Modelling jointly low, moderate and HEAVY rainfall intensities without a threshold selection

Raphaël Huser ©2016
(King Abdullah University of Science and Technology, SA)
KAUST

Univariate modeling of rainfall intensities ("wet rainfall events")

Naveau, Huser, Ribereau and Hannart (2016, to appear), Modeling jointly low, moderate and heavy rainfall intensities without a threshold selection, Water Resources Research

Discussion about spatial joint models for extreme and non-extreme data

Krupskii, Huser, Genton (2016, submitted), Factor copula models for spatial data
KAUST

- Newly founded (i.e., 6-year old) graduate university in Thuwal, Saudi Arabia;
- Located by the Red Sea;
- Highly multi-cultural (more than 100 nationalities on campus);
Current KAUST President is the former president of CalTech;

Currently, about 140 faculty, 150 research scientists, 400 postdocs, 850 students (academic population: \( \sim 1600 \); campus population: \( \sim 6000 \));

3 statistics groups:
- M. G. Genton (Spatio-Temporal Statistics and Data Science)
- Y. Sun (Environmental Statistics)
- R. Huser (Extreme Statistics), http://extstat.kaust.edu.sa

*We are looking for talented PhD students and postdocs.*
An extreme event is rare (by definition), but not necessarily intense.

We would like to characterize rainfall intensities from low extremes to high extremes.

Extreme-Value Theory is a natural starting point.
Let $X \sim F(x)$ be a random variable. Under mild assumptions, the tails of $F$ should be (asymptotically) Generalized Pareto (GP) distributed, i.e.,

\[
\Pr(X > x \mid X > u) \approx 1 - H_\xi \left( \frac{x-u}{\sigma} \right), \quad \text{for large } x > u,
\]

where

\[
1 - H_\xi(x) = \begin{cases} 
(1 + \xi x)^{-1/\xi}, & \xi \neq 0, \\
\exp(-x), & \xi = 0,
\end{cases}
\]

and $\xi \in \mathbb{R}$ is the upper tail shape parameter.
Let $X \sim F(x)$ be a random variable. Under mild assumptions, the tails of $F$ should be (asymptotically) Generalized Pareto (GP) distributed, i.e.,

\[ \Pr(X > x \mid X > u) \approx 1 - H_\xi \left( \frac{x-u}{\sigma} \right), \text{ for large } x > u, \text{ where} \]

\[ 1 - H_\xi(x) = \begin{cases} 
(1 + \xi x)^{-1/\xi}, & \xi \neq 0, \\
\exp(-x), & \xi = 0,
\end{cases} \]

and $\xi \in \mathbb{R}$ is the upper tail shape parameter.

\[ \Pr(-X > -x \mid -X > -u) \approx 1 - H_\xi \left( \frac{-x+u}{\sigma} \right), \text{ for large } -x > -u. \]
Extreme-Value Theory

Let $X \sim F(x)$ be a random variable. Under mild assumptions, the tails of $F$ should be (asymptotically) Generalized Pareto (GP) distributed, i.e.,

\[
\Pr(X > x \mid X > u) \approx 1 - H_\xi \left( \frac{x-u}{\sigma} \right), \text{ for large } x > u, \text{ where}
\]

\[
1 - H_\xi(x) = \begin{cases} 
(1 + \xi x)^{-1/\xi}, & \xi \neq 0, \\
\exp(-x), & \xi = 0,
\end{cases}
\]

and $\xi \in \mathbb{R}$ is the upper tail shape parameter.

\[
\Pr(-X > -x \mid -X > -u) \approx 1 - H_\xi \left( \frac{-x+u}{\sigma} \right), \text{ for large } -x > -u.
\]

With the constraint $X > 0$, this implies that for small $0 < x < u$,

\[
\Pr(X < x \mid X < u) \approx \text{Cst} \times x^\kappa
\]

where $\kappa > 0$ is the lower tail shape parameter.
Rainfall data in Lyon, France (1996–2011)

Spring data

Summer data

Fall data

Winter data
Classical approach in EVT

Spring data

TOY workshop, NCAR, Boulder CO
Classical approach in EVT

Spring data

TOY workshop, NCAR, Boulder CO  
April 26, 2016 – slide 10
Classical approach in EVT

Rainfall amount [mm] vs Time index

Spring data

Density

TOY workshop, NCAR, Boulder CO
Classical approach in EVT

Spring data

Rainfall amount [mm]

Time index

Density

Rainfall amounts [mm]
Modeling of rainfall intensities

LOW – 2011 European Drought

HIGH – 2005 European Floods
Modeling of rainfall intensities

LOW – 2011 European Drought

HIGH – 2005 European Floods

Our goal: To construct simple parametric models for the full range of rainfall intensities
Modeling of rainfall intensities

LOW – 2011 European Drought

HIGH – 2005 European Floods

Our goal: To construct simple parametric models for the full range of rainfall intensities, continuous (i.e., bypassing threshold selection to separate low, moderate and large rainfall)
Modeling of rainfall intensities

LOW – 2011 European Drought

HIGH – 2005 European Floods

Our goal: To construct simple parametric models for the full range of rainfall intensities, continuous (i.e., bypassing threshold selection to separate low, moderate and large rainfall), easy to simulate from
Modeling of rainfall intensities

LOW – 2011 European Drought
HIGH – 2005 European Floods

Our goal: To construct simple parametric models for the full range of rainfall intensities, continuous (i.e., bypassing threshold selection to separate low, moderate and large rainfall), easy to simulate from, Gamma-like
Our goal: To construct simple parametric models for the full range of rainfall intensities, continuous (i.e., bypassing threshold selection to separate low, moderate and large rainfall), easy to simulate from, Gamma-like, but heavy-tailed.
Modeling of rainfall intensities

LOW – 2011 European Drought  
HIGH – 2005 European Floods

Our goal: To construct simple parametric models for the full range of rainfall intensities, continuous (i.e., bypassing threshold selection to separate low, moderate and large rainfall), easy to simulate from, Gamma-like, but heavy-tailed, in compliance with Extreme-Value Theory on the left and right tails.
Modeling of rainfall intensities

LOW – 2011 European Drought

HIGH – 2005 European Floods

Our goal: To construct simple parametric models for the full range of rainfall intensities, continuous (i.e., bypassing threshold selection to separate low, moderate and large rainfall), easy to simulate from, Gamma-like, but heavy-tailed, in compliance with Extreme-Value Theory on the left and right tails. Furthermore, inference should be simple.
Modeling of rainfall intensities

LOW – 2011 European Drought

HIGH – 2005 European Floods

Our goal: To construct simple parametric models for the full range of rainfall intensities, continuous (i.e., bypassing threshold selection to separate low, moderate and large rainfall), easy to simulate from, Gamma-like, but heavy-tailed, in compliance with Extreme-Value Theory on the left and right tails. Furthermore, inference should be simple. And return levels easy to compute.
Literature review

- Carreau and Bengio (2006, Extremes): Mixtures of hybrid Gaussian–Pareto densities for moderate and extreme data;
- Frigessi et al. (2002, Extremes): Mixture of light-tailed and GP densities, with mixture proportion depending smoothly on the event “extremeness”;
- Carvalho et al. (2013, Tech Rep): Non-stationary extreme-value mixture model via P-splines;
Let $U \sim \text{Unif}(0, 1)$. Then, $X = \sigma H_\xi^{-1}(U)$ is distributed according the the GP distribution with scale $\sigma$ and shape $\xi$. 
Let $U \sim \text{Unif}(0, 1)$. Then, $X = \sigma H^{-1}_\xi(U)$ is distributed according the the GP distribution with scale $\sigma$ and shape $\xi$.

To enrich this family of distributions, let $X = \sigma H^{-1}_\xi(V)$, where $V = G^{-1}(U)$ and $G$ is a continuous distribution on $[0, 1]$.
Let $U \sim \text{Unif}(0, 1)$. Then, $X = \sigma H^{-1}_\xi(U)$ is distributed according the GP distribution with scale $\sigma$ and shape $\xi$.

To enrich this family of distributions, let $X = \sigma H^{-1}_\xi(V)$, where $V = G^{-1}(U)$ and $G$ is a continuous distribution on $[0, 1]$.

- The distribution $F$ of $X$ is $F(x) = G\{H_\xi(x/\sigma)\}$. 
Let $U \sim \text{Unif}(0, 1)$. Then, $X = \sigma H_{\xi}^{-1}(U)$ is distributed according to the GP distribution with scale $\sigma$ and shape $\xi$.

To enrich this family of distributions, let $X = \sigma H_{\xi}^{-1}(V)$, where $V = G^{-1}(U)$ and $G$ is a continuous distribution on $[0, 1]$.

- The distribution $F$ of $X$ is $F(x) = G\{H_{\xi}(x/\sigma)\}$.
- One can easily simulate from $F$, whenever $G^{-1}$ is available.
Let \( U \sim \text{Unif}(0, 1) \). Then, \( X = \sigma H_\xi^{-1}(U) \) is distributed according to the GP distribution with scale \( \sigma \) and shape \( \xi \).

To enrich this family of distributions, let \( X = \sigma H_\xi^{-1}(V) \), where \( V = G^{-1}(U) \) and \( G \) is a continuous distribution on \([0, 1]\).

- The distribution \( F \) of \( X \) is \( F(x) = G\{H_\xi(x/\sigma)\} \).
- One can easily simulate from \( F \), whenever \( G^{-1} \) is available.
- Similarly, quantiles (i.e., return levels) may be readily obtained.
Rainfall intensities modelling

- Let \( U \sim \text{Unif}(0, 1) \). Then, \( X = \sigma H_\xi^{-1}(U) \) is distributed according to the GP distribution with scale \( \sigma \) and shape \( \xi \).

- To enrich this family of distributions, let \( X = \sigma H_\xi^{-1}(V) \), where \( V = G^{-1}(U) \) and \( G \) is a continuous distribution on \([0, 1]\).
  - The distribution \( F \) of \( X \) is \( F(x) = G\{H_\xi(x/\sigma)\} \).
  - One can easily simulate from \( F \), whenever \( G^{-1} \) is available.
  - Similarly, quantiles (i.e., return levels) may be readily obtained.
  - \( G \) must be such that
    - The right-tail of \( F \) behaves like \( H_\xi (x \to \infty) \);
    - The left-tail of \( F \) behaves like \( x^\kappa (x \to 0) \).
Sufficient conditions on $G$

(A) For some finite $a > 0$, one has

$$
\lim_{v \downarrow 0} \frac{1 - G(1 - v)}{v} = a;
$$

(B) For any positive function $w(v)$ such that $w(v) = 1 + o(v)$ as $v \to 0$, one has for some finite $b > 0$

$$
\lim_{v \downarrow 0} \frac{G\{v \cdot w(v)\}}{G(v)} = b;
$$

(C) For some finite $c > 0$, one has

$$
\lim_{v \downarrow 0} \frac{v^\kappa}{G(v)} = c.
$$
Parametric models for $G$

- $G(v) = v^\kappa$, $\kappa > 0$: obeys conditions (A), (B), and (C).

- $F(x)$ has only three parameters ($\kappa$, $\sigma$ and $\xi$) and corresponds to the extended GP model Type III in Papastathopoukos and Tawn (2013, JSPI).

Model (i): Case $G(v) = v^\kappa$
Parametric models for $G$

- $G(v) = p v^{\kappa_1} + (1 - p) v^{\kappa_2}$, $p \in [0, 1]$, $\kappa_1, \kappa_2 > 0$: obeys conditions (A), (B), and (C) (with $\kappa = \min(\kappa_1, \kappa_2)$).

- $F(x)$ has five parameters ($p$, $\kappa_1$, $\kappa_2$, $\sigma$ and $\xi$).

Model (ii): Case $G(v) = (v^{\kappa_1} + v^{\kappa_2}) / 2$
Parametric models for $G$

- $G(v) = 1 - Q_\delta \{(1 - v^\delta)\}$, where $\delta > 0$ and $Q_\delta$ is the Beta distribution with parameters $1/\delta, 2$: obeys conditions (A), (B), and (C) (with $\kappa = 2$).
- $F(x)$ has only three parameters ($\delta, \sigma$ and $\xi$).
- One may write $f(x) = \text{Cst} \times \sigma^{-1} h_\xi(x/\sigma)[1 - \{1 - H_\xi(x/\sigma)\}^\delta]$, which makes a link with skewed distributions.
- One can generalize the model to $G(v) = [1 - Q_\delta \{(1 - v^\delta)\}]^{\kappa/2}$ with $\kappa > 0$.

Model (iii): Case $G(v) = 1 - Q_\delta \{(1 - v^\delta)\}$
Inference

- Maximum Likelihood Estimator (MLE)
  - based on the density \( f(x) = \sigma^{-1} g\{H_\xi(x/\sigma)\}h_\xi(x/\sigma), x > 0 \).
  - straightforward to apply whenever the density \( g(v) \) is available.
  - very efficient (especially for large sample sizes), but may suffer from model misspecification.

- Probability Weighted Moments (PWMs)
  - equate empirical and theoretical probability weighted moments of the form \( E[X\{1 - F(X)\}^s], s = 0, 1, \ldots \).
  - need explicit expression of theoretical moments.
  - usually fast to compute.
  - typically more robust than MLEs.
  - less efficient (in large samples) but have good small sample properties.
Simulation study

Setting: $X_1, \ldots, X_n \overset{iid}{\sim} F(x) = G\{H_\xi(x/\sigma)\}, \, x > 0.$

- **Sample size**: $n = 300$ (similar to the rainfall data)
- **GP scale parameter** $\sigma = 1$, **GP shape parameter** $\xi = 0.1, 0.2, 0.3$.
- **Four different models for $G$**: 
  
  (i) $G(v) = v^\kappa$, with lower tail parameter $\kappa \in \{1, 2, 5, 10\}$;
  
  (ii) $G(v) = 1 - Q_\delta \{(1 - v)\delta\}$, with skewness parameter $\delta \in \{0.5, 1, 2, 5\}$
  
  (iii) $G(v) = pv^\kappa_1 + (1 - p)v^\kappa_2$, with $p = 0.4$, $\kappa_1 \in \{1, 2, 5, 10\}$ and $\kappa_2 \in \{2, 5, 10, 20\}$, with $\kappa_1 < \kappa_2$;
  
  (iv) $G(v) = [1 - Q_\delta \{(1 - v)\delta\}]^{\kappa/2}$, with $\delta \in \{0.5, 1, 2, 5\}$, and $\kappa \in \{1, 2, 5, 10\}$.

- **Replicates**: $R = 10^5$
Selected simulation results

Model (i) $G(v) = v^\kappa$

\[ \kappa = 2 \quad \sigma = 1 \quad \xi = 0.2 \quad \text{0.99-quantile} \]
Application to the rainfall data in Lyon, France (1996–2011)

Spring data

Summer data

Fall data

Winter data
Hourly rainfall data (1996–2011)
Application to the rainfall data in Lyon, France (1996–2011)

- Hourly rainfall data (1996–2011)
- Temporal dependence
  ⇒ Declustering: keep every third observation
Application to the rainfall data in Lyon, France (1996–2011)

- Hourly rainfall data (1996–2011)
- Temporal dependence
  - Declustering: keep every third observation
- A high proportion of zeros!
  - Focus on wet rainfall events
Hourly rainfall data (1996–2011)

Temporal dependence

⇒ Declustering: keep every third observation

A high proportion of zeros!

⇒ Focus on wet rainfall events

Discretization (rainfall measured to the closest 0.1mm)

⇒ Use censored ML and PWMs (below 0.5mm)

⇒ About 400 fully informative observations (above this threshold)
Hourly rainfall data (1996–2011)

Temporal dependence

⇒ Declustering: keep every third observation

A high proportion of zeros!

⇒ Focus on wet rainfall events

Discretization (rainfall measured to the closest 0.1mm)

⇒ Use censored ML and PWMs (below 0.5mm)

⇒ About 400 fully informative observations (above this threshold)

We fit separate models for each season.
Application to the rainfall data in Lyon, France (1996–2011)

- **Hourly** rainfall data (1996–2011)
- **Temporal dependence**
  - Declustering: keep every third observation
- A high proportion of zeros!
  - Focus on wet rainfall events
- **Discretization** (rainfall measured to the closest 0.1mm)
  - Use censored ML and PWMs (below 0.5mm)
  - About 400 fully informative observations (above this threshold)
- We fit separate models for each season.
- **Uncertainty assessment**: using non-parametric bootstrap.
### Spring data

<table>
<thead>
<tr>
<th>Method</th>
<th>Estimated parameters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\kappa$</td>
<td>$\sigma$</td>
</tr>
<tr>
<td>ML</td>
<td>$1_{(0.6, 2)} \times 10^3$</td>
<td>0.00$_{(0.00, 0.00)}$</td>
</tr>
<tr>
<td>ML-c</td>
<td>0.59$_{(0.51, 0.73)}$</td>
<td>1.44$_{(1.09, 1.65)}$</td>
</tr>
<tr>
<td>PWMs</td>
<td>0.61$_{(0.56, 0.67)}$</td>
<td>1.47$_{(1.27, 1.61)}$</td>
</tr>
<tr>
<td>PWMs-c</td>
<td>0.50$_{(0.44, 0.57)}$</td>
<td>1.55$_{(1.31, 1.73)}$</td>
</tr>
</tbody>
</table>

### Summer data

<table>
<thead>
<tr>
<th>Method</th>
<th>Estimated parameters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\kappa$</td>
<td>$\sigma$</td>
</tr>
<tr>
<td>ML</td>
<td>$0.7_{(0.4, 2)} \times 10^3$</td>
<td>0.00$_{(0.00, 0.00)}$</td>
</tr>
<tr>
<td>ML-c</td>
<td>0.51$_{(0.41, 0.64)}$</td>
<td>1.94$_{(1.36, 2.65)}$</td>
</tr>
<tr>
<td>PWMs</td>
<td>0.56$_{(0.51, 0.63)}$</td>
<td>1.82$_{(1.46, 2.16)}$</td>
</tr>
<tr>
<td>PWMs-c</td>
<td>0.43$_{(0.37, 0.53)}$</td>
<td>2.16$_{(1.61, 2.66)}$</td>
</tr>
</tbody>
</table>
### Parameter estimates for model (i)

#### Fall data

<table>
<thead>
<tr>
<th>Method</th>
<th>$\kappa$</th>
<th>$\sigma$</th>
<th>$\xi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML</td>
<td>$1 \times 10^3$</td>
<td>0.00</td>
<td>0.85</td>
</tr>
<tr>
<td>ML-c</td>
<td>0.80</td>
<td>1.00</td>
<td>0.29</td>
</tr>
<tr>
<td>PWMs</td>
<td>0.66</td>
<td>1.32</td>
<td>0.20</td>
</tr>
<tr>
<td>PWMs-c</td>
<td>0.62</td>
<td>1.22</td>
<td>0.22</td>
</tr>
</tbody>
</table>

#### Winter data

<table>
<thead>
<tr>
<th>Method</th>
<th>$\kappa$</th>
<th>$\sigma$</th>
<th>$\xi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML</td>
<td>$3 \times 10^3$</td>
<td>0.00</td>
<td>0.67</td>
</tr>
<tr>
<td>ML-c</td>
<td>0.84</td>
<td>0.63</td>
<td>0.23</td>
</tr>
<tr>
<td>PWMs</td>
<td>0.59</td>
<td>1.07</td>
<td>0.04</td>
</tr>
<tr>
<td>PWMs-c</td>
<td>0.63</td>
<td>0.73</td>
<td>0.20</td>
</tr>
</tbody>
</table>
Diagnostics for model (i)

Spring data

Winter data
Diagnostics for model (i)

Spring

Empirical quantiles [mm]

Fitted quantiles [mm]

- MLE
- PWMs
- GPD

Summer

Empirical quantiles [mm]

Fitted quantiles [mm]

- MLE
- PWMs
- GPD
Towards spatial modeling of extreme and non-extreme data

- Skar's Theorem: \( F(x_1, \ldots, x_D) = C\{F_1(x_1), \ldots, F_D(x_D)\} \).
  
  \( C \) is the copula (uniquely defined if \( F \) is continuous).
Towards spatial modeling of extreme and non-extreme data

- **Skar’s Theorem:** \( F(x_1, \ldots, x_D) = C\{F_1(x_1), \ldots, F_D(x_D)\} \).
  
  \( C \) is the copula (uniquely defined if \( F \) is continuous).

- **Copula families:**
  - **Gaussian:** tractable but tail independent
Towards spatial modeling of extreme and non-extreme data

- **Skar’s Theorem:** $F(x_1, \ldots, x_D) = C\{F_1(x_1), \ldots, F_D(x_D)\}$. 
  $C$ is the copula (uniquely defined if $F$ is continuous).

- **Copula families:**
  - **Gaussian:** tractable but tail independent
  - **Trans-Gaussian:** same as Gaussian copula
Towards spatial modeling of extreme and non-extreme data

- **Skar’s Theorem:** \( F(x_1, \ldots, x_D) = C\{F_1(x_1), \ldots, F_D(x_D)\} \).
  
  \( C \) is the copula (uniquely defined if \( F \) is continuous).

- **Copula families:**
  
  - **Gaussian:** tractable but tail independent
  
  - **Trans-Gaussian:** same as Gaussian copula
  
  - **V-transformed Gaussian copula (\( \chi^2 \) processes):** intractable and tail independent
Towards spatial modeling of extreme and non-extreme data

- **Skar’s Theorem:** \( F(x_1, \ldots, x_D) = C\{F_1(x_1), \ldots, F_D(x_D)\} \).
  \( C \) is the copula (uniquely defined if \( F \) is continuous).

- **Copula families:**
  - **Gaussian:** tractable but tail independent
  - **Trans-Gaussian:** same as Gaussian copula
  - **V-transformed Gaussian copula (\( \chi^2 \) processes):** intractable and tail independent
  - **Student \( t \):** tail dependent but tail-symmetric
Towards spatial modeling of extreme and non-extreme data

- **Skar’s Theorem**: \( F(x_1, \ldots, x_D) = C\{F_1(x_1), \ldots, F_D(x_D)\} \).
  
  \( C \) is the copula (uniquely defined if \( F \) is continuous).

- **Copula families**:
  - **Gaussian**: tractable but tail independent
  - **Trans-Gaussian**: same as Gaussian copula
  - **V-transformed Gaussian copula (\( \chi^2 \) processes)**: intractable and tail independent
  - **Student \( t \)**: tail dependent but tail-symmetric
  - **Extreme-value copulas**: justified for extremes only, computationally demanding to fit
Towards spatial modeling of extreme and non-extreme data

- **Skar’s Theorem:** \( F(x_1, \ldots, x_D) = C\{F_1(x_1), \ldots, F_D(x_D)\} \). 
  \( C \) is the copula (uniquely defined if \( F \) is continuous).

- **Copula families:**
  - **Gaussian:** tractable but tail independent
  - **Trans-Gaussian:** same as Gaussian copula
  - **V-transformed Gaussian copula (\( \chi^2 \) processes):** intractable and tail independent
  - **Student \( t \):** tail dependent but tail-symmetric
  - **Extreme-value copulas:** justified for extremes only, computationally demanding to fit
  - **Vine copulas:** (too?) flexible but lack interpretability
Towards spatial modeling of extreme and non-extreme data

- **Skar’s Theorem:** \( F(x_1, \ldots, x_D) = C\{F_1(x_1), \ldots, F_D(x_D)\} \).
  
  \( C \) is the copula (uniquely defined if \( F \) is continuous).

- **Copula families:**
  - **Gaussian:** tractable but tail independent
  - **Trans-Gaussian:** same as Gaussian copula
  - **V-transformed Gaussian copula (\( \chi^2 \) processes):** intractable and tail independent
  - **Student \( t \):** tail dependent but tail-symmetric
  - **Extreme-value copulas:** justified for extremes only, computationally demanding to fit
  - **Vine copulas:** (too?) flexible but lack interpretability

- **Factor copula models:** combine tractability, interpretability, flexibility, and yield tail dependence if the factor is suitably chosen.
Factor copula models

- Model:

\[ X(s) = Z(s) + V, \quad s \in \mathbb{R}^d, \]

where \( Z(s) \sim \text{GP}(0, 1, \rho(h)) \perp \perp V \sim f_V. \)
Factor copula models

Model:

\[ X(s) = Z(s) + V, \quad s \in \mathbb{R}^d, \]

where \( Z(s) \sim \text{GP}(0, 1, \rho(h)) \perp \perp V \sim f_V. \)

The random variable \( V \) is the latent random factor.

\[ \Rightarrow \text{Interpretation: An unobserved random phenomenon affects the joint dependence of variables.} \]
Factor copula models

□ Model:

\[ X(s) = Z(s) + V, \quad s \in \mathbb{R}^d, \]

where \( Z(s) \sim \text{GP}(0, 1, \rho(h)) \perp V \sim f_V. \)

□ The random variable \( V \) is the latent random factor.

\[ \Rightarrow \text{Interpretation: An unobserved random phenomenon affects the joint dependence of variables.} \]

□ If suitably chosen, \( V \) yields tail dependence and tail asymmetry!
Factor copula models

- Model:
  \[ X(s) = Z(s) + V, \quad s \in \mathbb{R}^d, \]
  where \( Z(s) \sim \text{GP}(0, 1, \rho(h)) \perp V \sim f_V. \)

- The random variable \( V \) is the latent random factor.
  \( \Rightarrow \) Interpretation: An unobserved random phenomenon affects the joint dependence of variables.

- If suitably chosen, \( V \) yields tail dependence and tail asymmetry!

- Copula density for \( D \) locations \( s_1, \ldots, s_D \in \mathbb{R}^d \):
  \[ c(u_1, \ldots, u_D) = \frac{f_D^{X} \{ F_1^{-1}(u_1), \ldots, F_D^{-1}(u_D) \}}{f_1^{X} \{ F_1^{-1}(u_1) \} \times \cdots \times f_D^{X} \{ F_D^{-1}(u_D) \}}, \]
  where
  \[ f_D^{X} = \int_{-\infty}^{\infty} \phi_D(x_1 - v, \ldots, x_D - v; \Sigma) f_V(v) dv, \]
  and \( \phi_D \) is the multivariate standard Gaussian density with covariance matrix \( \Sigma = \{ \rho(s_i - s_j) \}_{i,j=1}^{D} \).
Factor copula models

- If $V$ is chosen to be Gaussian, $X(s)$ will be tail independent.
Factor copula models

- If $V$ is chosen to be Gaussian, $X(s)$ will be tail independent.
- **Proposition:** Let $\Pr(V > v) \sim K v^\beta \exp(-\theta v^\alpha)$, $0 \leq \alpha < 2$, $\theta > 0$, $K > 0$, as $v \to \infty$.
  - $0 < \alpha < 1$ or $\alpha = 0$, $\beta < 0$: compete (upper) tail dependence
  - $\alpha = 1$: (upper) tail dependence
  - $\alpha > 1$: (upper) tail independence
If $V$ is chosen to be Gaussian, $X(s)$ will be tail independent.

**Proposition:** Let $\Pr(V > v) \sim K v^\beta \exp(-\theta v^\alpha)$, $0 \leq \alpha < 2$, $\theta > 0$, $K > 0$, as $v \to \infty$.

- $0 < \alpha < 1$ or $\alpha = 0$, $\beta < 0$: compete (upper) tail dependence
- $\alpha = 1$: (upper) tail dependence
- $\alpha > 1$: (upper) tail independence

**Proposition:** if $V \sim \text{Exp}(\lambda)$, the process $X(s)$ is in the max-domain of attraction of the Brown-Resnick process. ($\Rightarrow X(s)$ lives in a penultimate world).
Factor copula models

- If $V$ is chosen to be Gaussian, $X(s)$ will be tail independent.
- **Proposition:** Let $\Pr(V > v) \sim Kv^\beta \exp(-\theta v^\alpha)$, $0 \leq \alpha < 2$, $\theta > 0$, $K > 0$, as $v \to \infty$.
  - $0 < \alpha < 1$ or $\alpha = 0$, $\beta < 0$: compete (upper) tail dependence
  - $\alpha = 1$: (upper) tail dependence
  - $\alpha > 1$: (upper) tail independence
- **Proposition:** if $V \sim \text{Exp}(\lambda)$, the process $X(s)$ is in the max-domain of attraction of the Brown-Resnick process.
  ($\Rightarrow X(s)$ lives in a penultimate world).
- By setting $V = V_1 - V_2$ with $V_1, V_2 \geq 0$, one can model tail asymmetry.
Factor copula models

- If $V$ is chosen to be Gaussian, $X(s)$ will be tail independent.
- Proposition: Let $\Pr(V > v) \sim K v^\beta \exp(-\theta v^\alpha)$, $0 \leq \alpha < 2$, $\theta > 0$, $K > 0$, as $v \to \infty$.
  - $0 < \alpha < 1$ or $\alpha = 0, \beta < 0$: compete (upper) tail dependence
  - $\alpha = 1$: (upper) tail dependence
  - $\alpha > 1$: (upper) tail independence
- Proposition: if $V \sim \text{Exp}(\lambda)$, the process $X(s)$ is in the max-domain of attraction of the Brown-Resnick process.
  ($\Rightarrow X(s)$ lives in a penultimate world).
- By setting $V = V_1 - V_2$ with $V_1, V_2 \geq 0$, one can model tail asymmetry.
- If $V_1 \sim \text{Exp}(\lambda_1) \perp \perp V_2 \sim \text{Exp}(\lambda_2)$, the likelihood can be obtained in closed-form!
Factor copula models

We have proposed univariate models for the full range of rainfall intensities, which bypass the need to select suitable thresholds distinguishing between low, moderate and heavy rainfall.

These models are in compliance with extreme-value theory on both tails.

Simulation and inference are simple.

ML performs slightly better when the model is well-specified, but PWMs are much more robust to model misspecification.

Spatial (or multivariate) modeling of extreme and non-extreme data may be achieved by combining our univariate models with factor copula models.

Open questions:
– Correlation between lower and upper shape parameters?
– More flexible models for the bulk (Doug’s approach?)
– Modeling of rainfall occurrences (wet/dry events)
– Non-stationarity