Examining Extremal Dependence in Continental USA Climate Data

Zhengjun Zhang

Department of Statistics
University of Wisconsin
Madison, WI 53706, USA

A presentation in
Extreme events in climate and weather — an interdisciplinary workshop at the Banff Centre in Banff, Alberta, Canada

World example 1 of spatial extreme dependence

Satellite image of China precipitation

Figure 1: *China heavy rain real time image*

World example 2 of temporal extreme dependence

Does CO_2 cause the increase of temperature?

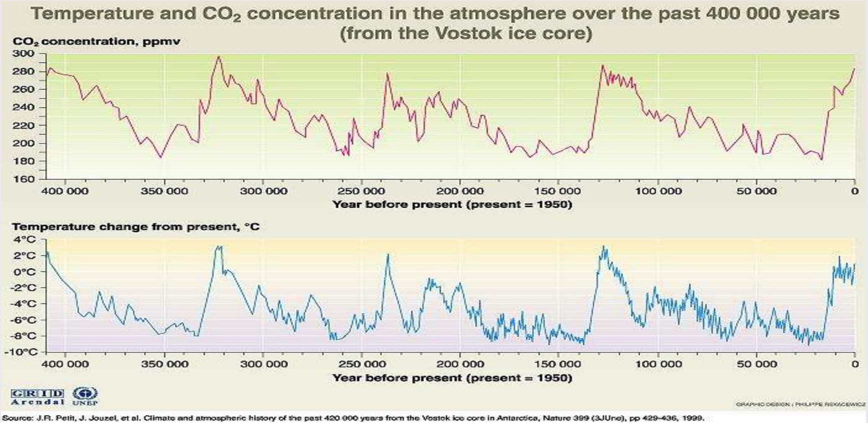


Figure 2: *Temperature and CO_2 concentration in the atmosphere over the past 400 000 years*

World example 3 of conditional temporal and spatial extreme dependence

Temperatures rocketed and rainfall reached extremes

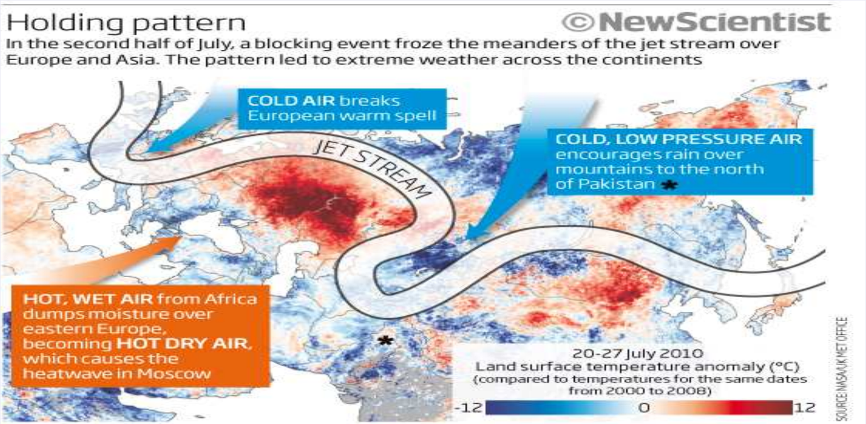
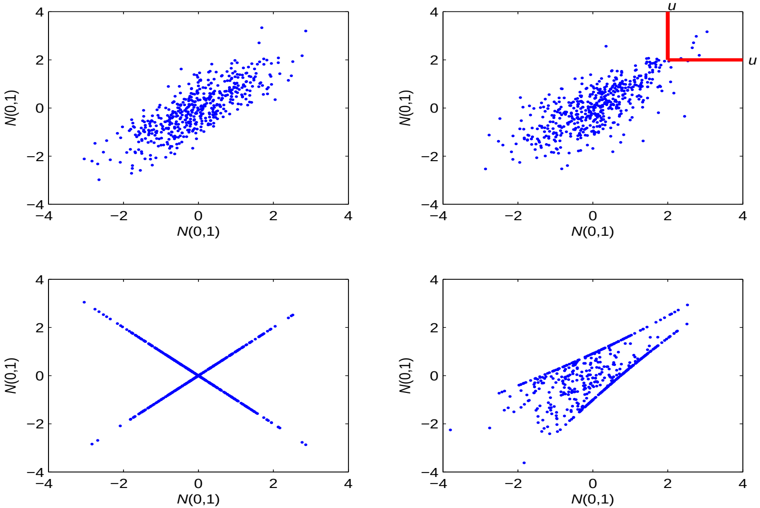


Figure 3: *Russia's Fires & Pakistan's Floods: The Result of a Stagnant Jet Stream?*

Illustration of bivariate tail (in)dependence



Purpose of this talk (1)

In theory and methodology

- Introduce a class of tail quotient correlation coefficients (TQCC) which allows the underlying threshold values to be random and diverge to infinity almost surely.
- Test statistics for extremal independence are constructed and shown to have asymptotic power one under the alternative hypothesis of extremal dependence and M4 approximation.
- Introduce a class of nonlinear quotient correlation coefficients (NQCC) for studying nonlinear dependency between random variables.

Purpose of this talk (2)

In application

- Apply TQCC and NQCC to study spatial extremal dependency and nonlinear dependency of daily precipitation during 1950–1999 recorded at 5873 stations from NCDC Rain Gauge Data.
- Our results suggest nonstationarity, asymmetry, spatial clusters, and extremal dependency in the data. They provide useful information for next generation climate models.

Extremal (in)dependence definition

Two identically distributed r.v. X and Y are called *extremely independent* if

$$\lambda = \lim_{u \to x_F} P(Y > u \mid X > u) \tag{1}$$

exists and equals 0, where $x_F = \sup\{x \in \mathbb{R} : P(X \leq x) < 1\}$. The quantity λ, if exists, is called the bivariate extremal dependence index. If $\lambda > 0$, then (X, Y) is called *extremely dependent* and we say there are extreme co-movements between X and Y.

Remark

The notion *extremal dependence*, also known as *tail dependence* or *asymptotic dependence*, between the components of a two-dimensional random vector, refers to the concurrence of extreme values in the components.

Extremal (in)dependence definition

Two identically distributed r.v. X and Y are called *extremely independent* if

$$\lambda = \lim_{u \to x_F} P(Y > u \mid X > u) \tag{1}$$

exists and equals 0, where $x_F = \sup\{x \in \mathbb{R} : P(X \leq x) < 1\}$. The quantity λ, if exists, is called the bivariate extremal dependence index. If $\lambda > 0$, then (X, Y) is called *extremely dependent* and we say there are extreme co-movements between X and Y.

Remark

The notion *extremal dependence*, also known as *tail dependence* or *asymptotic dependence*, between the components of a two-dimensional random vector, refers to the concurrence of extreme values in the components.

Importance of Studying Extreme (In)dependence

In theory

- Suppose $\{(X_i, Y_i), i = 1, \ldots, n\}$ is a random sample of (X, Y).
- If $\lambda = 0$, then the limit joint bivariate extreme value distribution is the product of the univariate limit distributions, i.e.

$$\lim_{n\to\infty} P\left\{a_n(\max_i X_i - b_n) < x,\ c_n(\max_i Y_i - d_n) < y\right\}$$
$$= \lim_{n\to\infty} P\left\{a_n(\max_i X_i - b_n) < x\right\} \lim_{n\to\infty} P\left\{c_n(\max_i Y_i - d_n) < y\right\}.$$

- When $\lambda = 0$, the limit theory of the joint maxima is simple and easy!
  - Example: a bivariate normal random variable with correlation coefficient $\rho \neq 1$.
- When $\lambda > 0$, the limit theory of the joint maxima does not show a unified parametric form!!
  - Example: a bivariate t random variable with correlation coefficient $\rho > 0$.

Importance of Studying Extreme (In)dependence: In financial risk management.

Did Gaussian copula cause Wall Street crash?

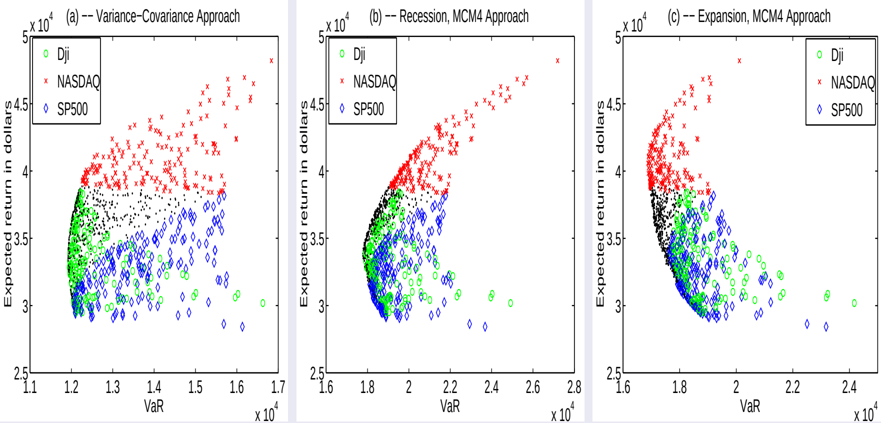


Figure 4: *VaR comparison of portfolios of different combinations. Zhang and Huang (2006), Zhang and Shinki (2006), Zhang and Zhao (2009)*

Importance of Studying Extreme (In)dependence: In extremal climatic conditions.

Notes

- Global warming causes severe storms.
- Increased ocean temperatures cause increasingly intense hurricanes
- Three major earthquakes struck within an hour and 10 minutes in the morning of October 8, 2009 near Vanuatu in the South Pacific, prompting a tsunami warning that was quickly lifted.

Reliability of climate models

- There are increasing concerns about the reliability of climate models.
- Climate models are used to predict climate changes, which draw the most attention and debate among politicians, environmentalists and even scientists.

Importance of Studying Extreme (In)dependence: In extremal climatic conditions.

Notes

- Global warming causes severe storms.
- Increased ocean temperatures cause increasingly intense hurricanes
- Three major earthquakes struck within an hour and 10 minutes in the morning of October 8, 2009 near Vanuatu in the South Pacific, prompting a tsunami warning that was quickly lifted.

Reliability of climate models

- There are increasing concerns about the reliability of climate models.
- Climate models are used to predict climate changes, which draw the most attention and debate among politicians, environmentalists and even scientists.

The specific question for forecasting future extreme weather events:

- How to account for historical records.

Two fundamental issues:

1. How to identify extremal dependency and nonlinear dependency between climatic variables.
2. How to develop statistical models dealing with extremal dependency and nonlinear dependency. Zhang (2008) AISM, studied wave heights in North Sea.

Our present focus:

- The first issue.

The specific question for forecasting future extreme weather events:

- How to account for historical records.

Two fundamental issues:

1. How to identify extremal dependency and nonlinear dependency between climatic variables.
2. How to develop statistical models dealing with extremal dependency and nonlinear dependency. Zhang (2008) AISM, studied wave heights in North Sea.

Our present focus:

- The first issue.

The specific question for forecasting future extreme weather events:

- How to account for historical records.

Two fundamental issues:

1. How to identify extremal dependency and nonlinear dependency between climatic variables.
2. How to develop statistical models dealing with extremal dependency and nonlinear dependency. Zhang (2008) AISM, studied wave heights in North Sea.

Our present focus:

- The first issue.

Hypotheses of extremal (in)dependence

$$H_0 : X \text{ and } Y \text{ are extremely independent}$$
$$\leftrightarrow H_1 : X \text{ and } Y \text{ are extremely dependent},$$

which can also be written as

$$H_0 : \lambda = 0 \longleftrightarrow H_1 : \lambda > 0. \tag{2}$$

In the remaining of the talk, we discuss how to test the null of (2) and how to estimate λ under the alternative hypothesis.

Remarks

- The null and alternative hypotheses in Ledford and Tawn (1996, 1997) are reversed in this talk, see also Peng (1999), Draisma et al. (2004), and others.
- Other significant tests include Falk and Michel (2006), Hüsler and Li (2009), Bacro, Bel, and Lantuéjoul (2010) etc.

Hypotheses of extremal (in)dependence

$$H_0 : X \text{ and } Y \text{ are extremely independent}$$
$$\leftrightarrow H_1 : X \text{ and } Y \text{ are extremely dependent},$$

which can also be written as

$$H_0 : \lambda = 0 \longleftrightarrow H_1 : \lambda > 0. \tag{2}$$

In the remaining of the talk, we discuss how to test the null of (2) and how to estimate λ under the alternative hypothesis.

Remarks

- The null and alternative hypotheses in Ledford and Tawn (1996, 1997) are reversed in this talk, see also Peng (1999), Draisma et al. (2004), and others.
- Other significant tests include Falk and Michel (2006), Hüsler and Li (2009), Bacro, Bel, and Lantuéjoul (2010) etc.

Motivations:

- In view of (1), the extremal dependence index λ is mainly relying on a high threshold value u and the dependence between tails of two random variables.

Examples of constructing extremal (in)dependence

- Let
$$\xi_1, \dots, \xi_n, \eta_1, \dots, \eta_n \tag{3}$$
be a sequence of independent unit Fréchet random variables with distribution function $F(x) = e^{-1/x}$, $x > 0$.

Motivations:

- In view of (1), the extremal dependence index λ is mainly relying on a high threshold value u and the dependence between tails of two random variables.

Examples of constructing extremal (in)dependence

- Let

$$\xi_1, \ldots, \xi_n, \eta_1, \ldots, \eta_n \tag{3}$$

be a sequence of independent unit Fréchet random variables with distribution function $F(x) = e^{-1/x}$, $x > 0$.

Example 2.1

Let $\{(U_{ni},\ Q_{ni}),\ i=1,\ldots,n\}$ be a sample of independent random pairs, where U_{ni} and Q_{ni} are correlated and both are supported on $(0,u_n]$ for a positive high threshold value u_n. Let

$$X_i = \xi_i\,\mathrm{I}_{\{\xi_i>u_n\}} + U_{ni}\,\mathrm{I}_{\{\xi_i\le u_n\}},\quad Y_i = \eta_i\,\mathrm{I}_{\{\eta_i>u_n\}} + Q_{ni}\,\mathrm{I}_{\{\eta_i\le u_n\}}$$

for $i=1,\ldots,n$. Then it follows that tail values $X_i\,\mathrm{I}_{\{X_i>u_n\}}$ $(=\xi_i\,\mathrm{I}_{\{\xi_i>u_n\}})$ and $Y_i\,\mathrm{I}_{\{Y_i>u_n\}}$ $(=\eta_i\,\mathrm{I}_{\{\eta_i>u_n\}})$ are independent, but $X_i\,\mathrm{I}_{\{X_i\le u_n\}}$ $(=U_{ni}\,\mathrm{I}_{\{\xi_i\le u_n\}})$ and $Y_i\,\mathrm{I}_{\{Y_i\le u_n\}}$ $(=Q_{ni}\,\mathrm{I}_{\{\eta_i\le u_n\}})$ are dependent. Furthermore,

$$(X_1,\ Y_1),\ldots,(X_n,\ Y_n) \tag{4}$$

is a sample of independent and identically distributed random pairs with extremely independent margins.

Elementary facts

- **Ways to measure relative positions:**
  - the difference $X - Y$;
  - the quotient X/Y for positive variables X and Y.
    - $X/Y = 0$ 'means' no relation.
    - $X/Y = 1$ means they are identical.
- **In a 'Normal' world:** Pearson correlation coefficient is the sum of the products of the Z scores,

$$r_n = \frac{1}{n} \sum Z_{x_i} Z_{y_i}.$$

- Quotient correlation coefficients are based on the maxima of the quotients of the Fréchet scores,

$$q_n = \frac{\max_{i \le n}\{Y_i/X_i\} + \max_{i \le n}\{X_i/Y_i\} - 2}{\max_{i \le n}\{Y_i/X_i\} \times \max_{i \le n}\{X_i/Y_i\} - 1} \tag{5}$$

A generalized extremal dependence measure: Tail quotient correlation coefficient (TQCC)

Suppose now X_i and Y_i are two dependent unit Fréchet random variables. Define a sample based tail dependence measure by

$$q_{u_n} = \frac{\max\limits_{1\le i\le n}\left\{\frac{\max(X_i,u_n)}{\max(Y_i,u_n)}\right\} + \max\limits_{1\le i\le n}\left\{\frac{\max(Y_i,u_n)}{\max(X_i,u_n)}\right\} - 2}{\max\limits_{1\le i\le n}\left\{\frac{\max(X_i,u_n)}{\max(Y_i,u_n)}\right\} \times \max\limits_{1\le i\le n}\left\{\frac{\max(Y_i,u_n)}{\max(X_i,u_n)}\right\} - 1}. \qquad (6)$$

- In the particular case of $u_n = u$ (a constant), definition (6) coincides with the one defined in Zhang (2008). In this talk, u_n is allowed to diverge to infinity.

Analytical properties of q_{u_n}

$$f(x,y) = \frac{x+y-2}{xy-1}, \quad \text{for } x \geq 1,\ y \geq 1,\ x+y > 2. \tag{7}$$

is a bounded and monotone function.

- $0 \leq f(x,y) \leq 1$, $f(1,y) = 1$, $f(x,1) = 1$
- $f(x_1,\ y_1) \leq f(x_2,\ y_2)$, where $x_1 \geq x_2$ and/or $y_1 \geq y_2$.

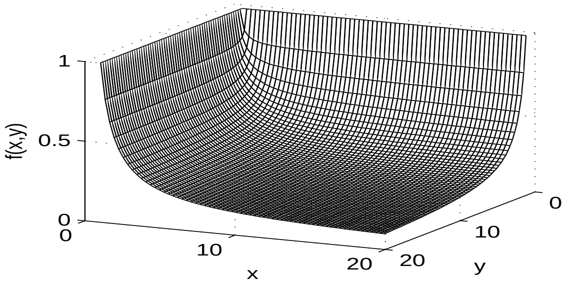


Figure 5: *Illustration of the function $f(x,y)$ in (7).*

Example 2.2

Using (3), we define

$$X_i^* = \max\{a\xi_i,\ (1-a)\eta_i\}, \quad Y_i^* = \max\{(1-b)\xi_i,\ b\eta_i\},$$

where $0 < a < 1$ and $0 < b < 1$.
Suppose $\{(\varepsilon_{1i},\ \varepsilon_{2i}),\ i = 1, \ldots, n\}$ is a sample of independent random pairs from a bivariate standard normal random variables $(\varepsilon_1,\ \varepsilon_2)$ with correlation coefficient ρ. For $i = 1, \ldots, n$, define

$$U_{ni} = -1/\log\{\Phi(\varepsilon_{1i})e^{-1/u_n}\}, \quad Q_{ni} = -1/\log\{\Phi(\varepsilon_{2i})e^{-1/u_n}\},$$

where $\Phi(\cdot)$ denotes N(0,1) distribution function, and define

$$X_i = X_i^* \mathrm{I}_{\{X_i^* > u_n\}} + U_{ni} \mathrm{I}_{\{X_i^* \le u_n\}}, \quad Y_i = Y_i^* \mathrm{I}_{\{Y_i^* > u_n\}} + Q_{ni} \mathrm{I}_{\{Y_i^* \le u_n\}}. \tag{8}$$

Then for $0 < a < 1$ and $0 < b < 1$,

$$\mathrm{q}_{u_n} \xrightarrow{a.s.} \lim_{u\to\infty} P(X_i > u \mid Y_i > u) = \lim_{u\to\infty} P(X_i^* > u \mid Y_i^* > u) = \lambda^* > 0,$$

Illustration of extremal dependence

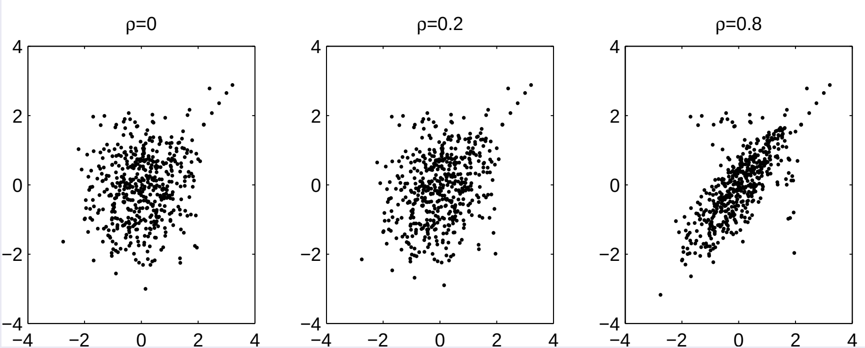


Figure 6: *Scatterplot of* $\{(\Phi^{-1}(\exp(-1/X_i)), \Phi^{-1}(\exp(-1/Y_i)))\}_{i=1}^{500}$ *in Example 2.2 (8). Values at lower regions in three panels are drawn from a bivariate standard normal random variable with correlation coefficient* ρ*.*

How to determine extremal (in)dependence: Random thresholds at work

Suppose X and Y are unit Fréchet distributed, $u_n = W_{n,t}$, where $W_{n,t}$ is distributed as e^{-n/w^t}, for $w > 0$ and $t > 1$, and $W_{n,t}$ is independent of (X, Y). Then

$$\lim_{u\to\infty} \frac{P(X > u,\ Y > u)}{P(X > u)} = \lim_{u\to\infty} \frac{P\{\max(X, W_{n,t}) > u,\ \max(Y, W_{n,t}) > u\}}{P\{\max(X, W_{n,t}) > u\}}. \tag{9}$$

Furthermore, suppose X and Y are extremely independent satisfying

$$\frac{P(X > u,\ Y > u)}{P(X > u)} = O\{u^{-(t_0-1)}\}$$

for a fixed $t_0 > 1$.

Remarks

- The existence of t_0 is guaranteed.

How to determine extremal (in)dependence: Random thresholds at work

Suppose X and Y are unit Fréchet distributed, $u_n = W_{n,t}$, where $W_{n,t}$ is distributed as e^{-n/w^t}, for $w > 0$ and $t > 1$, and $W_{n,t}$ is independent of (X, Y). Then

$$\lim_{u \to \infty} \frac{P(X > u,\ Y > u)}{P(X > u)} = \lim_{u \to \infty} \frac{P\{\max(X, W_{n,t}) > u,\ \max(Y, W_{n,t}) > u\}}{P\{\max(X, W_{n,t}) > u\}}. \tag{9}$$

Furthermore, suppose X and Y are extremely independent satisfying

$$\frac{P(X > u,\ Y > u)}{P(X > u)} = O\{u^{-(t_0 - 1)}\}$$

for a fixed $t_0 > 1$.

Remarks

- The existence of t_0 is guaranteed.

Random thresholds at work

Suppose X' and Y' are independent unit Fréchet random variables, and they are independent of $W_{n,t}$. Then

$$\frac{P(X > u,\ Y > u)}{P(X > u)} = O\Big(\frac{P\{\max(X', W_{n,t_0}) > u,\ \max(Y', W_{n,t_0}) > u\}}{P\{\max(X', W_{n,t_0}) > u\}}\Big) \tag{10}$$

and

$$\frac{P(X > u,\ Y > u)}{P(X > u)} = o\Big(\frac{P\{\max(X', W_{n,t}) > u,\ \max(Y', W_{n,t}) > u\}}{P\{\max(X', W_{n,t}) > u\}}\Big) \tag{11}$$

for $t \in (1, t_0)$.

Random thresholds at work

- Equation (9) tells that under the null hypothesis of extremal independence, testing for (X, Y) is equivalent to testing for $(\max(X, W_{n,t}),\ \max(Y, W_{n,t}))$ where $W_{n,t}$ can be simulated values.
- Equation (9) implies

$$\begin{aligned} &P\{\max(X, W_{n,s}) > u,\ \max(Y, W_{n,s}) > u\} \\ &\quad = O(P\{\max(X', W_{n,t}) > u,\ \max(Y', W_{n,t}) > u\}) \end{aligned}$$

  for all $s > 1$ and $t \in (1, t_0]$, which tells that the upper tail probability of $(\max(X, W_{n,s}),\ \max(Y, W_{n,s}))$ is 'equivalent' to that of $(\max(X', W_{n,t}),\ \max(Y', W_{n,t}))$.
- Equation (11) tells that if (10) holds, one can always theoretically choose a t such that $1 < t < t_0$ and construct a test statistic based on a bivariate random sample from $(\max(X', W_{n,t}),\ \max(Y', W_{n,t}))$.
- On the other hand, if X and Y are extremely dependent, (10) can never be true, i.e. the above test statistic (procedure) is an asymptotic power one test under the alternative hypothesis of extremal dependence.

Limit distribution and extremal independence test

Suppose that random variables X_i, Y_i, $W_{n,t}$, $i = 1, \ldots, n$, are independent, where X_i and Y_i are unit Fréchet random variables, $W_{n,t}$ is distributed as e^{-n/w^t}, for $w > 0$ and $t > 1$. Define

$$\mathrm{q}_{n,t} = \frac{\max_{i \le n} \frac{\max(X_i, W_{n,t})}{\max(Y_i, W_{n,t})} + \max_{i \le n} \frac{\max(Y_i, W_{n,t})}{\max(X_i, W_{n,t})} - 2}{\max_{i \le n} \frac{\max(X_i, W_{n,t})}{\max(Y_i, W_{n,t})} \times \max_{i \le n} \frac{\max(Y_i, W_{n,t})}{\max(X_i, W_{n,t})} - 1}.$$

Then $2n\{1 - e^{-1/W_{n,t}}\}\mathrm{q}_{n,t} \xrightarrow{\mathscr{L}} \chi_4^2$.

Random thresholds at work

- The limit distribution does not depend on the power transformation index t. This is an important property in practice. One does not need to deal with $W_{n,t}$.
- For example, suppose we simulate a value u_1 from W_{n,t_1} for a pre-specified t_1. Then for any $s_1 > 0$, $u_1^{s_1}$ can be used as a value simulated from $W_{n,s_1 * t_1}$, i.e. we can set $W_{n,t}$ as some pre-specified value u_n, for example the sample 100pth percentiles.

Limit distribution and extremal independence test

Suppose that random variables X_i, Y_i, $W_{n,t}$, $i = 1, \ldots, n$, are independent, where X_i and Y_i are unit Fréchet random variables, $W_{n,t}$ is distributed as e^{-n/w^t}, for $w > 0$ and $t > 1$. Define

$$\mathrm{q}_{n,t} = \frac{\max_{i\le n} \frac{\max(X_i, W_{n,t})}{\max(Y_i, W_{n,t})} + \max_{i\le n} \frac{\max(Y_i, W_{n,t})}{\max(X_i, W_{n,t})} - 2}{\max_{i\le n} \frac{\max(X_i, W_{n,t})}{\max(Y_i, W_{n,t})} \times \max_{i\le n} \frac{\max(Y_i, W_{n,t})}{\max(X_i, W_{n,t})} - 1}.$$

Then $2n\{1 - e^{-1/W_{n,t}}\}\mathrm{q}_{n,t} \xrightarrow{\mathscr{L}} \chi_4^2$.

Random thresholds at work

- The limit distribution does not depend on the power transformation index t. This is an important property in practice. One does not need to deal with $W_{n,t}$.
- For example, suppose we simulate a value u_1 from W_{n,t_1} for a pre-specified t_1 . Then for any $s_1 > 0$, $u_1^{s_1}$ can be used as a value simulated from $W_{n,s_1 * t_1}$, i.e. we can set $W_{n,t}$ as some pre-specified value u_p, for example the sample 100pth percentiles.

Testing procedure

- For a given significance level α and an appropriate chosen u_n, if $2n\{1 - \exp(-1/u_n)\}\mathrm{q}_{u_n} > \chi^2_{4;\alpha}$, H_0 of (2) is rejected, and we conclude there exists extremal dependence between two random variables of interest. Here $\chi^2_{4;\alpha}$ is the upper α percentile of a χ^2 distributed random variable with 4 degrees of freedom.
- If H_0 of (2) is rejected, (6) is an estimate of λ.

Remarks

- Pearson's sample correlation coefficient and TQCC are asymptotically independent. Zhang, Qi, and Ma (2010).
- TQCC is $\sqrt{n}$ convergence under the alternative hypothesis of bivariate Gumbel copula. Wang (2010).

Testing procedure

- For a given significance level α and an appropriate chosen u_n, if $2n\{1 - \exp(-1/u_n)\}\mathrm{q}_{u_n} > \chi^2_{4;\alpha}$, H_0 of (2) is rejected, and we conclude there exists extremal dependence between two random variables of interest. Here $\chi^2_{4;\alpha}$ is the upper α percentile of a χ^2 distributed random variable with 4 degrees of freedom.
- If H_0 of (2) is rejected, (6) is an estimate of λ.

Remarks

- Pearson's sample correlation coefficient and TQCC are asymptotically independent. Zhang, Qi, and Ma (2010).
- TQCC is $\sqrt{n}$ convergence under the alternative hypothesis of bivariate Gumbel copula. Wang (2010).

Model specifications

(1) Componentwise maxima of bivariate normal random variables with $\rho = 0.2, 0.4, 0.6, 0.8$.

(2) Example 2.1 in this talk revisited with $\rho = 0.2, 0.4, 0.6, 0.8$..

(3) Gumbel copula.

(4) Resnick's example $(1/U, 1/(1-U))$.

(5) Product of two random variables: $X_i = E_i * Z_i,\ Y_i = E_i * Z_i'$.

(6) Bivariate t distribution example. d.f.=4. $\rho = 0.8$.

Test statistics and threshold levels

- Hüsler and Li (2009) test (HLT), top 25% order statistics
- Bacro, Bel, and Lantuéjoul (2010) Madogram test (MaT), no levels specified.
- TQCC, at levels 80%:0.025:97.5%

Simulation and performance comparison

Model specifications

(1) Componentwise maxima of bivariate normal random variables with $\rho = 0.2, 0.4, 0.6, 0.8$.

(2) Example 2.1 in this talk revisited with $\rho = 0.2, 0.4, 0.6, 0.8$..

(3) Gumbel copula.

(4) Resnick's example $(1/U, 1/(1-U))$.

(5) Product of two random variables: $X_i = E_i * Z_i,\ Y_i = E_i * Z_i'$.

(6) Bivariate t distribution example. d.f.=4. $\rho = 0.8$.

Test statistics and threshold levels

- Hüsler and Li (2009) test (HLT), top 25% order statistics
- Bacro, Bel, and Lantuéjoul (2010) Madogram test (MaT), no levels specified.
- TQCC, at levels 80%:0.025:97.5%

Empirical Type I errors and powers

Sample size n=300										
Model	HLT	MaT	TQCC							
			.80	.825	.85	.875	.90	.925	.95	.975
(3)	16	100	98	98	97	97	96	94	86	54
(4)	6	0	1	1	1	1	1	1	1	1
(5)	4	100	7	7	7	6	5	4	4	2
(6)	12	100	99	99	99	99	99	97	97	95
Sample size n=500										
(3)	20	100	100	100	100	100	100	99	98	70
(4)	4	0	2	2	3	3	3	3	3	3
(5)	5	100	4	3	3	3	3	4	4	4
(6)	25	100	100	100	100	100	100	100	100	99

Empirical Type I errors and powers

Sample size n=1000										
Model	HLT	MaT	TQCC							
			.80	.825	.85	.875	.90	.925	.95	.975
(3)	52	100	100	100	100	100	100	99	98	96
(4)	3	0	2	2	3	3	3	3	3	3
(5)	5	100	5	5	5	6	5	4	4	4
(6)	39	100	100	100	100	100	100	100	100	100
Sample size n=2000										
(3)	93	100	100	100	100	100	100	100	100	98
(4)	1	0	4	4	4	5	6	6	7	7
(5)	7	100	10	9	8	6	5	5	4	2
(6)	87	100	100	100	100	100	100	100	100	100

Empirical Type I errors and powers

Model (1), Sample size n=300										
ρ	HLT	MaT	TQCC							
			.80	.825	.85	.875	.90	.925	.95	.975
0.2	2	4	3	3	2	2	2	2	2	1
0.4	6	4	2	1	1	1	1	1	1	1
0.6	2	18	3	3	2	2	2	2	1	1
0.8	4	90	9	9	8	8	7	5	2	2
Model (2), Sample size n=300										
0.2	3	9	5	5	5	6	6	4	4	3
0.4	5	25	2	2	2	2	3	2	0	0
0.6	1	66	3	3	3	3	3	2	1	1
0.8	4	88	3	2	2	2	2	2	2	1

Empirical Type I errors and powers

ρ	HLT	MaT	TQCC .80	.825	.85	.875	.90	.925	.95	.975
Model (1), Sample size n=500										
0.2	3	1	4	4	4	4	4	4	4	3
0.4	2	3	6	6	6	5	5	5	4	3
0.6	6	14	3	3	2	2	2	2	2	2
0.8	7	100	13	12	11	10	9	6	5	4
Model (2), Sample size n=500										
0.2	0	11	4	3	3	3	4	4	2	2
0.4	4	44	5	5	3	3	2	2	1	1
0.6	1	88	5	6	5	5	4	4	3	3
0.8	3	99	9	10	10	10	10	9	7	6

What do we learn from these three tests

- HL test is conservative.
- Madogram test is too aggressive. Type I errors are not controlled within the pre-specified nominal levels.
- TQCC based test seems acceptable.

The precipitation data

- The data are daily precipitation totals covering period 1950-1999 over 5873 stations in the continental USA (excluding Alaska and Hawaii). The data units are tenths of a millimeter.
- The data are the same as used by Smith, Grady, and Hegerl (2007), and by Shamseldin, Smith, Sain, Mearns, and Cooley (2008).
- The data are first fitted to GEVs, and then transformed to unit Fréchet margins.

$$H(x;\xi,\mu,\psi) = \exp[-\{1+\xi(x-\mu)/\psi\}_+^{-1/\xi}], \tag{12}$$

to local maxima of observations, where μ is a location parameter, $\psi > 0$ is a scale parameter, and ξ is a shape parameter

Tail indecies of precipitations

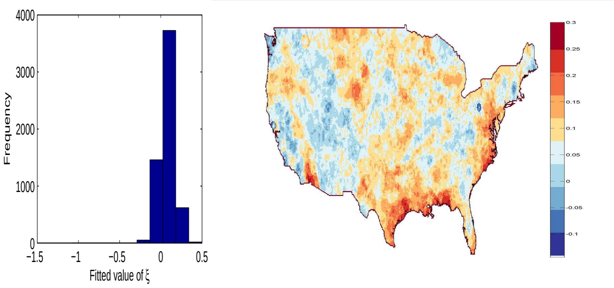


Figure 7: *Fitted tail shape parameter values for all 5873 station time series. The left panel shows the distribution of fitted shape parameter values. The right panel plots fitted shape parameter values to US map using krigging.*

- Precipitations appear to be non-stationarity, spatial clusters, and asymmetry over all stations.
- Precipitations over stations near Mexican bay region and stations near Atlantic ocean and along North Carolina coast have heavier tails then precipitations over other stations have.

Overall extremal dependency across all stations

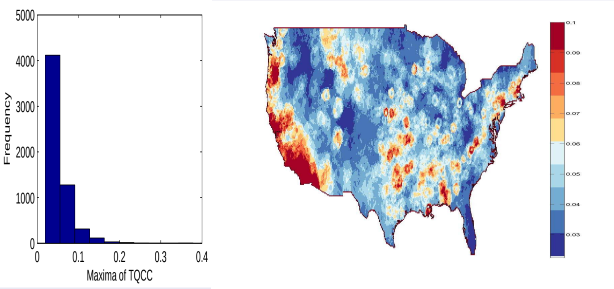


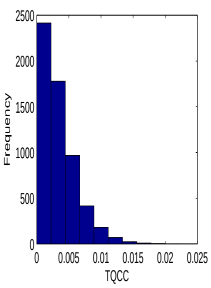
Figure 8: *Maximal extremal precipitation dependencies between one station and the rest of stations on the same day.*

- Each individual station s_i, at least 80% of paired stations (s_i, s_j), $j = 1, \ldots, 5873; j \neq i$ are rejecting the null hypothesis of extremal independence.
- The maximal extremal dependencies decay as time goes by.

Table 1: *The 10 largest tail quotient correlation coefficients and their corresponding stations information.*

Pair ID	TQCC	Latitude	Longitude	Elevation	Station name
(I)	.3787	33.92	-118.13	34	DOWNEY FIRE STN FC107D
		33.97	-118.02	128	WHITTIER CITY YD FC106C
(II)	.2652	44.40	-122.48	262	CASCADIA
		44.10	-122.68	206	LEABURG 1 SW
(III)	.2324	34.48	-119.50	633	JUNCAL DAM
		34.53	-119.78	312	LOS PRIETOS RANGER STN
(IV)	.2271	40.08	-99.20	610	HARLAN COUNTY LAKE
		40.07	-99.13	573	NAPONEE
(V)	.2208	39.35	-123.12	309	POTTER VALLEY P H
		39.13	-123.20	193	UKIAH
(VI)	.2206	42.48	-71.28	49	BEDFORD
		42.52	-71.13	27	READING
(VII)	.2200	34.52	-119.68	473	GIBRALTAR DAM 2
		34.48	-119.50	633	JUNCAL DAM
(VIII)	.2198	34.08	-117.87	175	COVINA NIGG FC193B
		33.97	-118.02	128	WHITTIER CITY YD FC106C
(IX)	.2128	29.95	-90.13	6	NEW ORLEANS WATER PLT
		29.98	-90.02	3	NEW ORLEANS D P S 5
(X)	.2092	33.53	-117.77	11	LAGUNA BEACH

Illustrations of an individual station



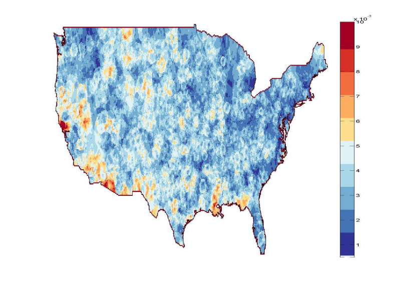


Figure 9: *Maximal extremal precipitation dependencies between Station Healdsburg and the rest of stations on the same day.*

- We can see that each of these plots itself can be viewed as a skewed and long tailed distribution. This phenomenon suggests flooding can be anywhere which shares smaller extremal dependencies with other locations.

The most recent flooding areas

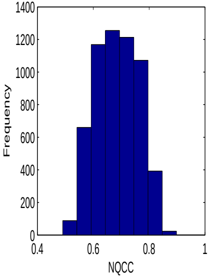
The largest tail quotient correlation coefficients and their corresponding stations information.

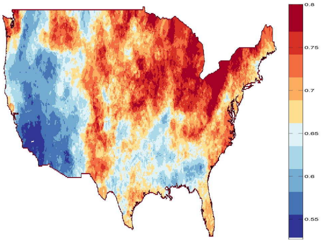
Pair ID	TQCC	Latitude	Longitude	Elevation	Station name
(I)	.0223	31.95 38.62	-112.80 -122.87	512 33	ORGAN PIPE CACTUS N M HEALDSBURG
(II)	.0239	47.55 47.62	-116.17 -117.52	680 718	KELLOGG AIRPORT SPOKANE WSO AIRPORT
(III)	.0363	41.25 41.63	-91.37 -91.52	204 195	COLUMBUS JUNCT 2 SSW IOWA CITY
(IV)	.0214	29.98 45.52	-90.25 -89.20	1 488	NEW ORLEANS WSCMO ARPT SOUTH PELICAN
(V)	.0326	47.93 47.92	-97.17 -97.08	256 253	GRAND FORKS FAA AP GRAND FORKS UNIV NWS
(VI)	.0769	33.98 34.32	-78.00 -77.92	6 12	SOUTHPORT 5 N WILMINGTON 7 N

Nonlinear quotient correlation for less extremes

$$q_n(g) = \frac{g\left(\frac{\max(u_n,X_1)}{\max(u_n,Y_1)},\ldots,\frac{\max(u_n,X_n)}{\max(u_n,Y_n)}\right) + g\left(\frac{\max(u_n,Y_1)}{\max(u_n,X_1)},\ldots,\frac{\max(u_n,Y_n)}{\max(u_n,X_n)}\right) - 2}{g\left(\frac{\max(u_n,X_1)}{\max(u_n,Y_1)},\ldots,\frac{\max(u_n,X_n)}{\max(u_n,Y_n)}\right) \times g\left(\frac{\max(u_n,Y_1)}{\max(u_n,X_1)},\ldots,\frac{\max(u_n,Y_n)}{\max(u_n,X_n)}\right) - 1},$$

where $g(z_1,\ldots,z_n)$ gives the kth largest value, or the pth percentile, of $\{z_1,\ldots,z_n\}$ such that $g(z_1,\ldots,z_n) \geq 1$.

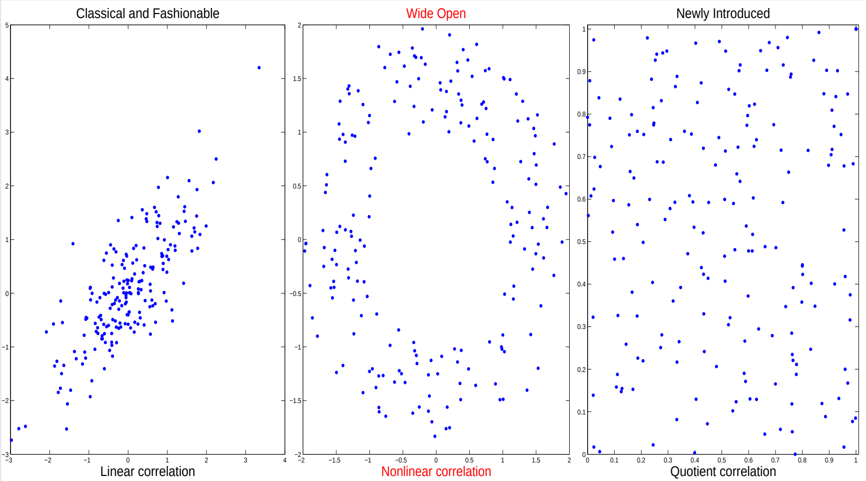




- Note that the shape of nonlinear dependence correlations show a bell shaped curve.

Linear versus nonlinear (extreme)

Asymptotic independence of Pearson's correlation and the quotient correlation: Zhang, Qi and Ma (2010)



We hope:

- TQCC is a sample based alternative to Pearson's correlation coefficient.
- TQCC and NQCC can be applied to many applications in which as long as one uses Pearson's correlation coefficient.
- TQCC may be used to evaluate climate model performance and to guide model building.
- TQCC may result in true sparsity in very large correlation matrices.

Thank You!