RANDOM THRESHOLD DRIVEN TAIL DEPENDENCE MEASURES WITH APPLICATION TO PRECIPITATION DATA ANALYSIS

Author(s): Zhengjun Zhang, Chunming Zhang and Qiurong Cui

Source: *Statistica Sinica*, April 2017, Vol. 27, No. 2 (April 2017), pp. 685-709

Published by: Institute of Statistical Science, Academia Sinica

Stable URL: https://www.jstor.org/stable/26383296

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms



is collaborating with JSTOR to digitize, preserve and extend access to *Statistica Sinica*

This content downloaded from 144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

Statistica Sinica **27** (2017), 685-709
doi: https://doi.org/10.5705/ss.202015.0421

# RANDOM THRESHOLD DRIVEN TAIL DEPENDENCE MEASURES WITH APPLICATION TO PRECIPITATION DATA ANALYSIS

Zhengjun Zhang, Chunming Zhang and Qiurong Cui

*University of Wisconsin*

*Abstract:* This paper first studies the theoretical properties of the tail quotient correlation coefficient (TQCC) which was proposed to measure tail dependence between two random variables. By introducing random thresholds in TQCC, an approximation theory between conditional tail probabilities is established. The new random threshold-driven TQCC can be used to test the null hypothesis of tail independence under which TQCC test statistics are shown to follow a Chi-squared distribution under two general scenarios. The TQCC is shown to be consistent under the alternative hypothesis of tail dependence with a general approximation of max-stable distribution. Second, we apply TQCC to investigate tail dependencies of a large scale problem of daily precipitation in the continental US. Our results, from the perspective of tail dependence, reveal nonstationarity, spatial clusters, and tail dependence from the precipitations across the continental US.

*Key words and phrases:* Climate extremes, conditional tail probability approximation, extreme value theory, hypothesis testing, nonlinear dependence.

## 1. Introduction

Extreme climatic conditions are more often observed in recent years. A natural question is: What is the likelihood of severe weather conditions in your area given that severe weather conditions have been observed/forcasted in some regions, or in the continent on a large scale? Answers to this question can be sought from historical extreme climatic records. Efforts have been made to understand and model extreme weather conditions. In a recent CCSP SAP 3.3 report, Easterling et al. (2008) listed twelve measures to improve our understanding of weather and climate extremes. In our view, two fundamental issues are: how to identify tail dependence between climatic variables, and how to develop statistical models dealing with tail dependence. The focus of this paper is on the first issue.

To address the basic question, we explore the notion of *tail dependence*, also known as *extremal dependence* or *asymptotic dependence*, between the compo-

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

nents of a two-dimensional random vector, which refers to the concurrence of extreme values in the component variables.

**Definition 1.** *Two identically distributed random variables $X$ and $Y$ with distribution function $F$ are called tail independent, if*

$$\lambda = \lim_{u \to x_F} P(Y > u \mid X > u) \tag{1.1}$$

*is 0, where $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$. The quantity $\lambda$, if it exists, is called the bivariate tail dependence index; it quantifies the amount of dependence of the bivariate upper tails. If $\lambda > 0$, $X$ and $Y$ are called tail dependent and we say that there are extreme co-movements between $X$ and $Y$. We further denote by $\mathcal{U}$ the class of bivariate unit Fréchet random variables with marginal distribution function, $F(x) = \exp(-1/x)$, for $x > 0$, such that $\lambda$ exists.*

Sibuya (1960) introduced the idea of asymptotic independence between two random variables with identical marginal distributions, and de Haan and Resnick (1977) extended it to the multivariate case, see also Coles, Heffernan and Tawn (1999). Examples of tail dependence indices of bivariate random variables were presented in Embrechts, McNeil and Straumann (2002). For instance, the tail dependence index of a bivariate normal (Gaussian) random vector is zero as long as the corresponding correlation coefficient is less than one; the tail dependence index of a bivariate $t$ random vector with a positive correlation is greater than zero. Many financial analysts, for example Salmon (2012), blamed a mathematical formula, the Gaussian copula, as the major cause of the 2007–2008 financial crisis mainly because Gaussian random variables are tail independent. This example indicates that tail (in)dependence modeling is of practical importance, see also Embrechts, McNeil and Straumann (2002) for properties and pitfalls of correlations and dependence measures.

In practice, dependent random variables are not necessarily tail dependent. It is thus of importance to check or test whether any two sequences of data are tail dependent or tail independent before choosing a certain class of models for the data. In statistical modeling of tail dependent variables, a significant step is due to Ledford and Tawn (1996, 1997). They introduced a class of models for tail dependence and near tail independence, and constructed test statistics for the null hypothesis of tail dependence. Contrary to their null hypothesis, Zhang (2008b) introduced an empirically efficient test statistic for the null hypothesis of tail independence based on the tail quotient correlation coefficient (TQCC), where the underlying threshold is a constant. We note that the null and alternative

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

hypotheses in Ledford and Tawn (1996, 1997) are reversed in Zhang (2008b). To our knowledge, there do not exist thorough results showing that the existing test statistics are consistent. We construct an alternative expression of Definition 1, and prove the consistency of the test statistic under certain conditions (see Section 2.5.)

In practice, data-driven thresholds are commonly used. This motivates us to consider random thresholds. In Section 2, Fréchet random variables $\{T_{n,t}\}$ are introduced as random thresholds in TQCC. We demonstrate why TQCC is intuitively appealing and easily interpreted. Based on the random variables $\{T_{n,t}\}$, we define a class $\mathcal{T}$ of bivariate random variables, that contains the class $\mathcal{U}$ in Definition 1. It is shown that the tail dependence index of a bivariate random vector in $\mathcal{U}$ is identical to that of a different pair of random variables in $\mathcal{T}$. Test statistics are constructed based on $\mathcal{T}$, using TQCC with random thresholds. These test statistics, facilitated by the approximation theory of tail dependent random variables, are shown to be consistent under the alternative hypothesis of tail dependence.

Computations of TQCC need not jointly model paired random variables, only marginally fitted distributions to exceedances are needed. This is particularly appealing and is comparable to computations of Pearson's correlation coefficients. In practice, Pearson's correlation coefficients are widely used to measure linear relationships among random variables as long as paired random variables are tested to be correlated. TQCC, as a sample-based alternative to Pearson's correlation coefficients, can also be used as a measure of tail dependence as long as paired random variables are tested to be tail dependent.

There is a fast-growing body of literature on modeling climate extremes in recent years. Data modeling and inference can be found in Smith (2003), Naveau et al. (2005), Falk and Michel (2006), Kunkel et al. (2007), Beniston et al. (2007), Cooley, Nychka and Naveau (2007), Bel, Bacro and Lantuejoul (2007), Elek and Márkus (2007), Yiou *et al.* (2008), Zhang (2008a), Smith and Stephenson (2009), Naveau, Zhang and Zhu (2011), and Gilleland, Brown and Ammann (2013), among others. Several papers, such as Peng (1999), Draisma et al. (2004) and Ferreria and de Haan (2004) focused on rare events modeling, particularly of Dutch coastal wind and water level extremes.

The rest of the paper is structured as follows. Section 2 introduces a generalized sample-based tail dependence measure, TQCC, which allows the threshold values to be random. Proposition 1 shows that the tail dependence index of the original pair of random variables is identical to that of a new pair of random vari-

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

ables which incorporates random thresholds. The limiting distributions of TQCC with random threshold values are derived based on a very mild condition (assuming the existence of $\lambda$), or on a random diverging random sequence. Section 3 discusses marginal transformations of non-Fréchet variables and reviews fitting of generalized extreme value distributions for our data analysis. Theorem 4 ensures asymptotic results for marginally transformed observations via estimated parameters and distributions. Section 4 tests tail independence and explores tail dependence of daily precipitation during 1950–1999 recorded at 5,873 stations from NCDC rain gauge data. Section 5 concludes with a brief discussion. Section S1 evaluates the performance of TQCC for testing tail independence in simulation studies. Technical details are postponed to Section S2 Appendix (in a supplementary file together with Section S1, available online).

## 2. Theory of TQCC with Random Thresholds

### 2.1. Hypothesis for tail independence

For the sake of discussion, the existence of $\lambda$ in (1.1) is assumed. Intuitively, tail dependence indices can be estimated using empirical probabilities above certain large threshold $u$ in (1.1), but this may give misleading results since, in general, the estimated values of $\lambda$ may well exceed zero even if $\lambda = 0$. A natural question then arises as to whether or not two random variables are tail dependent. This issue can be formulated as testing the hypotheses:

$$H_0 : X \text{ and } Y \text{ are tail independent} \leftrightarrow H_1 : X \text{ and } Y \text{ are tail dependent}, \quad (2.1)$$

which can also be written as

$$H_0 : \lambda = 0 \longleftrightarrow H_1 : \lambda > 0. \quad (2.2)$$

One major goal of this paper is to test (2.2). In Section 2.2, we use TQCC, a generalized measure of tail dependence, to characterize $\lambda$ and to construct test statistics.

### 2.2. TQCC and its properties

From (1.1), the estimation of the tail dependence index $\lambda$ mainly relies on choice of thresholds $u$ and the dependence mechanism between upper tails of two random variables $X$ and $Y$. This suggests that one can ignore values of $X$ and $Y$ below $u$ and construct sample-based measures of tail dependence or test statistics for testing the null hypothesis by using exceedances above $u$ only. We use two examples to explore this idea. Let

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

$$\xi_1, \ldots, \xi_n, \ \eta_1, \ldots, \eta_n \tag{2.3}$$

be a sequence of independent unit Fréchet random variables. Example 1 presents tail independent random variables $(X_i, Y_i)$ with upper tails equivalent to the upper tail of a unit Fréchet distribution.

**Example 1** (Tail independent case). *Let* $\{(L_{ni},\ Q_{ni})\}_{i=1}^n$ *be a sample of independent random pairs, where* $L_{ni}$ *is dependent of* $Q_{ni}$ *and both are supported on* $(0, u_n]$ *for a positive large value* $u_n$. *Let* $X_i = \xi_i \mathrm{I}(\xi_i > u_n) + L_{ni}\mathrm{I}(\xi_i \le u_n)$ *and* $Y_i = \eta_i \mathrm{I}(\eta_i > u_n) + Q_{ni}\mathrm{I}(\eta_i \le u_n)$, $i = 1, \ldots, n$, *where* $\mathrm{I}(\cdot)$ *denotes the indicator function. It follows that* $\max(X_i, u_n)$ $(= \max(\xi_i, u_n))$ *and* $\max(Y_i, u_n)$ $(= \max(\eta_i, u_n))$ *are independent. Furthermore,* $\{(X_i,\ Y_i)\}_{i=1}^n$ *is a random sample of tail independent random variables. In fact, for an arbitrary dependence structure between* $L_{ni}$ *and* $Q_{ni}$, *the constructed* $X_i$ *and* $Y_i$ *are tail independent.*

The sequence of thresholds $u_n$ in Example 1 is deterministic. In Sections 2.4 and 2.5, we allow $u_n$ to be in a class of random variables that diverge to infinity in probability. It can be seen that with such $u_n$, $\max(X_i, u_n)$ and $\max(Y_i, u_n)$ are no longer independent, but continue to be tail independent. More generally, we propose a sample-based measure of tail dependence.

**Definition 2.** *If* $\{(X_i,\ Y_i)\}_{i=1}^n$ *is a random sample of unit Fréchet random variables* $(X,\ Y)$,

$$\mathrm{q}_{u_n} = \frac{\max\limits_{1\le i\le n}\{\max(X_i, u_n)/\max(Y_i, u_n)\} + \max\limits_{1\le i\le n}\{\max(Y_i, u_n)/\max(X_i, u_n)\} - 2}{\max\limits_{1\le i\le n}\{\max(X_i, u_n)/\max(Y_i, u_n)\} \times \max\limits_{1\le i\le n}\{\max(Y_i, u_n)/\max(X_i, u_n)\} - 1} \tag{2.4}$$

*is the tail quotient correlation coefficient* (TQCC).

If $u_n \equiv u$ (a constant), the TQCC reduces to the one in Zhang (2008b). In the current paper, $u_n$ is allowed to be random.

REMARK 1. It is not necessary to require the marginal distribution to be unit Fréchet. Any positive scales (as long as $X_i$ and $Y_i$ have a tail equivalent scale) can be used at (2.4). With this relaxation, $\mathrm{q}_{u_n}$ can be regarded as a nonlinear dependence measure. One could use the second largest values, the third largest values, and the $p$th percentiles of the ratios etc. in the definition.

To study the property of $\mathrm{q}_{u_n}$, take $\zeta_n^{\mathrm{I}} = \max_{1\le i\le n}\{\max(X_i, u_n)\ /\max(Y_i, u_n)\}$ and $\zeta_n^{\mathrm{II}} = \max_{1\le i\le n}\{\max(Y_i, u_n)/\max(X_i, u_n)\}$. Then $\mathrm{q}_{u_n} = (\zeta_n^{\mathrm{I}} + \zeta_n^{\mathrm{II}} - 2)/(\zeta_n^{\mathrm{I}} \times \zeta_n^{\mathrm{II}} - 1)$. We see that $P(\zeta_n^{\mathrm{I}} > 1) > 0$ and $P(\zeta_n^{\mathrm{II}} > 1) > 0$ for any $n \ge 1$, and that both $\zeta_n^{\mathrm{I}}$ and $\zeta_n^{\mathrm{II}}$ are greater than or equal to 1 almost surely as $n \to \infty$. When

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

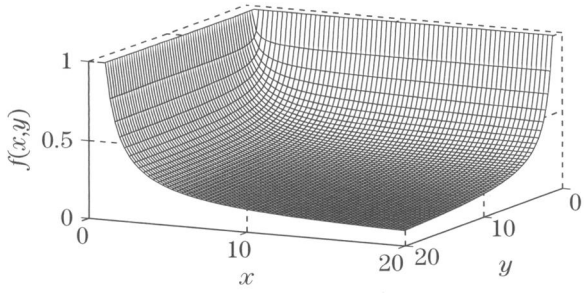



Figure 1. Illustration of the function $f(x, y)$ in (2.5).

both $\zeta_n^{\mathrm{I}}$ and $\zeta_n^{\mathrm{II}}$ are close to 1 for a sufficiently large sample size $n$, one can say that there exist co-movements between $X_i$ and $Y_i$, especially for their very large observations. A similar conclusion holds when the denominator of (2.4) is close to zero. Now $\mathrm{q}_{u_n} = f(\zeta_n^{\mathrm{I}}, \zeta_n^{\mathrm{II}})$, where

$$f(x, y) = \frac{x + y - 2}{xy - 1}, \quad \text{for } x \geq 1,\ y \geq 1,\ x + y > 2. \tag{2.5}$$

Figure 1 plots the function $f$. The case of $x + y = 2$ is trivial, since it implies that $x$ and $y$ are equal when they both are greater than or equal to 1. It is easy to see that $0 \leq f(x, y) \leq 1$, $f(1, y) = 1$, $f(x, 1) = 1$, and $\partial f/\partial x = -(y-1)^2/(xy-1)^2 < 0$, $\partial f/\partial y = -(x-1)^2/(xy-1)^2 < 0$, $\partial^2 f/(\partial x \partial y) = -2(x-1)(y-1)/(xy\ -1)^3 < 0$, which imply $f(x_1,\ y_1) \leq f(x_2,\ y_2)$, when $x_1 \geq x_2$ and $y_1 \geq y_2$. The properties of $\mathrm{q}_{u_n} = f(\zeta_n^{\mathrm{I}}, \zeta_n^{\mathrm{II}})$ with $f$ defined in (2.5) can be summarized as follows. First, $\mathrm{q}_{u_n}$ takes values between 0 and 1 for a sufficiently large $n$. Second, the monotonicity of $f(\zeta_n^{\mathrm{I}}, \zeta_n^{\mathrm{II}})$ in $\zeta_n^{\mathrm{I}}$ and/or $\zeta_n^{\mathrm{II}}$ suggests that for a fixed sample size $n$, the larger the $\zeta_n^{\mathrm{I}}$ or the $\zeta_n^{\mathrm{II}}$, the smaller the $\mathrm{q}_{u_n}$, hence the less the agreement of changing magnitudes at tails. Third, $f(\zeta_n^{\mathrm{I}}, 1) = 1$ and $f(1, \zeta_n^{\mathrm{II}}) = 1$. Fourth, as long as one of $\zeta_n^{\mathrm{I}}$ and $\zeta_n^{\mathrm{II}}$ is very large, $\mathrm{q}_{u_n}$ is largely determined by the smaller one. This property is very useful, especially when outliers exist in one of the two sequences.

The quantity $\mathrm{q}_{u_n}$ assesses the magnitude of tail dependence between $X$ and $Y$. If $\lambda = 0$, Theorem 5.1 of Zhang (2008b) shows that when $u_n \to u^*$ (a finite constant) as $n \to \infty$, $\mathrm{q}_{u_n}$ tends to zero almost surely. This is illustrated in Example 1. If $\mathrm{q}_{u_n}$ is close to 1, either $\zeta_n^{\mathrm{I}}$ or $\zeta_n^{\mathrm{II}}$ is close to 1, thus by the boundary property of $f$, there are tail co-movements between $\{X_i\}$ and $\{Y_i\}$, hence $X$ and $Y$ are nearly completely tail dependent.

**Example 2** (Tail dependent case)**.** *Using* (2.3)*, take* $X_i^* = \max\{a\xi_i,\ (1\ -a)\eta_i\}$

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

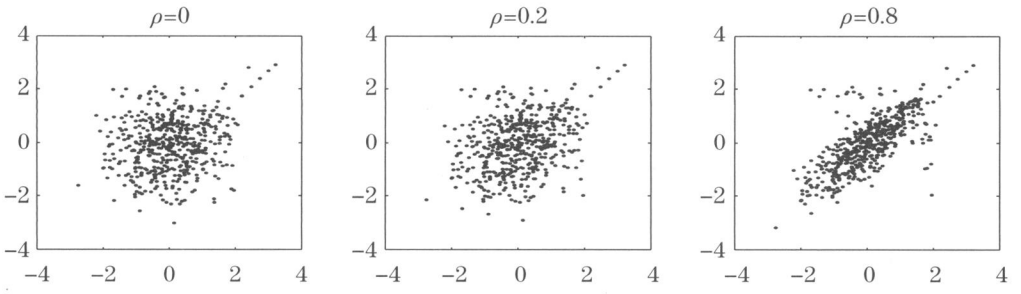



Figure 2. Scatterplot of $\{(\Phi^{-1}(\exp(-1/X_i)), \Phi^{-1}(\exp(-1/Y_i)))\}_{i=1}^{500}$ in Example 2. Values at lower left regions in three panels are drawn from a bivariate standard normal random variable with correlation coefficient $\rho$. Here $u_n$ is set to be the minimum of the 95th percentile of $\{X_i\}$ and the 95th percentile of $\{Y_i\}$.

*and* $Y_i^* = \max\{(1-b)\xi_i,\ b\eta_i\}$, *where* $0 < a < 1$ *and* $0 < b < 1$. *Suppose that* $\{(\epsilon_{1i},\ \epsilon_{2i})\}_{i=1}^n$ *is a sample of independent random pairs from a bivariate standard normal random variable with correlation coefficient* $\rho$, *and independent of* $\{\xi_i\}$ *and* $\{\eta_i\}$. *For* $i = 1, \ldots, n$, *let*

$$L_{ni} = -\frac{1}{\log\{\Phi(\epsilon_{1i})\exp(-1/u_n)\}}, \quad Q_{ni} = -\frac{1}{\log\{\Phi(\epsilon_{2i})\exp(-1/u_n)\}}, \tag{2.6}$$

*where* $\Phi(\cdot)$ *denotes the standard normal distribution function. If* $X_i = X_i^* \mathrm{I}(X_i^* > u_n) + L_{ni}\,\mathrm{I}(X_i^* \le u_n)$, *and* $Y_i = Y_i^* \mathrm{I}(Y_i^* > u_n) + Q_{ni}\,\mathrm{I}(Y_i^* \le u_n)$, *then for any* $a \in (0,1)$ *and* $b \in (0,1)$,

$$\mathrm{q}_{u_n} \xrightarrow{\mathrm{P}} \lim_{u\to\infty} P(X_1 > u \mid Y_1 > u) = \lim_{u\to\infty} P(X_1^* > u \mid Y_1^* > u) = \lambda^* > 0,$$

*provided that* $u_n \to u^*$ (*a finite constant*) *as* $n \to \infty$. *Here* $\xi_i$, $\eta_i$, $X_i^*$, $Y_i^*$, $X_i$ *and* $Y_i$ *all have the unit Fréchet distribution. Figure 2 displays the scatterplot of simulated tail dependent data* $\{(\Phi^{-1}(\exp(-1/X_i)),\ \Phi^{-1}(\exp(-1/Y_i)))\}_{i=1}^{500}$ *with different values of the correlation coefficient* $\rho$ *when* $a = .25$ *and* $b = .25$, *where* $\Phi^{-1}(p) = \inf\{x : \Phi(x) \ge p\}$. *It shows that when studying tail dependence, observations below thresholds* $u_n$ *can be discarded, whereas values of* $\mathrm{q}_{u_n}$ *evaluated from different models (different* $\rho$ *values) are identical for large values of* $u_n$.

### 2.3. Limit of $\mathrm{q}_{u_n}$ for random variables in the broader M4 model

The sequence $(X_i^*,\ Y_i^*)$ in Example 2 is a special case of a multivariate maxima of moving maxima process (called M4 processes) that have been studied in Smith and Weissman (1996) and Zhang and Smith (2004). Theorem 2 of Deheuvels (1983) establishes that, for any finite $m$, the joint distribution of $m$

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

consecutive values from a max-stable process (including M4) with Fréchet margins can be approximated by a weighted maximum, $Y_{id} = \max_{-\infty<k<\infty} \alpha_{idk} Z_k$, $d = 1, \ldots, D$, for suitable coefficients $\{\alpha_{idk}\}$, where $\{Z_k\}$ are i.i.d. unit Fréchet. Here "can be approximated" refers to convergence in distribution. See Smith and Weissman (1996) for detailed arguments. Heffernan, Tawn and Zhang (2007) and Zhang (2009) suggest that the M4 process model constitutes a rich family which can be used in modeling a wide range of multivariate dependence. For $X_i = Y_{id}$ and $Y_i = Y_{id'}$ belonging to the M4 model, it can be shown that the corresponding $\mathrm{q}_{u_n}$ tends to a positive quantity in $(0, 1]$ in probability, under Condition A1 (to be defined in Section 2.5), for the sequence $\{(X_i, Y_i)\}$ and $u_n \xrightarrow{\mathrm{P}} u$ (a finite constant or infinity).

These results suggest that as long as two random variables $X$ and $Y$ (not necessarily unit Fréchet) are tail dependent, the M4 model can be used to approximate tail values of $X$ and $Y$, after being marginally transformed to unit Fréchet scales, then one can use the approximate model to conduct statistical inference, and $\mathrm{q}_{u_n}$ can be used as a sample-based measure of tail dependence.

### 2.4. Approximation theory of conditional tail probabilities

Here $u_n$ is allowed to be random. We first show in Proposition 1 that the tail dependence index $\lambda_{X,Y}$ of $X$ and $Y$, where $(X, Y) \in \mathcal{U}$ in Definition 1, is equal to the tail dependence index $\lambda^*_{X,Y}$ of $\max(X, u_n)$ and $\max(Y, u_n)$ using a random threshold, $u_n = T_{n,t}$, where $T_{n,t}$ is a Fréchet variable with the distribution function $\exp(-n/x^t)$ for $x > 0$, $n \geq 0$ and $t > 0$. It is easy to see that the class of random vectors $\mathcal{T} = \{(\max(X, T_{n,t}),\ \max(Y,\ T_{n,t})) :\ n \geq 0,\ t > 0\}$ contains $\mathcal{U}$. This connection leads to a consistent test procedure for tail independence.

**Proposition 1.** *If $X$ and $Y$ are unit Fréchet random variables, $T_{n,t}$ has the distribution function $\exp(-n/x^t)$, for $x > 0$ and $t > 1$, and $T_{n,t}$ is independent of $(X, Y)$, then*

(i) *for any finite $n \geq 0$,*

$$\lim_{u\to\infty} \frac{P(X > u,\ Y > u)}{P(X > u)} = \lim_{u\to\infty} \frac{P\{\max(X, T_{n,t}) > u,\ \max(Y, T_{n,t}) > u\}}{P\{\max(X, T_{n,t}) > u\}}. \tag{2.7}$$

*Suppose $X$ and $Y$ are tail independent with*

$$g(u) = \frac{P(X > u)}{P(X > u,\ Y > u)}. \tag{2.8}$$

*If $X'$ and $Y'$ are independent unit Fréchet random variables, and $T_{n,t}$ is inde-*

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

*pendent of* $(X', Y')$, *with* $h(u) = \min(g(u), n^{-1}u^{t-1})$ *and*

$$g^*(u) = \frac{P\{\max(X', T_{n,t}) > h(u),\ \max(Y', T_{n,t}) > h(u)\}}{P\{\max(X', T_{n,t}) > h(u)\}},$$

*then as* $u \to \infty$,

(ii) *for any* $t > 1$,

$$\frac{P(X > u,\ Y > u)}{P(X > u)} = O\Big(g^*(u)\Big), \tag{2.9}$$

$$\frac{P\{\max(X, T_{n,t}) > u,\ \max(Y, T_{n,t}) > u\}}{P\{\max(X, T_{n,t}) > u\}} = O\Big(g^*(u)\Big). \tag{2.10}$$

Equation (2.7) is an alternative expression of (1.1), so $\lambda_{X,Y} = \lambda^*_{X,Y}$. If a test procedure based on $(X, Y)$ concludes that $X$ and $Y$ are tail independent, one can also conclude that $\max(X, T_{n,t})$ and $\max(Y, T_{n,t})$ with $t > 1$ are tail independent, and vice versa.

Equation (2.9) can be expressed as

$$\frac{P(X > u,\ Y > u)}{P(X > u)} = O\Big(\frac{P\{\max(h^{-1}(X'), h^{-1}(T_{n,t})) > u,\ \max(h^{-1}(Y'), h^{-1}(T_{n,t})) > u\}}{P\{\max(h^{-1}(X'), h^{-1}(T_{n,t})) > u\}}\Big). \tag{2.11}$$

REMARK 2. For any function $g(u)$ (not necessarily defined by (2.8)) such that $g(u) \to \infty$ as $u \to \infty$, (2.9) and (2.11) do not hold when $X$ and $Y$ are tail dependent. In such case, the left hand of (2.11) is greater than zero, while the right hand side is zero. When $X$ and $Y$ are tail independent, the $g(u)$ (defined by (2.8)) or the $1/g(u)$ function need not be $g(u) \sim L(u)u^{-1+1/\eta}$, $\eta \in (0, 1)$, where $L(u)$ is a slowly varying function, as defined in Ledford and Tawn (1997).

Equations (2.9)–(2.11) show that the conditional tail probabilities between tail independent random variables can be approximated by the conditional tail probability of a pair of random variables $(\max(h^{-1}(X'), h^{-1}(T_{n,t})), \max(h^{-1}(Y'), h^{-1}(T_{n,t})))$. Bearing this in mind, under $H_0$ the tail independence of $(X, Y)$ over a common random threshold $T_{n,t}$ is equivalent to the tail independence of $(\max(h^{-1}(X'), h^{-1}(T_{n,t})), \max(h^{-1}(Y'), h^{-1}(T_{n,t})))$ for all $t > 1$. Thus, we study the hypotheses (which is connected to (2.2)),

$$H_0^*: \text{ (2.10) holds } \longleftrightarrow H_1^*: \text{ (2.10) does not hold.} \tag{2.12}$$

If a test statistic calculated from a bivariate sample of $(\max(X, T_{n,t}), \max(Y, T_{n,t}))$ does not reject the null hypothesis of data being drawn from $(\max(X', T_{n,t}), \max(Y', T_{n,t}))$ with a chosen $t$ value and a connecting function $g(u) \to \infty$

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

as $u \to \infty$, within a pre-specified Type I error rate, one can first conclude that (2.9) is true, then conclude that $X$ and $Y$ are tail independent. In Section 2.5, we see that the choices of $t$ and $g(u)$ are not essential.

### 2.5. Limit distribution of TQCC with random thresholds

We derive the limiting distribution of TQCC, $\mathrm{q}_{u_n}$, in cases of two random thresholds $u_n$: $u_n = T_{n,t} \xrightarrow{\mathrm{P}} \infty$ in Theorem 3; $u_n = u_n^* a_n$ with $u_n^* \xrightarrow{\mathrm{P}} u^* \in (0, \infty)$, $a_n \to \infty$ and $a_n/n \to 0$ as $n \to \infty$ in Theorem 4.

**Assumption T1**: For $1 < t < 1 + \delta$, $\delta > 0$, paired tail independent random variables $(X_i, Y_i)$ satisfy

$$\frac{\max_{1\le i\le n} \max(X_i, T_{n,t})/\max(Y_i, T_{n,t})}{\max_{1\le i\le n} \max(X_i, T_{n,t})/T_{n,t}} = 1 + o_p(1),$$
$$\frac{\max_{1\le i\le n} \max(Y_i, T_{n,t})/\max(X_i, T_{n,t})}{\max_{1\le i\le n} \max(Y_i, T_{n,t})/T_{n,t}} = 1 + o_p(1).$$

REMARK 3. Assumption T1 is natural since the tail independence of $(X_i, Y_i)$ (also $(\max(X_i, T_{n,t}), \max(Y_i, T_{n,t}))$) implies $\max(X_i, T_{n,t})$, $i = 1, \ldots, n$ and $\max(Y_i, T_{n,t})$, $i = 1, \ldots, n$, will hug $T_{n,t}$ in each axis direction when the threshold value $T_{n,t}$ is sufficiently large.

**Proposition 2.** *If $g(u) \sim L(u)u^{-1+1/\eta}$, $\eta \in (0, 1]$, then T1 holds for $t\eta < 1$ when $\eta < 1$, T1 does not hold when $\eta = 1$.*

REMARK 4. Setting $u_n = T_{n,t}$, Example 1 satisfies T1.

**Theorem 3.** *Suppose for given $t > 1$, all random variables $X_1', \ldots, X_n'$, $Y_1', \ldots, Y_n'$, and $T_{n,t}$ are independent, where $X_i'$ and $Y_i'$ are unit Fréchet random variables, and $T_{n,t}$ has the distribution function $\exp(-n/x^t)$ for $x > 0$. If $A_{n,t} = n\{1 - \exp(-1/T_{n,t})\}$, then*

(i) *for $z > 0$,*

$$\lim_{n\to\infty} P\Big\{A_{n,t}^{-1} \max_{1\le i\le n} \frac{\max(X_i', T_{n,t})}{\max(Y_i', T_{n,t})} \le z\Big\} = \exp(-\frac{1}{z});$$

*for $z_1 > 0$ and $z_2 > 0$,*

$$\lim_{n\to\infty} P\Big\{A_{n,t}^{-1} \max_{1\le i\le n} \frac{\max(X_i', T_{n,t})}{\max(Y_i', T_{n,t})} \le z_1,\ A_{n,t}^{-1} \max_{1\le i\le n} \frac{\max(Y_i', T_{n,t})}{\max(X_i', T_{n,t})} \le z_2\Big\}$$
$$= \exp(-\frac{1}{z_1}) \exp(-\frac{1}{z_2}).$$

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

(ii) *Further,*

$$2n\{1-\exp(-\frac{1}{T_{n,t}})\}\mathrm{q}'_{T_{n,t}} \xrightarrow{\mathcal{L}} \chi^2_4, \tag{2.13}$$

*where $\chi^2_4$ is a chi-squared random variable with four degrees of freedom; $\mathrm{q}'_{T_{n,t}}$ is defined as $\mathrm{q}_{u_n}$ by replacing $u_n$ by $T_{n,t}$, $X_i$ by $X'_i$, and $Y_i$ by $Y'_i$, $i=1,\ldots,n$ respectively.*

**Corollary 1.** *If random variables $X_1,\ldots,X_n$, $Y_1,\ldots,Y_n$ satisfy T1, where $X_i$ and $Y_i$ are unit Fréchet random variables, under the null hypothesis of* (2.12),

$$2n\{1-\exp(-\frac{1}{T_{n,t}})\}\mathrm{q}_{T_{n,t}} \xrightarrow{\mathcal{L}} \chi^2_4, \tag{2.14}$$

*where $\chi^2_4$ is a chi-squared random variable with four degrees of freedom.*

**Theorem 4.** *Suppose $\{X'_1,\ldots,X'_n,\ Y'_1,\ldots,Y'_n\}$ are independent unit Fréchet random variables and $u_n = u^*_n a_n$ satisfies $u^*_n \xrightarrow{\mathrm{P}} u^*$, $a_n \to \infty$, and $a_n/n \to 0$ as $n\to\infty$, where $u^* \in (0,\infty)$ is a constant. Then $2n\{1-\exp(-1/u_n)\}\mathrm{q}'_{u_n} \xrightarrow{\mathcal{L}} \chi^2_4$.*

**Corollary 2.** *Suppose random variables $X_1,\ldots,X_n$, $Y_1,\ldots,Y_n$ satisfy T1, where $X_i$ and $Y_i$ are unit Fréchet random variables. Under the null hypothesis of* (2.12), $2n\{1-\exp(-1/u_n)\}\mathrm{q}_{u_n} \xrightarrow{\mathcal{L}} \chi^2_4$.

When applying (2.9)-(2.14) to data, one might think it obligatory to determine a joint tail region, choose a $t>1$ value, and simulate a threshold $T_{n,t}$ before applying (2.14). However, the process is much simplified due to the following. The inclusion of $T_{n,t}$ in the $\mathrm{q}_{u_n}$ definition naturally projects all joint observations into a tail region. As a result, the thresholds $u$ and $g(u)$ in both sides of (2.9) are not essential as long as a simulated $T_{n,t}$ value in the $\mathrm{q}_{T_{n,t}}$ computation is large enough. Then too, the limiting distribution in (2.14) does not depend on the power transformation index $t$. Suppose that $u_1(t_1) \sim T_{n,t_1}$ for a pre-specified $t_1$. For any $s_1>0$, it follows that $u_1^{1/s_1}(t_1) \sim T_{n,s_1\times t_1}$. One can use either the simulated value of $u_1(t_1)$ or $u_1^{1/s_1}(t_1)$ as a threshold in TQCC. For any given $t_1$ and a simulated value of $u_1(t_1)$ and any pre-specified threshold value $u_n$ that does not depend on $t$, one can solve $u_1^{1/s_1}(t_1)=u_n$ for $s_1$, and use $u_n$ to substitute for $T_{n,t}$. This is important for applications as one does not need to deal with the simulated threshold $T_{n,t}$. For a given sample size $n$, we can set $u_n$ as the minimum of the sample $100p$th percentile of $\{X_i\}$ and the sample $100p$th percentile of $\{Y_i\}$. Here $p$ can be chosen as, say, .8, .9, .95, .975 etc. Such choices of $u_n$ were treated as non-random in Zhang (2008b), and the resulting testing procedure had high empirical power under $H_1$ of (2.2). Employing exceedances

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

over a threshold is a common practice in extreme value theory and statistical inference, see Coles (2001). Theorem 4 supports these empirical arguments.

Applying Theorems 3 and 4, Corollaries 1 and 2 and the arguments above, we propose a testing procedure.

(1) For a given level $\alpha$, if $2n\{1-\exp(-1/u_n)\}\mathrm{q}_{u_n} > \chi^2_{4,\alpha}$, then $H_0$ of (2.1) is rejected, and we conclude that there exists tail dependence between two random variables of interest. Here $\chi^2_{4,\alpha}$ is the upper $\alpha$ percentile of the $\chi^2_4$ distribution. In practice, we just choose $u_n$ as the sample percentile.

(2) If $H_0$ of (2.1) is rejected, using the arguments in Section 2.3 and the results in Proposition 3, (2.4) can be used as an estimate of $\lambda$. This estimate may be conservative, as shown in the proof of our Theorem 5 and the results obtained in Wang (2012).

To derive the asymptotic power of the test statistics, we apply the distribution approximation discussed in Section 2.3. We use a condition to explore the discriminating power of the test.

**Condition A1**: Let $\{\xi_{i,j} : i \geq 1, -\infty < j < \infty\}$ be mutually independent unit Fréchet random variables, and suppose that tail dependent unit Fréchet random variables $X_i$ and $Y_i$ have the approximation representations

$$X_i = \max_{-\infty<j<\infty} \alpha_j \xi_{i,j}, \quad Y_i = \max_{-\infty<j<\infty} \beta_j \xi_{i,j}, \tag{2.15}$$

where $\alpha_j \geq 0$, $\beta_j \geq 0$, $\sum_{j=-\infty}^{\infty} \alpha_j = 1$, $\sum_{j=-\infty}^{\infty} \beta_j = 1$, and $\alpha_j > 0$ iff $\beta_j > 0$ for all $j$. Then there exist $0 < c_1 < c_2 < \infty$ such that $c_1 \leq \alpha_j/\beta_j \leq c_2$.

Based on Deheuvels (1983) and Smith and Weissman (1996), (2.15) can arbitrarily closely approximate (in distribution) jointly max-stable unit Fréchet random variables. Zhang (2009) further established approximation theory when the range of index $j$ is finite. Under A1, we establish Theorem 5 for the asymptotic power of the test. It is possible to explore the consistency under alternatives beyond (2.15).

**Theorem 5.** *Suppose that* $\{X_i\}_{i=1}^n$ *and* $\{Y_i\}_{i=1}^n$ *satisfy* A1. *Then for* $u_n = T_{n,t} \xrightarrow{\mathrm{P}} \infty$ *in Theorem* 3, *and* $u_n = u_n^* a_n$ *with* $u_n^* \xrightarrow{\mathrm{P}} u^* \in (0,\infty)$, $a_n \to \infty$ *and* $a_n/n \to 0$ *in Theorem* 4, $2n\{1-\exp(-1/u_n)\}\mathrm{q}_{u_n} \xrightarrow{\mathrm{P}} \infty$, *and the test is consistent.*

**Proposition 3.** *Suppose* $\alpha_1 + \alpha_2 = 1$ *and* $\beta_1 + \beta_2 = 1$ *in (2.15). Then* $\mathrm{q}_{u_n} \xrightarrow{\mathrm{P}} \lambda$.

The convergence rate of $\mathrm{q}_0$, using $\{u_n = 0\}$, is $1/\sqrt{n}$ when $X$ and $Y$ follow a joint bivariate Fréchet distribution (see Wang (2012)). As a side note, a result

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

gives the limiting distribution of $q_{u_n}$ for normally distributed random variables without marginally transforming them to unit Fréchet variables.

**Proposition 4.** *Suppose* $\{(\epsilon_{i1}, \epsilon_{i2}) : i = 1, \ldots, n\}$ *is a random sample from a standard bivariate normal random variable with correlation coefficient* $|\rho| < 1$. *If* $N\mathrm{q}_n = (\max_{1\le i\le n}\{\epsilon_{i1}/\epsilon_{i2}\} + \max_{1\le i\le n}\{\epsilon_{i2}/\epsilon_{i1}\} - 2)/(\max_{1\le i\le n}\{\epsilon_{i1}/\epsilon_{i2}\} \times \max_{1\le i\le n}\{\epsilon_{i2}/\epsilon_{i1}\} - 1)$, *then* $(2n\sqrt{1-\rho^2}/\pi)N\mathrm{q}_n \xrightarrow{\mathcal{L}} \chi^2_4$, *as* $n \to \infty$.

Comparing Proposition 4 with Theorems 3 and 4, we see the convergence rate, $1/n$, of the limiting distribution of $\mathrm{q}_{u_n}$ and $N\mathrm{q}_n$ under different null hypotheses, either tail independence or independence. Theorem 5 and the results in Wang (2012) suggest that the test based on Theorems 3 and 4 is consistent. The limiting distribution of $N\mathrm{q}_n$ under the bivariate normal assumption in Proposition 4 and the unit Fréchet marginal (after transformation) is still an open problem. Zhang, Qi and Ma (2011) proved that the sample-based Pearson correlation coefficient and TQCC with $u_n = 0$ are asymptotically independent. A combination of these two coefficients outperforms many existing test statistics in testing the hypothesis of independence. Clearly, Proposition 4 is not for testing tail independence as the threshold is taken at $-\infty$. Yet, the bivariate normal distribution satisfies the properties in Proposition 1 and (2.11). In Zhang (2008b), simulation examples of bivariate normal random variables with $\rho$ up to 0.8 showed that the Type I errors were controlled within nominal levels with a sample size of 300. We also conducted extensive simulation studies and found that with the values of $\rho$ and growing sample sizes, Type I errors are still controlled within the nominal levels. Draisma et al. (2004) studied a maximum likelihood approach in tail dependence estimation and reported that their approach worked well for a bivariate normal distribution with $\rho$ up to 0.6. In Section S1, we report new simulation examples and compare the results with the results from two recently published test statistics.

Before we close this section, we remark that TQCC is suitable for asymptotic dependence, while the coefficient of tail dependence, the $\eta$ of Ledford and Tawn (1996, 1997), or equivalently $\bar{\chi}$ (see (Coles, 2001, p.164)), is an excellent measure for the dependence in asymptotic independence; see some counter examples of the existence of $\eta$ in Schlather (2001). In practice, these two measures can both be pronounced. In this present work, we apply TQCC measure to test asymptotic independence for a large scale problem of precipitations, then report TQCC values as tail dependence indices once the hypotheses of asymptotic independence are rejected.

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

So far we have established the TQCC measure for examining Fréchet distributed random variables. In Section 3, we discuss its validity after marginal transformation, often used on data before applying the tests.

## 3. Marginal Transformation and Generalized Extreme Value Distributions

### 3.1. Limiting distribution of TQCC with estimated parameters in GEV

In obtaining unit Fréchet scales via marginal distribution transformations, the results in Section 2.5 are valid using the true marginal distributions. However, in practice true marginal distributions are unknown. In extreme value analysis, the generalized extreme value (GEV) distribution or the generalized Pareto distribution (GPD) is often first fitted to observed data, and then data are transformed to unit Fréchet scales, see Pickands (1975), Embretchts, Klüppelberg and Mikosch (1997), and Smith (2003) for details. The GEV has the form

$$H(x;\xi,\mu,\psi) = \exp[-\{1+\frac{\xi(x-\mu)}{\psi}\}_+^{-1/\xi}] \tag{3.1}$$

for local maxima of observations, where $\mu$ is a location parameter, $\psi > 0$ is a scale parameter, and $\xi$ is a shape parameter. In principle, unit Fréchet scales can be obtained using the transformation $-1/\log\{H(x;\xi,\mu,\psi)\}$. In this section, we establish results to show that the calculated TQCC, with estimated parameters in GEV, results in the same limiting distribution as if the true parameters in GEV were known. For simplicity, we assume $\mu = 0$ and $\psi = 1$ in (3.1).

**Theorem 4.** *Let $\{(X_i,\ Y_i)\}_{i=1}^n$ be a sample of T1 type tail independent random variables $(X,\ Y)$ whose marginal domains of attraction are GEV random variables with shape parameters $\xi_X = \xi_{0;X}$ and $\xi_Y = \xi_{0;Y}$, respectively. Suppose estimators of $\xi_X$ and $\xi_Y$, $\hat{\xi}_X = \xi_{n;X}(X_1,\ldots,X_n)$ and $\hat{\xi}_Y = \xi_{n;Y}(Y_1,\ldots,Y_n)$, satisfy $n^{\alpha_X}(\hat{\xi}_X - \xi_{0;X}) \xrightarrow{\mathcal{L}} W_X$ and $n^{\alpha_Y}(\hat{\xi}_Y - \xi_{0;Y}) \xrightarrow{\mathcal{L}} W_Y$, where $\alpha_X > 0$, $\alpha_Y > 0$ and $W_X$ and $W_Y$ are random variables. If $\hat{X}_i = -1/\log\{H(X_i;\hat{\xi}_X)\}$, $\hat{Y}_i = -1/\log\{H(Y_i;\hat{\xi}_Y)\}$ and*

$$\hat{\mathrm{q}}_{u_n} = \frac{\max\limits_{1\le i\le n}\{\max(\hat{X}_i,u_n)/\max(\hat{Y}_i,u_n)\} + \max\limits_{1\le i\le n}\{\max(\hat{Y}_i,u_n)/\max(\hat{X}_i,u_n)\} - 2}{\max\limits_{1\le i\le n}\{\max(\hat{X}_i,u_n)/\max(\hat{Y}_i,u_n)\} \times \max\limits_{1\le i\le n}\{\max(\hat{Y}_i,u_n)/\max(\hat{X}_i,u_n)\} - 1}, \tag{3.2}$$

*then under the conditions in Theorems 3 and 4, $2n\{1-\exp(-1/u_n)\}\hat{\mathrm{q}}_{u_n} \xrightarrow{\mathcal{L}} \chi_4^2$.*

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

For applications, Theorem 4 lends support to applying TQCC to transformed data in which those observed values larger than the chosen threshold value are transformed using the inverse of fitted GEV distribution functions, while observed values smaller than the threshold value are transformed based on rank transformation. As TQCC also uses a threshold, one can see that observations smaller than the threshold do not affect the inference using TQCC.

A broader Lemma 2 is needed to prove Theorem 4; it is a powerful tool for proving uniform convergence of transformed random sequences and could be applicable to other circumstances.

## 4. Data application

The data are daily precipitation totals covering the period 1950–1999 over 5,873 stations in the continental USA (excluding Alaska and Hawaii). The data units are tenths of a millimeter. The data were obtained from Dr. Pavel Groisman of the National Climatic Data Center (NCDC), the same as used by Smith, Grady and Hegerl (2007) and Shamseldin et al. (2008).

Despite extensive work in analyzing extreme precipitation events, especially detections of trend, extreme value theory gives a fresh impetus in analyzing climate changes. References include Smith, Grady and Hegerl (2007), Karl and Knight (1998), Groisman et al. (1999), and Hegerl et al. (2006), among others. Our main focus is on illustrating how variables are tail dependent on each other.

We fit GEV to each of 5,873 series, performed marginal transformations, conducted TQCC-based tail independence tests, and report the tail dependence measure (TQCC) after controlling the false discover rate (FDR) at level 0.05 using a nonparametric BH multiple testing procedure, referring to Benjamini and Hochberg (1995) and Benjamini and Yekutieli (2001).

### 4.1. GEV fitting and extreme precipitation comparison

Regarding fitting GEV to data, Smith (1989) showed how a process of exceedances over a high threshold can be modeled in terms of the limiting GEV distribution function of form (3.1). This procedure was used to fit the data above a certain threshold. In each fitting, approximately 10% of observations were above the chosen threshold value. For the purpose of illustration, in Table 1 we present fitted values $\hat{\xi}$ of parameters in (3.1) for six selected stations. The top three stations have the smallest $\hat{\xi}$, whereas the bottom three have the largest.

From Table 1, one can see that the three stations whose precipitations have shortest tails are around latitude 40°N. Stations with longest tailed precipitations

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

Table 1. Information of six selected stations. The top three stations have the smallest $\hat{\xi}$, and the bottom three have the largest.

| Station ID | Shape $\hat{\xi}$ (s.e.) | Latitude | Longitude | Elevation | City name |
|---|---|---|---|---|---|
| (I) | −0.2316 (0.0604) | 40.45 | −111.70 | 1,720 | Timpanogos, UT |
| (II) | −0.2300 (0.0711) | 40.15 | −79.03 | 558 | Boswell, PA |
| (III) | −0.2248 (0.0703) | 42.02 | −86.25 | 265 | Eau Claire, MI |
| (IV) | 0.4383 (0.1065) | 29.15 | −95.45 | 8 | Angleton, TX |
| (V) | 0.4409 (0.0960) | 31.30 | −86.52 | 76 | Andalusia, AL |
| (VI) | 0.4977 (0.0970) | 47.55 | −116.17 | 680 | Kellogg, ID |

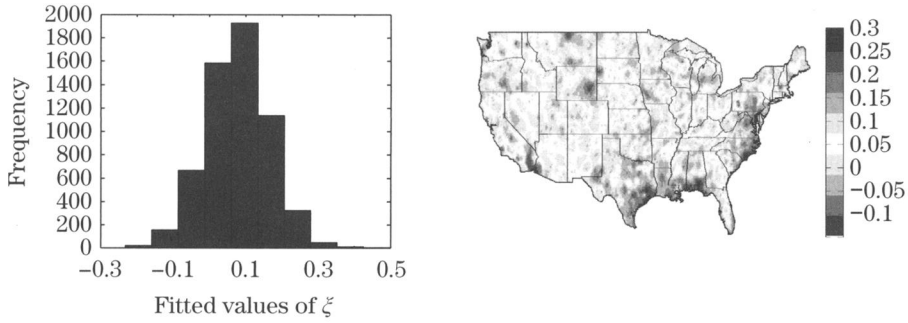



Figure 3. Fitted values of $\xi$ from time series at each of 5,873 stations. The left panel shows the distribution of $\hat{\xi}$. The right panel plots $\hat{\xi}$ to US map using the inverse distance method. Note that the values larger than 0.3 are truncated to 0.3 in order to display an overall visual impression.

are spread to different latitudes, though two of them are around latitude 30°N. The standard errors suggest that all shape parameters are significantly different from zero. We note that only two stations have the estimated $\hat{\xi}$ values less than $-1$ which may be a problem for asymptotic inference using a limiting normal distribution. A two-step estimation procedure may be helpful and lead to relevant asymptotics. In this study, we kept these two stations as they won't affect the main findings given that we are calculating thousands of TQCC test statistics.

In Figure 3, the right panel plots $\hat{\xi}$ to a US color map using the inverse distance method in a neighborhood of 10 points. The left panel plots the distribution of $\hat{\xi}$. One can see that precipitations over stations near the Gulf of Mexico region and stations near the Atlantic ocean and along North Carolina coast have heavier tails than precipitations over other stations.

These results suggest that precipitations are spatially nonstationary, clustered, and asymmetric over all stations. From the left panel of Figure 3, one can also see that the number of stations having positive values of $\hat{\xi}$ is more than the number of those having negative values. We note that similar findings regarding

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

precipitation data have been reported in the literature, and our viewing angle is an extreme value perspective.

### 4.2. Spatial tail dependence in precipitations across the continental US

We assumed marginal transformations being performed, and used observations running from January 1, 1970 to December 31, 1999 to avoid unequal sequence lengths at different stations. We first calculated all pairwise (17,243,128 pairs) TQCCs among all 5,873 stations. We included all days with at least one non-zero precipitation value from one of the paired stations within this time window. In Figure 4, the horizontal axis depicts the distance (in miles) between a chosen station and the corresponding station with the largest TQCC value within it, and the vertical axis depicts the distance between the chosen station and the corresponding station with the smallest TQCC value within it. The corresponding point displays these two types of distances. The scale associated with the horizontal axis is much smaller than that on the vertical axis. The left panel is drawn for all stations. The right panel filters stations to those with smallest TQCC value greater than 0.05. One can see that, in general, the closer the two stations, the larger the TQCC value. Again, this figure indicates, from the perspective of extreme weather condition, nonstationarity, spatial clusters, and tail dependence in the precipitations across the continental USA.

For each station among 5,873 stations, we performed pairwise (144,224) tail independence tests for stations within a distance of 84.7275 miles to the chosen station. The distance of 84.7275 miles was chosen to be the 95th percentile of distances between pairs with TQCC values greater than 0.05. As such, we only consider regional paired stations for each of 5,873 stations.

We used TQCC value as an estimate of the tail dependence index of the paired stations if the null hypothesis of tail independence was rejected by a TQCC-based test. For the transformed sample $\{(X_i, Y_i)\}$ in each paired stations, the threshold was chosen to be the smaller one of the two 95th percentiles of $\{X_i\}$ and $\{Y_i\}$. With the distance of 84.7275, some of TQCC values less than 0.05 still tested as significant. Using BH multiple testing to control the FDR at level 0.05, the total number of rejections of the null hypotheses of tail independence was 141,006 for testing the same day tail dependence.

### 4.3. Overall tail dependence across regional-wise stations

For comparison purpose, at each station we obtained significant values of

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

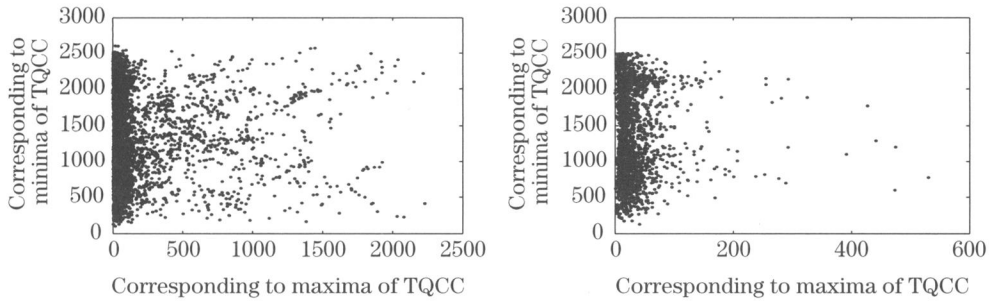



Figure 4. Scatterplots of two types of distances in miles. Left: calculated at all stations; right: calculated at those stations with the smallest TQCC value being greater than 0.05.

TQCC after controlling FDR at level 0.05, and took the maximum of them. We then plotted a total of 5,873 maximal values of TQCC at all stations in Figures 5, 6 and 7, corresponding to the same day (with 141,006 significant tests), 1-lagged day (today's data at station $s_i$ was combined with tomorrow's data at station $s_j$ into a pair of $(X_t, Y_{t+1})$, with 138,778 significant tests) and 7-lagged day (with 137,889 significant tests), respectively. The tail dependence map should be interpreted as the conditional probability of observing heavy precipitations at $s_i$ given heavy precipitations having been observed at $s_j$ within a time window and in the region where both stations are located. For different stations $s_{j'}$, heavy precipitations may or may not be observed within that time window.

From the right panels of Figures 5, 6 and 7, we can see that the maximal tail dependence decays as time lag increases, and that the dependence patterns change. These patterns can be seen in the color maps and color bars. The histograms in the left panels show that there is a decreasing trend, since the spread of the histograms shrinks. These empirical findings are useful in guiding climate model development and weather forecast in the perspective of extreme weather co-movements.

Table 2 reports 10 largest values of TQCC on the same day (as in Figure 5) and names of the corresponding paired stations. The paired stations are close to each other: the differences of latitudes are less than 0.2 degrees, and the same is true for longitudes. The TQCC in the second column of the table indicates that the probability of one location having large precipitation, conditional on another location concurring, is at least as high as 20%.

### 4.4. Illustrations of several individual stations

In this section, we choose two stations, Kellogg, Idaho and Wilmington,

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

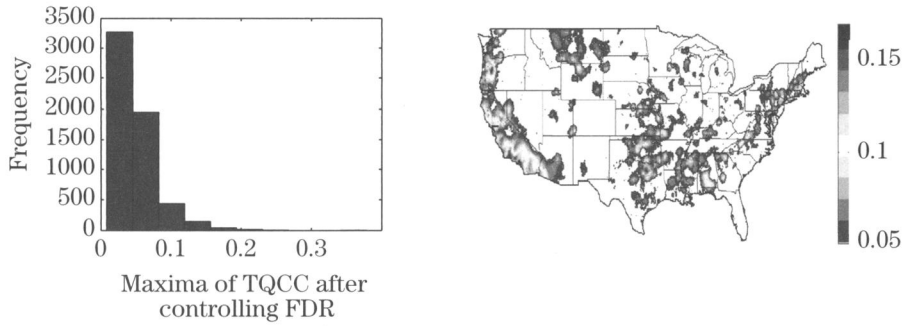



Figure 5. Maximal precipitation tail dependence between one station and the remaining stations on the same day. The colorbar on the right panel has been adjusted to reflect the left panel with TQCCs being larger than 0.05, but the very large values are truncated to the colorbar upper limit as they are just a few points.

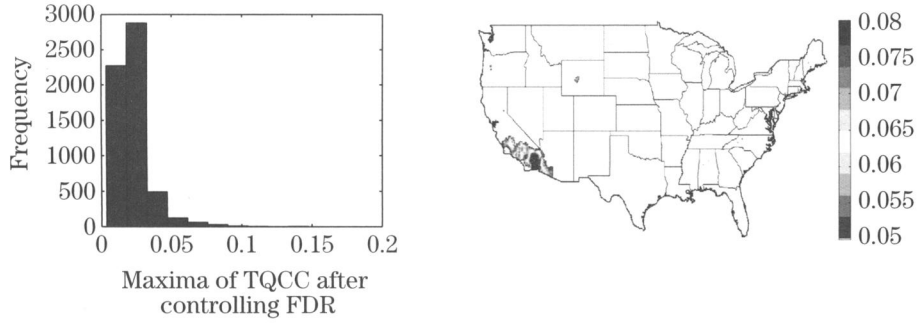



Figure 6. Maximal precipitation tail dependence between one station and the remaining stations on the lagged-1 day. The colorbar on the right panel has been adjusted to reflect the left panel with TQCCs being larger than 0.05, but the very large values are truncated to the colorbar upper limit as they are just a few points.

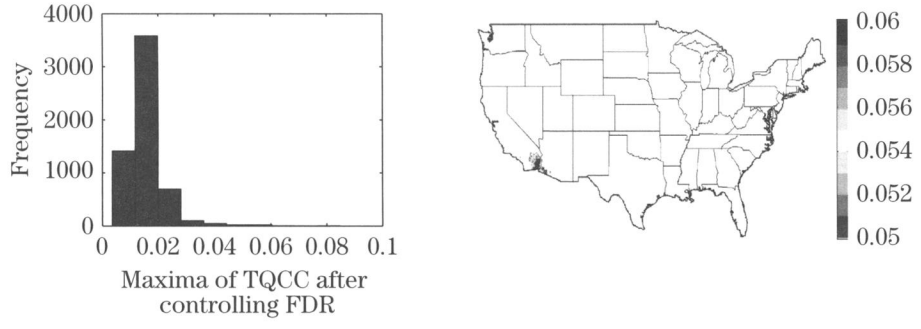



Figure 7. Maximal precipitation tail dependence between one station and the remaining stations on the lagged-7 day. The colorbar on the right panel has been adjusted to reflect the left panel with TQCCs being larger than 0.05, but the very large values are truncated to the colorbar upper limit as they are just a few points.

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

Table 2. The 10 largest values of TQCC and information of corresponding stations.

| Pair ID | TQCC | Latitude | Longitude | Elevation | City name |
|---|---|---|---|---|---|
| (I) | 0.3787 | 33.92 | −118.13 | 34 | Downey, CA |
| | | 33.97 | −118.02 | 128 | Whittier, CA |
| (II) | 0.2652 | 44.40 | −122.48 | 262 | Cascadia, OR |
| | | 44.10 | −122.68 | 206 | Leaburg, OR |
| (III) | 0.2324 | 34.48 | −119.50 | 633 | Juncal Dam, CA |
| | | 34.53 | −119.78 | 312 | Los Prietos Ranger, CA |
| (IV) | 0.2271 | 40.08 | −99.20 | 610 | Harlan County Lake, NE |
| | | 40.07 | −99.13 | 573 | Naponee, NE |
| (V) | 0.2208 | 39.35 | −123.12 | 309 | Potter Valley, CA |
| | | 39.13 | −123.20 | 193 | Ukiah, CA |
| (VI) | 0.2206 | 42.48 | −71.28 | 49 | Bedford, MA |
| | | 42.52 | −71.13 | 27 | Reading, MA |
| (VII) | 0.2200 | 34.52 | −119.68 | 473 | Gibraltar Dam, CA |
| | | 34.48 | −119.50 | 633 | Juncal Dam, CA |
| (VIII) | 0.2198 | 34.08 | −117.87 | 175 | Covina Nigg, CA |
| | | 33.97 | −118.02 | 128 | Whittier, CA |
| (IX) | 0.2128 | 29.95 | −90.13 | 6 | New Orleans, LA |
| | | 29.98 | −90.02 | 3 | New Orleans, LA |
| (X) | 0.2092 | 33.53 | −117.77 | 11 | Laguna Beach, CA |
| | | 33.73 | −117.87 | 41 | Santa Ana, CA |

Table 3. Fitted values of GEV parameters and standard errors.

| Station ID | Shape $\hat{\xi}$ (s.e.) | Scale $\hat{\psi}$ (s.e.) | Location $\hat{\mu}$ (s.e.) | City Name |
|---|---|---|---|---|
| (I) | 0.4977 (0.0970) | 5.0767 (0.1654) | 377.4 (24.75) | Kellogg, ID |
| (II) | 0.3110 (0.0820) | 5.8077 (0.1343) | 920.5 (49.12) | Southport, NC |

North Carolina, to illustrate the patterns of tail dependence. The former is also displayed in Table 1 with the largest $\hat{\xi}$, while the latter is located in one of the regions suffering from severe flooding in the last fifteen years.

Table 3 exhibits the estimated values $(\hat{\xi}, \hat{\psi}, \hat{\mu})$ in fitting the GEV distribution.

For each of the two chosen stations in Table 3, Table 4 illustrates the station name achieving the largest value of TQCC with that station. The tail dependence magnitudes (in the second column of Table 4) are significantly smaller than those presented in Table 2, but they are statistically significant. In an extreme value context, a tail dependence index as small as 2% may be already practically significant and an effective test statistic should be able to detect such a small tail dependence index given the relatively large sample size in our precipitation data. As an illustration, let us consider Pair ID (II) which has 2,416 concurrent

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

Table 4. The largest TQCC (p-values in parenthese) between each of the two chosen stations and the rest of stations.

| Pair ID | TQCC | Latitude | Longitude | Distance | City name |
|---|---|---|---|---|---|
| (I) | 0.0239 | 47.55 | −116.17 | 0.9132 | Kellogg, ID |
| | (0.0000) | 47.62 | −117.52 | | Spokane, WA |
| (II) | 0.0769 | 33.98 | −78.00 | 0.3464 | Southport, NC |
| | (0.0000) | 34.32 | −77.92 | | Wilmington, NC |

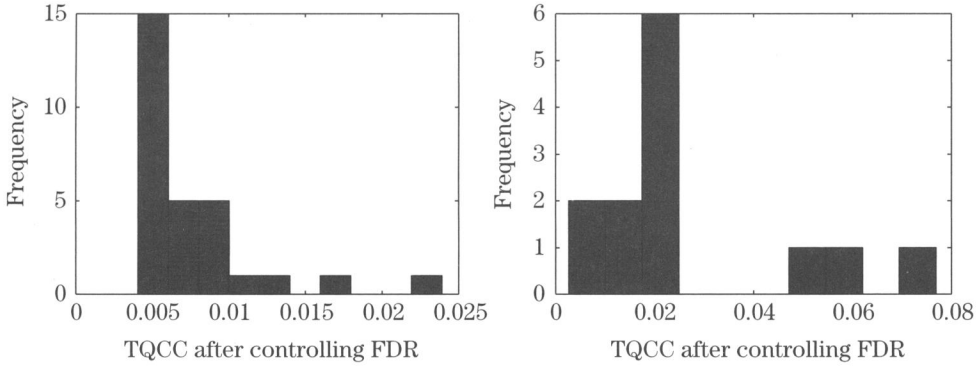



Figure 8. Left panel: Precipitation tail dependence between Kellogg, ID and the remaining stations on the same day. Precipitation tail dependence between South Port, NC and the remaining stations on the same day.

days of precipitations and 4,377 precipitation days at either station, respectively. There are occasions that the two stations have strong precipitation concurrently, flooding does occur in one place given that another place has a strong precipitation, even though the probability is low generally. Here a strong precipitation should be understood on a relative scale from station to station. These empirical findings from historical records indicate that weather forecasting models need to describe such tail dependence more accurately in extreme weather conditions.

Figure 8 displays distributions of TQCC between the selected station and the remaining ones. Each of these plots itself can be viewed as a skewed and long tailed distribution. This phenomenon suggests that flooding can be anywhere which shares smaller tail dependence with other locations. These plots show stronger tail dependence existing near the chosen stations. For instance, South Port station and its neighboring regions once suffered from severe flooding. This phenomenon, under extreme weather conditions, suggests potential flooding in certain regions that show similar patterns of tail dependence, hence analysis results may be helpful in regional flooding prevention and strategic policy making.

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

## 5. Conclusions

In this study, generalizing a recently introduced tail dependence measure, tail quotient correlation coefficient (Zhang (2008b)), we evaluated spatial tail dependence of precipitations recorded at selected stations from NCDC rain gauge data, including daily precipitation from 5,873 stations during 1950–1999. Much work has been based on Ledford and Tawn (1996) tail dependence measure and Coles (2001) $\bar{\chi} = 2\eta - 1$. There are various ways of estimating $\bar{\chi}$ and $\eta$, and some reported results based on different estimators do not agree with each other. The TQCC is directly related to tail values, is intuitively appealing, and enjoys a nice geometric property. In our Figures 5–8 in Section 4, TQCC demonstrates its unique feature of being a meaningful measure of tail dependence: skewed to the right and with heavy tailed pattern, which may not be seen with other dependence measures. The TQCC has a relatively high empirical power in detecting tail dependence. In this paper, we further show that the test is consistent under $H_1$ of (2.2). In addition, one can easily calculate all pairwise TQCC among all 5,873 stations. We have conducted a large number of pairwise tests for tail independence with BH multiple testing correction.

Our results, from the perspective of tail dependence, reveal nonstationarity, asymmetry, spatial clusters, and tail dependence in the data. Although various studies have documented that there are tail dependencies in precipitation data, the results obtained could be very useful in enhancing our understanding of weather extremes and in guiding new climate model development. Our TQCC-based analysis is for regionally paired stations only. A seasonal-based analysis can be done for temperature dependence patterns, snowfall dependence patterns, and smog spread patterns. We intend to study these applications.

The TQCC uses the largest values of two sequences of quotients. It is possible to extend the definition of TQCC using either the first $k$ largest quotients (e.g. the geometric mean or other forms), or the $100p$th percentiles of the two quotient sequences. Such newly extended definitions are useful in measuring nonlinear dependence between observations. Detailed investigation of this topic will be given in a future study.

A final note is that the asymptotics of TQCC, under the alternative hypothesis of tail dependent, depends on the distributions of $X_i/Y_i$ and $Y_i/X_i$, precluding any unified form of limit distributions. A project here is to provide a dictionary of limit distributions of TQCCs under different joint dependencies. In practice, a bootstrap approach can be applied to construct a bootstrap confidence interval.

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

## Supplementary Materials

Supplementary materials include simulation examples and all technical derivations. These materials are presented in Sections S1 and S2 respectively, and are available online.

## Acknowledgment

The authors thank editors and reviewers for their valuable comments. The work by Zhengjun Zhang was partially supported by NSF-DMS-1505367 and NSF-CMMI-1536978. The work by Chunming Zhang was partially supported by NSF-DMS–1308872 and DMS–1521761, and Wisconsin Alumni Research Foundation.

# References

Bacro, J. N., Bel, L. and Lantuejoul, Ch. (2010). Testing the independence of maxima: from bivariate vectors to spatial extreme fields. *Extremes* **13**, 155–175.

Bel, L., Bacro, J. N. and Lantuejoul, Ch. (2007). Assessing extremal dependence of environmental spatial fields. *Environmetrics* **19**, 163–182.

Beniston M., Stephenson, D. B., Christensen, O. B., Ferro, C. A. T., Frei, C., Goyette, S., Halsnaes, K., Holt, T., Jylhä, K., Koffi, B., Palutikof, J., Schöll, R., Semmler, T. and Woth, K. (2007). Future extreme events in European climate: an exploration of regional climate model projections. *Climatic Change*, DOI 10.1007/s10584-006-9226-z.

Benjamini Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. Roy Stat Soc B* **57**, 289–300.

Benjamini Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics* **29**, 1165–1188.

Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values.* Spinger-Verlag London.

Coles, S., Heffernan, J. and Tawn, J. (1999). Dependence measures for extreme value analyses. *Extremes* **2**, 339–365.

Cooley, D., Nychka, D. and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association* **102**, 824–840.

de Haan, L. and Resnick, S. I. (1977). Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie und verw. Gebiete* **40**, 317–337.

Draisma, G., Drees, H., Ferreira, A. and de Haan, L. (2004). Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli* **10**, 251–280.

Deheuvels, P. (1983). Point processes and multivariate extreme values. *Jour. of Multi. Anal.* **13**, 257–272.

Easterling, D. R., Anderson, D. M., Cohen, S. J., Gutowski, W. J. Jr., Holland, G. J., Kunkel, K. E., Peterson, T. C., Pulwarty, R. S., Stouffer, R. J. and Wehner, M. F. (2008). Measures to improve our understanding of weather and climate extremes. In *Weather and Climate Extremes in a Changing Climate. Regions of Focus: North America, Hawaii, Caribbean,*

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

*and U.S. Pacific Islands.* (Karl, T. R., Meehl, G. A., Miller, C. D., Hassol, S. J., Waple, A. M. and Murray, W. L. (eds.)) National Climatic Data Center, Washington, D.C., USA.

Elek, P. and Márkus, L. (2007). A light-tailed conditionally heteroscedastic model with applications to river flows. *Journal of Time Series Analysis* **29**, 14–36.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance.* Springer.

Embrechts, P., McNeil, A. and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In *Risk Management: Value at Risk and Beyond,* (ed. Dempster, M. A. H.), 176–223. Cambridge University Press, Cambridge.

Falk, M. and Michel, R. (2006). Testing for tail independence in multivariate extreme value models. *Annals. Inst. Statis. Math.* **58**, 261–290.

Ferreria, A. and de Haan, L. (2004). On the consitency of the estimation of the probability of a failure set. The 3rd International Syposium on Extreme Value Ananlysis: Theory and Practice. Aveiro, Portugal.

Gilleland, E., Brown, B. G. and Ammann, C. M. (2013). Spatial extreme value analysis to project extremes of large-scale indicators for severe weather. `http://www.rap.ucar.edu/staff/ericg/GillelandEtAlV2ca.pdf`.

Groisman, P. Y., Karl, T. R., Easterling, D. R., Knight, R. W., Jamason, P. F., Hennessy, K. J., Suppiah, R., Page, C. M., Wibig, J., Fortuniak, K., Razuvaev, V. N., Douglas, A., FØrland, E. and Zhai, P. (1999). Changes in the probability of heavy precipitation: Important indicators of climatic change. *Climatic Change* **42**, 243–283.

Heffernan, J. E., Tawn, J. A. and Zhang, Z. (2007). Asymptotically (in)dependent multivariate maxima of moving maxima processes. *Extremes* **10**, 57–82.

Hegerl, G. C., Karl, T. R., Allen, M., Bindoff, N. L., Gillett, N., Karoly, D., Zhang, X. and Zwiers, F. (2006). Climate change detection and attribution: Beyond mean temperature signals. *Journal of Climate* **19**, 5058–5077.

Hüsler, J. and Li, D. (2009). Testing asymptotic independence in bivariate extremes. *Jour. of Statist. Planning & Inference* **139**, 990–998.

Karl, T. R. and Knight, R. W. (1998). Secular trends of precipitation amount, frequency, and intensity in the United States. *Bulletin of the American Meteorological Society* **79**, 231–241.

Kunkel, K. E., Palecki, M. A., Hubbard, K. G., Robinson, D. A., Redmond, K. T. and Easterling, D. R. (2007). Trend identification in twentieth-century U.S. snowfall: The challenges. *Journal of Atmospheric and Oceanic Technology* **24**, 64–73.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* **55**, 169–187.

Ledford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society B* **59**, 475–499.

Naveau, P., Nogaj, M., Ammann, C., Yiou, P., Cooley, D. and Jomelli, V. (2005). Statistical methods for the analysis of climate extremes. *Comptes Rendus Geoscience* **337**, 1013–1022.

Naveau, P., Zhang, Z. and Zhu, B. (2011). An extension of max autoregressive models. *Statistics and Its Interface* **4**, 253–266.

Peng, L. (1999). Estimation of the coefficient of tail depdendence in bivariate extremes. *Statist. Probab. Letters* **43**, 399–409.

Pickands, J. III (1975). Statistical inference using extreme order statistics. *Ann. of Statist.* **3**, 119–131.

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms

Salmon, F. (2012). The formula that killed Wall Street. *Significance* **9**, 16–20.

Schlather, M. (2001). Examples for the coefficient of tail dependence and the domain of attraction of a bivariate extreme value distribution. *Statistics & Probability Letters* **53**, 325–329.

Shamseldin E. C., Smith, R. L., Sain, S. R., Mearns, L. and Cooley, D. (2008). Downscaling extremes: A comparison of extreme value distributions in point-source and gridded precipitation data. SAMSI Risk Program Presentation.

Sibuya, M. (1960). Bivariate extreme statistics, I. *Ann. Inst. Statist. Math.* **11**, 195–210.

Smith, E. L. and Stephenson, A. G. (2009). An extended Gaussian max-stable process model for spatial extremes. *Journal of Statistical Planning and Inference* **139**, 1266–1275.

Smith, R. L. (1989). Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. (with discussion) *Statistical Science* **4**, 367–393.

Smith, R. L. (2003). Statistics of extremes, with applications in the environment, insurance and finance. In *Extreme Value in Finance, Telecommunications and the Environment.* (Finkenstadt, B. and Rootzén, H. Editors). Chapman and Hall/CRC.

Smith, R. L. and Weissman, I. (1996). Characterization and estimation of the multivariate extremal index. Unpublished manuscript, University of North Carolina. `http://www.stat.unc.edu/postscript/rs/extremal.pdf`.

Smith, R. L., Grady, A. M. and Hegerl, G. C. (2007). Extreme precipitation trends over the continental United States. 15th Aha Hulikoa Hawaiian Winter Workshop, Honolulu.

Wang, Y. (2012). On the regular variation of ratios of jointly Frechet random variables. *Extremes* **15**, 175–196.

Yiou, P., Goubanova, K., Li, Z. X. and Nogaj, M. (2008). Weather regime dependence of extreme value statistics for summer temperature and precipitation. *Nonlin. Processes Geophys.* **15**, 365–378.

Zhang, Z. (2008a). The estimation of M4 processes with geometric moving patterns. *Ann. Instit. Stat. Math.* **60**, 121–150.

Zhang, Z. (2008b). Quotient correlation: A sample-based alternative to Pearson's correlation. *Annals of Statistics* **36**, 1007–1030.

Zhang, Z. (2009). On approximating max-stable processes and constructing extremal copula functions. *Statistical Inference for Stochastic Processes* **12**, 89–114.

Zhang, Z., Qi, Y. and Ma, X. W. (2011), Asymptotic independence of correlation coefficients with application to testing hypothesis of independence. *Electronic Journal of Statistics* **5**, 342–372.

Zhang, Z. and Smith, R. L. (2004). The behavior of multivariate maxima of moving maxima processes. *Journal of Applied Probability* **41**, 1113–1123.

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

E-mail: zjz@stat.wisc.edu

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

E-mail: cmzhang@stat.wisc.edu

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

E-mail: treasa916@gmail.com

(Received December 2015; accepted March 2016)

This content downloaded from
144.92.38.237 on Thu, 13 Aug 2026 23:54:54 UTC
All use subject to https://about.jstor.org/terms