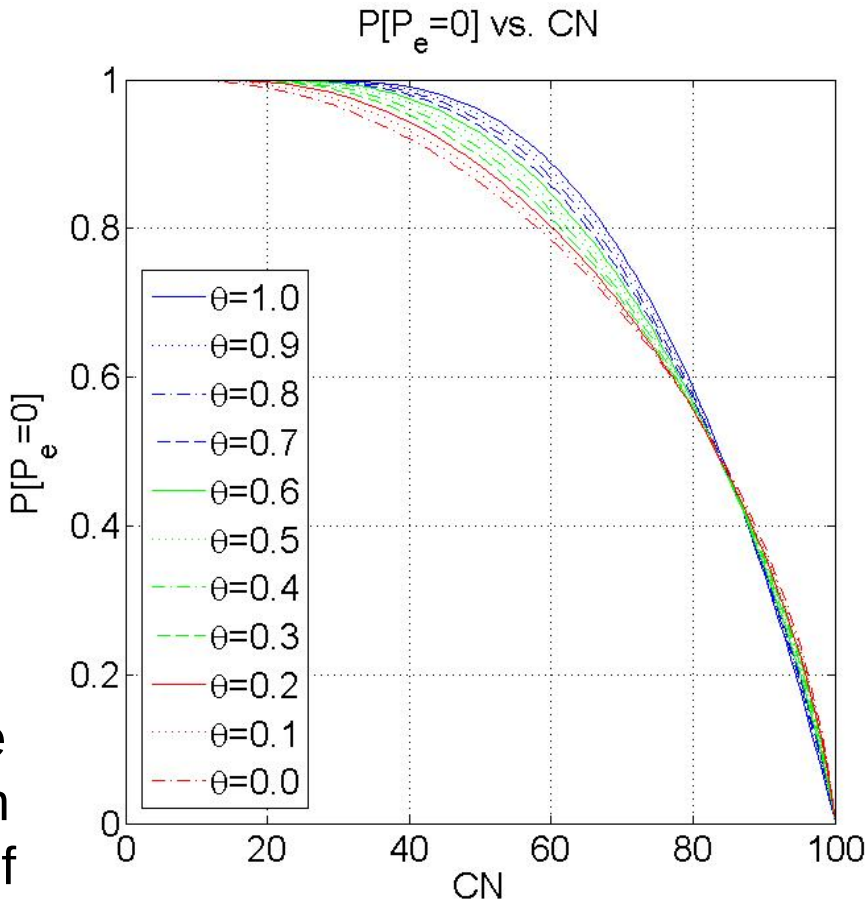
- Precipitation
  - NCDC hourly precipitation dataset
    - 53 stations with record length greater than 50 years
  - NCDC daily precipitation dataset
    - 73 stations with record length greater than 80 years
- Streamflow
  - USGS unregulated daily mean flow
    - 36 stations with record length greater than 50 years



Topic 1

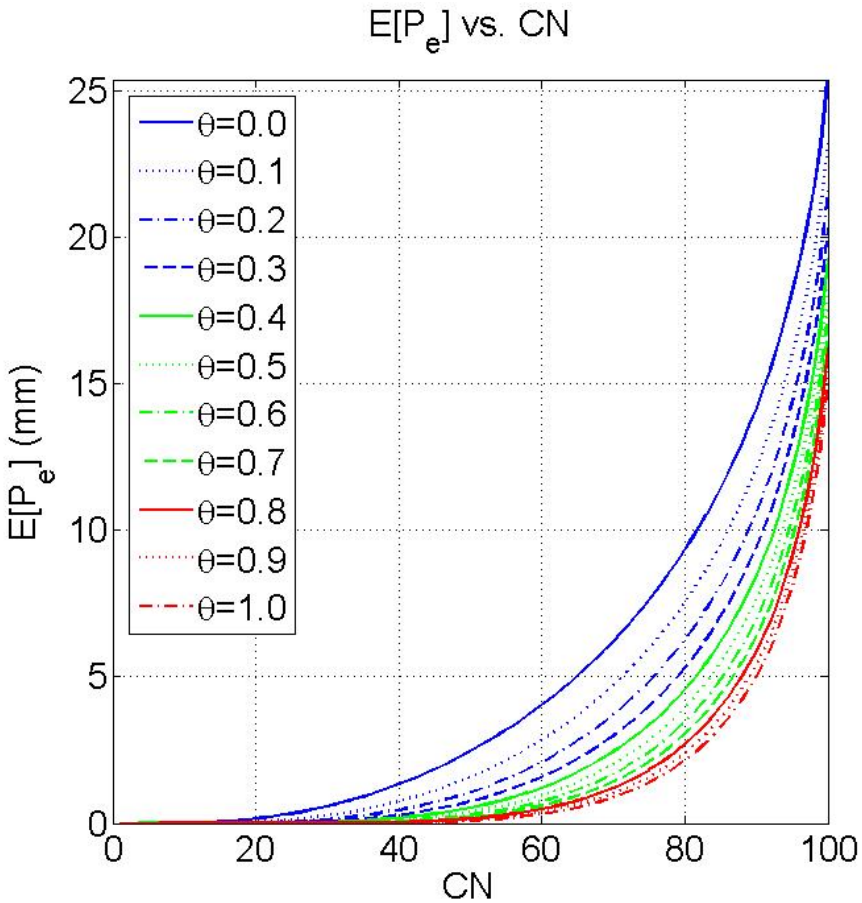
Probabilistic Structure of Surface Runoff (I)

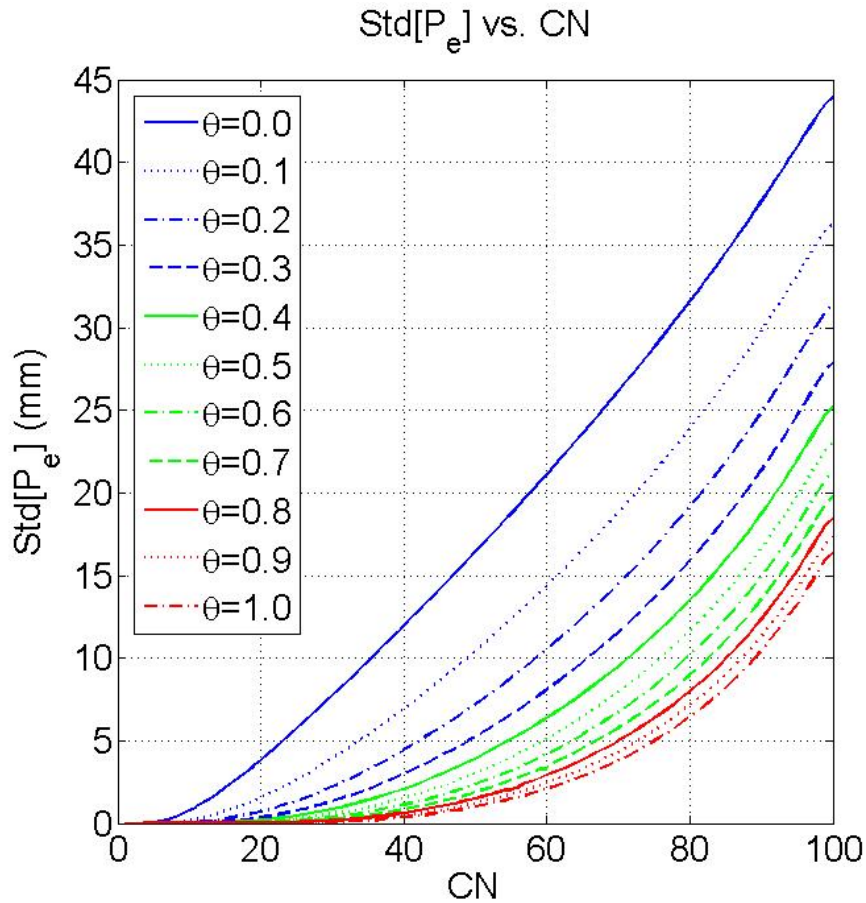
- Classical problem in derived flood frequency analysis
  - For regular rainfall events, duration (D) and average intensity (I) are assumed to be exponentially distributed
  - Eagleson (1972) assumed independence between D & I
  - Córdova and Rodríguez-Iturbe (1985) assumed positive dependence between D & I
  - Copulas are found to be a more mathematical efficient approach in solving probabilistic feature of rainfall excess (P_e)



Topic 1

Probabilistic Structure of Surface Runoff (II)



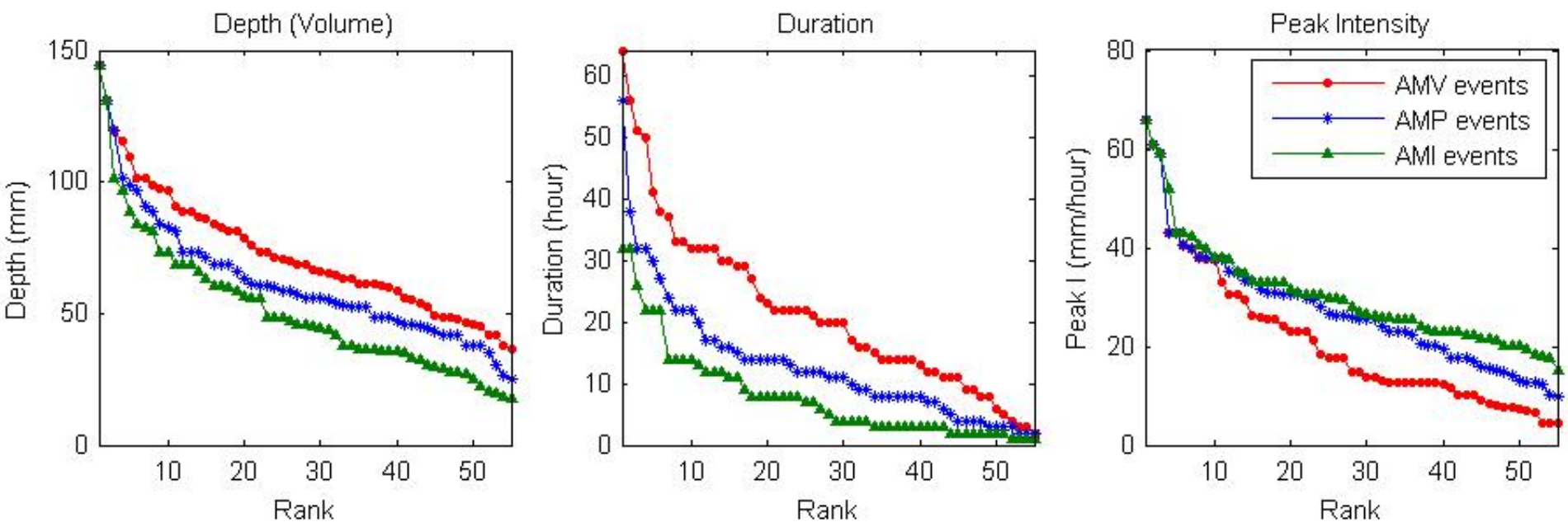


- Dependence between D & I cannot be neglected

Topic 2

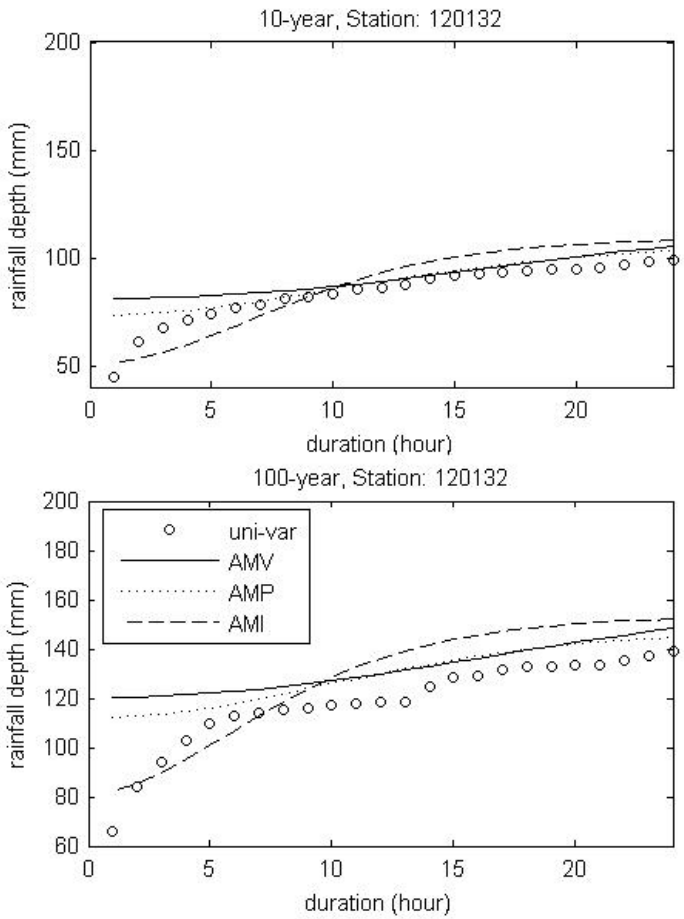
Extreme Rainfall Frequency Analysis

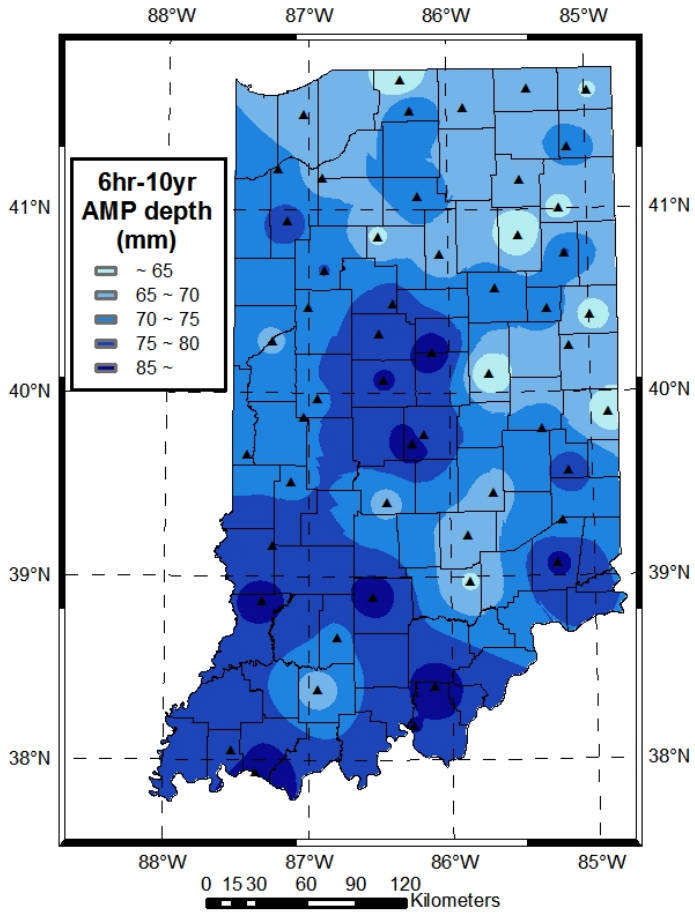
- Definitions of Extreme Rainfall Events
  - Hydrologic designs are usually governed by depth (volume) or peak intensity
  - Annual maximum volume (AMV) events
    - Longer duration
  - Annual maximum peak intensity (AMI) events
    - Shorter duration
  - Annual maximum cumulative probability (AMP) events
    - 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

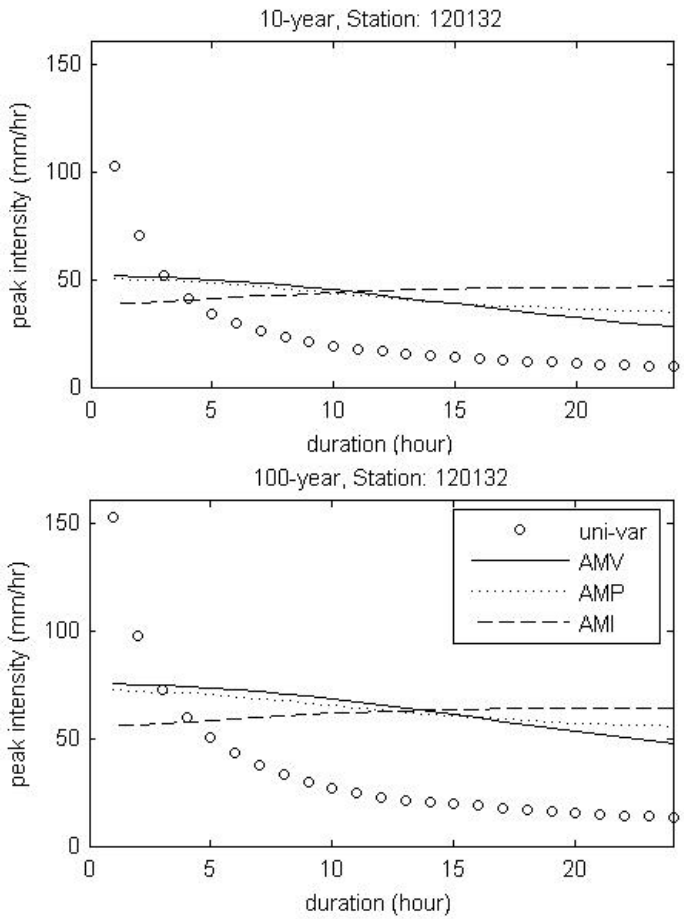


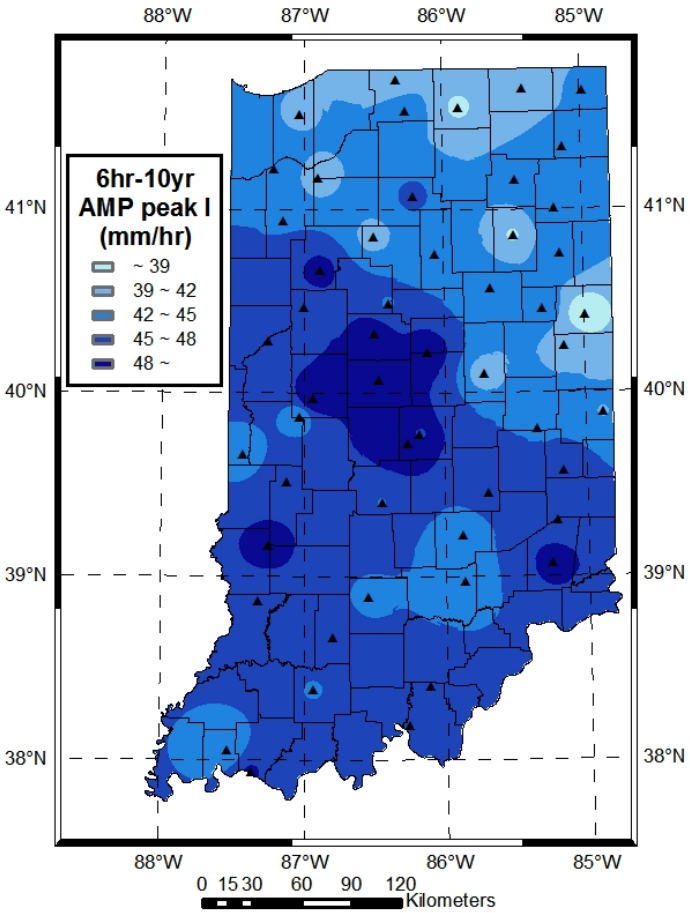


- The estimates of depth are similar for durations larger than 10 hours
- AMP definition seems to be an appropriate indicator for defining extreme events

Estimate of peak intensity for known duration

T-year peak intensity i_T given duration d

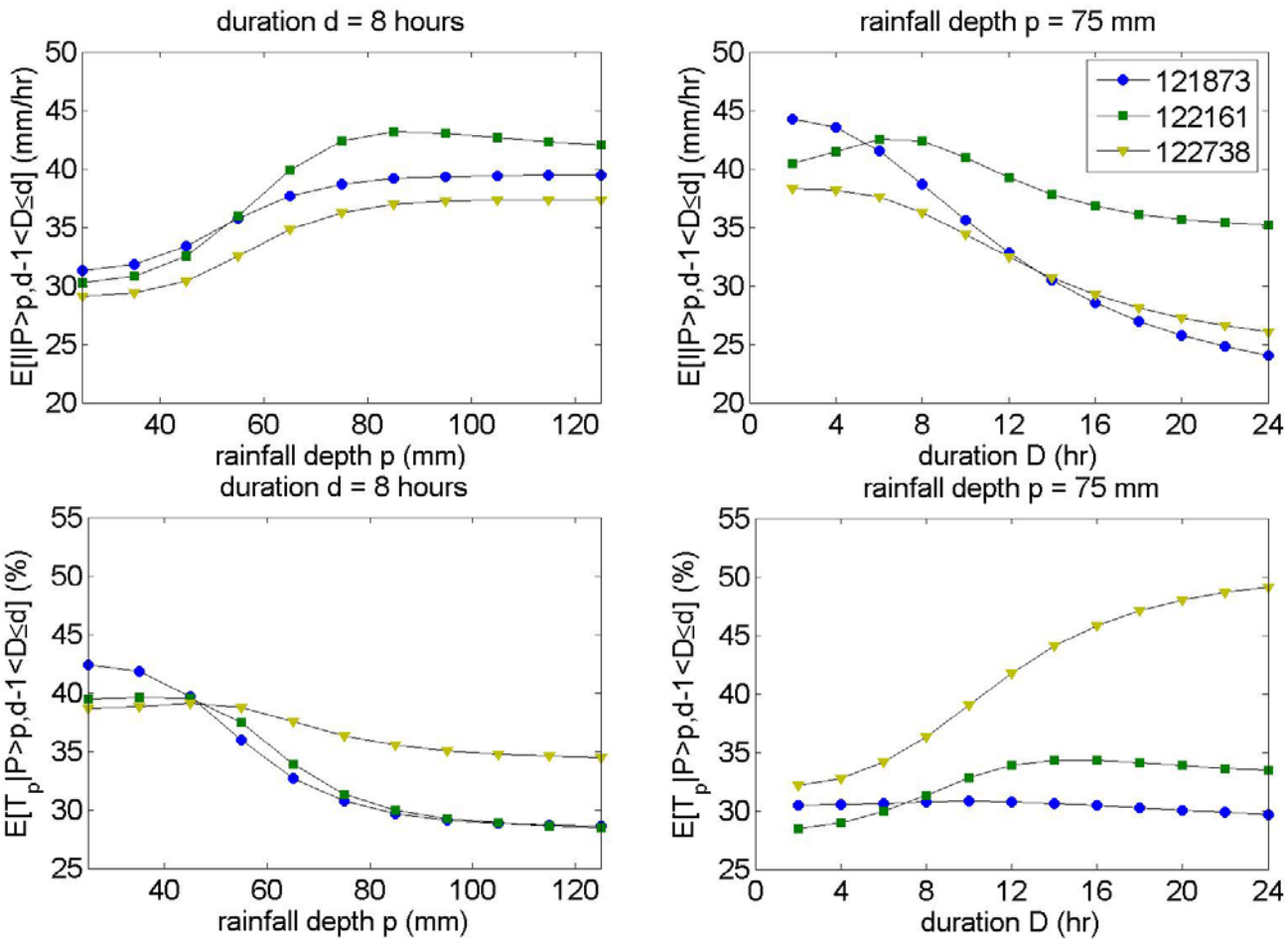




- Conventional approach fails to capture the peak intensity

Rainfall Peak Attributes

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

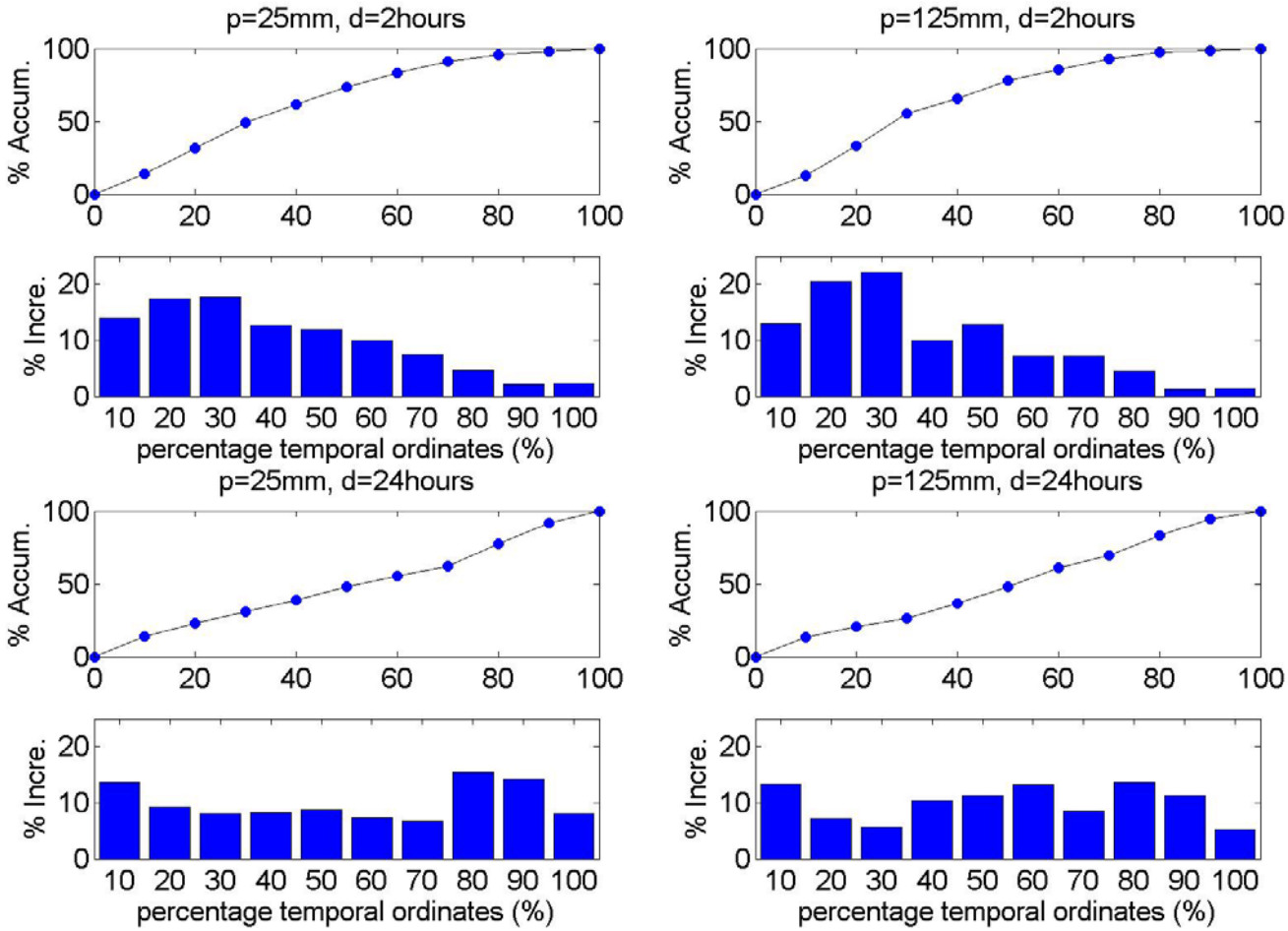


Expectation of peak intensity given P & D

Expectation of time to peak (%) given P & D

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

Topic 3

Drought Frequency Analysis

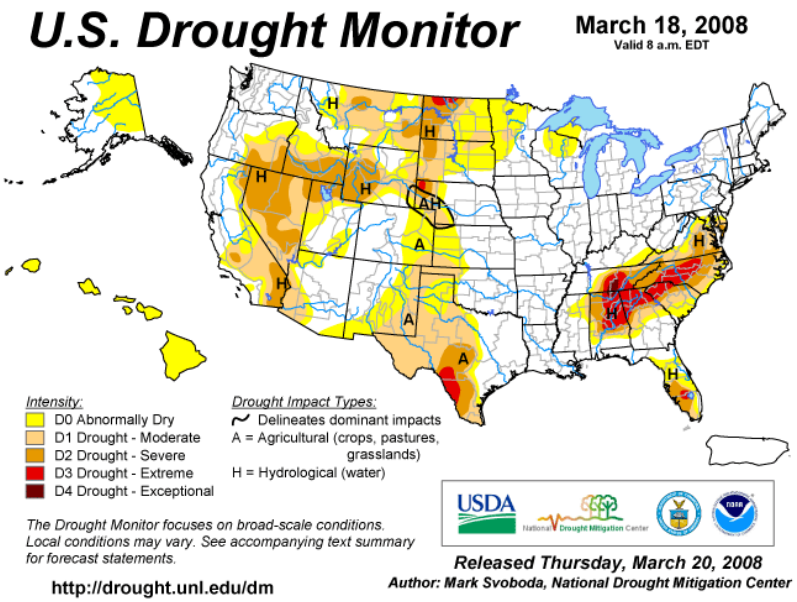
- Challenges in characterizing droughts
  - No clear (scientific) definition: deficit of water for prolonged time
  - Phenomenon dependent in time, space, and between various variables such as precipitation, streamflow, and soil moisture

- 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



http://drought.unl.edu/dm/monitor.html

- 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

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

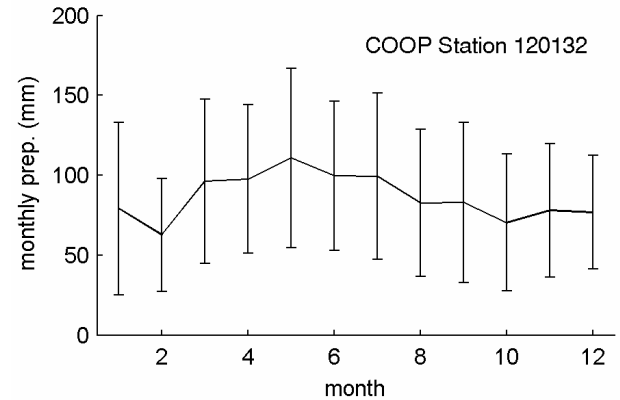
Probabilities of Occurrence (%)	SI Values	Drought Monitor Category	Drought Condition
20 ~ 30	-0.84 ~ -0.52	D0	Abnormally dry
10 ~ 20	-1.28 ~ -0.84	D1	Drought - moderate
5 ~ 10	-1.64 ~ -1.28	D2	Drought - severe
2 ~ 5	-2.05 ~ -1.64	D3	Drought - extreme
< 2	< -2.05	D4	Drought - exceptional

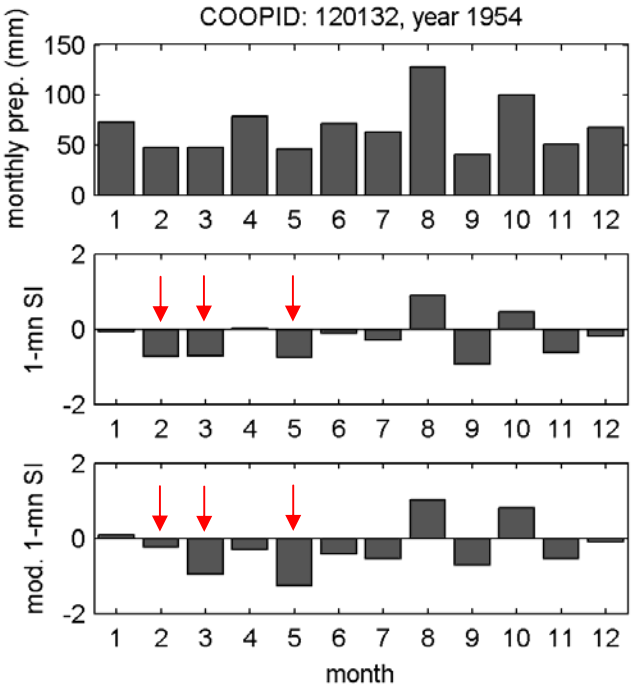
- Though SIs for different windows are dependent, no representative window can be determined

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

	Precipitation	Streamflow	
	G2	G2	GEV
SI	142 / 876	287 / 432	163 / 432
mod. SI	122 / 10512	190 / 5184	11 / 5184





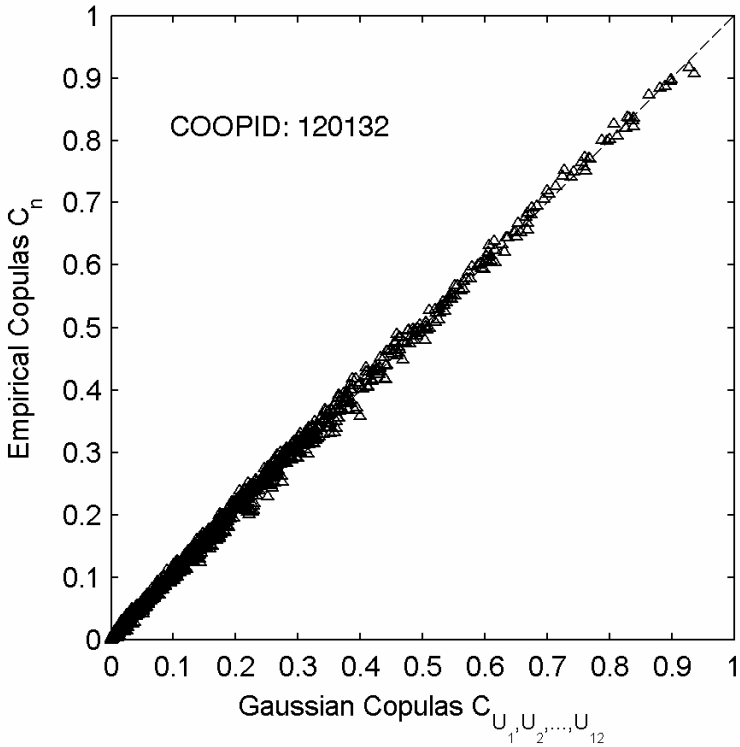
Dependence Structure

- 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

Spearman's $r_{i,j}$ between v_i and v_j (j)	Spearman's $r_{i,j}$ between u_i and u_j (i) 1	2	3	4	5	6	7	8	9	10	11	12
1		0.71	0.57	0.48	0.41	0.38	0.37	0.36	0.35	0.33	0.31	0.30
2	0.89		0.82	0.70	0.61	0.55	0.53	0.51	0.49	0.47	0.44	0.42
3	0.80	0.93		0.87	0.76	0.69	0.64	0.61	0.59	0.56	0.54	0.51
4	0.73	0.85	0.94		0.90	0.81	0.75	0.70	0.67	0.65	0.62	0.60
5	0.67	0.78	0.87	0.95		0.92	0.85	0.79	0.75	0.72	0.69	0.67
6	0.63	0.72	0.81	0.89	0.96		0.93	0.87	0.82	0.78	0.75	0.73
7	0.59	0.68	0.75	0.83	0.90	0.96		0.94	0.89	0.85	0.81	0.78
8	0.57	0.64	0.72	0.79	0.85	0.91	0.97		0.95	0.90	0.86	0.83
9	0.55	0.62	0.69	0.75	0.81	0.87	0.93	0.97		0.96	0.91	0.88
10	0.53	0.60	0.66	0.72	0.78	0.83	0.89	0.94	0.98		0.96	0.92
11	0.51	0.58	0.64	0.70	0.75	0.81	0.85	0.90	0.94	0.98		0.96
12	0.50	0.56	0.62	0.68	0.73	0.78	0.83	0.87	0.91	0.95	0.98	

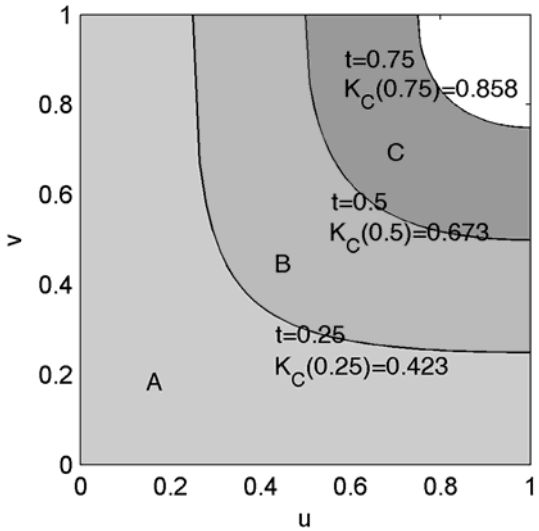
Higher Dimensional Copulas

- 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 Σ
  - Empirical copulas
    - Multi-dimensional rank-based probabilities
    - Treated as the observed probabilities when performing model verification
- Empirical copulas were adopted in this study.

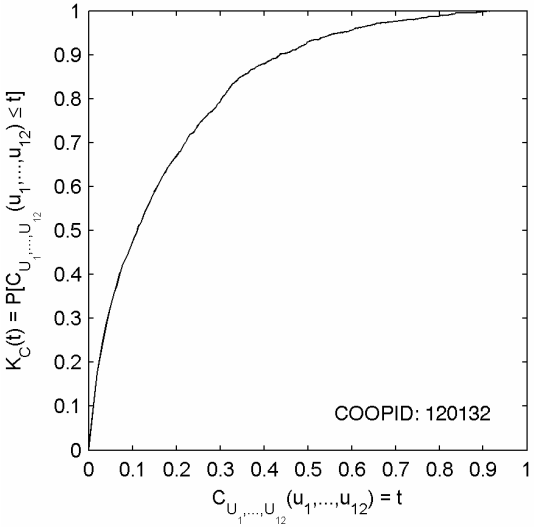


Joint Deficit Index (I)

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

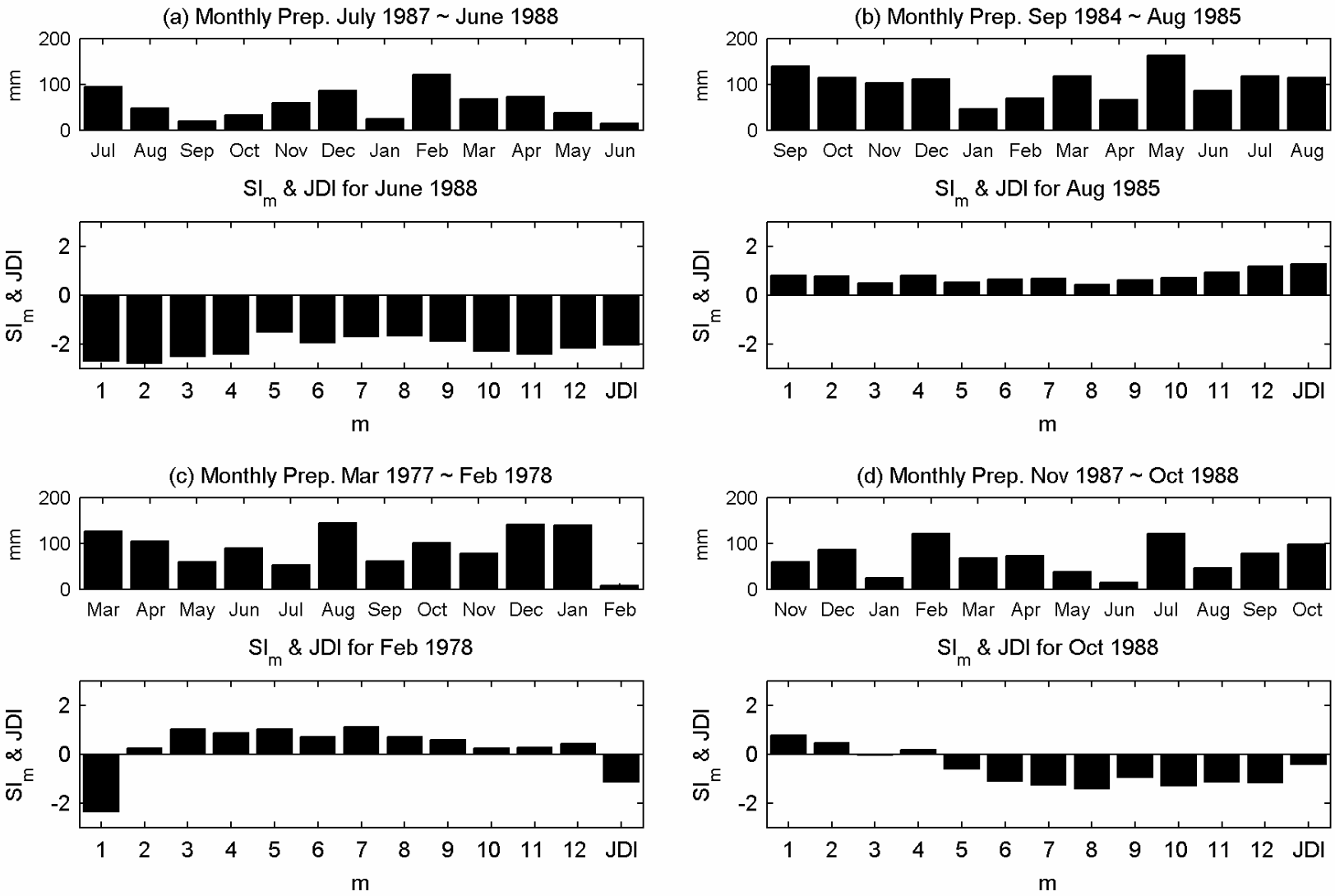


- Distribution function of copulas $K_C(t)$
  - Give probability measure for events with $C(u_1, u_2, \ldots, u_{12}) \leq t$



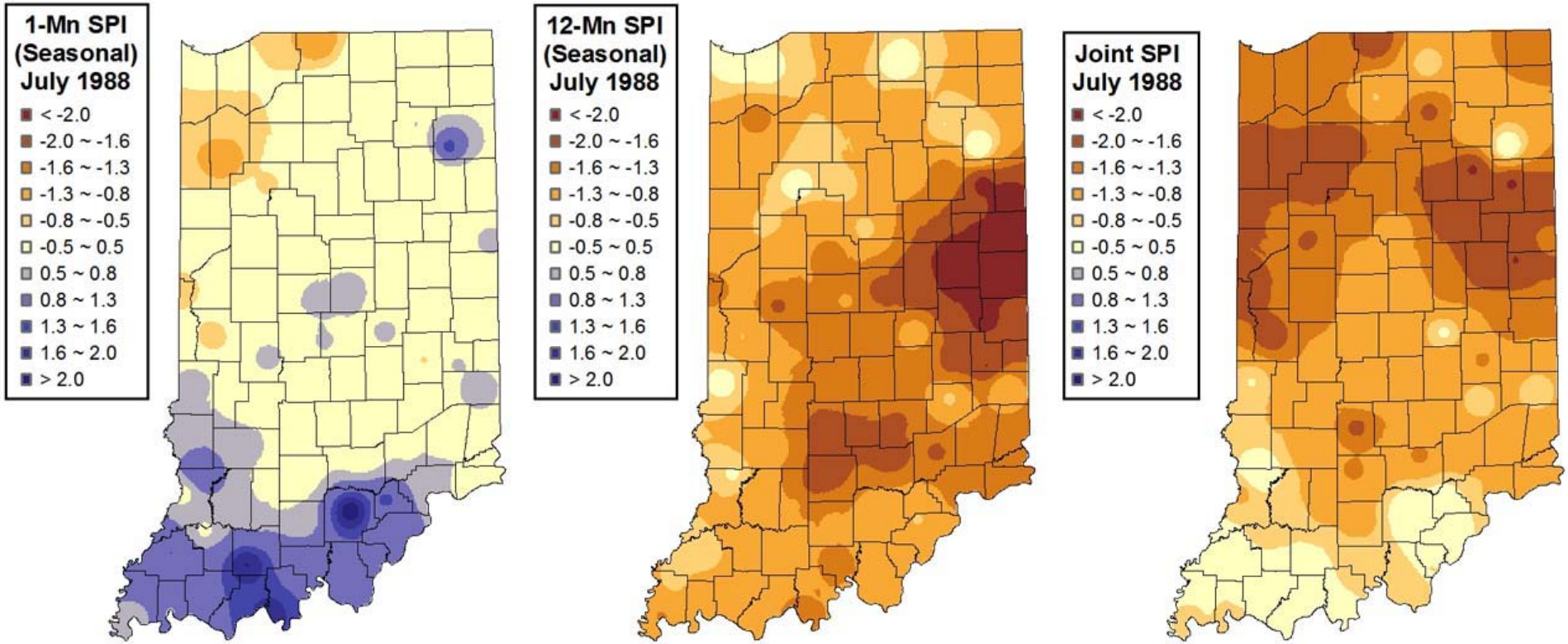
- Joint deficit index (JDI)
  - $JDI = \Phi^{-1}(K_C)$
  - Share the same classification with SI

Joint Deficit Index (II)

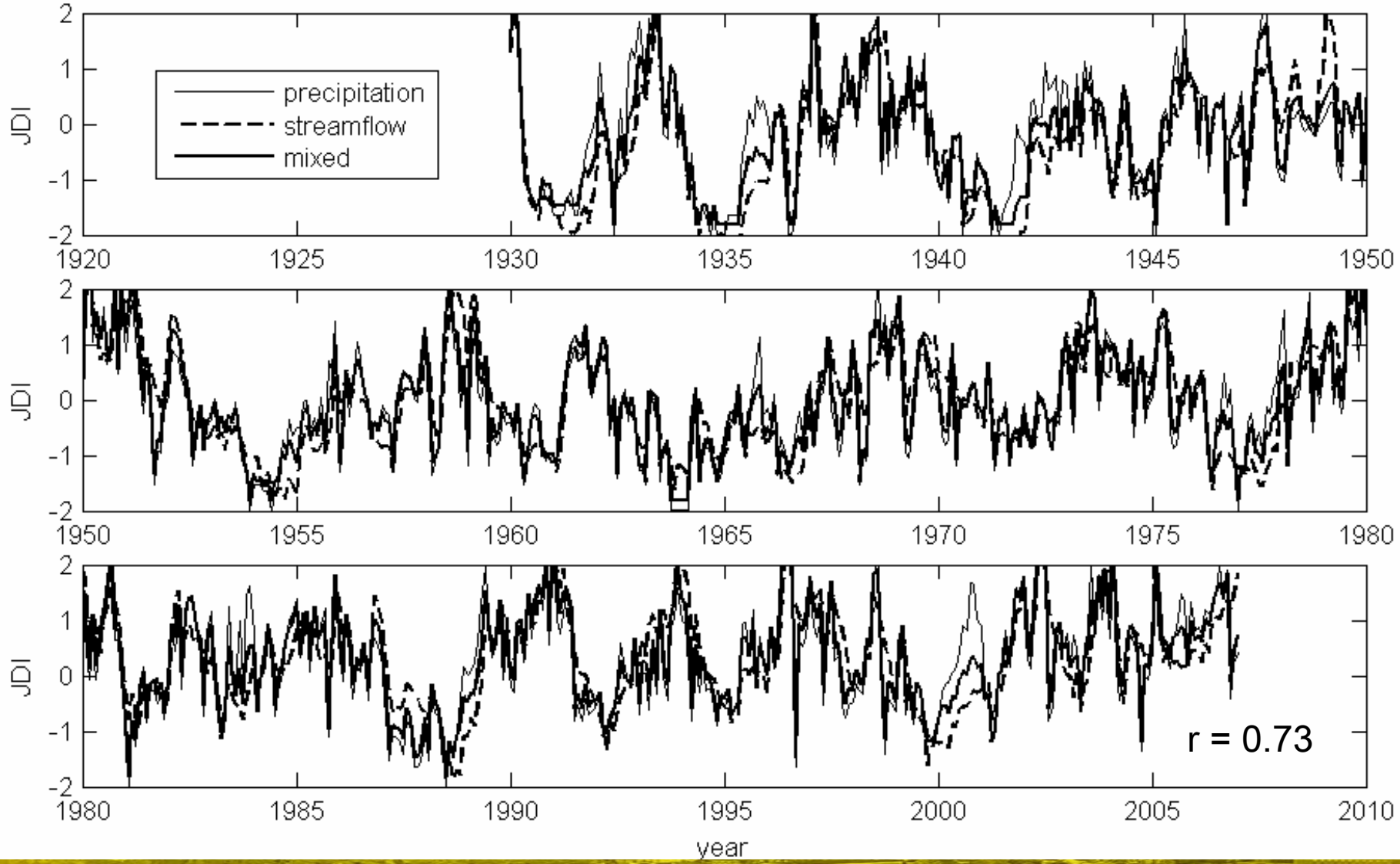


Joint Deficit Index (III)

- 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

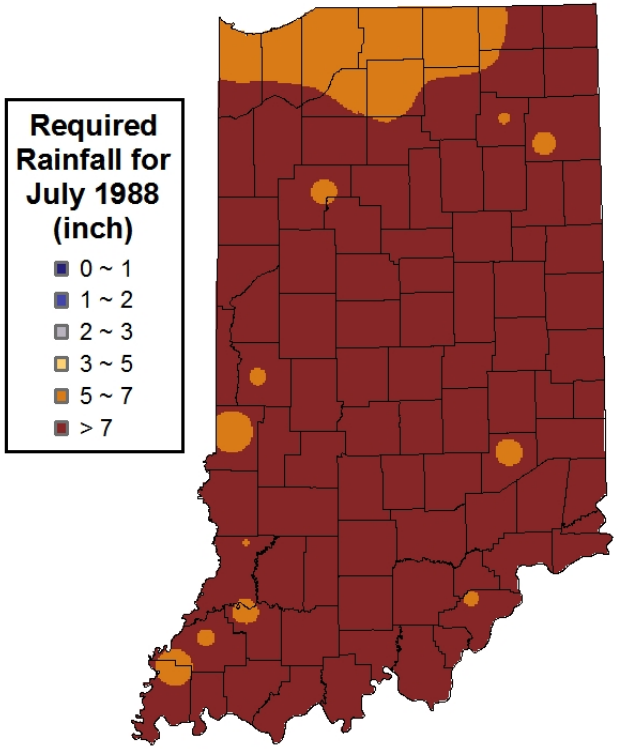


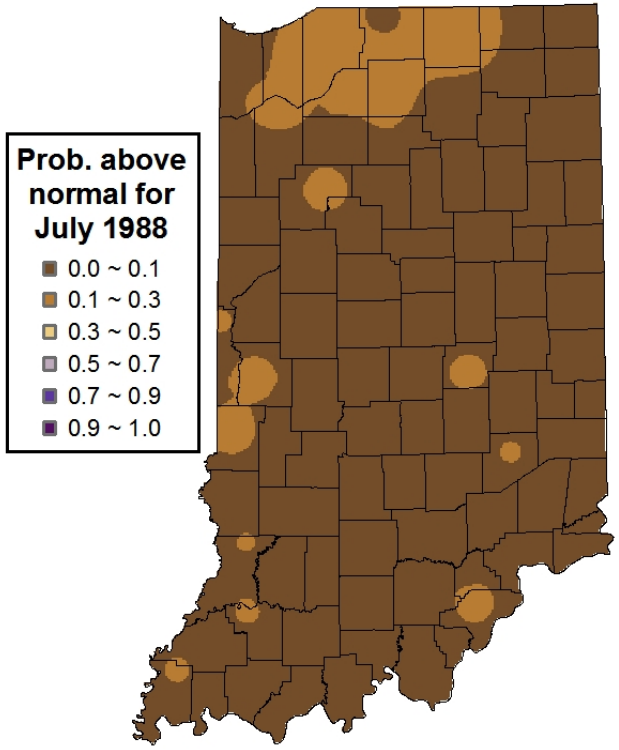
Precipitation vs. Streamflow



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?