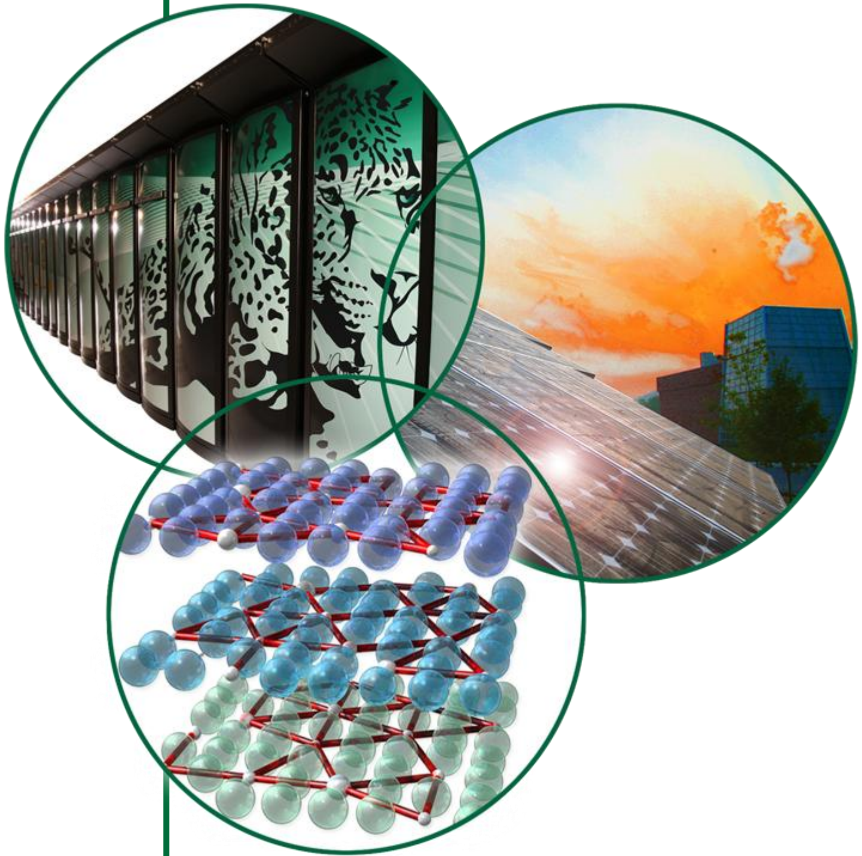
Multivariate Flood Frequency Analysis through Copulas in a Partially Gauged Watershed

Presented at
World Environmental & Water Resources Congress
20rd – 24th May, 2012, Albuquerque, NW

Shih-Chieh Kao (kaos@ornl.gov)
Environmental Sciences Division
Oak Ridge National Laboratory

Ni-Bin Chang (nchang@ucf.edu)
Department of Civil, Environmental, and Construction Engineering
University of Central Florida



U.S. DEPARTMENT OF ENERGY

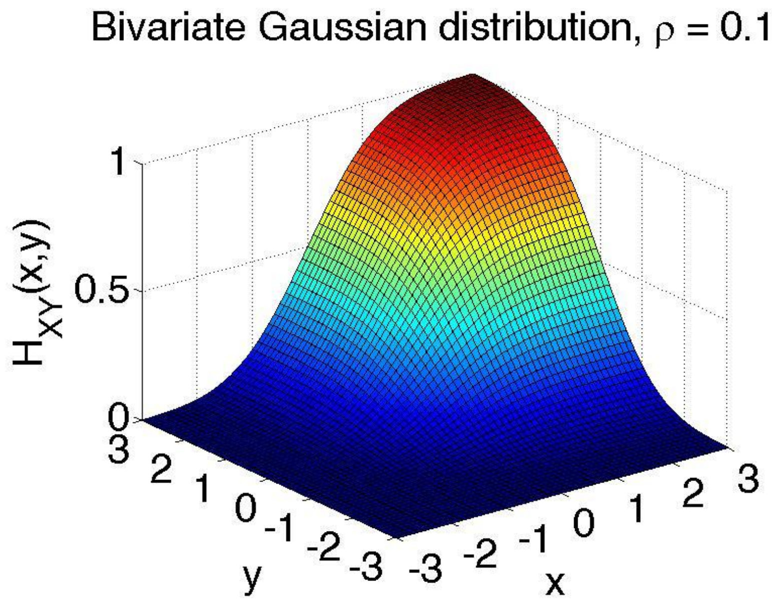
OAK RIDGE NATIONAL LABORATORY
MANAGED BY UT-BATTELLE FOR THE DEPARTMENT OF ENERGY

Background

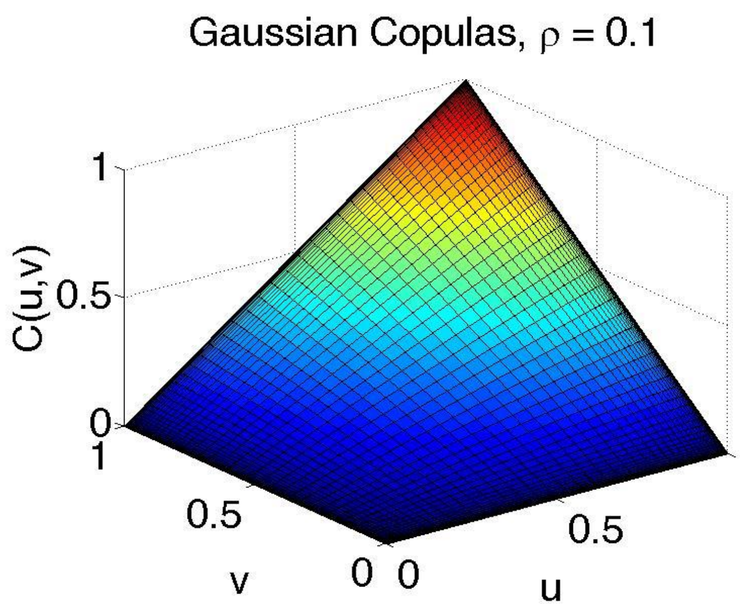
- 25000 USGS Gauge Stations
  - Enough?
- Estimate Flood Frequency at Ungauged Locations
  - Modeling Approach
  - Statistical Approach
- Limitations of the Univariate Flood Frequency Analysis
  - How to account for river confluences?
  - What if a river has been partially regulated?
  - How to account for major land use and land cover change?
- Multivariate Flood Frequency Analysis Could be a Solution
  - *But, can we make it easier?*

Joint Distribution and Copulas

- One may formulate any joint distribution in terms of copulas and marginals
  - $H_{XYZ}(x,y,z) = C_{UVW}(u,v,w)$
    $u = F_X(x)$, $v = F_Y(y)$, $w = F_Z(z)$
  - Copulas is a "distribution-free" dependence structure

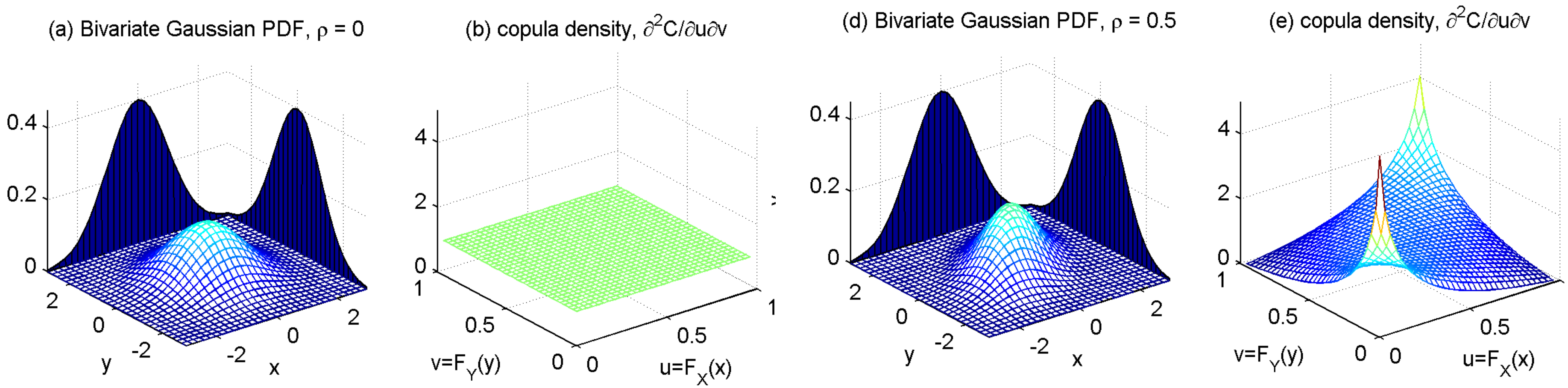


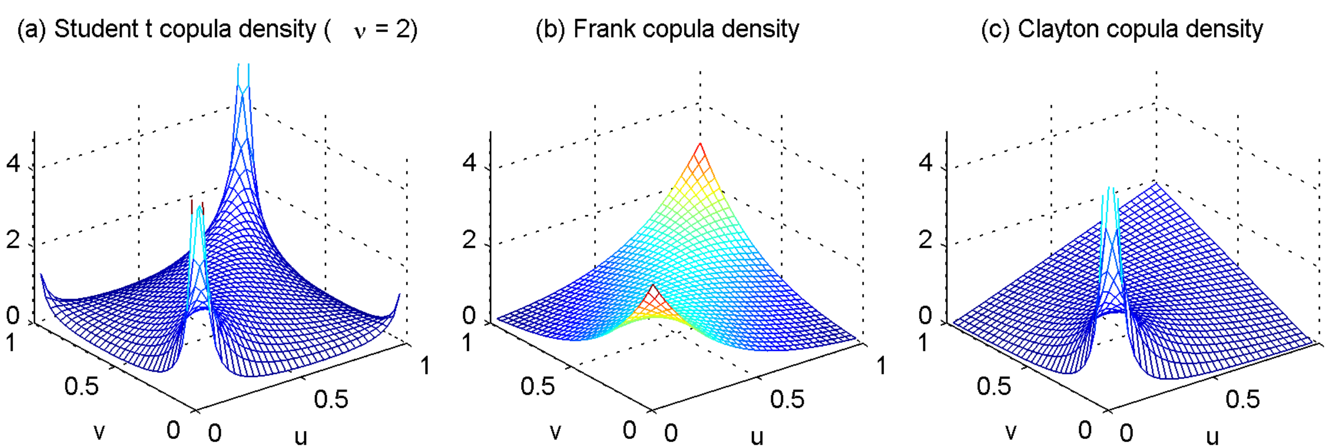
- Use copulas to construct joint distributions
  - Marginal distributions => selecting suitable PDFs
  - Dependence structure => selecting suitable copulas
  - Together they form joint distribution with no specific marginals



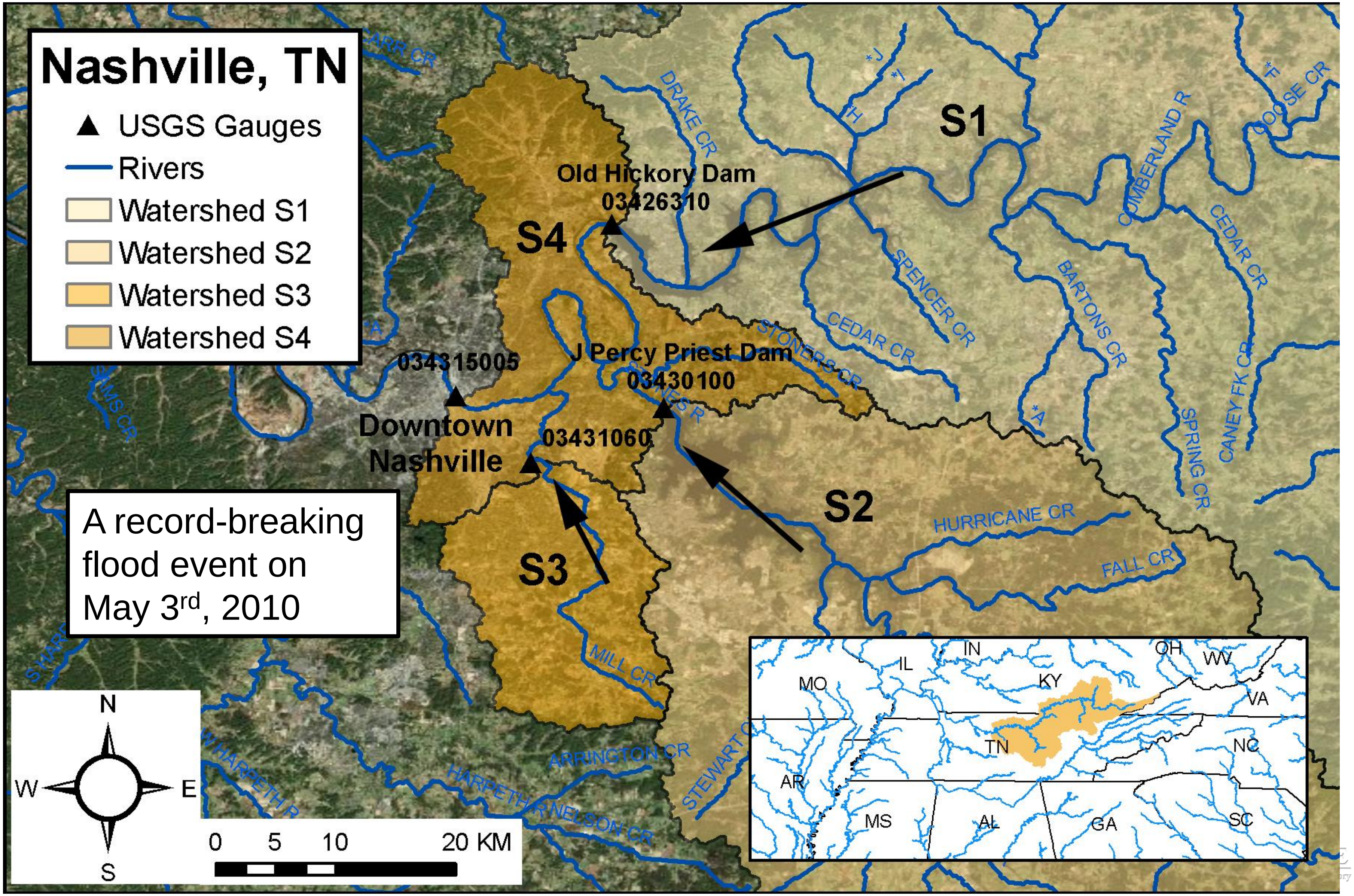
Copula Density

- **Joint PDF versus copula density**
  - Positive dependence: main diagonal (u = v)
  - Independence: flat surface
  - Negative dependence: secondary diagonal (u - v = 1)





Case Study

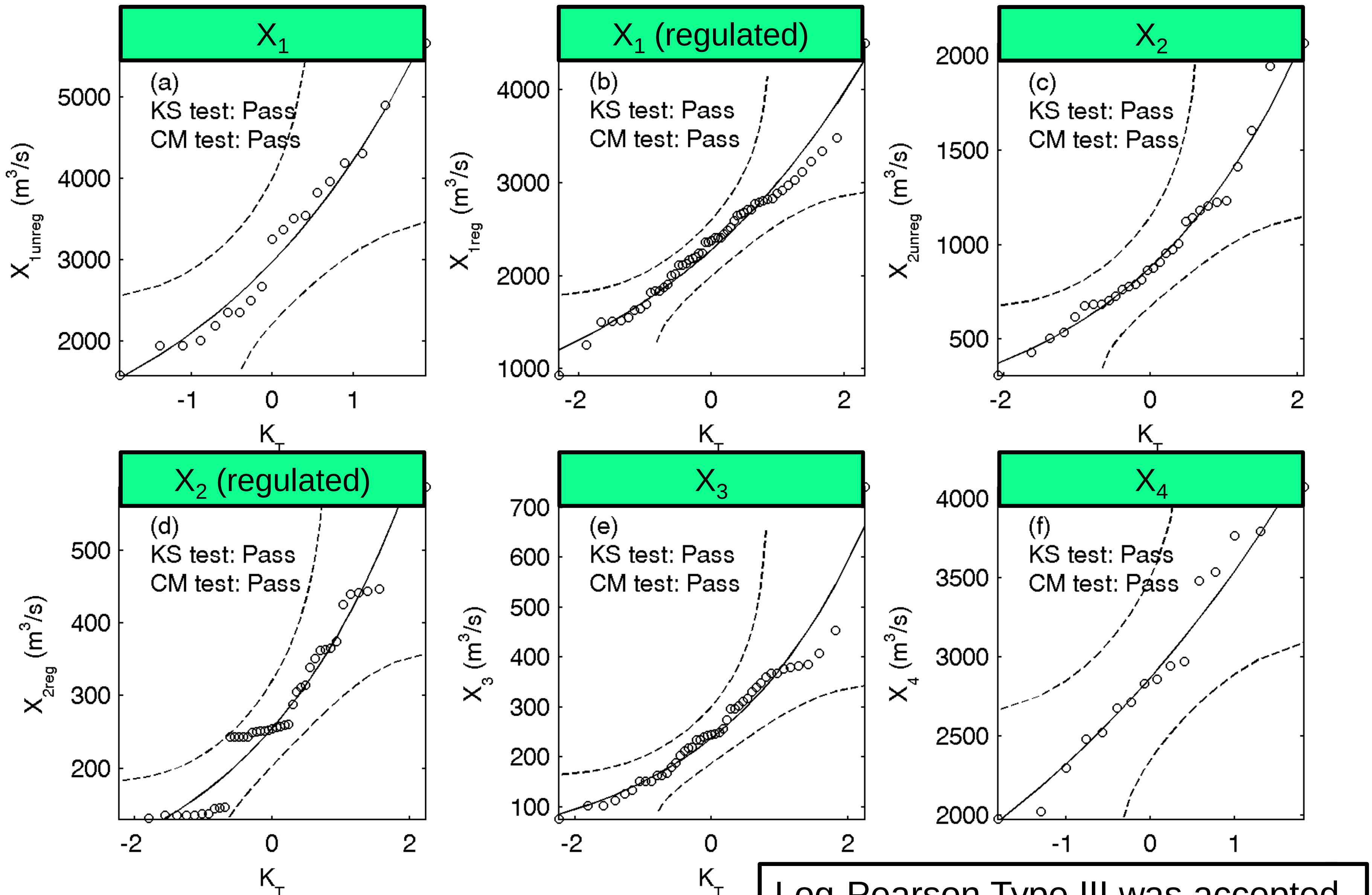


Nashville, TN
USGS Gauges
Rivers
Watershed S1
Watershed S2
Watershed S3
Watershed S4
S1
S2
S3
S4
Old Hickory Dam
03426310
J Percy Priest Dam
03430100
03431060
034315005
Downtown Nashville
A record-breaking flood event on May 3rd, 2010
N
S
E
W
0 5 10 20 KM
DRAKE CR
SPENCER CR
CEDAR CR
STONERS CR
STONES R
BARTONS CR
CUMBERLAND R
CEDAR CR
CANEY FK CR
SPRING CR
GOOSE CR
HURRICANE CR
FALL CR
MILL CR
ARRINGTON CR
STEWART CR
HARPETH R
NELSON CR
W HARPETH R
SIMS CR
CARR CR
IN
IL
OH
WV
MO
KY
VA
TN
NC
AR
MS
AL
GA
SC

Data Availability

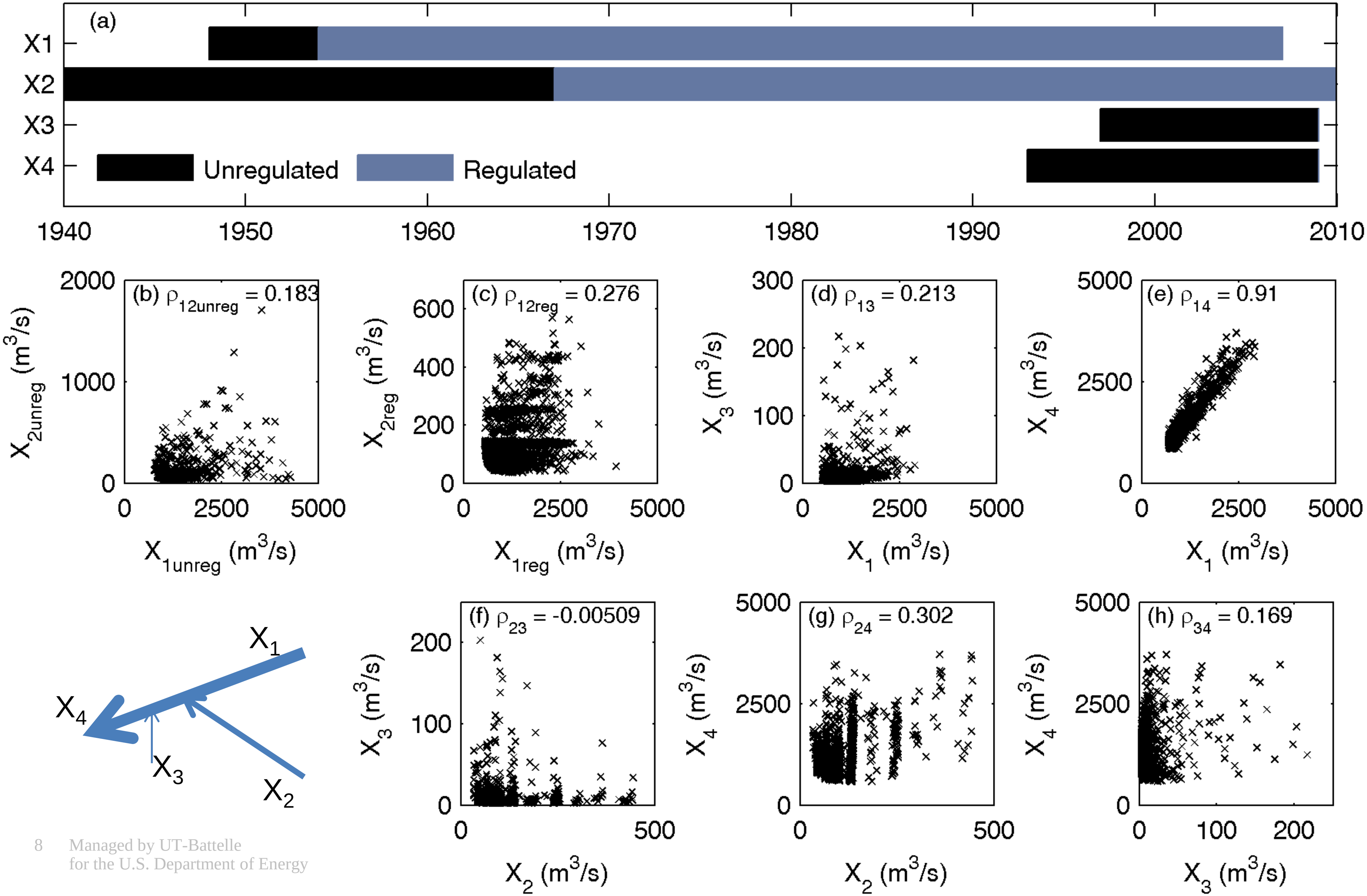
	X_1	X_2	X_3	X_4
USGS ID	03426310	03430100	03431060	034315005
Gage Name	Cumberland River at Old Hickory Dam	Stones River below J Percy Priest Dam	Mill Creek at Thompson Lane near Woodbine	Cumberland River at Woodland St at Nashville
Drainage Area (km^2)	30233	2310	241.9	33307
Corresponding Watersheds	S_1	S_2	S_3	S_1, S_2, S_3, and S_4
Data Coverage	WY1948~WY2007	WY1940~WY2010	WY1997~WY2009	WY1993~WY2009
# of Annual Peak Flow	19 (pre-regulated) 53 (pos-regulated)	30 (pre-regulated) 43 (post-regulated)	45 (peak flow since WY1965)	16
Mean Annual Flow (m^3/s)	526.12	39.93	4.08	589.98
Note	Old Hickory Dam regulated at 1954	J Percy Priest Dam regulated at 1967		

Fitting of Marginal Distributions



Log-Pearson Type III was accepted

Correlation between High Flow Pairs



Flow Synthesization through Copulas

- Gaussian copulas is chosen for simplification
  - Multivariate normal distribution (MVN)

$$\Phi_d(\phi^{-1}(u_1),...,\phi^{-1}(u_d)) = \int_{-\infty}^{\phi^{-1}(u_1)} \cdots \int_{-\infty}^{\phi^{-1}(u_d)} (2\pi)^{-\frac{d}{2}} |\mathbf{\Sigma}|^{-\frac{1}{2}} \exp(-\mathbf{z}\mathbf{\Sigma}^{-1}\mathbf{z}^T / 2) dz_1 \cdots dz_d$$

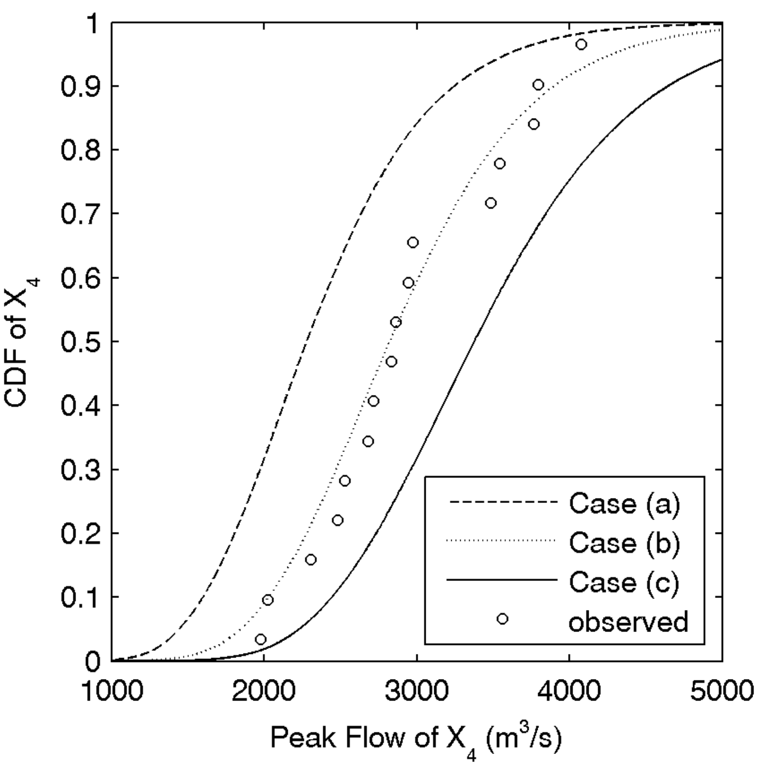
  - Gaussian Copulas

$$C_{U_1,...,U_d}(u_1,...,u_d) = \Phi_d(\phi^{-1}(u_1),...,\phi^{-1}(u_d)) == \Phi_d(\phi^{-1}(F_{X_1}(x_1)),...,\phi^{-1}(F_{X_d}(x_d)))$$

  - Existing MVN generators are easy to use
- Procedures
  - (1) Calculate the correlation matrix for MVN
  - (2) Generate 100,000 MVN samples
  - (3) Transform the MVN samples to Gaussian copulas, and then to different marginals
  - (4) The synthesized (x_1, x_2, x_3) flow are then used to estimate the flood frequency at downstream reaches.

Evaluation

- Three synthesizing functions were tested:
  - (1) $X_4 = X_1$
  - (2) $X_4 = X_1 + X_2 + X_3$
  - (3) $X_4 = X_1 + X_2 + w * X_3$
- Validate by observed X_4 flow
  - Function (2) works the best
  - More suitable function can be considered in the future

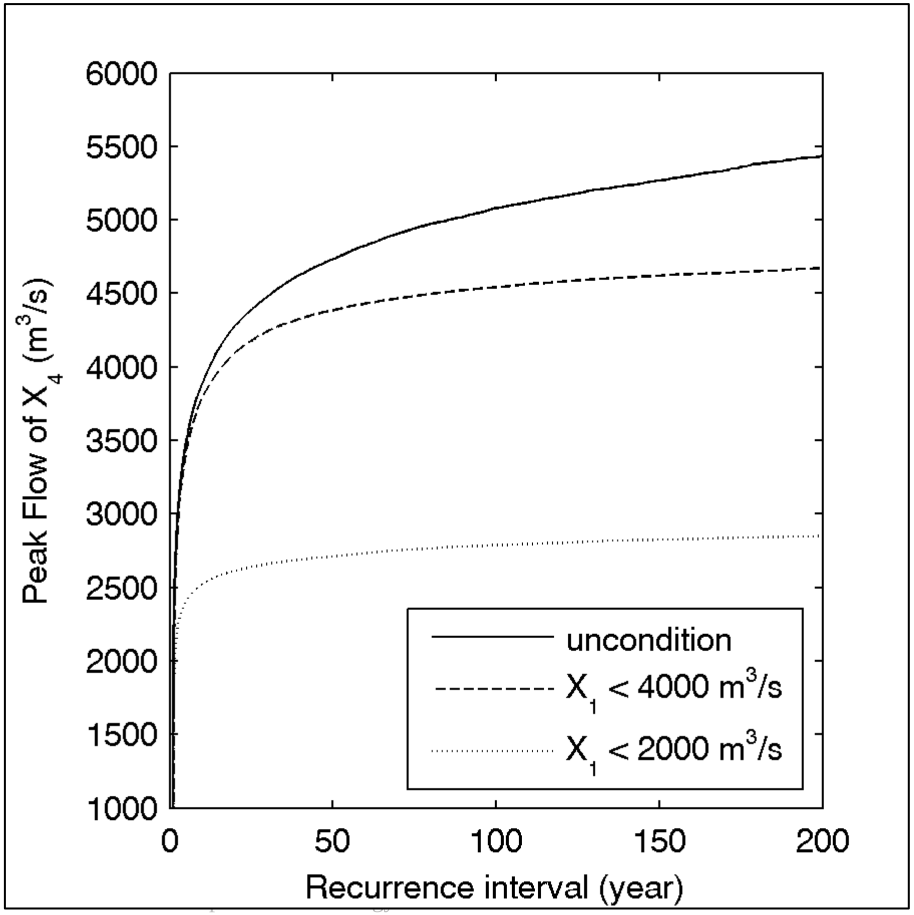


	KS (5%)	CM (5%)	Nash E	R^2	RMSE (m^3/s)	PE (%)	May 3rd, 2010
(a) X_1	Reject	Reject	0.0703	0.961	592.9	-20.5	885.3 yrs
(b) $X_1+X_2+X_3$	Accept	Accept	0.9363	0.964	155.2	-1.8	**166.3 yrs**
(c) $X_1+X_2+w^*X_3$	Reject	Reject	0.0923	0.963	585.9	17.5	29.5 yrs

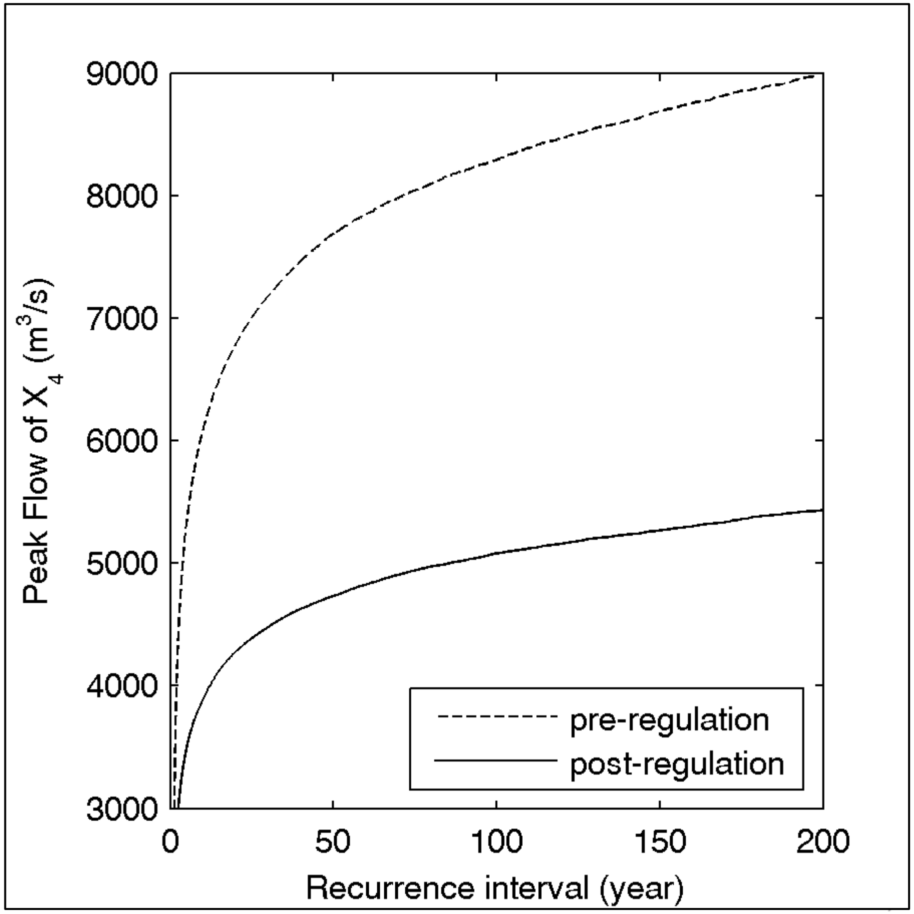
Flexibility of the Multivariate Approach

Frequency of Flood Considering Dam Regulation

$$H_{X_2,X_3|X_1\le a} = C_{U_2,U_3|U_1\le F_{X1}(a)}$$



Flood Frequency before and after dam regulation



Conclusions

- The multivariate flood frequency is more flexible, especially for river confluences considering dependence structure (comparing to the univariate statistical approach)
- It requires minimal data and moderate computation efforts (comparing to the modeling approach)
- Challenges and Future Works
  - What will be the best criteria to construct dependence structure for multivariate flood frequency analysis?
  - Dimensionality remains a major challenge. Gaussian (and t) copulas are the easiest but may not be the best solution.
  - How can we consider climate and land use / land cover change into this framework?

Thank you
Questions?

Shih-Chieh Kao
kaos@ornl.gov; http://www.ornl.gov/~5v1/