On the Probabilistic Structure of Storm Surface Runoff for Indiana

Shih-Chieh Kao
Purdue University
Sept 2006

Outline

- Background and Motivation
- Introduction to Copulas
- Literature Review of Copulas in Hydrology
- Data used in study
- Probabilistic Structure of Surface Runoff
- Conclusion

Background and Motivation

- Need to study the joint behavior between random variables, e.g. rainfall intensity and duration
- Single-variate frequency analysis is not capable of addressing multi-variate problems
- Assessing dependence structure
- Explore use of copulas
- Uncertainty of rainfall events induces uncertainty in storm surface runoff

Basic Probability Definitions

- For two random variables X & Y, we can define:
  - Joint Cumulative Density Function (j-CDF)

$$H_{XY}(x,y) = P[X \le x, Y \le y]$$

  - Joint Probability Density Function (j-PDF)

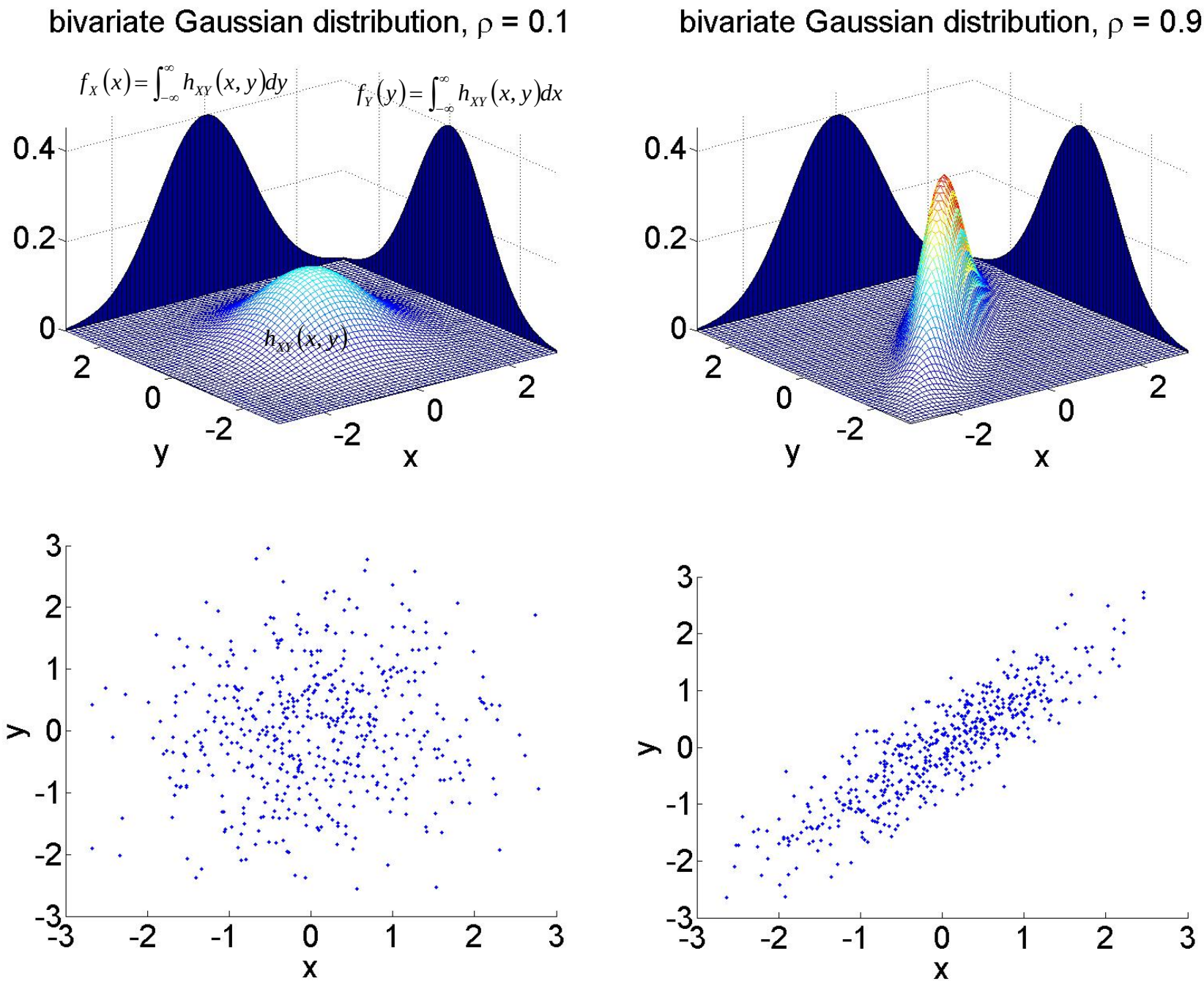
$$h_{XY}(x,y) = \frac{\partial^2 H_{XY}(x,y)}{\partial x \partial y}$$

- The Probability Density Function (PDF) of X can be obtained by integrating $h_{XY}(x,y)$ for the domain of Y

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

- Margins u & v of H_{XY} are defined as the Cumulative Density Functions (CDF) of X and Y

$$u = F_X(x) = \int_{-\infty}^{x} f_X(x)dx \qquad v = F_Y(y) = \int_{-\infty}^{y} f_Y(y)dy \qquad 0 \le u, v \le 1$$

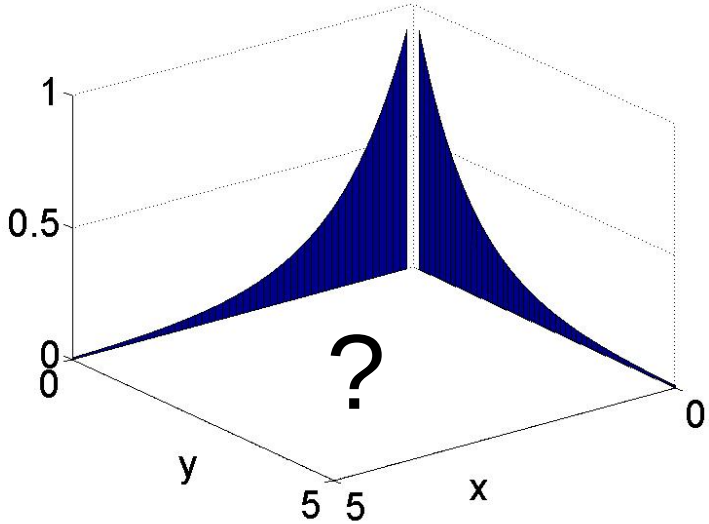

bivariate Gaussian distribution, ρ = 0.1
$f_X(x) = \int_{-\infty}^{\infty} h_{XY}(x,y)dy$
$f_Y(y) = \int_{-\infty}^{\infty} h_{XY}(x,y)dx$
$h_{XY}(x,y)$
0.4
0.2
0
2
0
-2
y
-2
0
2
x
bivariate Gaussian distribution, ρ = 0.9
0.4
0.2
0
2
0
-2
y
-2
0
2
x
3
2
1
0
-1
-2
-3
y
-3
-2
-1
0
1
2
3
x
3
2
1
0
-1
-2
-3
y
-3
-2
-1
0
1
2
3
x

Difficulties in Constructing joint Distribution

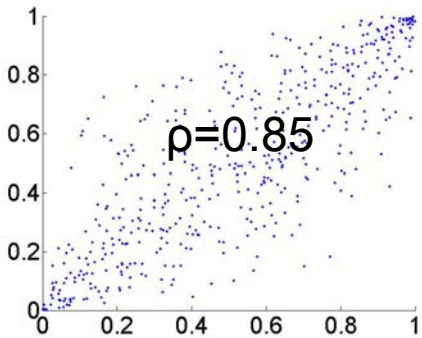
- Given marginal distributions and correlation coefficient, can joint-distribution be constructed?

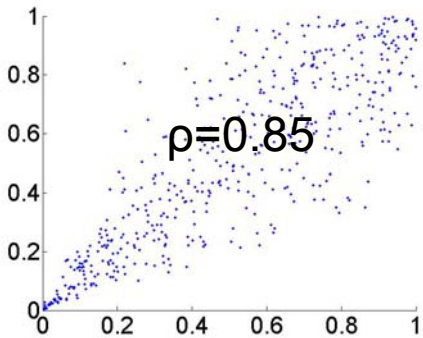
  ex: exponential margins

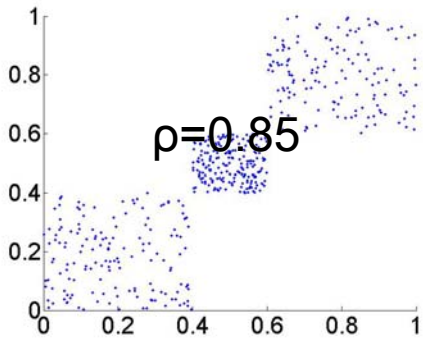
  => Unlimited choices!!



- Correlation coefficient can not correctly describe association between variables





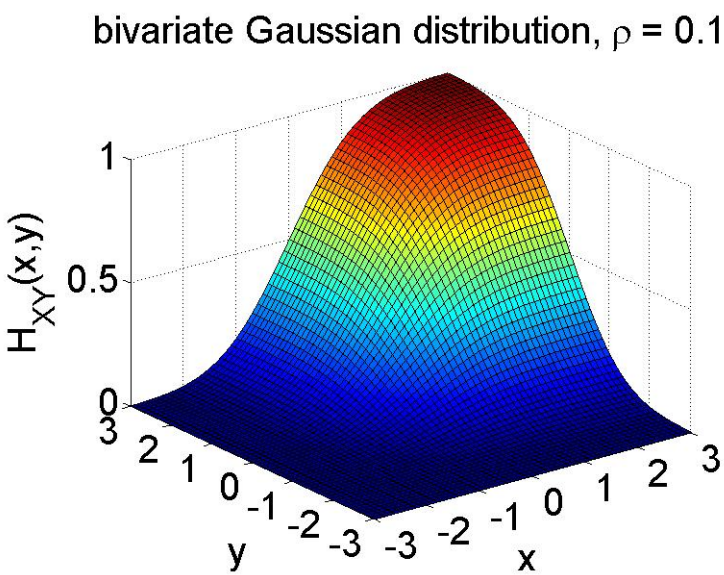


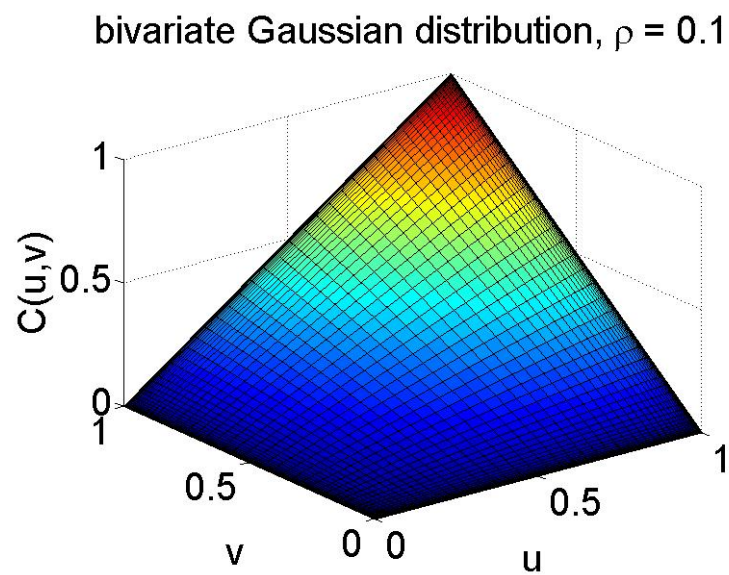
Introduction to Copulas (I)

- A Copula C(u,v) is a function comprised of margins u & v from [0,1]×[0,1] to [0,1].
- By Sklar's (1959) theorem, there exists a copula C such that

$$H_{XY}(x,y) = C(u,v)$$

If u and v are continuous, then C is unique.





Introduction to Copulas (II)

- Properties of C(u,v):

  (1) $C(u,0) = C(0,v) = 0$

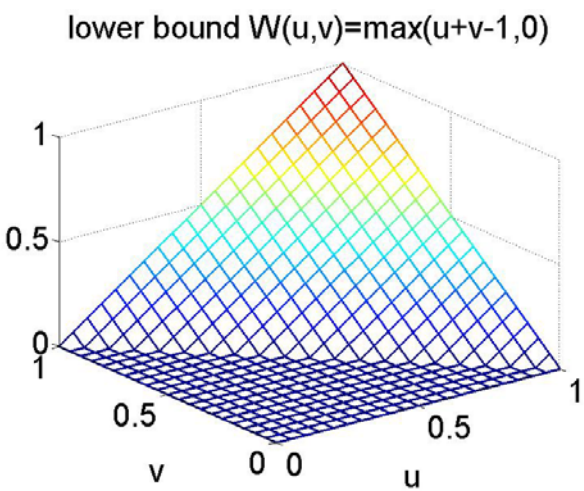
  (2) $C(u,1) = u$; $C(1,v) = v$

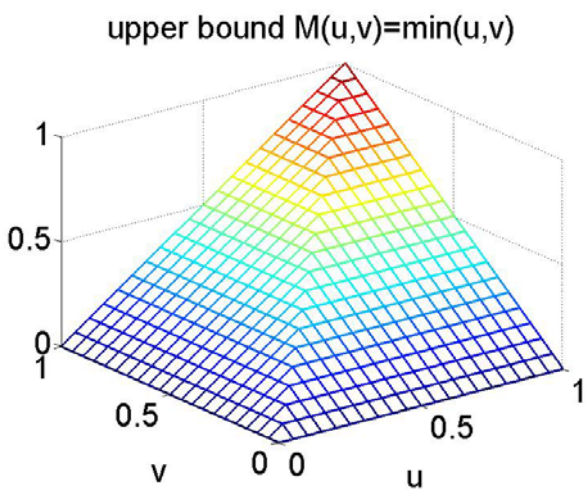
  (3) C(u,v) is a 2-increasing function

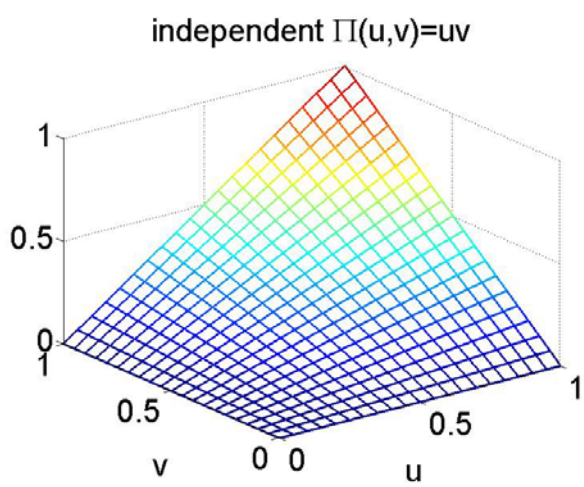
  For every $u_1 \le u_2$, $v_1 \le v_2$,

  $C(u_2,v_2)-C(u_1,v_2)-C(u_2,v_1)+C(u_1,v_1) \ge 0$

- Boundaries of copulas







Introduction to Copulas (III)

- Archimedean Copulas
  - There exists a generator φ(t), such that

$$\varphi(C(u,v)) = \varphi(u) + \varphi(v)$$

  - When φ(t) = ln(t), C(u,v) = uv. (Independent case)
  - Commonly used 1-parameter Archimedean Copulas:
    - Gumbel-Barnett family

$$\varphi(t) = \ln(1 - \theta \ln t) \qquad C(u,v) = uv \exp(-\theta \ln u \ln v)$$

    - Ali-Mikhail-Haq family

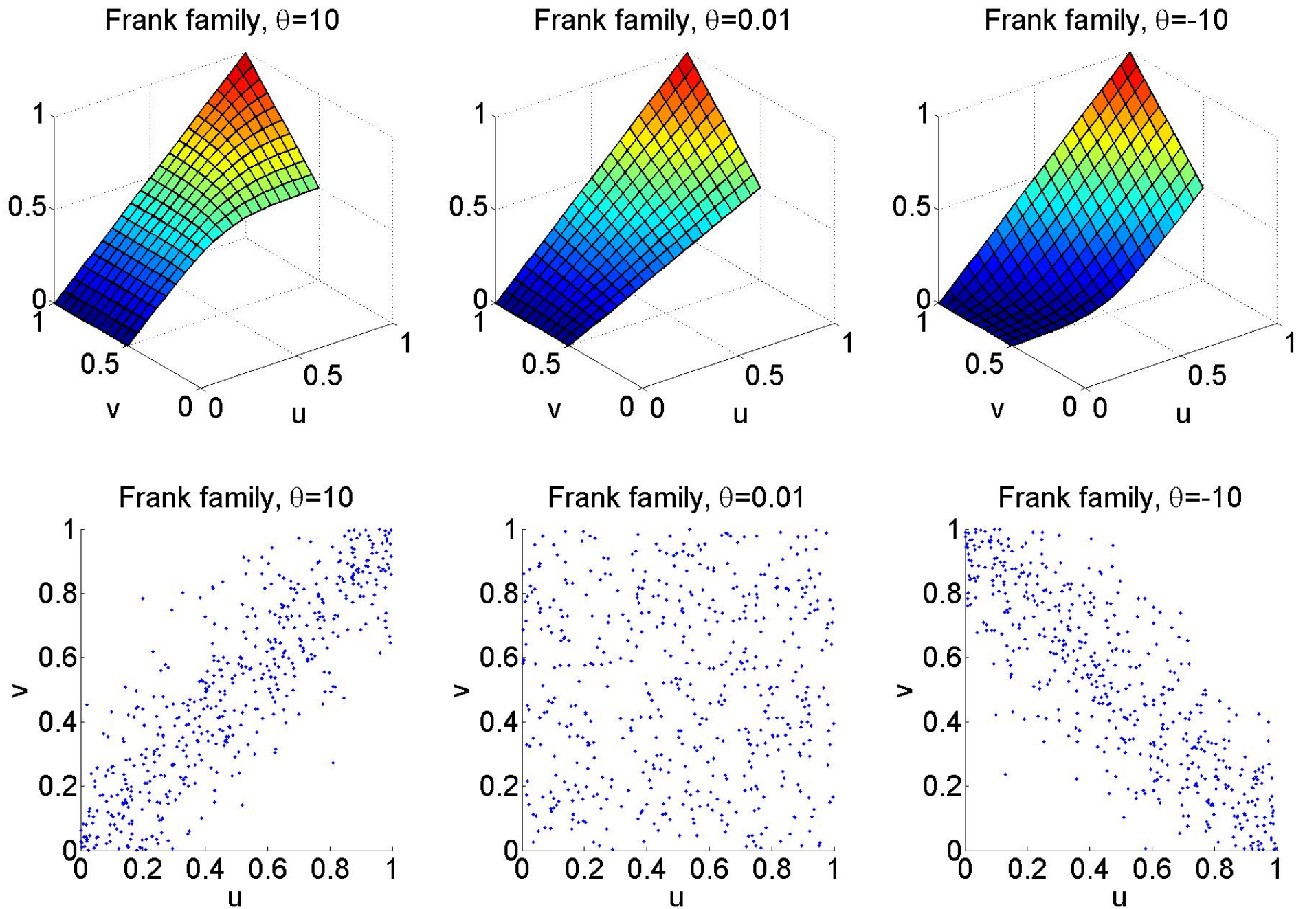
$$\varphi(t) = \ln\{[1 - \theta(1-t)]/t\} \qquad C(u,v) = \frac{uv}{1 - \theta(1-u)(1-v)}$$

    - Clayton family

$$\varphi(t) = (t^{-\theta} - 1)/\theta \qquad C(u,v) = \max\left(\left[u^{-\theta} + v^{-\theta} - 1\right]^{-\frac{1}{\theta}}, 0\right)$$

    - Frank family

$$\varphi(t) = -\ln\left[(e^{-\theta t} - 1)/(e^{-\theta} - 1)\right] \qquad C(u,v) = -\frac{1}{\theta}\ln\left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right)$$


Frank family, θ=10
Frank family, θ=0.01
Frank family, θ=-10
1
0.5
0
v
u
Frank family, θ=10
Frank family, θ=0.01
Frank family, θ=-10
0
0.2
0.4
0.6
0.8
1
v
u

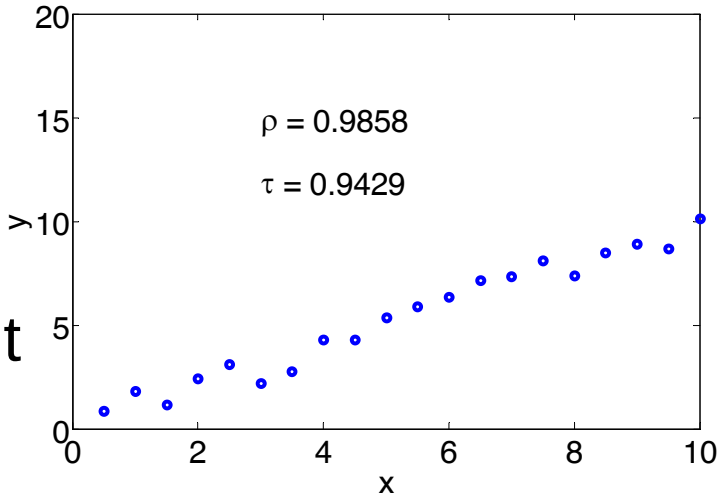
Measure of Dependence

- Concordance Measurement
  - Kendall's tau τ

$$\tau_{X,Y} = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0]$$

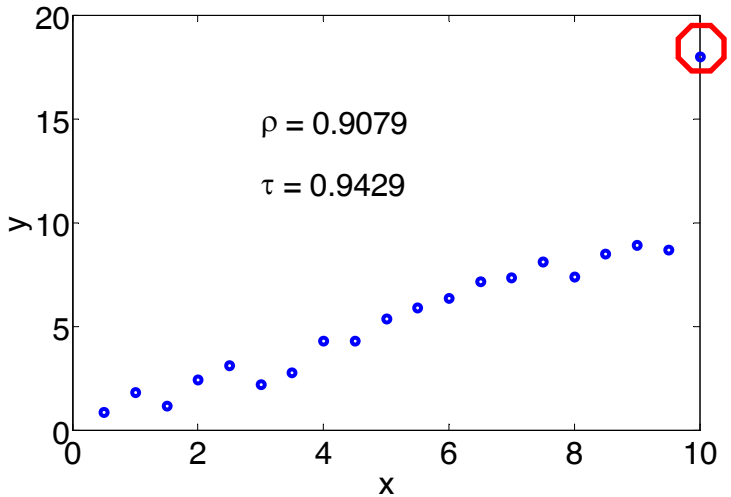
  - Sample estimator (c: concordant pairs, d: disconcordant pairs)

$$\hat{\tau}_{XY} = \frac{c-d}{c+d}$$



- Non-parametric Estimation of Copula Parameter θ
  - For Archimedean Copulas

$$\tau = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt$$



Literature Review of Copulas in Hydrology (I)

- De Michele and Salvadori (2003)
  - Bivariate rainfall frequency analysis using intensity and duration
  - Frank family of Archimedean copulas with Generalized Pareto distributed (GP) margins
  - Explain regular rainfall events
  - 2 rainfall stations in Italy with 7 years data

- Favre *et al*. (2004)
  - Bivariate flood frequency analysis using peak flow and volume of two watersheds in Canada
  - Assessment of combined risk
  - Conditional return period

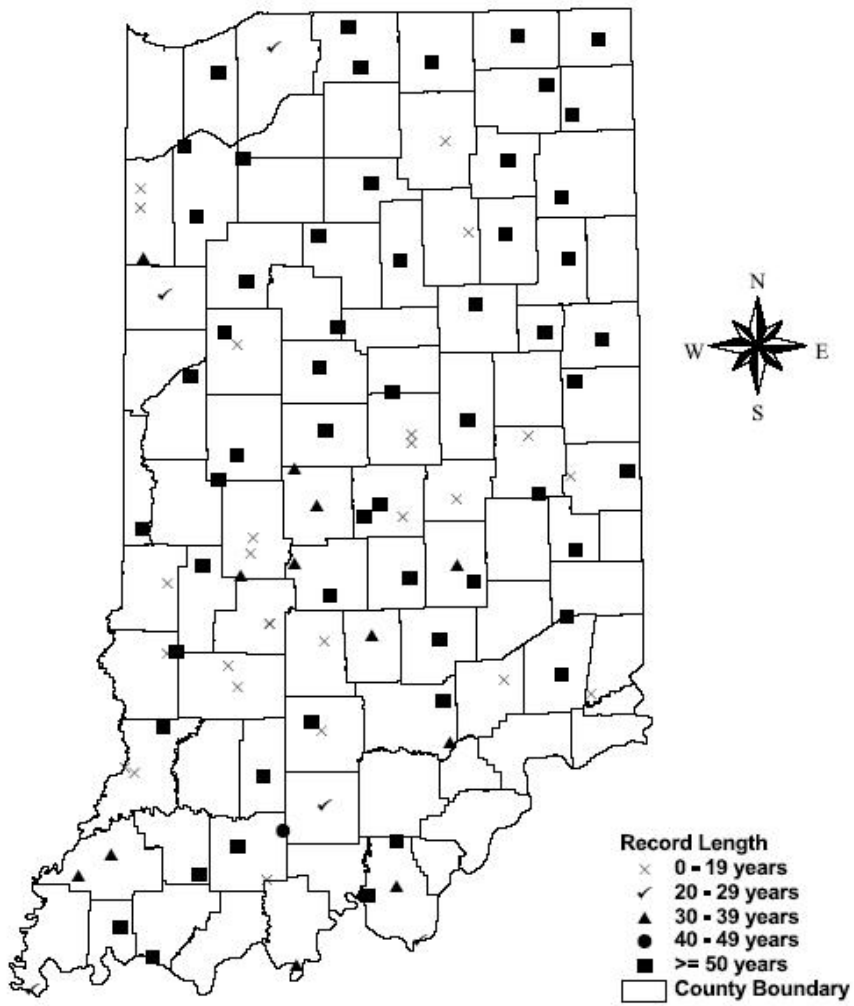
- Salvadori and De Michele (2004)
  - Return period assessment using bivariate model
  - Concept of secondary return period using Archimedean copulas

Literature Review of Copulas in Hydrology (II)

- De Michele *et al*. (2005)
  - Bivariate flood frequency analysis using peak flow and volume
  - Gumbel family of Archimedean copulas with GEV margins
  - Dam spillway adequacy assessment of Ceppo Modrelli dam in Northern Italy
- Zhang and Singh (2006)
  - Flood frequency analysis using peak flow, volume, and duration
  - Gumbel-Hougaard family of Archimedean copulas with EVI and LP(3) margins
- Salvadori and De Michele (2006)
  - Trivariate rainfall frequency analysis using intensity, duration, and dry period
  - Frank family of Archimedean copulas with GP margins
  - Bivariate copula functions were connected by a simple function to form trivariate copula function

Data Source & Study Area

- Data obtained from Nation Climate Data Center, Hourly Precipitation Dataset (NCDC, TD 3240 dataset)
- 74 Co-operative Rainfall Stations in Indiana with record length greater than 25 years
- 56 stations with record length greater than 50 years



Probabilistic Structure of Surface Runoff

- Estimate the variability of computed surface runoff based on design or observed rainfall
- Examine how dependence between rainfall variables (eg: rainfall duration, average intensity) affects surface runoff.
- Construct bivariate stochastic model to characterize storm and surface runoff
- Utilize copulas to assist in the mathematical development

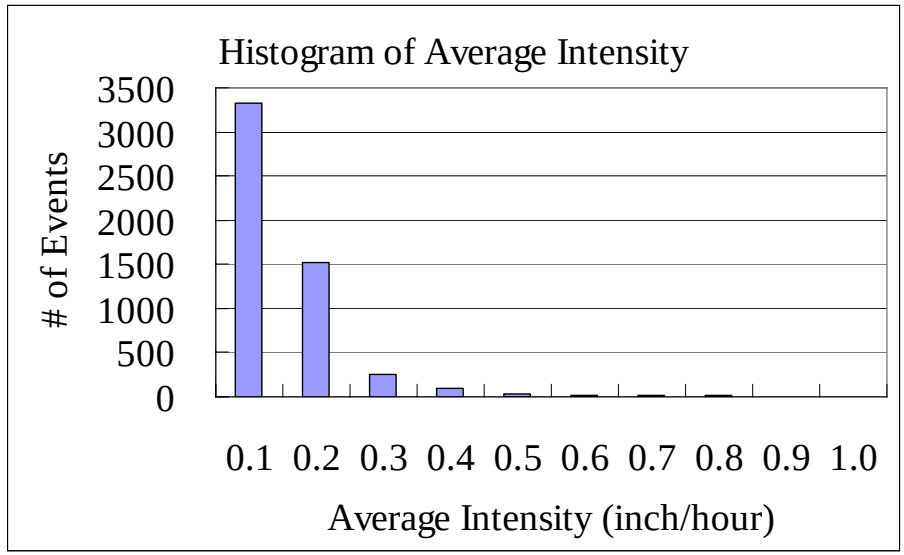
Previous Work on Probabilities of Storm Runoff

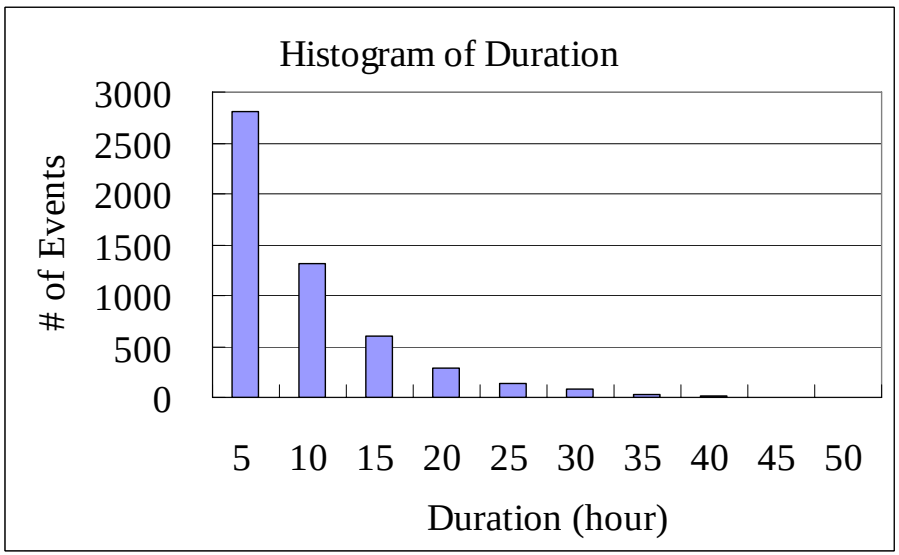
- Eagleson (1978)
  - Used rainfall duration and average intensity to characterize rainfall
  - Duration and intensity were assumed to be exponentially-distributed and independent of each other
  - Used Phillip's (1957) infiltration formula

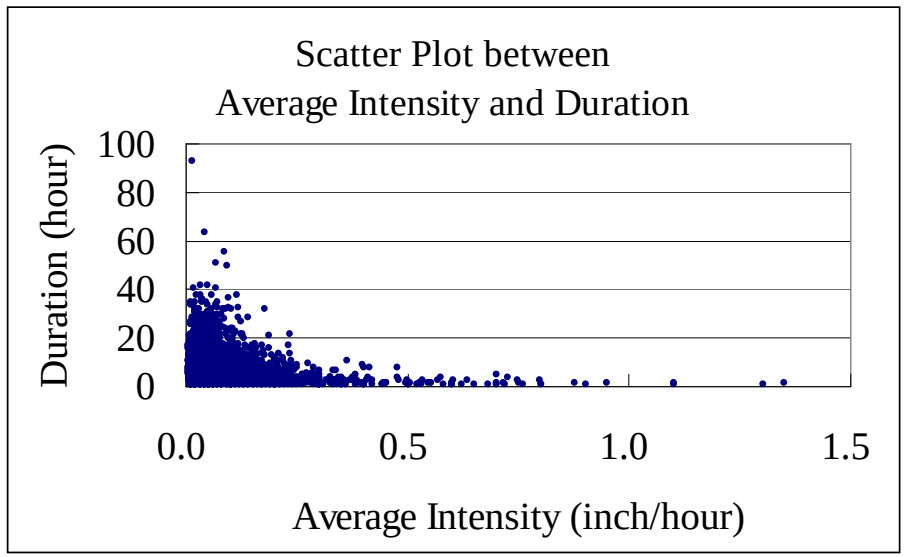
- Cordova and Rodriguez-Iturbe (1985)
  - Adopted one type of bivariate exponential distribution
  - Correlation between duration and intensity has non-negligible effect on storm surface runoff
  - Positive correlation assumption contrary to reality

Sample Plot

- Histogram and scatter plot from Station 120132
- 55 years data with 5278 divided events
- Marginal distributions decay exponentially
- Negative correlation between intensity and duration







Bivariate Exponential Distribution and Copulas

- Margins are assumed to be exponential distributed

$$f_I(i) = \lambda_I e^{-\lambda_I i} \qquad f_{T_r}(t_r) = \lambda_{T_r} e^{-\lambda_{T_r} t_r} \qquad I, T_r > 0$$

  - I denotes average intensity, T_r denotes rainfall duration
  - λ_I, λ_{Tr} are parameters of exponential distributions

- Gumbel (1960) bivariate exponential distribution

$$H_{IT_r}(i, t_r) = 1 - e^{-\lambda_I i} - e^{-\lambda_{T_r} t_r} + e^{-\left(\lambda_I i + \lambda_{T_r} t_r + \theta \lambda_I \lambda_{T_r} i t_r\right)}$$

  - H is a j-CDF, and θ is a parameter related to correlation
  - Valid for $0 \geq \rho \geq -0.40365$ ($0 \leq \theta \leq 1$)

- The survival function can be expressed as an Archimedean Copula (Gumbel-Barnett family)

$$\overline{H}_{IT_r}(i, t_r) = P[I > i, T_r > t_r] = \hat{C}(u, v) = uve^{-\theta \ln u \ln v} \qquad \varphi_\theta(t) = \ln(1 - \theta \ln t)$$

Parameter Estimation

- Marginal parameters λ_I, λ_{Tr} : Method of Moment & Maximum likelihood estimators

$$\hat{\lambda}_I = \frac{1}{\bar{I}} \qquad \hat{\lambda}_{T_r} = \frac{1}{\bar{T}_r}$$

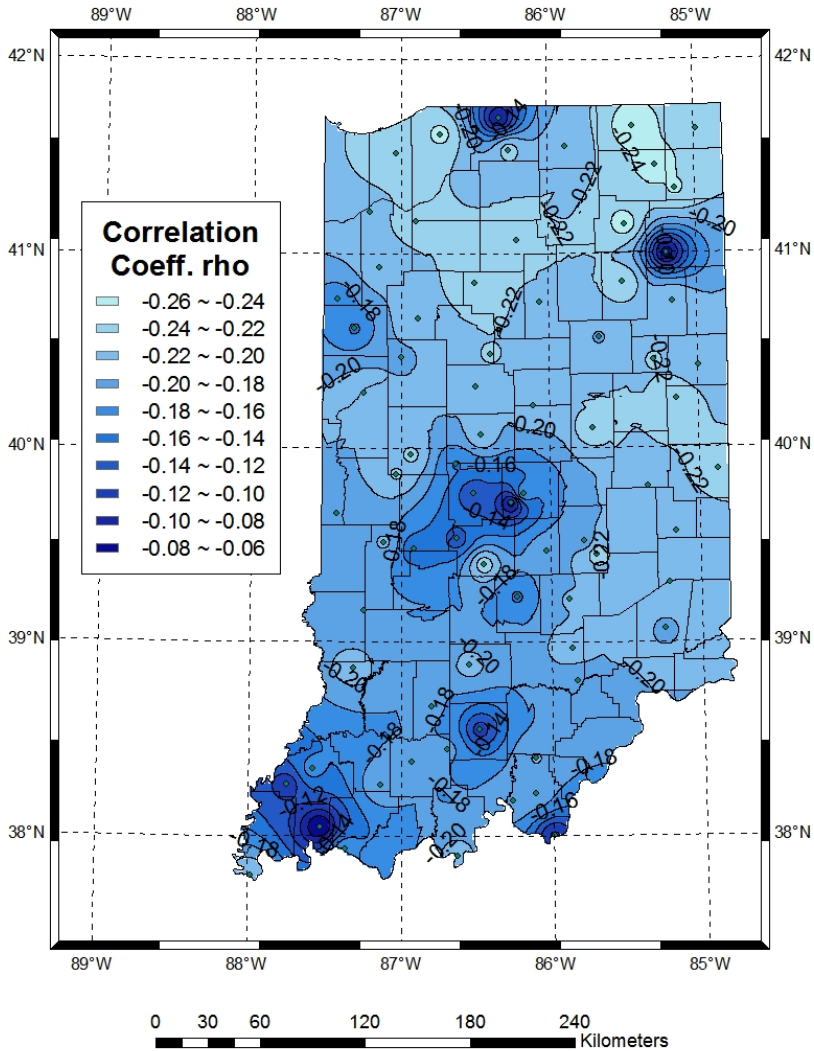
- Dependence parameter θ
  - Parametric Approach: Maximum likelihood

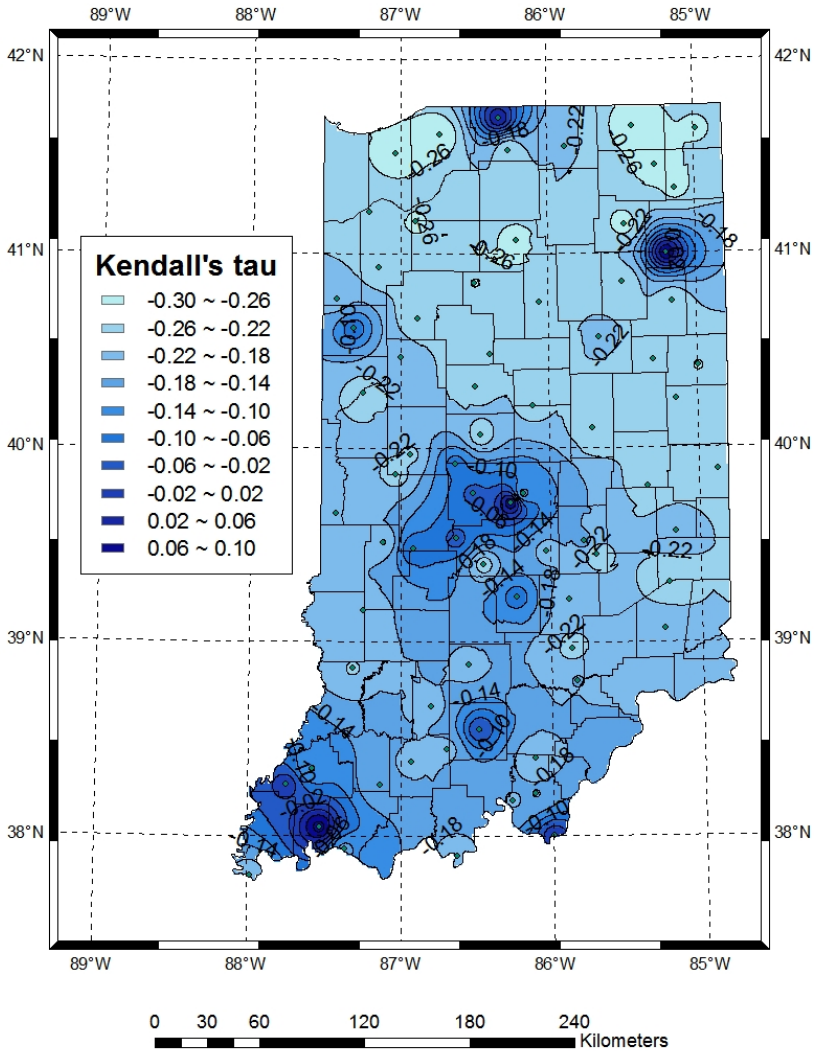
$$\frac{\partial l\left(\theta \middle| i_j, t_{r,j}\right)}{\partial \theta} = \sum_{j=1}^{N}\left[-\lambda_I \lambda_{T_r} i_j t_{r,j} + \frac{-1+\lambda_I i_j + \lambda_{T_r} t_{r,j} + 2\theta\lambda_I \lambda_{T_r} i_j t_{r,j}}{\left(1+\theta\lambda_I i_j\right)\left(1+\theta\lambda_{T_r} t_{r,j}\right)-\theta}\right] = 0$$

  - Non-parametric Approach: Copula Generator

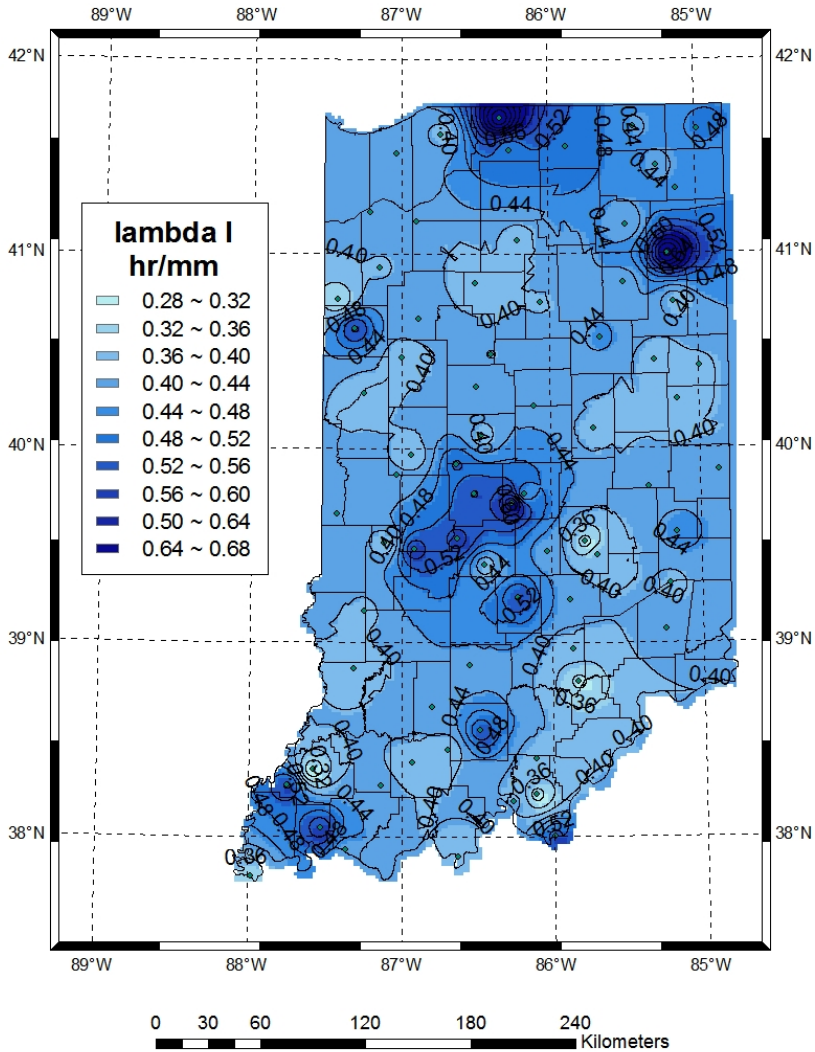
$$\tau(\theta) = -\exp\left(\frac{2}{\theta}\right)\int_{\frac{2}{\theta}}^{\infty}\frac{\exp(-t)}{t}dt$$

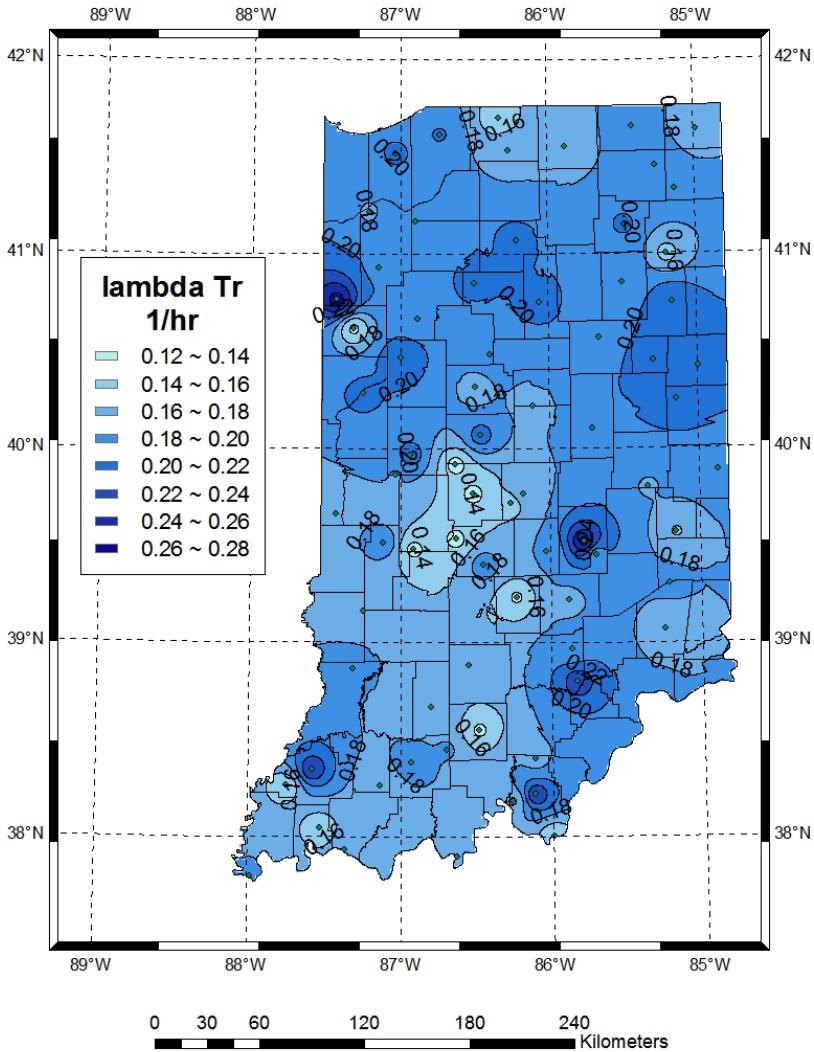
Correlation Coefficient and Kendall's Tau



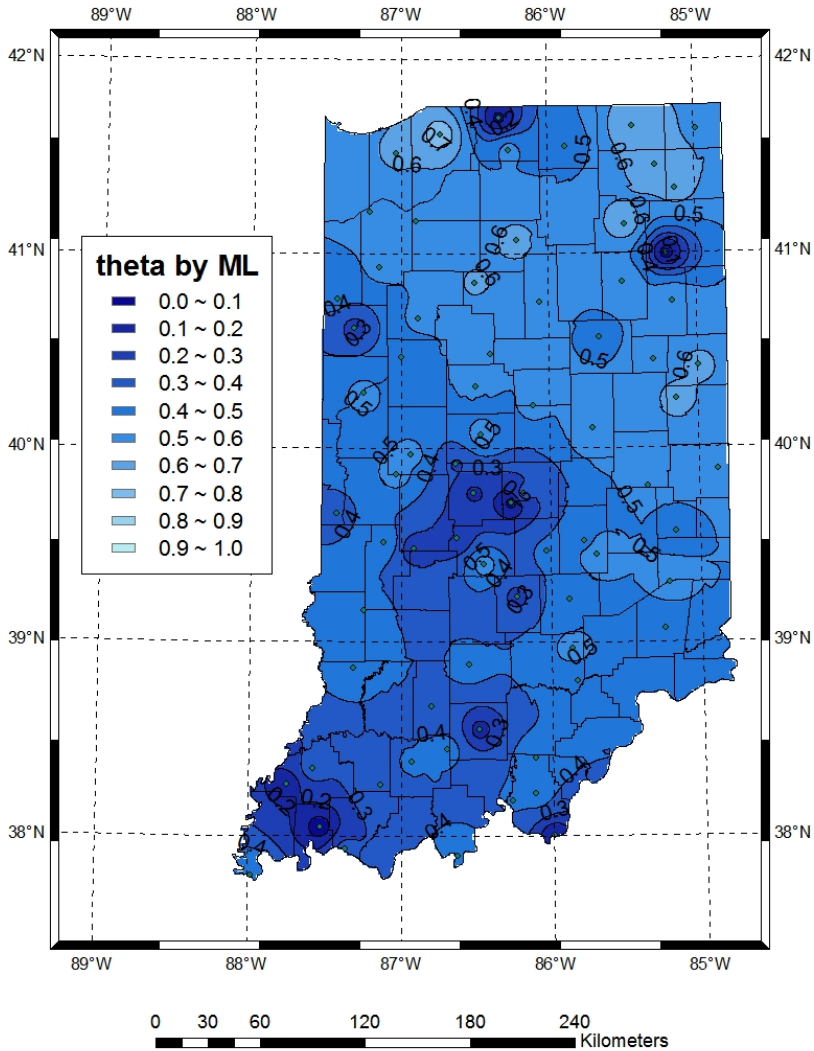


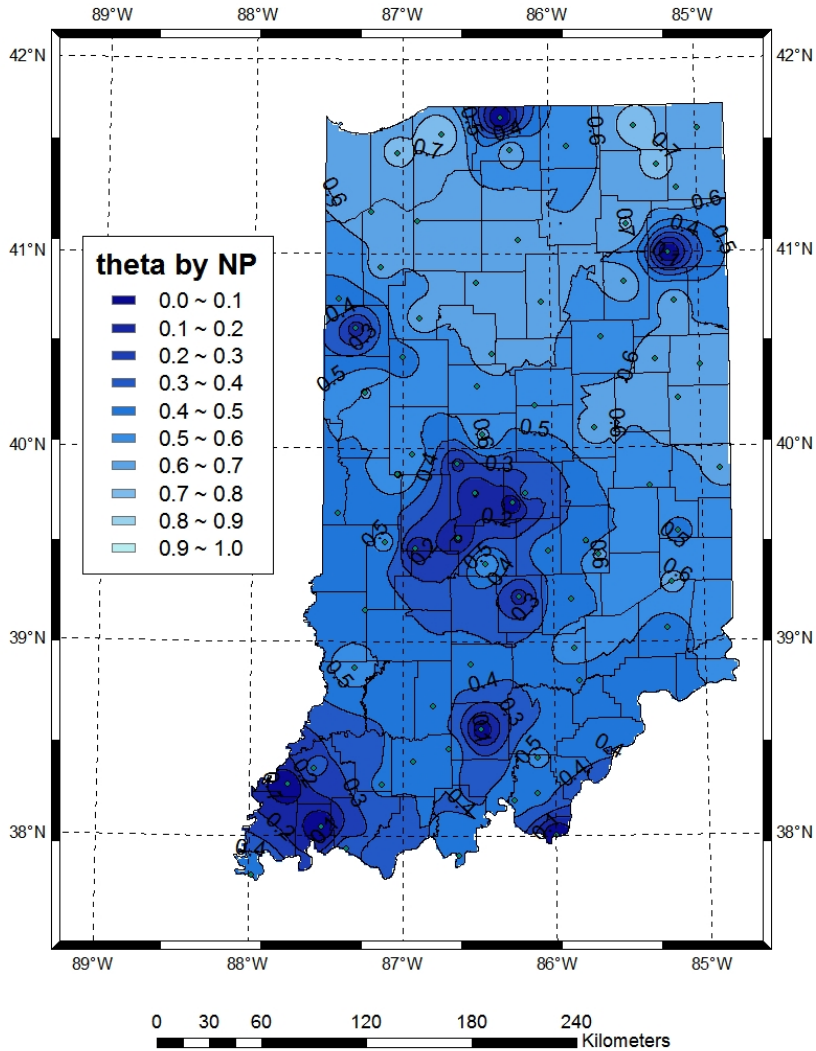
Marginal Parameters





Dependence Parameters





SCS CN Method

- Developed by the Soil Conservation Service (1972)
- Use Curve Number (CN) to describe land use, hydrologic soil type, and antecedent moisture condition
- Potential maximum retention S (inch):

$$S = \frac{1000}{CN} - 10$$

- Initial abstraction I_a (inch):

$$I_a = 0.2S$$

- For a given rainfall depth D, the Excess Rainfall Depth P_e (inch):

$$P_e = \frac{(D - 0.2S)^2}{D + 0.8S} = \frac{(IT_r - 0.2S)^2}{IT_r + 0.8S} \quad , D \geq 0.2S$$

Zero Runoff Probability - Conventional Approach

- Transformation

$$\begin{cases} Z = \lambda_I \lambda_{T_r} I T_r \\ W = \lambda_{T_r} T_r - \lambda_I I \end{cases} \qquad h_{IT_r}(i, t_r) \Rightarrow h_{ZW}(z, w) \Rightarrow f_Z(z)$$

$$f_Z(z) = \theta e^{-\theta z} \int_{-\infty}^{\infty} e^{-\sqrt{w^2+4z}} dw + \left(1 - \theta + \theta^2 z\right) e^{-\theta z} \int_{-\infty}^{\infty} \frac{e^{-\sqrt{w^2+4z}}}{\sqrt{w^2+4z}} dw$$

- Zero runoff condition: D ≤ 0.2S

$$P[P_e = 0] = P[D \le 0.2S] = P[IT_r \le 0.2S] = P\left[Z \le 0.2S\lambda_I \lambda_{T_r}\right]$$

- Equivalent initial abstraction $z^* = 0.2S\lambda_I\lambda_{Tr}$

$$P[P_e = 0] = P[Z \le z^*] = \int_0^{z^*} f_Z(z) dz = \int_0^{\infty} \varphi(w | z^*, \theta) dw$$

$$\varphi(w) = -\frac{1}{2} e^{-\sqrt{w^2+4z^*} - \theta z^*} \left(2 + \theta\sqrt{w^2+4z^*}\right) + \frac{1}{2} e^{-w}(2+w)$$

$$-\frac{1}{4} e^{\frac{1}{\theta} + \frac{w^2\theta}{4}} \sqrt{\pi\theta}\left(2 + w^2\theta\right)\left(Erf\left(\frac{2+\theta\sqrt{w^2+4z^*}}{2\sqrt{\theta}}\right) - Erf\left(\frac{2+w\theta}{2\sqrt{\theta}}\right)\right)$$

Zero Runoff Probability - Copulas Approach

- Inverse functions of survival margins

$$\overline{F}_I^{-1}(u) = \frac{\ln u}{-\lambda_I} \qquad \overline{F}_{T_r}^{-1}(v) = \frac{\ln v}{-\lambda_{T_r}}$$

- Transformation from actual domain to marginal domain

$$P[IT_r \le 0.2S] = P\left[\overline{F}_I^{-1}(U)\overline{F}_{T_r}^{-1}(V) \le 0.2S\right] = P\left[\overline{F}_I^{-1}(U) \le \frac{0.2S}{\overline{F}_{T_r}^{-1}(V)}\right]$$

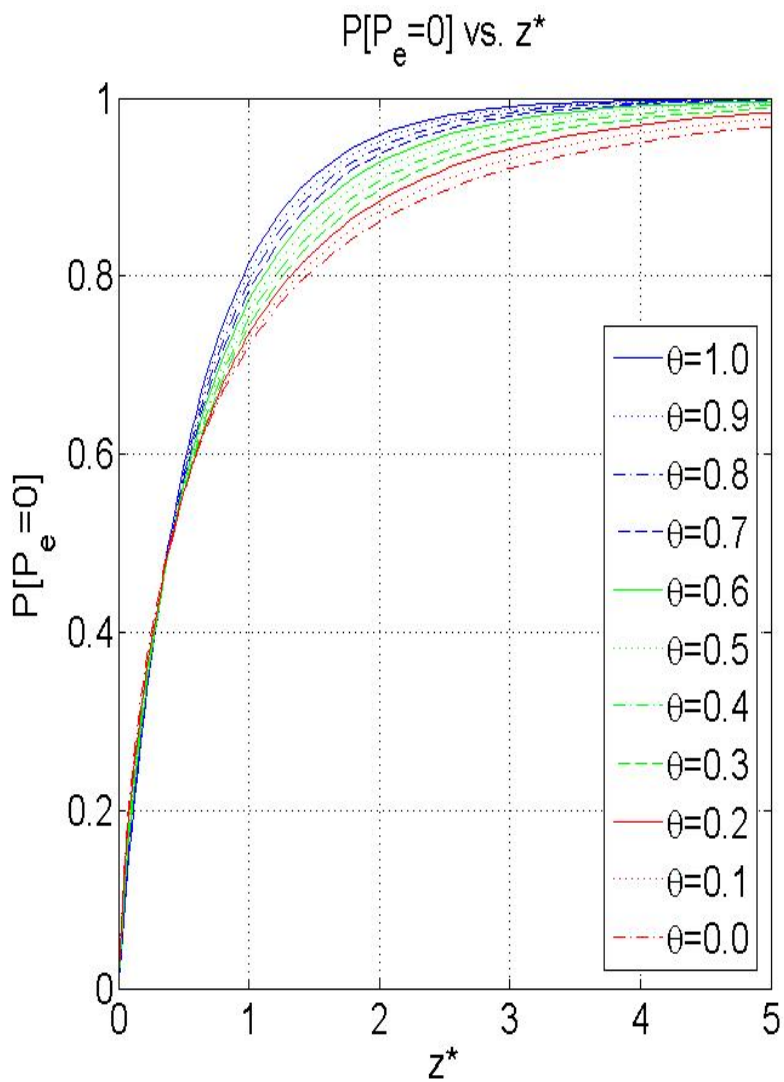
$$= P\left[U \le \overline{F}_I\left(\frac{0.2S}{\overline{F}_{T_r}^{-1}(V)}\right)\right] = P[U \le g(V)]$$

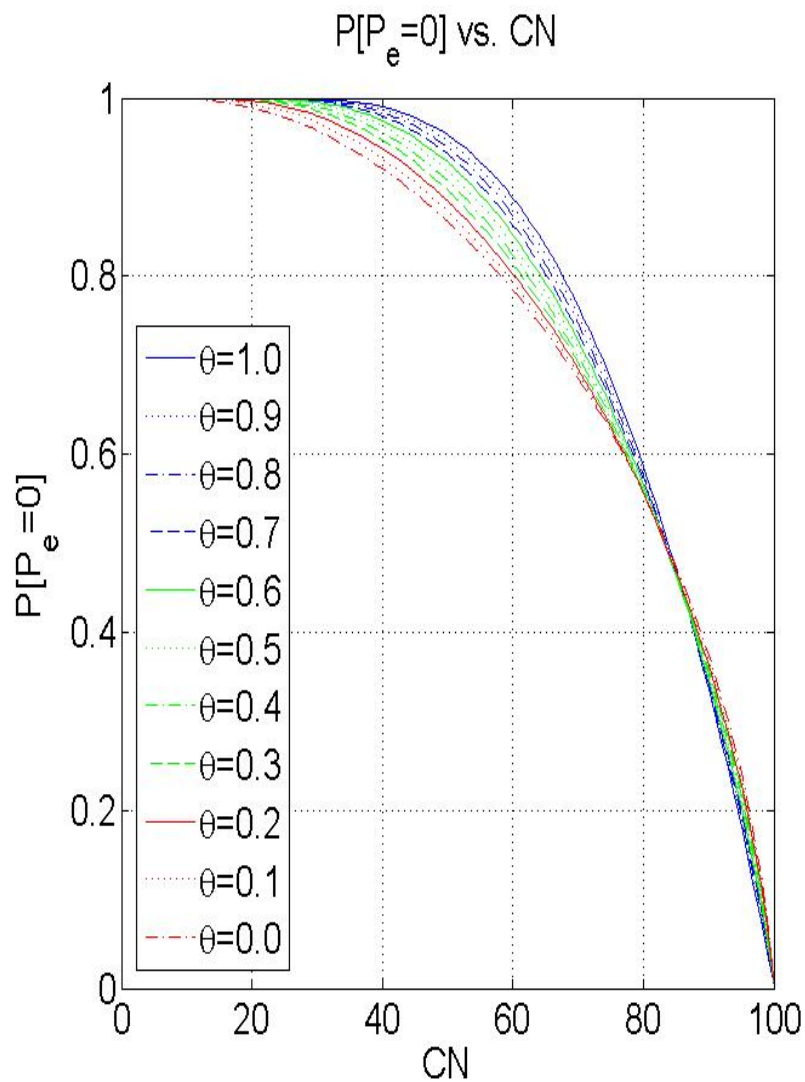
$$g(v) = \overline{F}_I\left(\frac{0.2S}{\overline{F}_{T_r}^{-1}(v)}\right) = \exp\left(\frac{0.2S\lambda_I\lambda_{T_r}}{\ln v}\right) = \exp\left(\frac{z^*}{\ln v}\right)$$

$$P[P_e = 0] = 1 - \int_0^1 \left(1 - \theta\frac{z^*}{\ln v}\right) e^{z^*\left(-\theta + \frac{1}{\ln v}\right)} dv$$

- Two different approaches yield the same results

Zero Runoff Probability





Mean and Standard Deviation of Rainfall Excess

- Conventional approach

$$\begin{cases} E[P_e] = E\left[\dfrac{(IT_r - 0.2S)^2}{IT_r + 0.8S}\right] = \int_0^\infty \int_0^\infty \dfrac{(it_r - 0.2S)^2}{it_r + 0.8S} h_{IT_r}(i, t_r)\, di\, dt_r \\ E[P_e^2] = E\left[\dfrac{(IT_r - 0.2S)^4}{(IT_r + 0.8S)^2}\right] = \int_0^\infty \int_0^\infty \dfrac{(it_r - 0.2S)^4}{(it_r + 0.8S)^2} h_{IT_r}(i, t_r)\, di\, dt_r \end{cases}$$

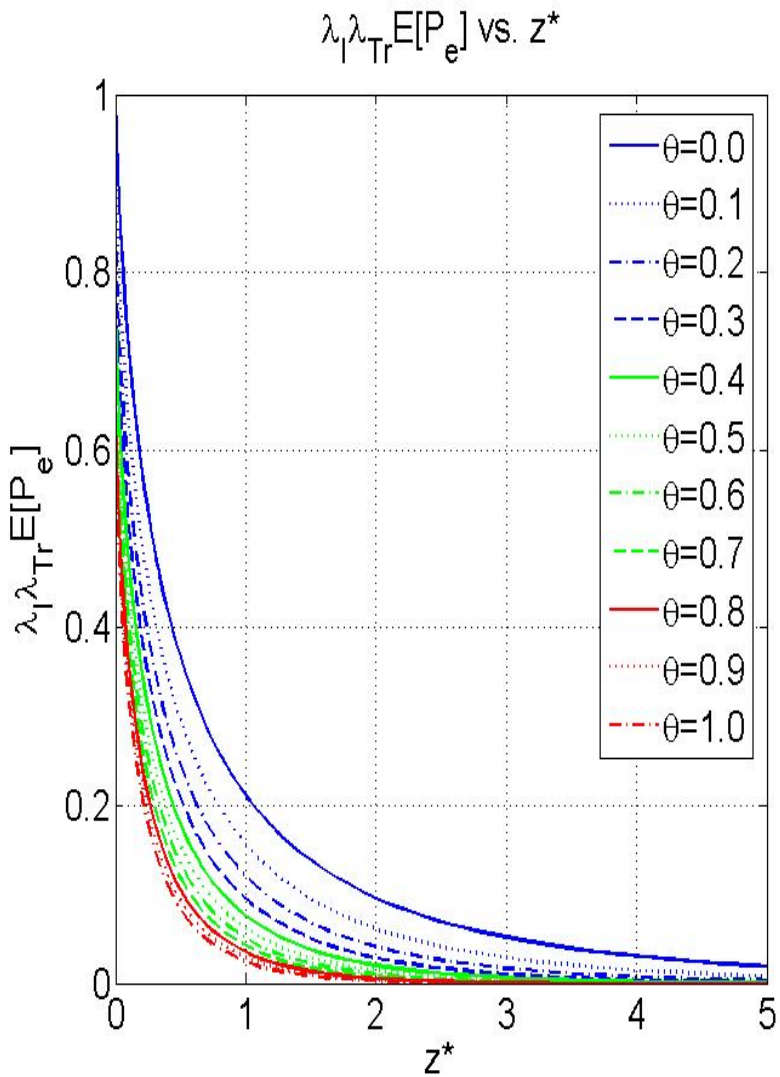
- By using transformation

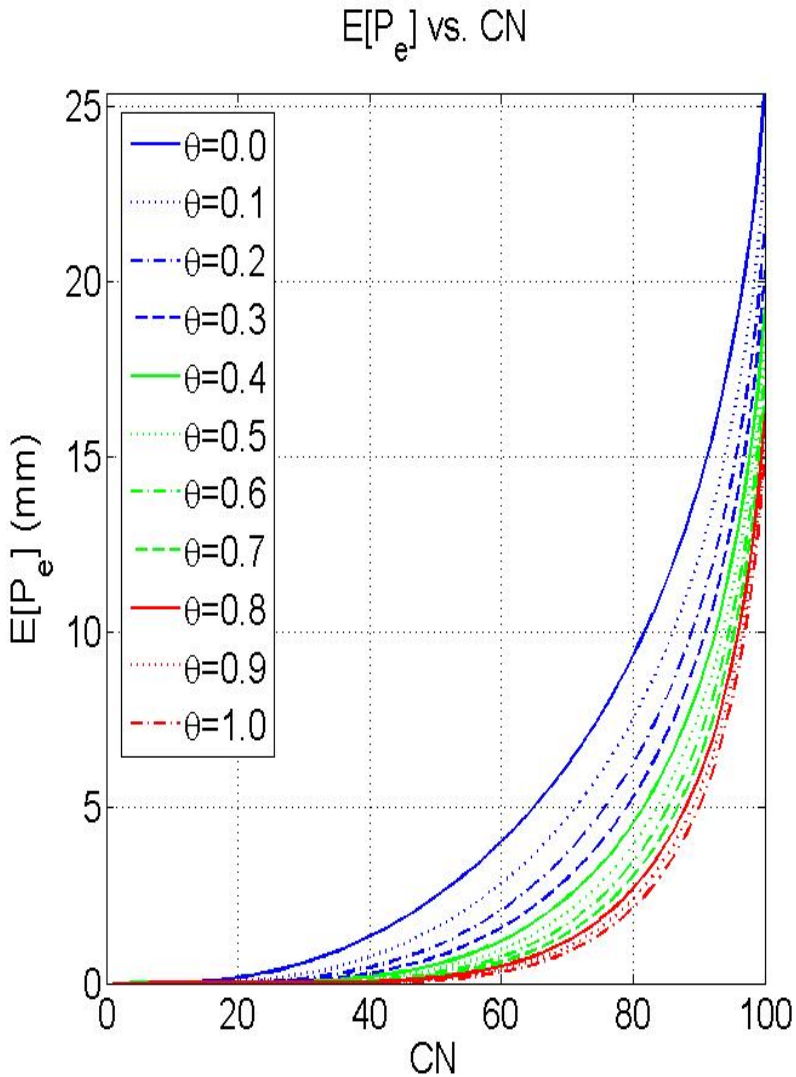
$$\begin{cases} \lambda_I \lambda_{T_r} E[P_e] = \int_0^\infty \dfrac{(z - z^*)^2}{z + 4z^*} f_Z(z)\, dz \\ (\lambda_I \lambda_{T_r})^2 E[P_e^2] = \int_0^\infty \dfrac{(z - z^*)^4}{(z + 4z^*)^2} f_Z(z)\, dz \end{cases}$$

$$\lambda_I \lambda_{T_r} Std[P_e] = \left((\lambda_I \lambda_{T_r})^2 E[P_e^2] - (\lambda_I \lambda_{T_r} E[P_e])^2\right)^{1/2}$$

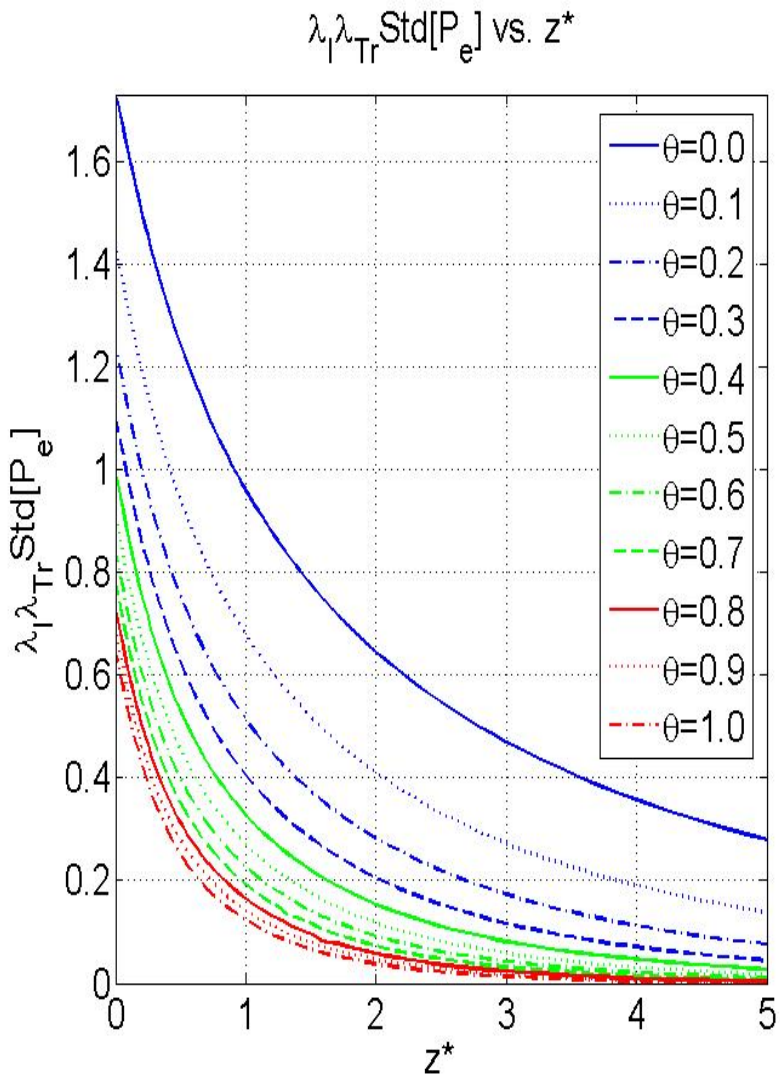
- Applying copulas in this case does not simplify the problem significantly

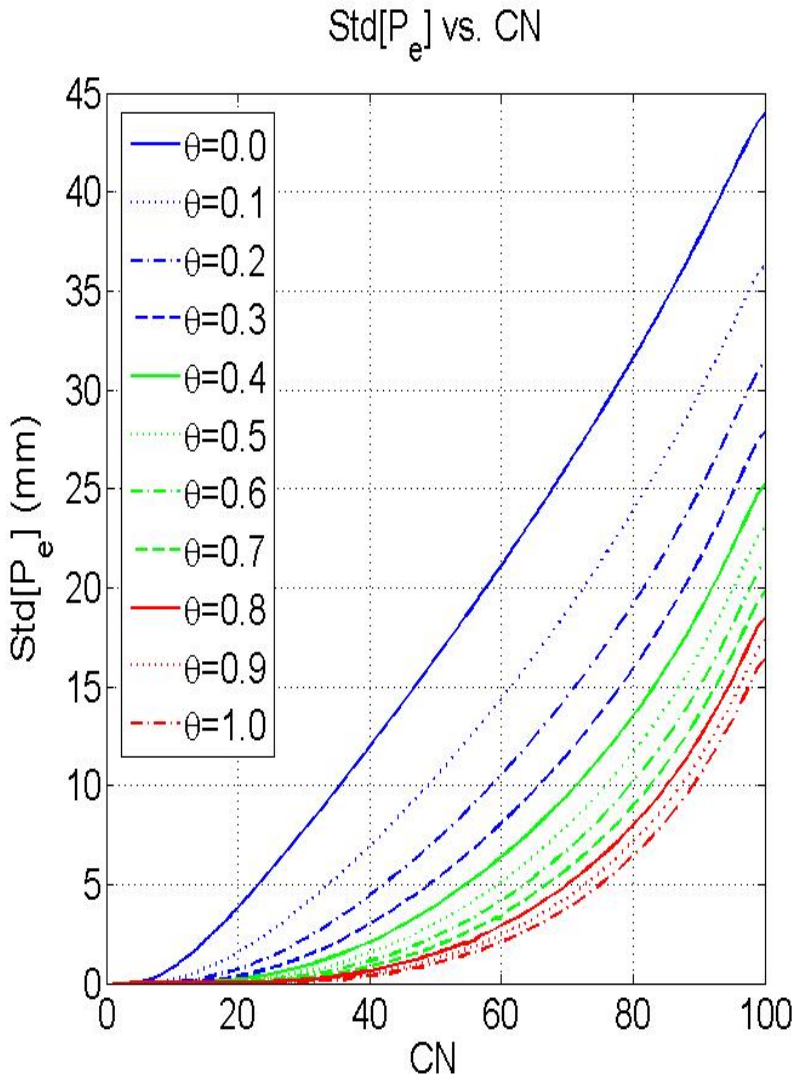
Mean of Rainfall Excess



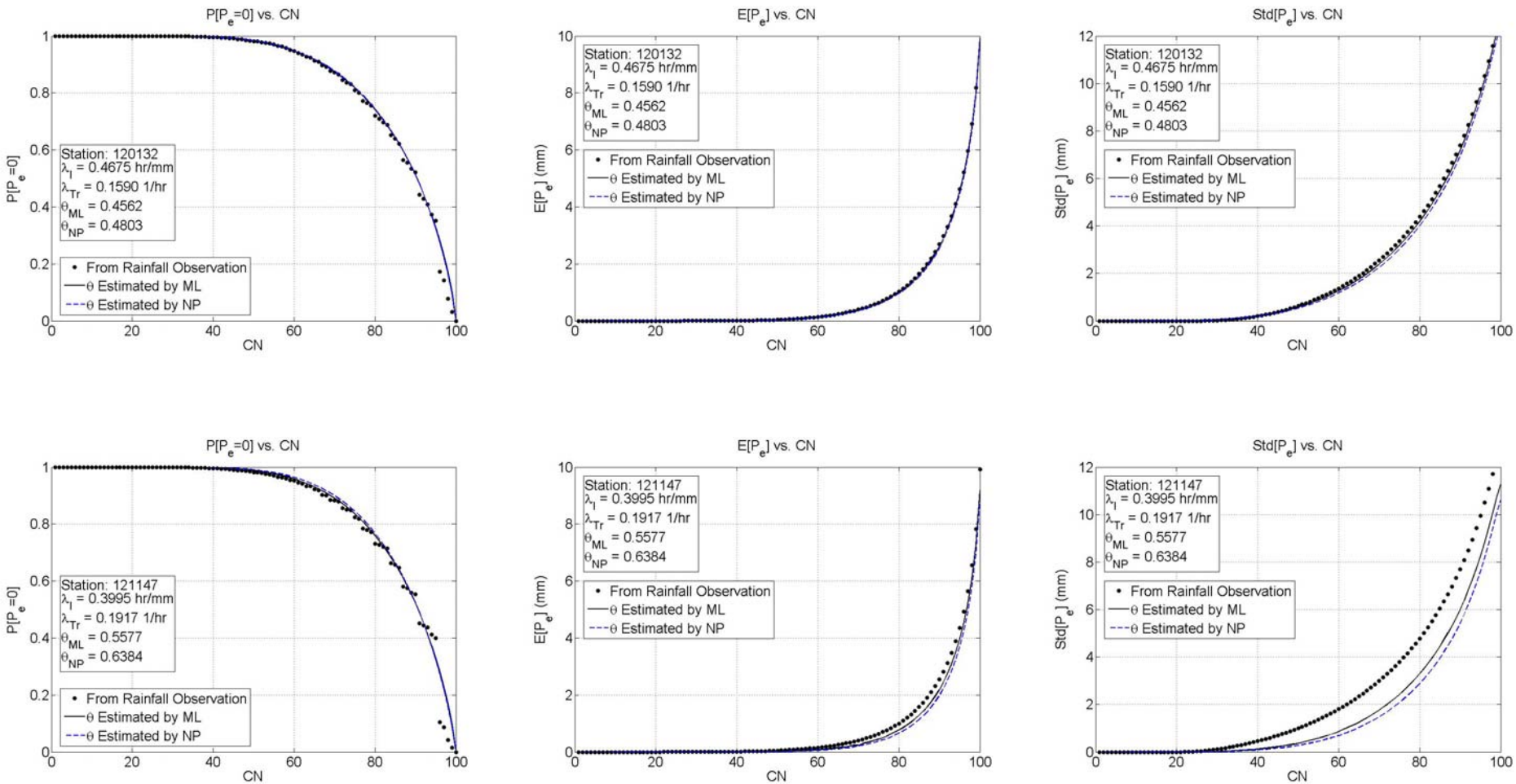


Standard Deviation of Rainfall Excess



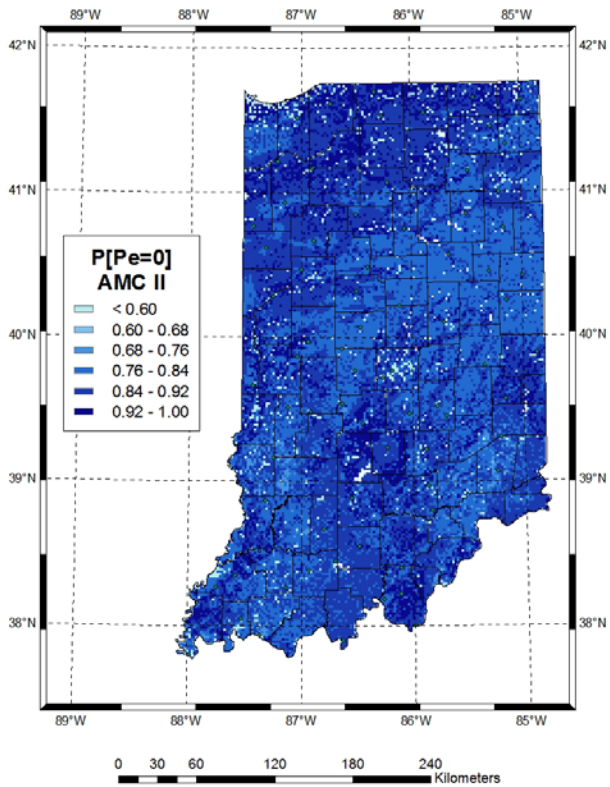


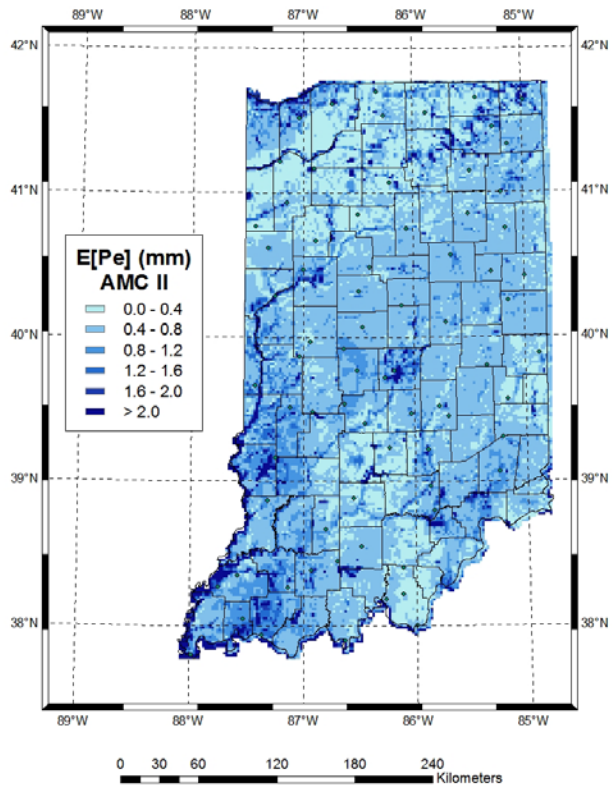
Model Validation

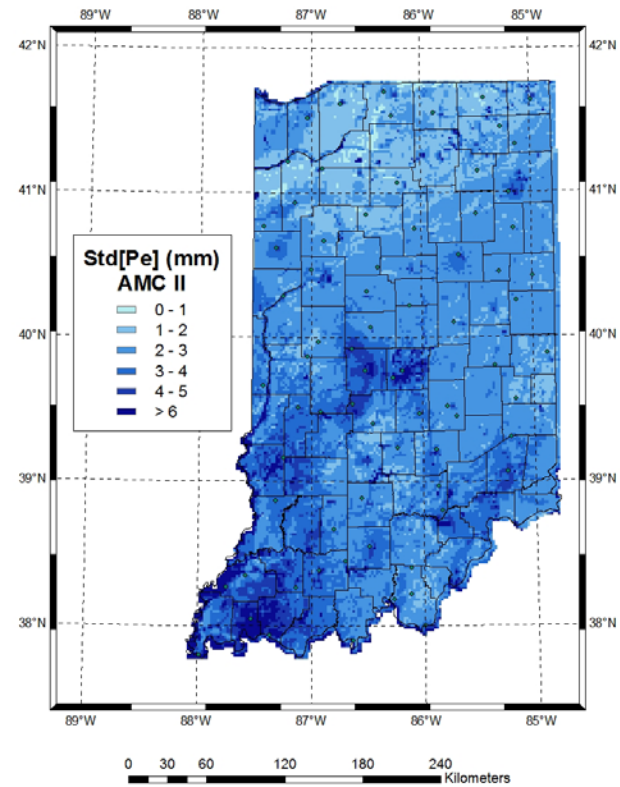


GIS Application

- CN values determined by incorporating USGS landuse information and STATGO soil database
- Reflect the effect of urbanization







Conclusion (I)

- Gumbel's (1960) bivariate exponential distribution is found suitable for describing average intensity and duration for the rainfall events in Indiana. These two quantities exhibit negative correlation with respect to each other.

- The dependence parameter can be estimated either by the ML method or the non-parametric estimation procedure relied on copulas. The use of copulas results in a simpler and more elegant estimator.

- The probability structure of storm surface runoff is sensitive to the dependence parameter, especially for the mid CN range ($20 < CN < 80$). The CN values for a large portion of Indiana fall in this range.

Conclusion (II)

- The dimensionless parameter z*, equivalent initial abstraction, is found to depend on both rainfall and soil properties.

- Based on the constant relationship between dimensionless statistical properties of rainfall excess versus z*, the probability structure of surface runoff can be derived for any location of interest. Such maps can be used to identify regions that are prone to development of surface runoff.

- For complicated multivariate stochastic models, the potential simplification offered by copulas for various rainfall and runoff statistics was demonstrated.

Thank you for your listening.
Questions?