Statistica Sinica **31** (2021), 1619-1641
doi:https://doi.org/10.5705/ss.202018.0420

A BAYESIAN SEMI-PARAMETRIC MIXTURE MODEL FOR BIVARIATE EXTREME VALUE ANALYSIS WITH APPLICATION TO PRECIPITATION FORECASTING

Yuan Tian and Brian J. Reich

North Carolina State University

Abstract: We propose a novel mixture generalized Pareto model for calibrating extreme precipitation forecasts. This model is able to describe the marginal distribution of observed precipitation and capture the dependence between climate forecasts and observed precipitation under suitable conditions. In addition, the full range distribution of precipitation can be estimated, conditional on grid forecast ensembles. Unlike the classical generalized Pareto distribution that can only model points over a hard threshold, our model takes the threshold as a latent parameter. We study the tail behavior of both univariate and bivariate models. The utility of our model is evaluated using a Monte Carlo simulation study. Lastly, we apply the model to US precipitation data, showing that it outperforms competing methods.

Key words and phrases: Bivariate extreme value model, generalized Pareto distribution, hierarchical model, tail dependence.

1. Introduction

Heavy precipitation events are frequent and widespread weather hazards. Precise forecasts of extreme precipitation are important to mitigating the risk of these dangerous events. Numerical weather predictions (NWP) produced by weather centers across the world enable us to model and predict precipitation extremes. However, NWP can be biased; thus, they require statistical calibration to account for the bias and to quantify the forecast uncertainty (Wilks (2011)).

Extreme value theory (EVT) provides elegant tools to model rare events. According to EVT, the conditional distribution of the independent and identically distributed (i.i.d.) observations that exceed a sufficiently high threshold can be well approximated by the generalized Pareto distribution (GPD) under some mild regularity conditions (Pickands III and others (1975); de Haan and Ferreira (2006)). However, the selection of an optimal threshold is a difficult task, in practice (see Caeiro and Gomes (2015) and the references therein for an overview).

Corresponding author: Brian J. Reich, Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA. E-mail: bjreich@ncsu.edu.

In addition, the GPD is only an approximate model for the tail, while the rest of the data (lower than the threshold) are typically discarded. To fully use the data, mixture models have been proposed to fit the tail and the bulk separately. The tail is usually assumed to follow the GPD, whereas the bulk component can be modeled using parametric (Behrens, Lopes and Gamerman (2004); Cabras and Castellanos (2011); MacDonald (2011); Hu (2013); Zheng, Ismail and Meng (2014)), semiparametric, (Cabras and Castellanos (2011); do Nascimento, Gamerman and Lopes (2012); Lee, Li and Wong (2012)) or nonparametric (Tancredi, Anderson and O'Hagan (2006); MacDonald (2011)) methods. The mixture distribution model with Pareto tails is shown to be useful in practice (Bentzien and Friederichs (2012)). However, these methods require specifying a hard threshold that separates the bulk and the tail. In addition, the probability density function of the mixture model is not necessarily continuous (smooth) at the threshold (though the cumulative distribution is continuous), which may lead to biased estimators (Scarrott and MacDonald (2012)). Extensions of the extreme value mixture models that exhibit a smoother transition from the bulk to the tail include the dynamically weighted mixture model (Frigessi, Haug and Rue (2002); Vrac and Naveau (2007)) and the interval transition mixture model (Holden and Haug (2009)). As an alternative to the mixture models, Naveau et al. (2016) propose an extended GPD (EGPD) model that can be applied to heavy, moderate, and low precipitation amounts, while avoiding the potential difficulties associated with threshold selection. Rather than considering all points over the threshold, the data are converted to block maxima and fit using a generalized extreme value (GEV) distribution (Scheuerer (2014)). In addition, Bjørnar Bremnes (2004) proposes a quantile regression approach, showing that it is useful when the training sample size is relatively large.

In this study, we propose a finite mixture generalized Pareto (MIXGP) model. The proposed MIXGP is an extension of the classical GPD. Here, we treat the threshold as a random variable and use a semiparametric Bayesian prior for its mixing distribution. By marginalizing over the threshold, we have a full range model that applies to both the bulk and the tail. The idea of using a random threshold that varies among observations was first proposed in Pigeon and Denuit (2011). However, their model is built upon the composite lognormal-Pareto model (Cooray and Ananda (2005); Scollnik (2007)), a type of extreme value mixture model, on which discussions of its tail behavior are limited.

Rather than specifying a single parametric model, as in Pigeon and Denuit (2011), the proposed MIXGP allows for more flexibility. We discuss the tail behavior of our proposed model in detail, providing an empirically based guideline for choosing suitable mixing distribution functions. With a normal mixing distribution, our proposed model has nice tail properties and its estimating procedure can be easily implemented.

Motivated by a joint analysis of extreme precipitation data and forecasts, we extend the proposed MIXGP to a bivariate model, which is necessary when modeling the dependence between extremes from different sources. We use a bivariate semi-parametric mixing distribution, and show that our proposed bivariate MIXGP is able to capture the tail dependence under mild conditions.

The conditional distribution of precipitation given a forecast is a byproduct of our bivariate model, and serves as a calibrated forecast of extreme precipitation. In the context of quantile regression, Taddy and Kottas (2010) develop a Dirichlet process mixture (DPM) model for the bivariate distribution of the response and the covariate in order to indirectly estimate the conditional quantiles, given the covariate. Bivariate DPM models are easy to fit, span the entire class on continuous joint density functions, and provide a consistent nonparametric estimate under general regularity conditions (Ghosal and Van der Vaart (2017)). Therefore, this indirect approach provides a convenient and flexible way to estimate a conditional distribution. In our application to extreme value modeling, the indirect approach allows for an arbitrarily flexible model for the effect of the covariate on the response, Furthermore, including different effects on the bulk and the tail of the distribution because the bivariate model allows for asymptotic dependence, the effect of the covariate on the conditional distribution need not dissipate for extreme quantiles.

The rest of the paper is organized as follows. Section 2 describes the MIXGP model and its tail behavior. Specifically, the tail index of the univariate model and the tail dependence of the bivariate model are discussed in depth. The method is evaluated using a simulation study in Section 3. We apply both the univariate and the bivariate MIXGP models to analyze observed and forecast daily precipitation data in Section 4. The proposed MIXGP is shown to give well-calibrated forecasts, with greater precision than competing methods. Finally, we conclude the work with a discussion section.

2. The MIXGP Model

In this section, we introduce our proposed MIXGP model and study its tail properties. EVT dictates that the distribution of Y conditional on $Y > y + \mu$ converges to the generalized Pareto distribution (GPD) as $\mu \to +\infty$, under some mild regularity conditions (Pickands III and others (1975); de Haan and Ferreira

(2006)), justifying its use as a model for threshold exceedances. The cumulative distribution function (CDF) of the GPD is

$$H(y) = \left(1 + \xi \frac{y-\mu}{\sigma}\right)_+^{-1/\xi}, \; y > \mu, \tag{2.1}$$

where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter, $\xi \in \mathbb{R}$ is the shape parameter, and $y_+ = \max(y, 0)$. We denote the density function of the GPD(μ, σ, ξ) as $h(\cdot|\mu, \sigma, \xi)$. When $\xi \geq 0$, the support is (μ, ∞); otherwise, the support is $(\mu, \mu - \sigma/\xi)$. Therefore, the shape parameter ξ determines the tail behavior. A challenge when applying the GPD is to choose an optimal threshold μ. Instead of specifying a deterministic μ, we extend the model by treating μ as a latent variable.

2.1. Univariate model

In this section, we present a univariate model for the marginal distribution of Y. To capture the tail behavior more flexibly, we specify a MIXGP model with latent location variable

$$f(y|\Theta, \mathbf{p}) = \sum_{k=1}^{K} p_k \int h(y|\mu, \sigma_k, \xi_k) dG_k(\mu), \tag{2.2}$$

where $\Theta = \{\sigma_1, \ldots, \sigma_K, \xi_1, \ldots, \xi_K\}$ denotes the collection of scale and shape parameters, and $\mathbf{p} = \{p_1, \ldots, p_K\}$ denotes the mixture probabilities. The model can be written hierarchically as

$$\begin{aligned} Y_i|\mu^{(i)}, \Theta, Z_i = k &\sim \text{GPD}(\mu^{(i)}, \sigma_k, \xi_k) \\ \mu^{(i)}|Z_i = k &\sim G_k(\cdot) \\ Z_i|\mathbf{p} &\sim \text{Cat}(\mathbf{p}), \end{aligned}$$

where Cat($\mathbf{p}$) denotes the categorical distribution, with Prob$(Z_i = k) = p_k$, for $k = 1, \ldots, K$, $\Theta = \{\sigma_1, \ldots, \sigma_K, \xi_1, \ldots, \xi_K\}$ denotes the collection of scale and shape parameters of the GPD, $\mathbf{p} = \{p_1, \ldots, p_K\}$ denotes the mixture probabilities, $Z_i \in \{1, \ldots, K\}$ is the cluster label, and the mixing distribution functions $\{G_k(\mu) : k = 1, \ldots, K\}$ are CDFs of the latent location variables. The choice of $G_k(\cdot)$'s is discussed later. Because $K \to \infty$, the resulting model is sufficiently flexible to approximate any true density function (Ghosal and Van der Vaart (2017)).

Extreme value analysis focuses on studying the upper-tail behavior of a random variable. A commonly used measure to characterize the tail is the right-tail index.

Definition 1 (Right-tail index)**.** For a CDF $F(y)$ with density $f(y)$ defined on $\mathbb{R}$, the right-tail index is given by

$$\alpha_+(F) = \liminf_{y \to +\infty} \frac{-\log\{1 - F(y)\}}{\log y}.$$

If $\alpha_+(F) = \infty$, the corresponding distribution function is thin-tailed (e.g., exponential or normal distribution). The tail index of the Student t-distribution is strictly positive and finite. As a result, it has a heavier tail than those of the exponential and the normal distributions. If $\alpha_+(F) = 0$, the distribution has a heavy tail (e.g., log-Pareto distribution) (Li, Lin and Dunson (2019)).

Below, we calculate the tail index of our proposed MIXGP model. The proof of Theorem 1 is given in the online Supplementary Material.

Theorem 1 (Tail index)**.** *Let a_k be the tail index of the mixture density $\int h(y|\mu, \sigma_k, \xi_k) dG_k(\mu)$. Then, we have*

$$a_k = \begin{cases} \alpha_+(G_k), & \xi_k \le 0, \\ \min\{\alpha_+(G_k), \xi_k^{-1}\}, & \xi_k > 0. \end{cases}$$

and the tail index of the model (2.2) *is $\alpha_+(F) = \min(a_1, \ldots, a_K)$.*

When $\xi_k < 0$, the GPD is short-tailed and a_k is completely determined by $G_k(\cdot)$. For example, if we choose G_k to be a normal distribution, then the kth component is a thin-tailed distribution and $a_k = \infty$. If we choose G_k to be a log-Pareto, then the kth component is a super heavy-tailed distribution and $a_k = 0$. When $\xi_k \ge 0$, the resulting GPD's domain is unbounded. Then, a_k is determined jointly by the tail behavior of the mixing distribution function $G_k(\cdot)$ and ξ_k. For example, if we set G_k to be a normal distribution and $\xi_k > 0$, then $a_k = \xi_k^{-1}$. If we set G_k to be a t-distribution and $\xi_k = 0$, then $a_k = \alpha_+(G_k)$. In addition, it can be seen from Theorem 1 that the tail index of the mixture distribution $\alpha_+(F)$ is determined by the mixture component that has the heaviest tail, and does not depend on the mixture probabilities $\mathbf{p}$.

As an example, consider $K = 1$, $\mu \sim U[0, b]$. Then, $G(\mu|b) = 0$ if $\mu < 0$, and $G(\mu|b) = 1$ if $\mu > b$; otherwise, if $0 \le \mu \le b$, $G(\mu|b) = \mu/b$. Therefore,

$$\bar{F}(y|\sigma,\xi,b)=\begin{cases}1+\frac{\sigma}{b(\xi-1)}\left\{1+\xi\frac{y-\min(b,y)}{\sigma}\right\}^{-1/\xi+1}I\left(y<b-\frac{\sigma}{\xi}\right)-\frac{\min(b,y)}{b}, & \xi<0,\\ 1+\frac{\sigma}{b}\left[\exp\left\{-\frac{y-\min(b,y)}{\sigma}\right\}-\exp\left\{-\frac{y}{\sigma}\right\}\right]-\frac{\min(b,y)}{b}, & \xi=0,\\ 1+\frac{\sigma}{b(\xi-1)}\left[\left\{1+\xi\frac{y-\min(b,y)}{\sigma}\right\}^{-1/\xi+1}-\left\{1+\xi\frac{y}{\sigma}\right\}^{-1/\xi+1}\right]-\frac{\min(b,y)}{b}, & \xi>0.\end{cases}$$

As a result, Y has a polynomial tail when $\xi > 0$, and an exponential decaying tail when $\xi = 0$. If $\xi < 0$, Y is short-tailed and bounded.

2.2. Bivariate model

In the bivariate case, we model the joint distribution of (X, Y) as a finite mixture of bivariate GPDs (BGPDs) with random location parameters. The BGPD CDF proposed in Smith (1994) is given by

$$H(x,y|\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\xi},\gamma) = \exp\left[-V\{-\log H_X(x|\mu_1,\sigma_1,\xi_1), -\log H_Y(y|\mu_2,\sigma_2,\xi_2);\gamma\}\right], \tag{2.3}$$

where $H(x,y|\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\xi},\gamma) = P(X \le x, Y \le y|\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\xi},\gamma)$, $\boldsymbol{\mu} = (\mu_1,\mu_2), \boldsymbol{\sigma} = (\sigma_1,\sigma_2)$, and $\boldsymbol{\xi} = (\xi_1,\xi_2)$. The structure function $V(s,t;\gamma)$ can take a variety of forms in general BGPDs. Here, we focus on the logistic dependence structure, where $V(s,t;\gamma) = (s^{1/\gamma} + t^{1/\gamma})^\gamma$ (Smith (1994)). The marginal distribution functions $H_X(x|\mu_1,\sigma_1,\xi_1) = H(x,\infty|\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\xi},\gamma)$ and $H_Y(y|\mu_2,\sigma_2,\xi_2) = H(\infty,y|\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\xi},\gamma)$ are the CDFs of the univariate GPD(μ_1,σ_1,ξ_1) and GPD(μ_2, σ_2,ξ_2), respectively. The parameter $\gamma \in [0,1]$ is referred to as the dependence parameter. Two extremal cases, where $\gamma = 0$ and $\gamma = 1$, correspond to the settings where X and Y are completely dependent and independent, respectively.

We consider the following bivariate density function of (X, Y):

$$f(x,y|\boldsymbol{\Theta},\mathbf{p},\boldsymbol{\gamma}) = \sum_{k=1}^{K} p_k \int h(x,y|\boldsymbol{\mu},\boldsymbol{\sigma}_k,\boldsymbol{\xi}_k;\gamma_k)dG_k(\boldsymbol{\mu}), \tag{2.4}$$

where $\boldsymbol{\Theta} = \{\boldsymbol{\sigma}_1,\ldots,\boldsymbol{\sigma}_K,\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_K\}$ denotes the collection of scale and shape parameters; $\mathbf{p} = \{p_1,\ldots,p_K\}$ are the mixture probabilities; $\boldsymbol{\sigma}_k = (\sigma_{k,1},\sigma_{k,2}), \boldsymbol{\xi}_k = (\xi_{k,1},\xi_{k,2})$, and $\boldsymbol{\mu} = (\mu_1,\mu_2)$; $\boldsymbol{\gamma} = \{\gamma_1,\ldots,\gamma_K\}$ is the set of dependence parameters; and $h(\cdot,\cdot|\boldsymbol{\mu},\boldsymbol{\sigma}_k,\boldsymbol{\xi}_k;\gamma_k)$ denotes the BGPD$(\boldsymbol{\mu},\boldsymbol{\sigma}_k,\boldsymbol{\xi}_k;\gamma_k)$ density function. To reduce the computational complexity in the model fitting procedure, we take $G_k(\mu_1,\mu_2)$ to be the bivariate normal distribution with cluster-specific mean vectors $\boldsymbol{u}_k = (u_{k,1},u_{k,2})$ and shared covariance matrix Σ. The marginal distributions $f(x|\Theta,\mathbf{p})$ and $f(y|\Theta,\mathbf{p})$ reduce to the univariate MIXGP model with normal mixing distributions. The dependence within the joint tail regions is modeled by the dependence parameter $\boldsymbol{\gamma}$.

Definition 2 (Tail dependence)**.** Suppose F_X and F_Y are the marginal distribution functions of X and Y, respectively. Then, the upper-tail dependence coefficient is defined as

$$\chi = \lim_{t \to 1} \Pr\{F_Y(Y) > t | F_X(X) > t\},$$

or equivalently,

$$\chi = \lim_{t \to 1} \Pr\{Y > F_Y^{-1}(t) | X > F_X^{-1}(t)\}.$$

In the following, we show that the upper tail dependence of X and Y can be determined jointly by the tails of G_k and $\boldsymbol{\xi}_k$ and the dependence parameter γ_k. Under certain conditions, our model is able to capture the tail dependence between Y and X.

Theorem 2 (Tail dependency)**.** *Assume $G_k(\mu_1, \mu_2)$ is a bivariate normal distribution with mean vector $\mathbf{u}_k$ and covariance matrix Σ. Suppose $\exists\ o \in \{1, \ldots, K\}$, s.t. $\gamma_o \in [0, 1), \xi_{o,1} = \max_{k=1,\ldots,K}\{\xi_{k,1}\}$, and $\xi_{o,2} = \max_{k=1,\ldots,K}\{\xi_{k,2}\}$, and $\xi_{o,1} > 0, \xi_{o,2} > 0$. Furthermore, p_o is the proportion of the o-th cluster. Then, we have $\chi > \chi_L > 0$, where $\chi_L = 2p_o / \sum_k \omega_{k,1}(2\xi_{o,1}/\sigma_{o,1})^{\xi_{o,1}^{-1}}(1-2^{\gamma_o - 1})$. Here, ξ_o, σ_o, and γ_o are the* BGPD *parameters, and the definition of ω_k, for $k = 1, \ldots, K$, can be found in the Supplementary Material.*

Under the conditions of Theorem 2, the marginal distributions of X and Y are univariate MIXGP models with normal mixing distributions. The marginal distributions of the oth mixture component have the heaviest upper tail. Specifically, the upper-tail indices are $\xi_{o,1}^{-1}$ and $\xi_{o,2}^{-1}$, respectively. The oth dependence parameter $\gamma_o \in [0, 1)$ indicates that X and Y are dependent in the oth component. Consequently, X and Y are also asymptotically dependent. In other words, the bivariate MIXGP model is able to capture this tail dependence if the oth component is tail dependent. Theorem 2 states that the tail dependence of the bivariate MIXGP model is also determined by the mixture component that has the heaviest-tailed marginal distribution.

We use a Metropolis-within-Gibbs Markov Chain Monte Carlo (MCMC) to implement the model. The prior distributions and details of the MCMC sampler used for the estimation are given in the Supplementary Material. For the simulation study in Section 3, we generate 10,000 MCMC iterations and discard the first 2,500 as burn-in. For the data analysis in Section 4, we generate 20,000 MCMC iterations and discard the first 5,000 as burn-in. Convergence is monitored using trace plots of the representative parameters. Determining the choice of K in a finite mixture model is an interesting research question. In general, a model with a large K will gain flexibility at the expense of computational burden. The number of clusters K can be viewed as an upper bound for the number of active mixture components, because some clusters may have a small probability (Gelman et al. (2013)). In our applications, we feel $K = 5$ is sufficiently large to make the model flexible. In both the simulation study and the real-data analysis, we find that some clusters often have low probability, which implies $K = 5$ is a proper upper bound for the number of mixture components. Therefore, we take $K = 5$ throughout this paper.

3. Simulation Study

In this section, we evaluate the performance of the bivariate MIXGP and compare it with that of the gamma distribution model ("GAM"), a mixture model with a gamma bulk and a GPD tail ("MIX-t"), the GEV model ("GEV"), and two EGPD (Naveau et al. (2016)) models, namely, the generalized linear extended Pareto distribution (GEPD) and the clustered extended GPD (CGPD). We use the maximum likelihood to fit the GAM model, and follow Bentzien and Friederichs (2012) and Scheuerer (2014) to fit the MIX-t and the GEV model, respectively. Because the original EGPD proposed in Naveau et al. (2016) is a univariate model with density EGPD(σ, ξ, δ), we modify it to be suitable for the bivariate case by transforming the scale parameter as $\log(\sigma) = \beta_0+\beta_1 X$ and using maximum likelihood estimators for $\beta_0, \beta_1, \xi, \delta$. This transformation is commonly used to incorporate covariates parametrically when fitting extreme value models (Wang and Li (2015)). For the CGPD, we first use k-means to partition the observed covariate X into K clusters, and then fit the univariate extended GPD model within each cluster using the maximum likelihood. We compare $K = 3$ and $K = 5$ clusters.

The potential advantage of the MIXGP is that it weights the location parameter in a flexible way. Therefore, it is possible to capture the tail behavior of any bivariate distribution. In this simulation study, we consider distributions with bounded and unbounded domains. For the bounded cases, we generate data from bivariate beta distributions, such that $0 \leq x \leq 1, 0 \leq y \leq 1$ (Gupta and Wong (1985); Nadarajah and Kotz (2005); Olkin and Trikalinos (2015)). For unbounded cases, we generate data from a left-truncated bivariate t distribution, such that $0 \leq x \leq \infty, 0 \leq y \leq \infty$ (Ho et al. (2012)). Because our primary interest is in studying the tail behavior, we estimate the extremal proportions $\hat{P}(Y \leq Y_{t_y}^C | X_{t_x})$ for covariate quantile levels $t_x = 0.9$ and conditional response quantile levels $t_y = 0.95, 0.99, 0.999$, where

$$P(Y \le Y^C_{t_y}|X) = t_y \quad \text{and} \quad P(X \le X_{t_x}) = t_x.$$

If the model fits the data well, then $\hat{P}(Y \le Y^C_{t_y}|X_{t_x})$ should be approximately t_y. The bias and MSE of $\hat{P}(Y \le Y^C_{t_y}|X_{t_x})$ are presented. Additional results can be found in the Supplementary Material. Under each setting, $n = 1{,}000$ samples are generated and the results are averaged over 100 replicates.

3.1. Bounded cases

For the bivariate beta simulation, we follow Olkin and Trikalinos (2015) to generate the data. We have the following four settings:

S1: Bivariate Beta (2,2,2,0.5)

S2: Bivariate Beta (0.5,0.5,0.5,3.5)

S3: Bivariate Beta (2.5,0.5,2.5,2.5)

S4: Bivariate Beta (2.5,2.5,0.5,2.5)

In S1 and S2, X and Y have the same marginal decaying rate, whereas in S3 and S4, X and Y have different marginal decaying rates. Figure 1 presents the densities for S1–S4.

Table 1 shows the bias of the estimators. MIXGP outperforms the others with the smallest bias in S2 and S4, and is competitive in S1 and S3. For the CGPD model, the performance varies with the number of clusters K, but $K = 5$ performs better than $K = 3$. In general, MIX-t outperforms GAM in the tail, but MIXGP has better or comparable performance than MIX-t across all scenarios. Table 2 summarizes the MSE of the estimated proportion. Here, MIXGP has the smallest MSE among the seven models in S2 and S4. In S1 and S4, MIXGP has a similar MSE to those of the top performing models.

3.2. Unbounded cases

The bivariate t distribution is a suitable choice for evaluating the aforementioned methods because it has a polynomial marginal decaying rate and its tail dependence is determined jointly by the degrees of freedom df and the correlation ρ (Schmidt (2002)). Because the GEPD and CGPD are limited to nonnegative distributions, we generate data from the truncated bivariate t distribution $TBT\{\mu = (2,2), \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, df = 4, x \ge 0, y \ge 0\}$ with different ρ. We also

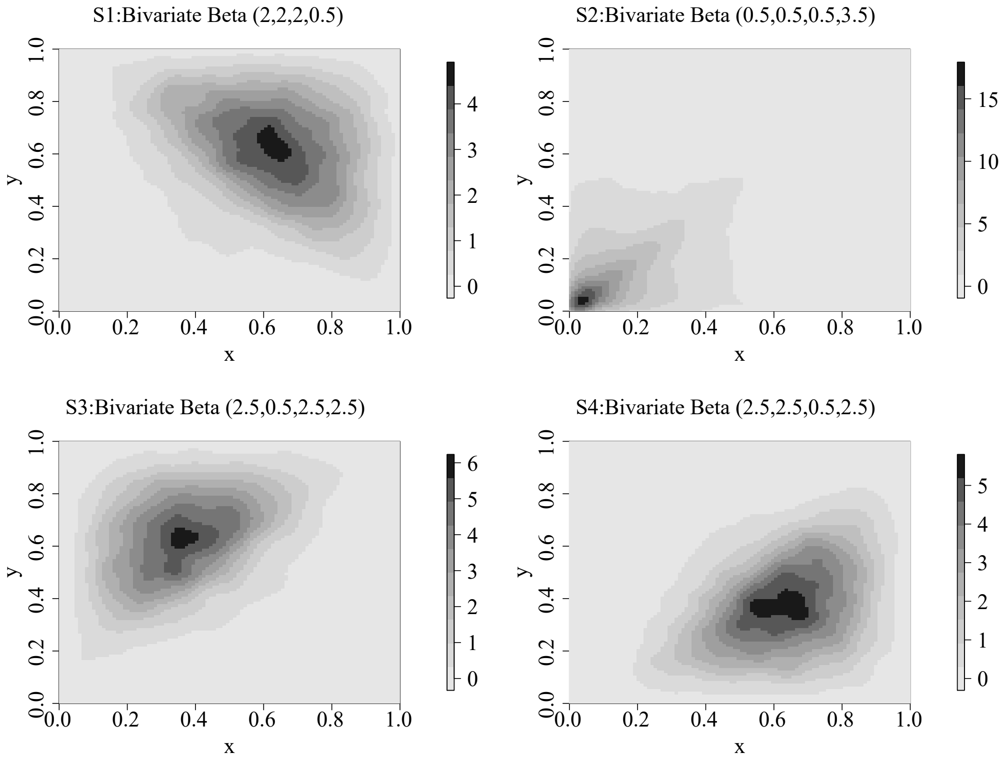

Figure 1. Density plot of the bivariate beta distributions used in S1–S4.

consider data from the GEPD to determine how much is lost by fitting an over-complex semiparametric model compared with fitting a correctly-specified parametric model. The four unbounded scenarios are:

S5: Truncated bivariate t with $\rho = 0$

S6: Truncated bivariate t with $\rho = 0.5$

S7: Truncated bivariate t with $\rho = 0.8$

S8: $X \sim Gamma(10, 10), Y|X \sim EGPD(\delta = 5, \sigma = \exp(X), \xi = 0.1)$

The densities for S5–S8 is presented in Figure 2.

Table 3 presents the results for S5–S8. The MIXGP, in general, outperforms better than competing models in S6 and S7 when the true model is correlated truncated bivariate t, and is among the top models in S5 and S8 when the true model is independent bivariate t and GEPD. The GEPD and CGPD perform well in S8 and S5 when $\rho = 0$. However, these two models show a greater discrepancy

Table 1. S1–4: $(E[\hat{P}(Y \leq Y_{t_y}^C | X_{t_x})] - t_y) \times 1{,}000$ (Bias) for covariate quantile level $t_x = 0.9$ and conditional response quantile levels $t_y = 0.95, 0.99, 0.999$. Standard errors are given in parentheses. Bold represents the smallest Bias among the four models under the same setting. "MIXGP" stands for our proposed MIXGP model, "GEPD" stands for the GEPD model, "C3" stands for the CGPD model with three clusters, "C5" stands for the CGPD model with five clusters, "GAM" stands for the gamma model, "GEV" stands for the GEV model, and "MIX-t" stands for the tail mixture model.

Setting	t_x	t_y	MIXGP	GEPD	C3	C5	GAM	GEV	MIX-t
S1	0.9	0.95	6.14	-77.10	-85.87	-78.44	**1.12**	-182.16	4.97
			(1.44)	(0.45)	(0.48)	(0.66)	(0.61)	(15.47)	(0.98)
		0.99	-8.88	-28.75	-31.75	-29.21	-15.47	-163.11	**1.32**
			(1.05)	(0.15)	(0.16)	(0.23)	(0.38)	(14.23)	(0.31)
		0.999	-10.21	-5.43	-138.11	-5.51	-15.68	-144.05	**-2.07**
			(0.83)	(0.03)	(0.64)	(0.04)	(0.28)	(13.43)	(0.14)
S2	0.9	0.95	**-12.98**	-90.05	-99.52	-65.05	-49.30	-89.92	-49.30
			(1.98)	(0.96)	(2.12)	(2.60)	(0.91)	(1.36)	(0.91)
		0.99	**-2.32**	-79.25	-80.98	-54.61	-39.59	-72.08	-38.08
			(0.94)	(0.78)	(2.04)	(2.22)	(0.58)	(1.03)	(0.75)
		0.999	**-1.63**	-56.43	-55.12	-36.00	-23.65	-50.26	-13.97
			(0.61)	(0.60)	(1.78)	(1.75)	(0.36)	(0.80)	(0.41)
S3	0.9	0.95	**-56.12**	-106.65	-107.57	-104.18	-87.14	-188.52	-87.14
			(1.61)	(0.64)	(0.86)	(1.92)	(1.16)	(15.79)	(1.16)
		0.99	-50.88	-48.20	-48.59	**-47.26**	-81.63	-180.42	-81.63
			(1.34)	(0.27)	(0.36)	(0.81)	(0.89)	(16.01)	(0.89)
		0.999	-39.04	-12.57	-232.69	**-12.35**	-67.53	-163.76	-67.53
			(1.20)	(0.07)	(0.72)	(0.20)	(0.73)	(15.99)	(0.73)
S4	0.9	0.95	**10.76**	-117.37	-101.34	-116.72	-26.04	-58.10	-26.04
			(1.23)	(0.86)	(0.84)	(1.25)	(0.78)	(3.47)	(0.78)
		0.99	**-4.15**	-118.79	-104.94	-118.25	-32.25	-59.02	-25.85
			(0.66)	(0.75)	(0.72)	(1.09)	(0.52)	(3.16)	(0.87)
		0.999	**-3.35**	-94.91	-83.37	-94.48	-20.30	-250.45	-8.81
			(0.33)	(0.63)	(0.59)	(0.91)	(0.32)	(33.34)	(0.38)

in estimating the extremal proportion in S6 and S7 when $\rho > 0$. GAM and MIX-t perform well in the tail when $t_y = 0.999$, but inevitably perform less well for less extreme quantile levels ($t_y = 0.95, 0.99$). Table 4 summarizes the MSE of the methods. In general, GAM and MIX-t have the smallest MSE among the seven models in S5–S7 and GEPD has the smallest MSE among the seven models in S8. MIXGP also has a competitive MSE in S5–S8, as expected. This indicates that MIXGP delivers a reliable estimation.

Table 2. S1–4: $E[\hat{P}(Y \leq Y^C_{t_y}|X_{t_x}) - t_y]^2 \times 1{,}000$ (MSE) for covariate quantile level $t_x = 0.9$ and conditional response quantile levels $t_y = 0.95, 0.99, 0.999$. Standard errors are given in parentheses. Bold represents the smallest MSE among the four models under the same setting. Standard errors are given in parentheses. "MIXGP" stands for our proposed MIXGP model, "GEPD" stands for the GEPD model, "C3" stands for the CGPD model with three clusters, "C5" stands for the CGPD model with five clusters, "GAM" stands for the gamma model, "GEV" stands for the GEV model, and "MIX-t" stands for the tail mixture model.

Setting	t_x	t_y	MIXGP	GEPD	C3	C5	GAM	GEV	MIX-t
S1	0.9	0.95	0.25	5.97	7.40	6.20	**0.04**	57.34	0.12
			(0.07)	(0.07)	(0.08)	(0.11)	(0.01)	(13.90)	(0.02)
		0.99	0.19	0.83	1.01	0.86	0.25	47.05	**0.01**
			(0.07)	(0.01)	(0.01)	(0.01)	(0.01)	(11.67)	(<0.01)
		0.999	0.18	0.03	19.11	0.03	0.25	38.96	**0.01**
			(0.05)	(<0.01)	(0.18)	(<0.01)	(0.01)	(10.14)	(<0.01)
S2	0.9	0.95	**0.57**	8.20	10.36	4.91	2.51	8.27	2.51
			(0.08)	(0.18)	(0.42)	(0.37)	(0.09)	(0.25)	(0.09)
		0.99	**0.09**	6.34	6.98	3.48	1.60	5.30	1.51
			(0.03)	(0.12)	(0.33)	(0.27)	(0.05)	(0.15)	(0.06)
		0.999	**0.04**	3.22	3.36	1.60	0.57	2.59	0.21
			(0.03)	(0.07)	(0.20)	(0.15)	(0.02)	(0.08)	(0.01)
S3	0.9	0.95	**3.41**	11.42	11.65	11.23	7.73	60.72	7.73
			(0.22)	(0.14)	(0.19)	(0.44)	(0.21)	(13.48)	(0.21)
		0.99	2.77	2.33	2.37	**2.30**	6.74	58.43	6.74
			(0.17)	(0.03)	(0.04)	(0.08)	(0.15)	(13.52)	(0.15)
		0.999	1.67	0.16	54.19	**0.16**	4.61	52.63	4.61
			(0.12)	(0.00)	(0.34)	(0.01)	(0.10)	(12.96)	(0.10)
S4	0.9	0.95	**0.27**	13.85	10.34	13.78	0.74	4.59	0.74
			(0.04)	(0.20)	(0.17)	(0.30)	(0.04)	(1.01)	(0.04)
		0.99	**0.06**	14.17	11.07	14.10	1.07	4.49	0.74
			(0.01)	(0.18)	(0.15)	(0.26)	(0.03)	(0.89)	(0.05)
		0.999	**0.02**	9.05	6.99	9.01	0.42	174.98	0.09
			(0.01)	(0.12)	(0.10)	(0.18)	(0.01)	(34.16)	(0.01)

4. Calibration of Extreme Precipitation Forecasts

The US daily precipitation data are downloaded from the United States Historical Climatology Network (USHCN, available at `http://cdiac.ess-dive.lbl.gov/epubs/ndp/ushcn/ushcn.html`). These data include the daily precipitation from 1,218 monitor stations across the contiguous 48 US states for the pe-

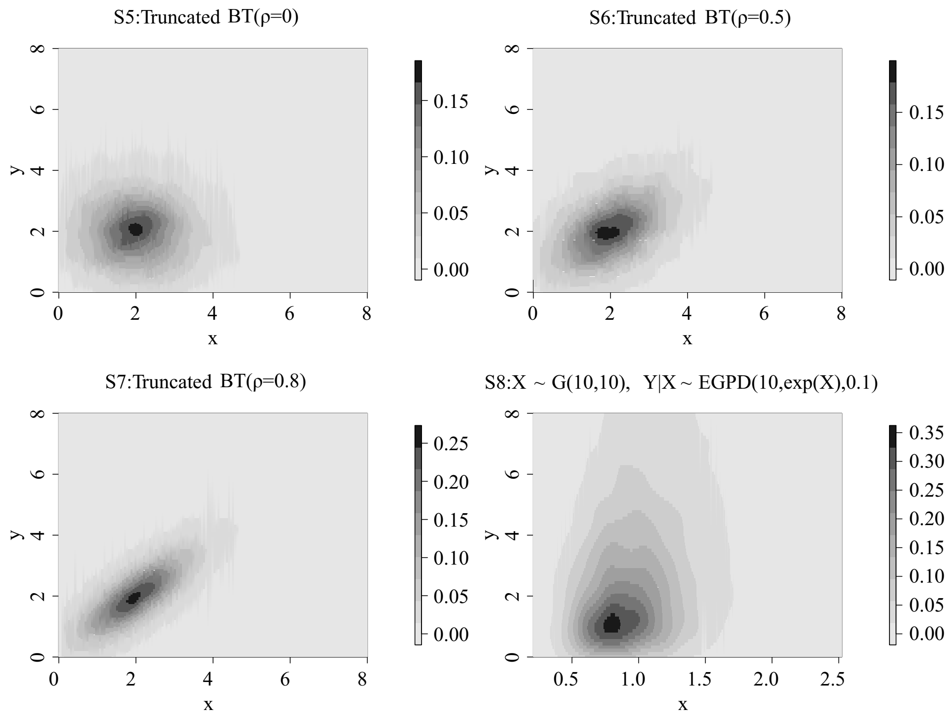

Figure 2. Density plot of the truncated bivariate t and the GEPD model used in S5–S8.

riod 1905 to 2014. The THORPEX Interactive Grand Global Ensemble (TIGGE) data set is a key part of THORPEX project, a world weather research program for the period 2006 to 2014 (Bougeault et al. (2010)). The TIGGE data consist of ensemble forecast data from 10 global numerical weather prediction centers. Each NWP center delivers forecasts every six hours on a 0.5×0.5 degree global grid. We use only the eight forecasts that deliver daily predictions for the continental US (CMA, CPTEC, ECCC, ECMWF, JMA, KMA, NCEPM, UKMO). The station data are matched with forecasts made at midnight (one-day prior forecast) for the grid cell closest to the station, and missing observations (either forecast or station data) are discarded. We use data for 2014 only, because this year has the most integrated data.

Figure 3 presents a QQ plot of the observed daily precipitation, eight forecast models, and ensemble mean and maximum at several stations. All forecast models show discrepancies from the reference line at the tail. The maximum forecast is the closest to the reference line. In addition to the QQ plot, we compared the Pearson correlation between the observed data and the ensemble mean/max forecast. Because we are interested in extremes, we restrict the computation of the

Table 3. S5–8: $(E[\hat{P}(Y \leq Y_{t_y}^C | X_{t_x})] - t_y) \times 1{,}000$ (Bias) for covariate quantile level $t_x = 0.9$ and conditional response quantile levels $t_y = 0.95, 0.99, 0.999$. Standard errors are given in parentheses. Bold represents the smallest Bias among the four models under the same setting. "MIXGP" stands for our proposed MIXGP model, "GEPD" stands for the GEPD model, "C3" stands for the CGPD model with three clusters, "C5" stands for the CGPD model with five clusters, "GAM" stands for the gamma model, "GEV" stands for the GEV model, and "MIX-t" stands for the tail mixture model.

Setting	t_x	t_y	MIXGP	GEPD	C3	C5	GAM	GEV	MIX-t
S5	0.9	0.95	**-7.62**	-49.30	-58.03	-50.96	18.59	-343.41	15.10
			(1.71)	(0.99)	(1.89)	(1.94)	(0.81)	(31.40)	(0.79)
		0.99	**-2.64**	-22.06	-26.52	-23.13	2.73	-300.21	-3.50
			(0.66)	(0.46)	(0.93)	(0.96)	(0.24)	(31.60)	(0.59)
		0.999	-0.90	-2.42	-3.33	-2.71	**0.56**	-93.32	-2.30
			(0.21)	(0.08)	(0.18)	(0.19)	(0.02)	(7.19)	(0.26)
S6	0.9	0.95	**-4.60**	-94.27	-119.57	-106.32	-42.21	-168.24	-42.21
			(2.14)	(0.90)	(5.41)	(2.76)	(1.69)	(28.09)	(1.69)
		0.99	**-2.66**	-57.77	-76.36	-66.00	-13.70	-144.61	-11.48
			(1.04)	(0.57)	(3.49)	(1.77)	(0.64)	(29.18)	(0.81)
		0.999	-1.46	-14.24	-22.22	-17.40	**-0.70**	-122.88	-2.86
			(0.32)	(0.19)	(1.25)	(0.62)	(0.07)	(6.82)	(0.32)
S7	0.9	0.95	**-4.16**	-167.65	-178.79	-179.73	-56.22	-180.11	-56.22
			(2.06)	(0.93)	(9.28)	(4.24)	(1.86)	(27.84)	(1.86)
		0.99	**-4.59**	-130.99	-144.65	-141.70	-32.73	-164.21	-31.17
			(1.09)	(0.75)	(7.32)	(3.47)	(0.99)	(28.90)	(1.20)
		0.999	**-3.03**	-59.01	-71.77	-66.34	-5.49	-37.21	-5.89
			(0.61)	(0.44)	(4.23)	(2.08)	(0.23)	(6.48)	(0.42)
S8	0.9	0.95	-6.51	**-0.89**	-18.98	7.59	29.39	-23.67	21.68
			(2.63)	(0.90)	(1.67)	(1.64)	(0.59)	(13.18)	(0.60)
		0.99	-1.24	**-0.40**	-7.62	1.43	9.01	-29.60	4.39
			(0.92)	(0.32)	(0.72)	(0.54)	(0.05)	(13.66)	(0.24)
		0.999	-1.18	**-0.15**	-2.00	-0.11	1.00	-15.60	0.38
			(0.34)	(0.07)	(0.24)	(0.13)	(<0.01)	(9.84)	(0.05)

correlation to observations with observed forecasts greater than their 0.95 quantiles. The correlation between the observations and the ensemble mean forecast is 0.267, with $[0.253, 0.282]$ as the 95% confidence interval. The correlation between the observations and the ensemble maximum forecast is 0.288, with $[0.274, 0.302]$ as the 95% confidence interval. This indicates that the max forecast is more correlated with observations in the tail. Therefore, the maximum forecast is used throughout as the one-number summary of the ensemble.

Table 4. S5–8: $E[\hat{P}(Y \leq Y^C_{t_y}|X_{t_x}) - t_y]^2 \times 1{,}000$ (MSE) for covariate quantile level $t_x = 0.9$ and conditional response quantile levels $t_y = 0.95, 0.99, 0.999$. Standard errors are given in parentheses. Bold represents the smallest MSE among the four models under the same setting. Standard errors are given in parentheses. "MIXGP" stands for our proposed MIXGP model, "GEPD" stands for the GEPD model, "C3" stands for the CGPD model with three clusters, "C5" stands for the CGPD model with five clusters, "GAM" stands for the gamma model, "GEV" stands for the GEV model, and "MIX-t" stands for the tail mixture model.

Setting	t_x	t_y	MIXGP	GEPD	C3	C5	GAM	GEV	MIX-t
S5	0.9	0.95	0.35	2.53	3.73	2.98	0.41	217.52	**0.29**
			(0.08)	(0.11)	(0.25)	(0.26)	(0.03)	(29.99)	(0.03)
		0.99	0.05	0.51	0.79	0.63	**0.01**	191.02	0.05
			(0.01)	(0.02)	(0.06)	(0.06)	(<0.01)	(33.58)	(0.01)
		0.999	0.01	0.01	0.01	0.01	<**0.01**	13.93	0.01
			(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(1.20)	(<0.01)
S6	0.9	0.95	**0.48**	8.97	17.25	12.07	2.07	108.03	2.07
			(0.09)	(0.17)	(1.18)	(0.59)	(0.15)	(27.07)	(0.15)
		0.99	**0.12**	3.37	7.06	4.67	0.23	106.93	0.20
			(0.03)	(0.07)	(0.50)	(0.24)	(0.02)	(29.37)	(0.02)
		0.999	0.01	0.21	0.65	0.34	<**0.01**	19.80	0.02
			(<0.01)	(0.01)	(0.06)	(0.02)	(0.00)	(1.21)	(0.00)
S7	0.9	0.95	**0.45**	28.19	40.67	34.12	3.51	110.75	3.51
			(0.10)	(0.31)	(2.97)	(1.52)	(0.21)	(27.08)	(0.21)
		0.99	**0.14**	17.22	26.34	21.29	1.17	111.33	1.12
			(0.04)	(0.20)	(1.95)	(0.99)	(0.07)	(29.34)	(0.07)
		0.999	0.05	3.50	6.96	4.84	**0.04**	5.63	0.05
			(0.01)	(0.05)	(0.57)	(0.29)	(<0.01)	(1.47)	(0.01)
S8	0.9	0.95	0.74	**0.08**	0.64	0.33	0.90	18.12	0.51
			(0.16)	(0.01)	(0.08)	(0.05)	(0.03)	(12.70)	(0.03)
		0.99	0.09	**0.01**	0.11	0.03	0.08	19.71	0.02
			(0.02)	(<0.01)	(0.01)	(<0.01)	(<0.01)	(13.79)	(<0.01)
		0.999	0.01	<**0.01**	0.01	<0.01	<0.01	10.01	<0.01
			(<0.01)	(<0.01)	(<0.01)	(<0.01)	(0.00)	(9.98)	(<0.01)

Previous studies recommend using a power transformation (Sloughter et al. (2007); Hamill, Whitaker and Wei (2004); Berrocal et al. (2008)), and so we transform both X and Y using square root transformations. The data are zero-inflated, with 71% of the Y samples equal to zero and 7% of X samples equal to zero. The individual forecasts have a higher proportion of zeros, but because X is the ensemble maximum, it is rarely zero. Therefore, we split the data according

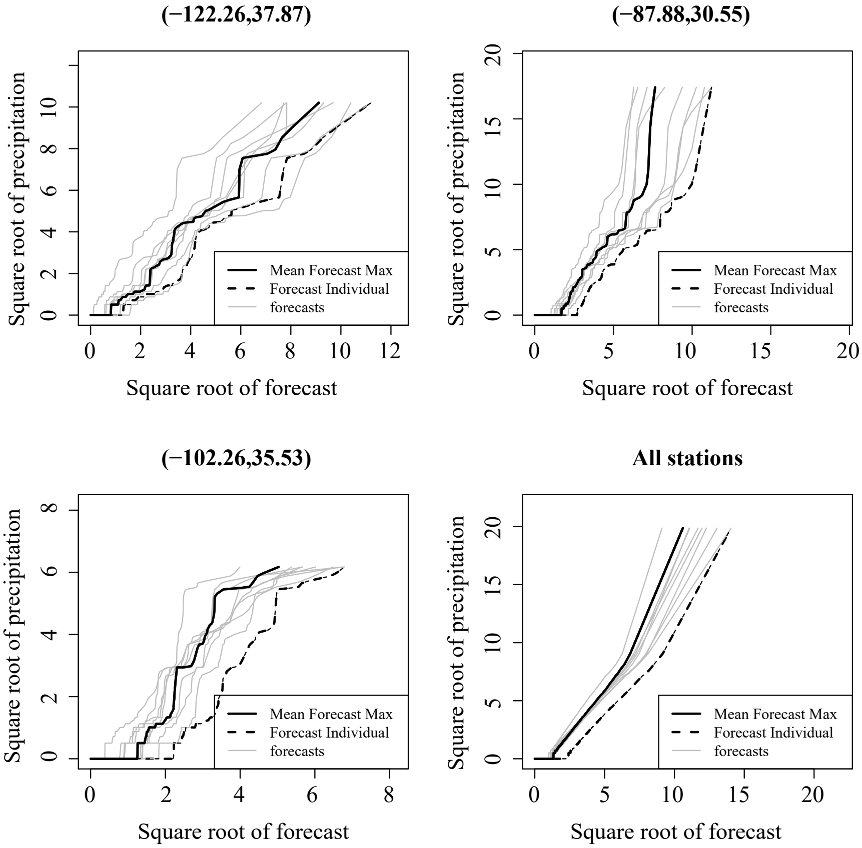

Figure 3. Sample QQ plots between the square root of the observed daily precipitation ($\mathrm{mm}^{1/2}$), eight individual forecast models, and square root of the ensemble mean and maximum forecasts ($\mathrm{mm}^{1/2}$). The longitude and latitude of three stations are given in the captions and the final plot pools data for all stations.

to whether $X = 0$ and fit the two categories of data separately:

$$\begin{aligned} f(Y=y|X=0) &= P_{00}I(Y=0|X=0) \\ &\quad + (1-P_{00})g_1(Y=y|X=0)I(Y>0|X=0), \\ f(Y=y|X>0) &= P_{10}I(Y=0|X>0) \\ &\quad + (1-P_{10})g_2(Y=y|X>0)I(Y>0|X>0), \end{aligned} \tag{4.1}$$

where $P_{00} = P(Y=0|X=0)$ and $P_{10} = P(Y=0|X>0)$. For both categories, we apply a two-stage model: First, we predict whether Y is positive or zero. We estimate $\hat{P}_{00}$ using the sample proportion of $I(Y=0|X=0)$, and estimate $\hat{P}_{10}$ by fitting a simple logistic regression model of $I(Y=0|X>0)$ regressed onto X.

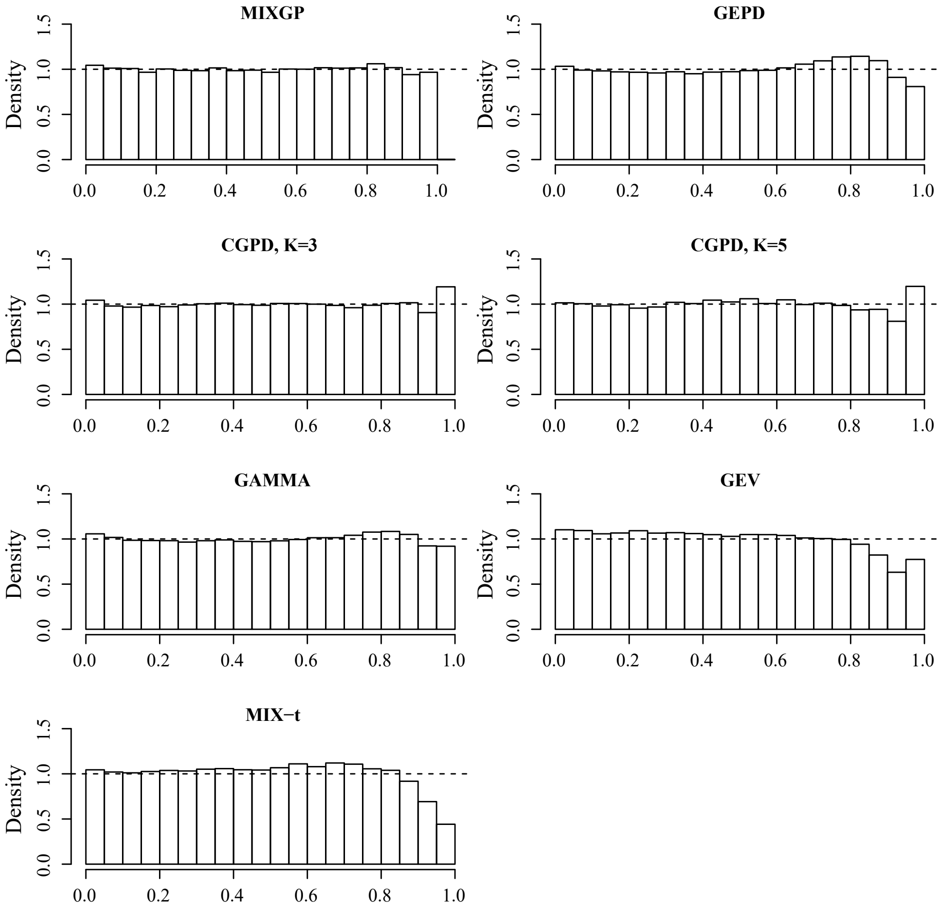

Figure 4. PIT histogram for the MIXGP, EGPD, CGPD model with cluster number K = 3; 5, GAM, GEV, and MIX-t model. "MIXGP" stands for our proposed MIXGP model, "CGPD,K=3" stands for the CGPD model with three clusters, "CGPD, K=5" stands for the CGPD model with five clusters, "GEPD" stands for the GEPD model, "GAM" stands for the gamma model, "GEV" stands for the GEV model, "MIX-t" stands for the tail mixture model.

Second, we fit univariate MIXGP or EGPD models to estimate $g_1(Y = y|X = 0)$, given positive Y and zero X, and fit the models described in Section 3 for the joint distribution of (X, Y), given that both are positive. Here $g_2(Y = y|X > 0)$ is obtained as a byproduct of the joint distribution of (X, Y). We perform a five-fold cross validation. Each time, we randomly sample 80% of the data as a training set, and evaluate the model on the remaining 20%.

To evaluate the performance for each model, we estimate the probability integral transform (PITs): $\omega_i = P(Y \le Y_i|X_i)$. If the model fits well, the PITs

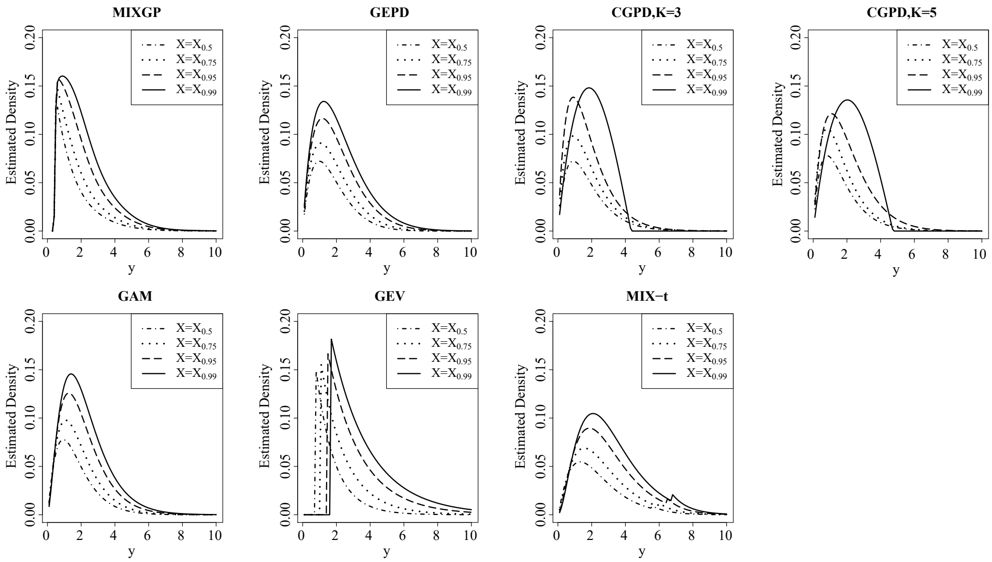

Figure 5. Estimated density comparison among different conditions by given $X_{0.5}, X_{0.75}, X_{0.95}, X_{0.99}$. "MIX" stands for our proposed MIXGP model, "CGPD,K=3" stands for the CGPD model with three clusters, "CGPD, K=5" stands for the CGPD model with five clusters, "GEPD" stands for the GEPD model, "GAM" stands for the gamma model, "GEV" stands for the GEV model, and "MIX-t" stands for the tail mixture model.

will be very close to the uniform distribution. PITs close to zero and one measure the fit of the lower and upper tails, separately. The estimated PIT plots for the different models are shown in Figure 4. All models perform well in the lower tail. However, the GEPD models are questionable for the upper tail because the PIT plots show departures from uniformity on the upper 50% of the data. The MIXGP performs well when estimating the upper tail. For the CGPD model with $K = 5$ clusters, the upper tail looks slightly lighter, but this drawback is remedied by choosing $K = 3$ clusters. The GAM model has a roughly uniform PIT with a slight departure in the tail. The GEV and MIX-t models fit poorly, especially in the upper tail.

To further evaluate the performance, we compare the fitted PDF of $Y|X$ for covariate quantile levels $t_x = 0.5, 0.95, 0.99$ in Figure 5. The density plots correspond to the second row of (4.1). Conditional on $X = X_{99\%}$, CGPD gives a bounded distribution, and the upper bound is smaller than $X_{99\%}$. This precludes the observed precipitation being greater than the forecast. The MIXGP and GEPD estimators have heavier tails, consistent with our intuition. The GAM model is close to the EGPD when the covariate quantile level is not extreme

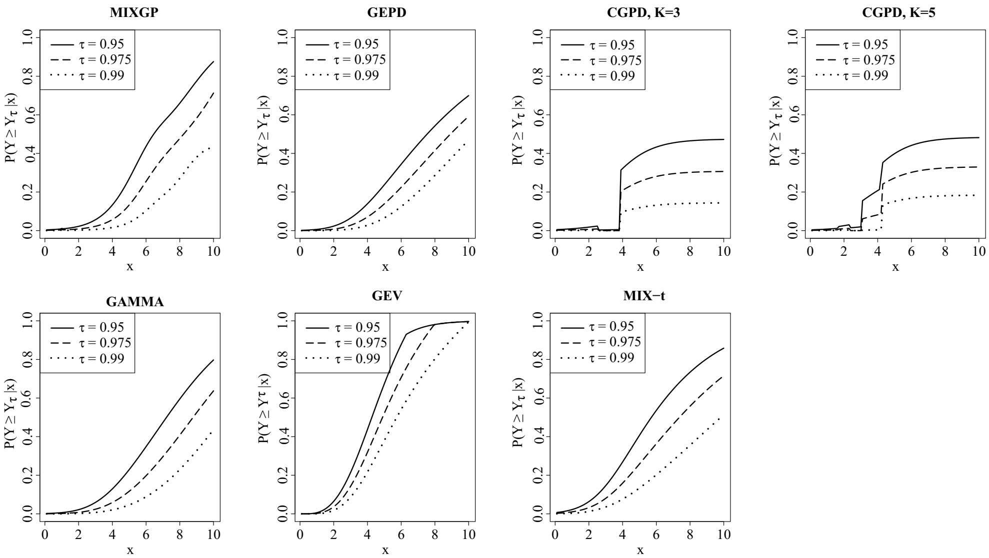

Figure 6. $\hat{P}(Y \geq Y_\tau | X = x)$ for different values of x by given $\tau = 0.95, 0.975, 0.99$. "MIX" stands for our proposed MIXGP model, "CGPD,K=3" stands for the CGPD model with three clusters, "CGPD, K=5" stands for the CGPD model with five clusters, "GEPD" stands for the GEPD model, "GAM" stands for the gamma model, "GEV" stands for the GEV model, and "MIX-t" stands for the tail mixture model.

($t_x = 0.5, 0.95$), and becomes closer to our MIXGP when $t_x = 0.99$. The GEV and the MIX-t have heavier tails than those of the other models, which is consistent with our findings in the PIT plot. Figure 6 presents the probability of Y being greater than a given value, conditional on different X. As expected, for a given Y_τ, the probability of $Y \geq Y_\tau$ increases with X.

We also consider the Brier score (BS, Brier (1950)) and quantile score (QS, Koenker and Machado (1999); Friederichs and Hense (2007); Gneiting and Raftery (2007)) to further evaluate the model fitting performance. The BS is appropriate for binary or categorical outcomes. The response is binarized by recording whether it exceeds $Y_{95\%}$ and comparing it to $P(Y_i \geq Y_{95\%} | X_i)$, the probability of being an extreme event. For QS, we consider the quantiles $\tau = 0.5, 0.95, 0.99$. The results are shown in Table 5. For the extreme quantile, GEV achieves the smallest QS. MIXGP achieves competitive results to those of GEV. For the nonextreme quantile, MIXGP achieves the smallest BS and QS. The performance of GAM is close to that of MIXGP.

Table 5. Brier scores (BS) and quantile scores (QS) with $\tau = 0.5, 0.95, 0.99$. Standard deviations are given in parentheses. "MIX" stands for our proposed MIXGP model, "CGPD,K=3" stands for the CGPD model with 3 clusters, "CGPD, K=5" stands for the CGPD model with 5 clusters and "GEPD" stands for the GEPD model,"GAM" stands for the gamma model, "GEV" stands for the GEV model, "MIX-t" stands for the tail mixture model.

	BS	QS($\tau = 0.5$)	QS($\tau = 0.95$)	QS($\tau = 0.99$)
MIXGP	0.0878	0.3007	0.1529	0.0475
	(0.0009)	(0.0019)	(0.0012)	(0.0007)
GEPD	0.0884	0.3018	0.1567	0.0499
	(0.0009)	(0.0016)	(0.0013)	(0.0007)
CGPD,K=3	0.0903	0.3070	0.1731	0.0682
	(0.0221)	(0.0038)	(0.0216)	(0.0234)
CGPD,K=5	0.0929	0.3150	0.1925	0.0866
	(0.0617)	(0.0127)	(0.0522)	(0.0594)
GAM	0.0884	0.3006	0.1536	0.0477
	(0.0009)	(0.0018)	(0.0016)	(0.0008)
GEV	0.0922	0.3373	0.1500	0.0430
	(0.0063)	(0.0408)	(0.0017)	(0.0038)
MIX-t	0.0879	0.3083	0.1721	0.0515
	(0.0010)	(0.0024)	(0.0012)	(0.0002)

5. Conclusion

In this paper, we have proposed a finite mixture univariate and bivariate models for precipitation extremes. Based on our proposed bivariate model, the conditional distribution of precipitation can be derived, which in turn gives us a calibrated forecast of the precipitation. Empirically, we show our proposed method outperforms other state-of-the-art methods on a large precipitation data set.

One of the main advantages of our proposed MIXGP model is that it allows us to flexibly capture the tail behavior in both univariate and bivariate cases. The flexibility is derived from two main sources. First, we randomize the threshold parameter in the GPD. Second, we choose a finite mixture instead of specifying a single parametric model.

It remains to explore the tail behavior of the conditional distribution. Our models can be made more flexible using different choices of structure function V in general BGPDs. We have restricted our discussion to univariate and bivariate models. It would be interesting to extend our proposed model to multivariate and spatial cases, where there is a more complicated dependence structure. For example, Oesting, Schlather and Friederichs (2017) proposed a parametric bivariate spatial model to calibrate forecasts; extending this to more flexible models is left to future work.

Supplementary Material

The proofs of Theorems 1 and 2, computational details, and additional simulation results are given in the online Supplementary Material.

Acknowledgments

This material is based upon work partially supported by the National Science Foundation under grant DMS-1638521 to the Statistical and Applied Mathematical Sciences Institute, as well as by the National Institutes of Health under grant R01-ES027892.

References

Behrens, C. N., Lopes, H. F. and Gamerman, D. (2004). Bayesian analysis of extreme events with threshold estimation. *Statistical Modelling* **4**, 227–244.

Bentzien, S. and Friederichs, P. (2012). Generating and calibrating probabilistic quantitative precipitation forecasts from the high-resolution nwp model cosmo-de. *Weather and Forecasting* **27**, 988–1002.

Berrocal, V. J., Raftery, A. E. and Gneiting, T. (2008). Probabilistic quantitative precipitation field forecasting using a two-stage spatial model. *The Annals of Applied Statistics* **2**, 1170–1193.

Bjørnar Bremnes, J. (2004). Probabilistic forecasts of precipitation in terms of quantiles using nwp model output. *Monthly Weather Review* **132**, 338–347.

Bougeault, P., Toth, Z., Bishop, C., Brown, B., Burridge, D., Chen, D. H., Ebert, B., Fuentes, M., Hamill, T. M., Mylne, K., Nicolau, J., Paccagnella, T., Park, Y.-Y., Parsons, D., Raoult, B., Schuster, D., Dias, P. S., Swinbank, R., Takeuchi, Y., Tennant, W., Wilson, L. and Worley, S. (2010). The thorpex interactive grand global ensemble. *Bulletin of the American Meteorological Society* **91**, 1059–1072.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthey Weather Review* **78**, 1–3.

Cabras, S. and Castellanos, M. E. (2011). A Bayesian approach for estimating extreme quantiles under a semiparametric mixture model. *ASTIN Bulletin: The Journal of the IAA* **41**, 87–106.

Caeiro, F. and Gomes, M. I. (2015). Threshold selection in extreme value analysis. In *Extreme Value Modeling and Risk Analysis: Methods and Applications* (Edited by K. D. Dey and J. Yan), 69–82. Chapman-Hall/CRC.

Cooray, K. and Ananda, M. M. (2005). Modeling actuarial data with a composite lognormal-pareto model. *Scandinavian Actuarial Journal* **2005**, 321–334.

de Haan, L. and Ferreira, A. (2006). Estimation of the extreme value index and testing. In *Extreme Value Theory: An Introduction*, 65–126. Springer.

do Nascimento, F. F., Gamerman, D. and Lopes, H. F. (2012). A semiparametric Bayesian approach to extreme value estimation. *Statistics and Computing* **22**, 661–675.

Friederichs, P. and Hense, A. (2007). Statistical downscaling of extreme precipitation events using censored quantile regression. *Monthly Weather Review* **135**, 2365–2378.

Frigessi, A., Haug, O. and Rue, H. (2002). A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes* **5**, 219–235.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC.

Ghosal, S. and Van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*, volume 44. Cambridge University Press.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* **102**, 359–378.

Gupta, A. K. and Wong, C. (1985). On three and five parameter bivariate beta distributions. *Metrika* **32**, 85–91.

Hamill, T. M., Whitaker, J. S. and Wei, X. (2004). Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts. *Monthly Weather Review* **132**, 1434–1447.

Ho, H. J., Lin, T.-I., Chen, H.-Y. and Wang, W.-L. (2012). Some results on the truncated multivariate t distribution. *Journal of Statistical Planning and Inference* **142**, 25–40.

Holden, L. and Haug, O. (2009). A multidimensial mixture model for unsupervised tail estimation. *NR-notat SAMBA/09/09*, 29.

Hu, Y. (2013). *Extreme Value Mixture Modelling with Simulation Study and Applications in Finance and Insurance*. Master's Thesis.

Koenker, R. and Machado, J. A. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association* **94**, 1296–1310.

Lee, D., Li, W. K. and Wong, T. S. T. (2012). Modeling insurance claims via a mixture exponential model combined with peaks-over-threshold approach. *Insurance: Mathematics and Economics* **51**, 538–550.

Li, C., Lin, L. and Dunson, D. B. (2019). On posterior consistency of tail index for Bayesian kernel mixture models. *Bernoulli* **25**, 1999-2028.

MacDonald, A. E. (2011). Extreme value mixture modelling with medical and industrial applications. PhD Theses. University of Canterbury.

Nadarajah, S. and Kotz, S. (2005). Some bivariate beta distributions. *Statistics* **39**, 457–466.

Naveau, P., Huser, R., Ribereau, P. and Hannart, A. (2016). Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research* **52**, 2753–2769.

Oesting, M., Schlather, M. and Friederichs, P. (2017). Statistical post-processing of forecasts for extremes using bivariate brown-resnick processes with an application to wind gusts. *Extremes* **20**, 309–332.

Olkin, I. and Trikalinos, T. A. (2015). Constructions for a bivariate beta distribution. *Statistics & Probability Letters* **96**, 54–60.

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

Pigeon, M. and Denuit, M. (2011). Composite lognormal–Pareto model with random threshold. *Scandinavian Actuarial Journal* **2011**, 177–192.

Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT–Statistical Journal* **10**, 33–60.

Scheuerer, M. (2014). Probabilistic quantitative precipitation forecasting using ensemble model output statistics. *Quarterly Journal of the Royal Meteorological Society* **140**, 1086–1096.

Schmidt, R. (2002). Tail dependence for elliptically contoured distributions. *Mathematical Methods of Operations Research* **55**, 301–327.

Scollnik, D. P. (2007). On composite lognormal-Pareto models. *Scandinavian Actuarial Journal* **2007**, 20–33.

Sloughter, J. M. L., Raftery, A. E., Gneiting, T. and Fraley, C. (2007). Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Monthly Weather Review* **135**, 3209–3220.

Smith, R. L. (1994). Multivariate threshold methods. In *Extreme Value Theory and Applications*, 225–248. Springer.

Taddy, M. A. and Kottas, A. (2010). A Bayesian nonparametric approach to inference for quantile regression. *Journal of Business & Economic Statistics* **28**, 357–369.

Tancredi, A., Anderson, C. and O'Hagan, A. (2006). Accounting for threshold uncertainty in extreme value estimation. *Extremes* **9**, 87.

Vrac, M. and Naveau, P. (2007). Stochastic downscaling of precipitation: From dry events to heavy rainfalls. *Water Resources Research* **43**.

Wang, H. and Li, D. (2015). Estimation of extreme conditional quantiles. In *Extreme Value Modeling and Risk Analysis: Methods and Applications* (Edited by K. D. Dey and J. Yan), 319–337. Chapman-Hall/CRC.

Wilks, D. S. (2011). *Statistical Methods in the Atmospheric Sciences*, volume 100. Academic press.

Zheng, L., Ismail, K. and Meng, X. (2014). Shifted gamma-generalized Pareto distribution model to map the safety continuum and estimate crashes. *Safety Science* **64**, 155–162.

Yuan Tian

Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA.

E-mail: ytian11@ncsu.edu

Brian J. Reich

Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA.

E-mail: bjreich@ncsu.edu

(Received October 2018; accepted November 2019)