Statistical Analysis of Extreme Rainfall Events over Indiana, USA

Shih-Chieh Kao & Rao S. Govindaraju
Purdue University
May 17, 2007

Background and Motivation

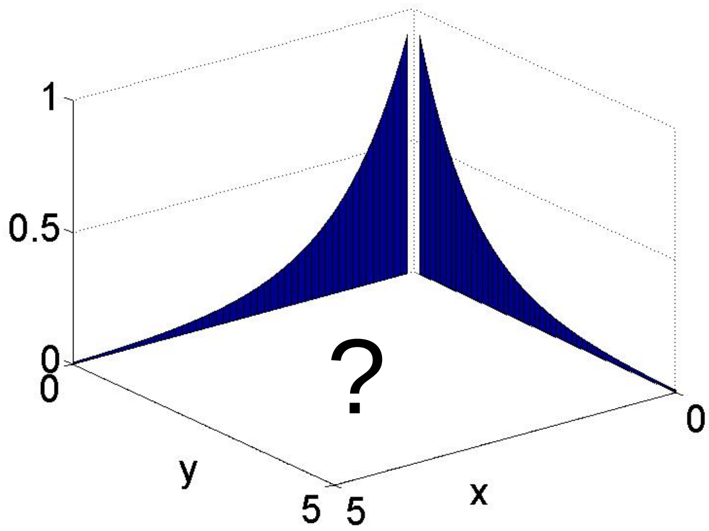
- Extreme rainfall behavior
  - Basis for hydrologic design
  - Conventionally analyzed only by “depth”
  - Pre-specified artificial duration (filter), *not* the real duration of extreme rainfall event
  - Hard to represent other rainfall characteristics, e.g. peak intensity
- Definition of extreme event in multi-variate sense is not clear
- Dependence exists between rainfall characteristics (e.g. volume(depth), duration, peak intensity)
- Explore the use of copulas

Difficulties in Constructing joint Distribution

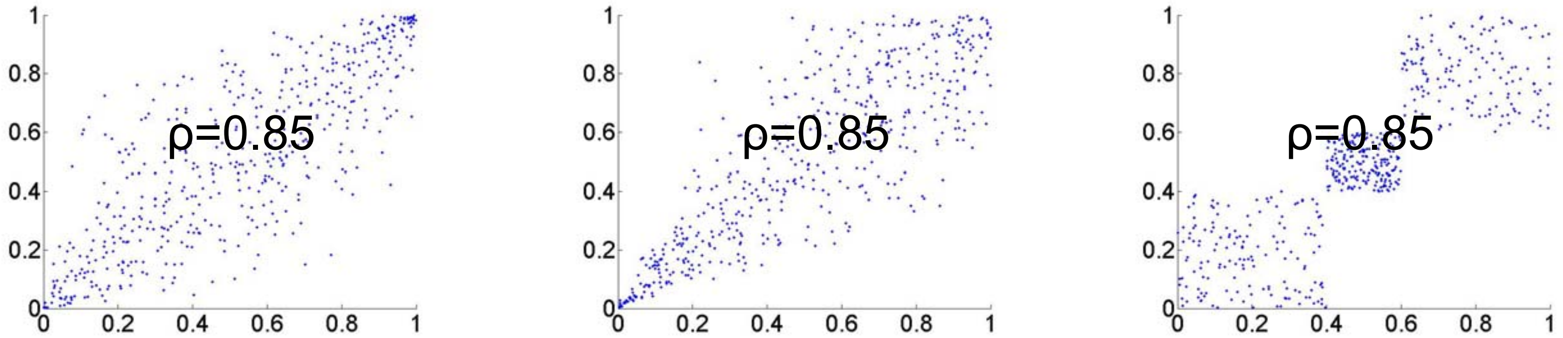
- Given marginal distributions and correlation coefficient, can joint-distribution be constructed?

  ex: exponential margins

  => Unlimited choices!!



- Correlation coefficient can not correctly describe association between variables

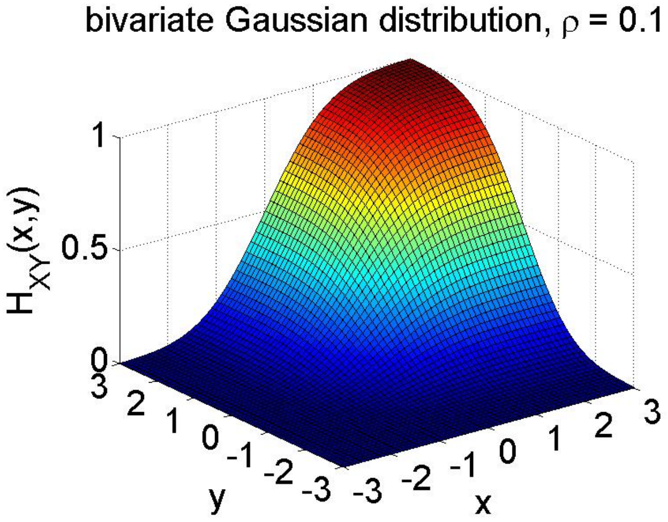


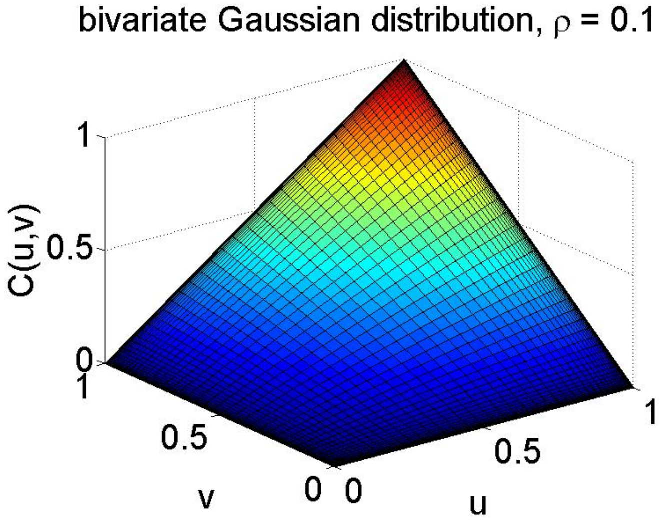
Introduction to Copulas

- A copula C(u,v) is a function comprised of margins $u=F_X(x)$ & $v=F_Y(y)$ from $[0,1]\times[0,1]$ to $[0,1]$.
  - Sklar (1959) showed that for continuous marginals u and v, there exists a unique copula C such that

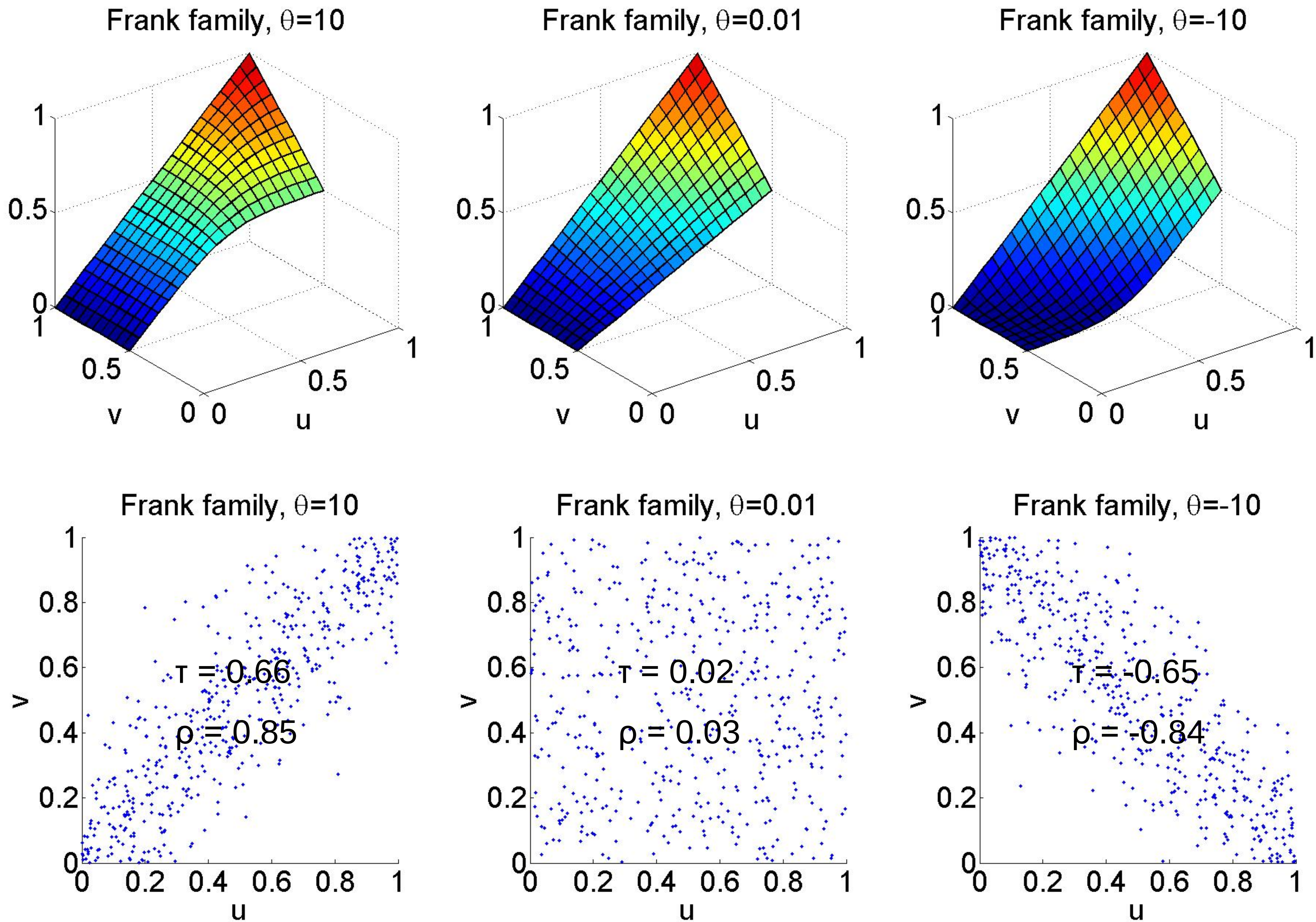
$$H_{XY}(x,y) = C_{UV}\left(F_X(x), F_Y(y)\right) = C_{UV}(u,v)$$

  - Transformation from $[-\infty,\infty]^2$ to $[0,1]^2$





  - Provides a complete description of dependence structure



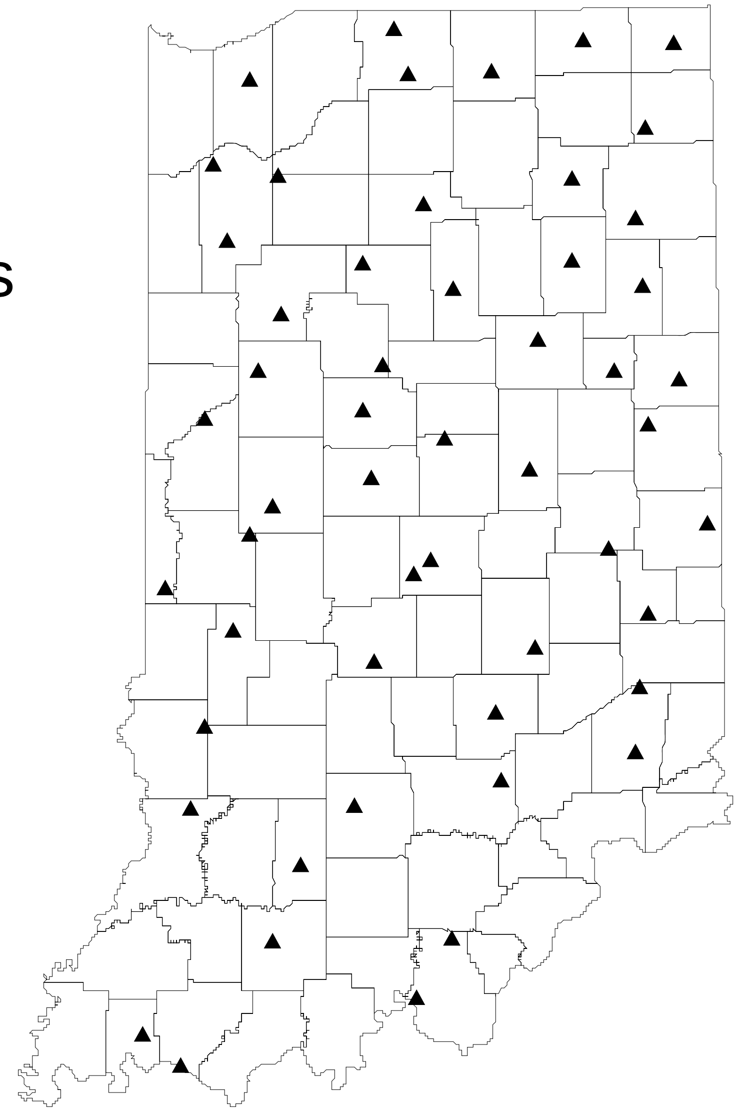
Frank family, θ=10
Frank family, θ=0.01
Frank family, θ=-10
u
v
Frank family, θ=10
Frank family, θ=0.01
Frank family, θ=-10
τ = 0.66
ρ = 0.85
τ = 0.02
ρ = 0.03
τ = -0.65
ρ = -0.84
u
v

Applications of Copulas in Hydrology

- Flood frequency analysis
  - Favre *et al*. (2004), De Michele *et al*. (2005), Grimaldi and Serinaldi (2006), Zhang and Singh (2006)
- Return period assessment using bivariate model
  - Salvadori and De Michele (2004)
- Probablistic structure of storm surface runoff
  - Kao and Govindaraju (2007)
- Rainfall frequency analysis
  - De Michele and Salvadori (2003, 2006), Grimaldi and Serinaldi (2006), Zhang and Singh (2006)
- Unanswered questions in rainfall frequency analysis
  - Data used for analysis may not be sufficient
  - Definition of “extreme events” in multi-variate sense?
  - Can results be applied for a large region?

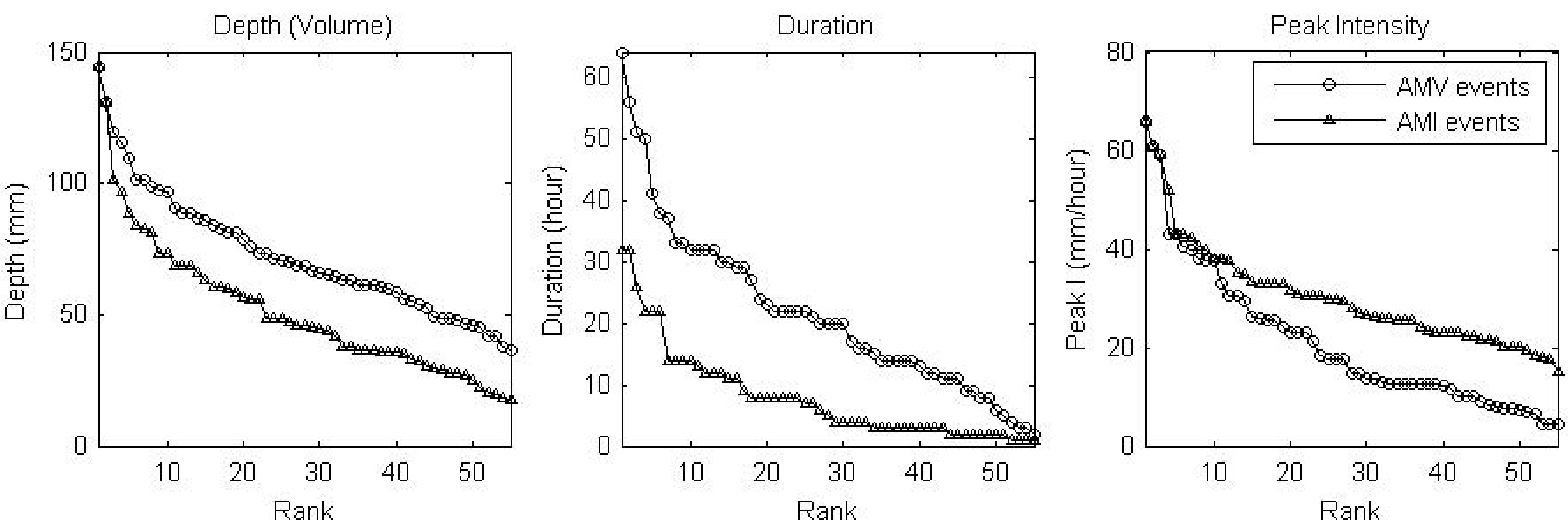
Data Source & Study Area

- Nation Climate Data Center, Hourly Precipitation Dataset (NCDC, TD 3240 dataset)
- 53 Co-operative Rainfall Stations in Indiana with record length greater than 50 years
- Minimum rainfall hiatus: 6 hours
- About 4800 events per station
- Selected variables for analysis:
  - Depth (volume), P (mm)
  - Duration, D (hour)
  - Peak Intensity, I (mm/hour)
- Marginals:
  - $u=F_P(p)$, $v=F_D(d)$, $w=F_I(i)$



Definitions of Extreme 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



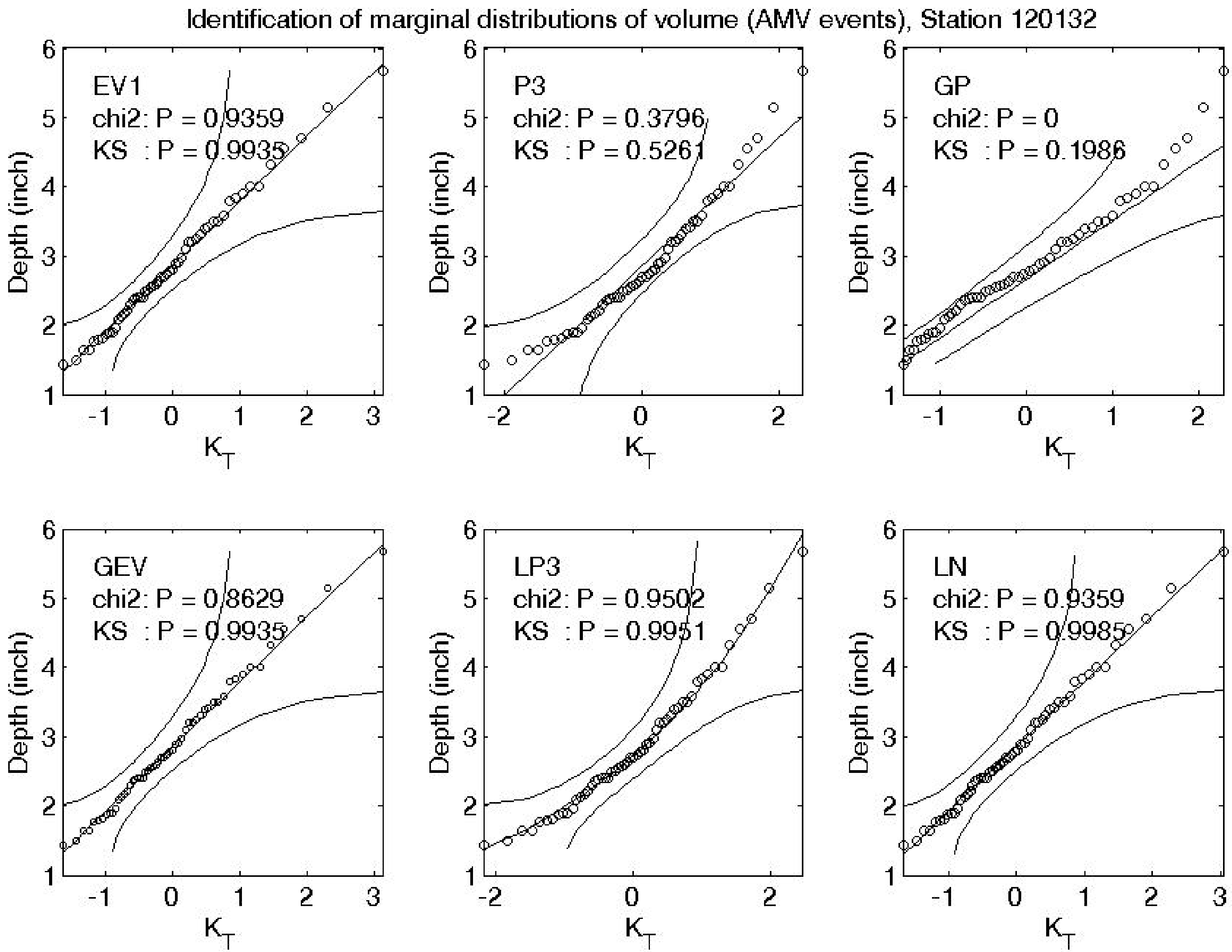
Analysis of Marginal Distributions (I)

- Candidate distributions
  - Extreme value type I (EV1)
  - Generalized extreme value (GEV)
  - Pearson type III (P3)
  - Log-Pearson type III (LP3)
  - Generalized Pareto (GP)
  - Log-normal (LN)

- Parameters estimated primarily by maximum likelihood (ML) or method of moments (MOM)

- Gringorton formula for empirical probabilities

- Chi-square and Kolmogorov-Smirnov (KS) test with 10% significance level


Identification of marginal distributions of volume (AMV events), Station 120132
EV1
chi2: P = 0.9359
KS : P = 0.9935
P3
chi2: P = 0.3796
KS : P = 0.5261
GP
chi2: P = 0
KS : P = 0.1986
GEV
chi2: P = 0.8629
KS : P = 0.9935
LP3
chi2: P = 0.9502
KS : P = 0.9951
LN
chi2: P = 0.9359
KS : P = 0.9985
Depth (inch)
K_T

Analysis of Marginal Distributions (II)

AMV events	Rejection rate (%) of Chi-square test						Rejection rate (%) of KS test					
	EV1	GEV	P3	LP3	GP	LN	EV1	GEV	P3	LP3	GP	LN
Depth, P	13.2	17.0	41.5	17.0	100	13.2	0.0	0.0	7.5	0.0	52.8	0.0
Duration, D	13.2	15.1	24.5	37.7	100	22.6	1.9	0.0	7.5	0.0	22.6	0.0
Intensity, I	15.1	17.0	45.3	20.8	100	11.3	0.0	0.0	1.9	0.0	54.7	0.0

AMI events	Rejection rate (%) of Chi-square test						Rejection rate (%) of KS test					
	EV1	GEV	P3	LP3	GP	LN	EV1	GEV	P3	LP3	GP	LN
Depth, P	5.7	3.8	62.3	3.8	100	1.9	0.0	0.0	11.3	0.0	45.3	0.0
Duration, D	60.4	39.6	88.7	37.7	100	28.3	15.1	0.0	45.3	0.0	45.3	0.0
Intensity, I	15.1	15.1	34.0	18.9	100	15.1	0.0	0.0	5.7	0.0	71.7	0.0

- EV1, GEV, LP3, LN provided better fit. GP provided the worst.
- Fitting for duration of AMI events did not yield very good result due to the limited temporal precision (hour)
- EV1 and LN could be recommended for use

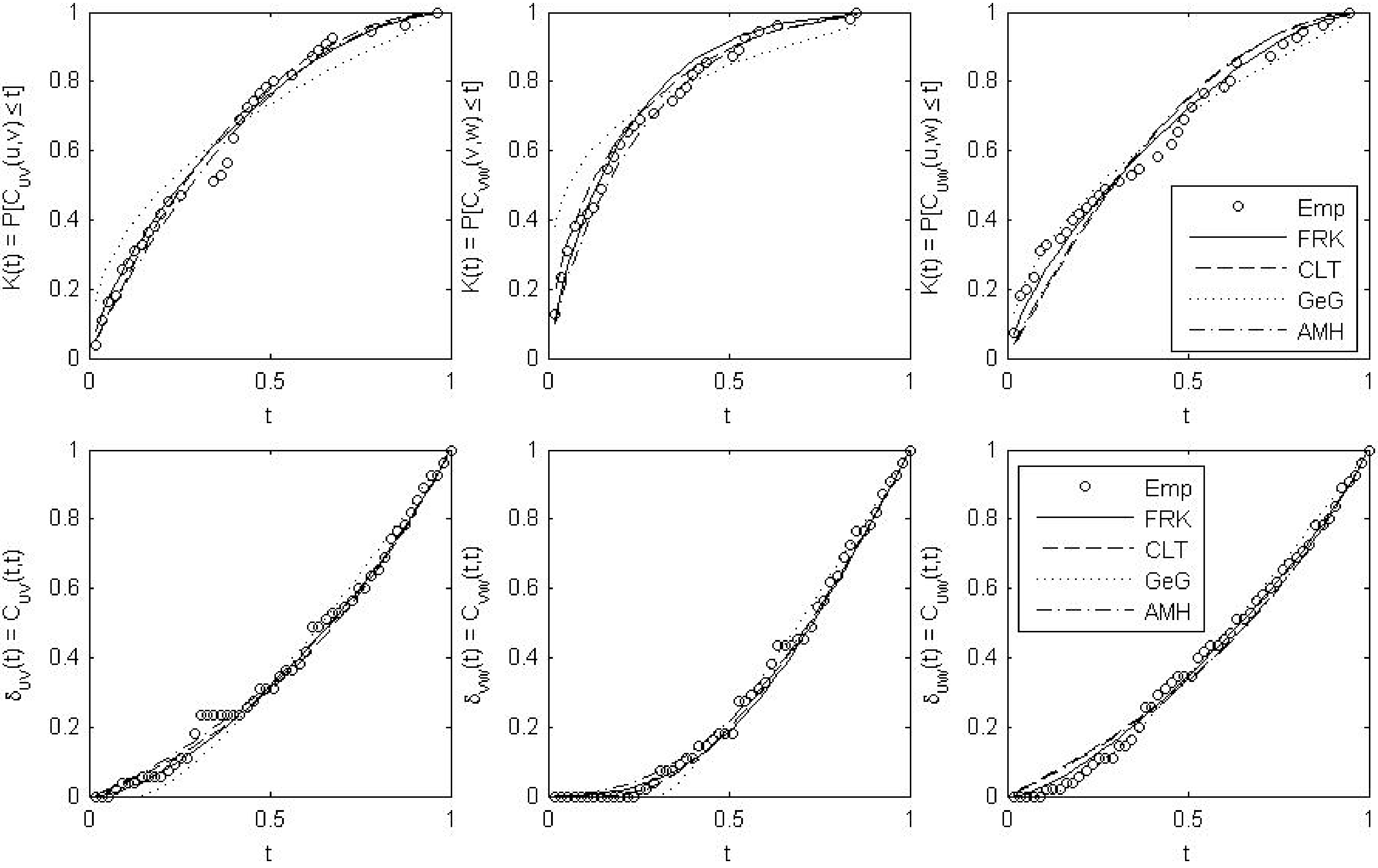
Analysis of Dependence Structure (I)

- Candidate Archimedean copulas
  - Frank family
  - Clayton family
  - Genest-Ghoudi family
  - Ali-Mikhail-Haq family

- Non-parametric procedure for estimating dependence parameter

- Examination of Goodness-of-fit
  - Distribution function $K_C(t)=P[C(u,v) \leq t]$
  - Diagonal section of copulas $\delta(t)=C(t,t)$
  - Multidimensional KS test (Saunders and Laud, 1980)

Assessment of Copula Performance



Analysis of Dependence Structure (II)

- Statistics of Kendall's tau

	τ_{PD}		τ_{DI}		τ_{PI}	
	mean	stdev	mean	stdev	mean	stdev
AMV events	0.183	0.084	-0.370	0.068	0.260	0.097
AMI events	0.407	0.070	-0.011	0.096	0.405	0.070

- The distribution function $K_C(t)$ provides the strictest examination of copulas
- Clayton and Ali-Mikhail-Haq families performed well for positive dependence cases (C_{UV} and C_{UW})
- Frank family of Archimedean copulas
  - performed well for both positive and negative dependence
  - passed the KS test for entire Indiana at the 10% significant level
  - recommended for use in practice

Construct Joint Distribution via Copulas

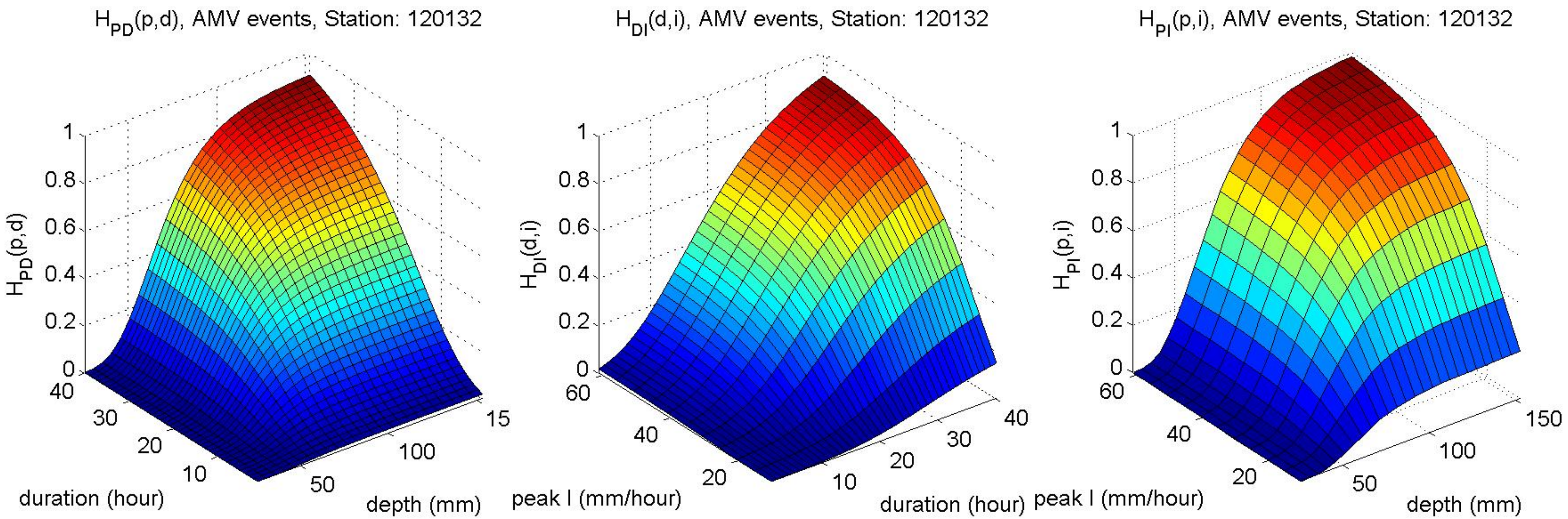
- Bivariate stochastic models

$$H_{PD}(p,d)=C_{UV}(F_P(p),F_D(d))=C_{UV}(u,v)$$

$$H_{DI}(d,i)=C_{VW}(F_D(d),F_I(i))=C_{VW}(v,w)$$

$$H_{PI}(p,i)=C_{UW}(F_P(p),F_I(i))=C_{UW}(u,w)$$

- Examples using Frank family and EV1 marginals



Estimate of depth for known duration (I)

- For a known (or measured) d-hour event

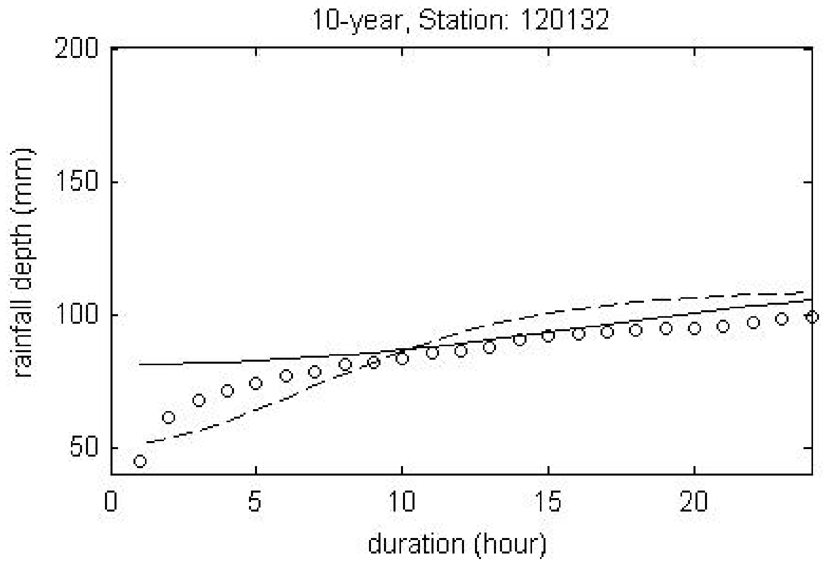
$$F_P\left(p \middle| d-1 < D \le d\right) = \frac{H_{PD}(p,d) - H_{PD}(p,d-1)}{F_D(d) - F_D(d-1)}$$
$$= \frac{C_{UV}\left(F_P(p), F_D(d)\right) - C_{UV}\left(F_P(p), F_D(d-1)\right)}{F_D(d) - F_D(d-1)}$$

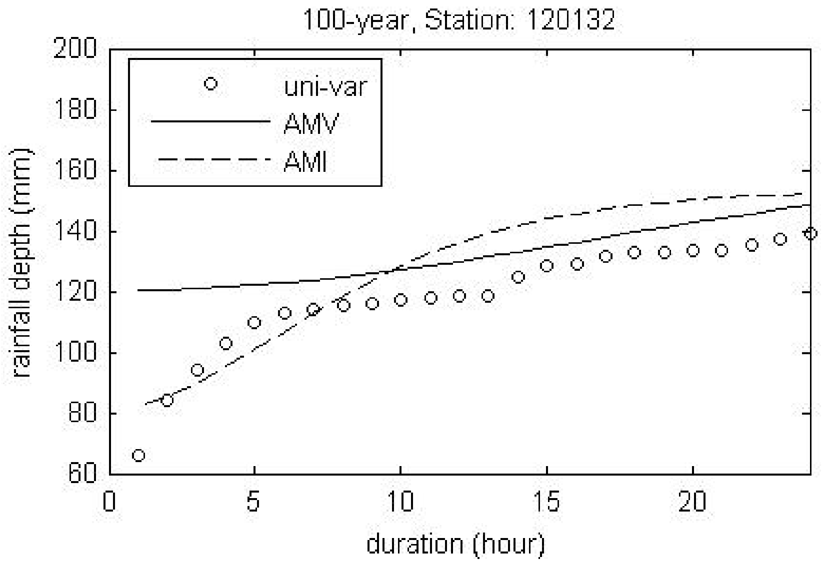
- Given return period T, the T-year, d-hour rainfall estimate p_T will satisfy

$$F_P\left(p_T \middle| d-1 < D \le d\right) = 1 - 1/T$$

- Comparison between bivariate and univariate depth estimates
  - Bivariate using EV1 marginals and Frank family
  - Univariate counterpart using GEV distribution (Rao and Kao, 2006)

Estimate of depth for known duration (II)





- Similar trends were observed for durations greater than 10-hour, close to the univariate counterpart
- For durations less than 10-hour
  - AMV estimates gave higher values
  - AMI estimates should be better, but fitting problem existed
- Overall, less parameters required for bivariate estimates

Conclusions (I)

- Definition of extreme events
  - AMV events are generally of longer duration than AMI events
  - AMV estimates should be better for longer duration, while AMI estimates being better for shorter duration
  - AMI results could be improved by adopting data with finer temporal resolutions

- Marginal distributions
  - EV1, GEV, LP3, LN were found to be appropriate marginal models for extreme rainfall
  - GP was found not appropriate for Indiana extreme rainfall data
  - EV1 and LN are recommended

Conclusions (II)

- Dependence structure
  - Between P and D, positive correlated
  - Between D and I, generally negative correlated
  - Between P and I, positive correlated
  - Frank family is recommended

- Estimate of depth for known duration
  - Similar results for durations larger than 10 hours
  - AMI estimates should be better for durations less than 10 hours
  - Bivariate approach is more flexible than univariate approach

Thank you for listening.
Questions?