Bridging asymptotic independence and dependence in spatial extremes using Gaussian scale mixtures

Emeric Thibaud

Colorado State University

Joint work with: Raphaël Huser (KAUST), Thomas Opitz (INRA)

STATMOS Workshop on Climate and Weather Extremes

Oct 25, 2016

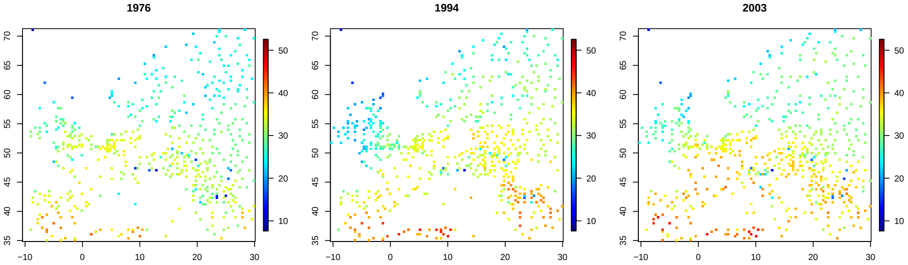
Outline

1. Modeling spatial extremes: asymptotic dependence and independence
2. Gaussian scale mixtures
3. Inference for threshold exceedances
4. Application to hourly wind speed in the Pacific Northwest
5. Conclusion

Modeling spatial extremes: asymptotic dependence and independence

Spatial modeling

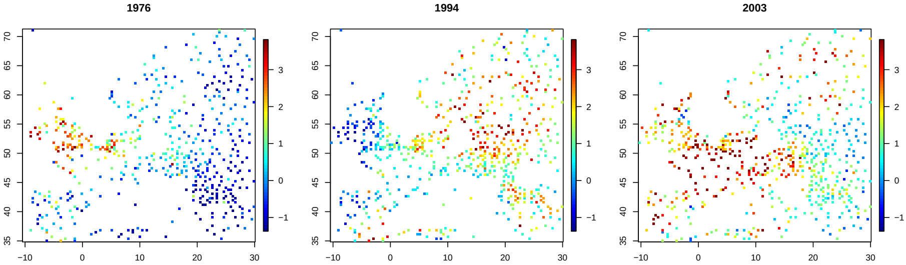
- Two spatial aspects: marginal distributions vary spatially (climate) and there is spatial dependence in the 'residuals' (weather).
- In this talk we focus on the residual dependence.



Annual maximum temperatures (^{o}C) over Europe

Spatial modeling

- Two spatial aspects: marginal distributions vary spatially (climate) and there is spatial dependence in the 'residuals' (weather).
- In this talk we focus on the residual dependence.



Annual maximum temperatures over Europe: Gumbel marginals

Spatial EVT

- Extreme value theory (EVT) motivates asymptotic models:
  - renormalized pointwise block maxima of spatial processes converge to max-stable processes (de Haan, 1984);
  - threshold exceedances of spatial processes 'converge' to Pareto processes (threshold stable; Ferreira & de Haan, 2014).
- Marginal distributions are 'easy' (parametric forms) but extremal dependence is complex (spectral measure). We have parametric models (Brown–Resnick, extremal-t) but inference is difficult.
- Do we really need asymptotic models?
  - for the marginals? It depends... (Morris et al.,2016)
  - for the dependence? There is at least one case you don't want to use the asymptotic model...

Asymptotic independence

- For (Y_1, Y_2) a random vector with marginal distributions F_1, F_2, define

$$\chi = \lim_{u \to 1} \Pr\{F_1(Y_1) > u \mid F_2(Y_2) > u\}.$$

We say (Y_1, Y_2) are asymptotically independent (AI) if $\chi = 0$, and asymptotically dependent (AD) otherwise.

- Gaussian vectors with $\rho < 1$ are AI.
- For AI processes, max-stable and Pareto limits are 'white-noise'... but dependence may be present at sub-asymptotic levels.

$\Rightarrow$ Asymptotic models are useless for AI processes.

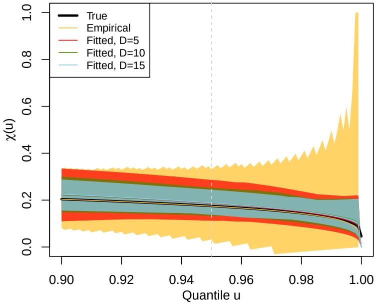
- Many environmental data seem to be AI (or at least the 'observed extremes' are not stable).

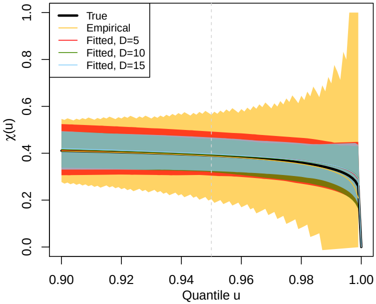
How to decide between AI and AD models?

- In practice we estimate

$$\chi(u) \approx \Pr\{F_1(Y_1) > u \mid F_2(Y_2) > u\}, \quad u \approx 1.$$

- Large variability! In spatial problems, can we borrow strength across locations to decide on AI/AD? Yes but we need flexible spatial models that can cover AI and AD cases.





These graphs show $\chi(u)$ for an AI and an AD process.

Copula

- In this talk we focus on the residual dependence. We use a copula framework to separate marginal and dependence modeling:

By Sklar's theorem, any continuous joint distribution $G(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^D$, with univariate margins $G_1, \ldots, G_D$ may be uniquely represented as

$$G(\boldsymbol{x}) = C\{G_1(x_1), \ldots, G_D(x_D)\}, \quad \boldsymbol{x} \in \mathbb{R}^D,$$

where

$$C(\boldsymbol{u}) = G\{G_1^{-1}(u_1), \ldots, G_D^{-1}(u_D)\}, \quad \boldsymbol{u} \in (0,1)^D,$$

is the copula associated to G.

- Copula $=$ multivariate distribution with Unif$(0,1)$ marginals.

Gaussian scale mixtures

Gaussian scale mixtures

- Gaussian scale mixtures = Gaussian processes with random variances:

Definition:

$$X(\boldsymbol{s}) = RW(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathcal{S} \subset \mathbb{R}^2,$$

where $W = \{W(\boldsymbol{s})\}_{\boldsymbol{s}}$ is a standard Gaussian process and $R \sim F(r)$ is a positive random variable independent of W.

- Conditional on R, X is Gaussian with variance R^2.
- If $R = r_0$ a.s., X is Gaussian.
- We will use the copula associated to X to model dependence in high threshold exceedances.
- EVT: looks a bit like a Pareto process...

First properties

- Finite dimensional distributions are 'easy': let $\boldsymbol{X} = R\boldsymbol{W} \in \mathbb{R}^D$ where $R \sim F(r)$ has a density $f(r)$, and $\boldsymbol{W} \sim N_D(\mathbf{0}, \boldsymbol{\Sigma})$. The distribution G and the density g of $\boldsymbol{X}$:

$$G(\boldsymbol{x}) = \int_0^\infty \Phi_D(\boldsymbol{x}/r; \boldsymbol{\Sigma}) f(r) \mathrm{d}r, \quad g(\boldsymbol{x}) = \int_0^\infty \phi_D(\boldsymbol{x}/r; \boldsymbol{\Sigma}) r^{-D} f(r) \mathrm{d}r.$$

  Marginal distributions G_k and their corresponding densities g_k:

$$G_k(x_k) = \int_0^\infty \Phi(x_k/r) f(r) \mathrm{d}r, \quad g_k(x_k) = \int_0^\infty \phi(x_k/r) r^{-1} f(r) \mathrm{d}r.$$

  Partial derivatives of G:

$$G_I(\boldsymbol{x}) = \int_0^\infty \Phi_{|I^c|} \left\{ (\boldsymbol{x}_{I^c} - \boldsymbol{\Sigma}_{I^c;I} \boldsymbol{\Sigma}_{I;I}^{-1} \boldsymbol{x}_I)/r; \boldsymbol{\Sigma}_{I^c|I} \right\} \phi_{|I|}(\boldsymbol{x}_I/r; \boldsymbol{\Sigma}_{I;I}) r^{-|I|} f(r) \mathrm{d}r.$$

- Numerical methods can be used to estimate the previous expressions.
- X is an elliptical process.
- We derived an algorithm for conditional simulation.

Asymptotic properties of Gaussian scale mixtures

- We characterize the bivariate joint tail decay of Gaussian scale mixtures with the coefficients χ and $\overline{\chi}$ (Coles et al., 1999) defined as $\chi := \lim_{u\to 1} \chi(u)$ and $\overline{\chi} := \lim_{u\to 1} \overline{\chi}(u)$, where

$$\chi(u) = 2 - \frac{\log C(u,u)}{\log(u)}, \quad \overline{\chi}(u) = \frac{2\log(1-u)}{\log \overline{C}(u,u)} - 1,$$

  and $C(u_1, u_2)$ is the copula associated to $(X_1, X_2)^T$ and $\overline{C}(u_1, u_2) = 1 - u_1 - u_2 + C(u_1, u_2)$.
- AD $\Rightarrow \chi > 0$ and $\overline{\chi} = 1$.
- AI $\Rightarrow \overline{\chi} \in [-1, 1]$ and $\chi = 0$.
- To understand the asymptotic dependence of Gaussian scale mixtures we study the asymptotic properties of $R(W_1, W_2)^T$, where the Gaussian vector $(W_1, W_2)^T$ has correlation $\rho \in (-1, 1)$. AI/AD depends on the tail of R...

AI for Gaussian scale mixtures

Suppose that R is a Weibull-type distribution, i.e.,

$$\Pr(R \geq r) = 1 - F(r) \sim \alpha r^{\gamma} \exp(-\delta r^{\beta}), \qquad r \to \infty, \tag{1}$$

for some constants $\alpha > 0$, $\beta > 0$, $\gamma \in \mathbb{R}$ and $\delta > 0$. Then $\chi = 0$ and

$$\overline{\chi} = 2\left\{(1+\rho)/2\right\}^{\beta/(\beta+2)} - 1.$$

The joint tail can be written as

$$\overline{C}\{1 - 1/x, 1 - 1/x\} = \mathcal{L}(x) x^{-1/\eta}, \qquad x \to \infty, \tag{2}$$

where $\eta = (1 + \overline{\chi})/2$ is the coefficient of tail dependence (Ledford & Tawn, 1996), $\mathcal{L}(x) \sim K \log(x)^{(1-1/\eta)\frac{2\gamma+\beta}{2\beta}+1/(2\eta)-1}$ is a slowly varying function as $x \to \infty$ and K is a positive constant depending on α, β, γ and δ.

- Note: The case where R is deterministic or upper-bounded a.s. can be interpreted as a limit of (1) as $\beta \to \infty$ (and in this case $\overline{\chi} = \rho$).

AD for Gaussian scale mixtures

Suppose that R is a Pareto-type distribution, i.e., R is regularly varying at infinity,

$$\frac{\Pr(R \geq tr)}{\Pr(R \geq t)} = \frac{1 - F(tr)}{1 - F(t)} = r^{-\gamma}, \qquad r > 0, \quad t \to \infty, \tag{3}$$

for some $\gamma > 0$. Then $\overline{\chi} = 1$ and

$$\chi = 2\left[1 - T\left\{(1+\gamma)^{1/2}(1-\rho)(1-\rho^2)^{-1/2}; \gamma + 1\right\}\right], \tag{4}$$

where $T(\cdot; \mathrm{Df})$ is the univariate Student-t distribution with $\mathrm{Df} > 0$ degrees of freedom. The joint tail can be written as

$$\overline{C}(1 - 1/x, 1 - 1/x) \sim \chi \times \Pr\{G_1(X_1) > 1 - 1/x\} \sim \chi/x, \quad x \to \infty. \tag{5}$$

- Intuition: AD is obtained when the tail of R dominates the tail of X_1.
- EVT: 'close' to an elliptical Pareto process (Thibaud & Opitz, 2015).

EVT: limits for Gaussian scale mixtures

- AD case: extremal-t process (Opitz, 2013) and elliptical Pareto process (Thibaud & Opitz, 2015).
- AI case: 'white-noise'. (But a Brown–Resnick process limit for block maxima can be obtained using triangular arrays of Gaussian scale mixtures with increasing correlation.)

Bridging asymptotic independence and dependence

We propose for R a two-parameter distribution with support $[1, \infty)$:

$$F(r) = \begin{cases} 1 - \exp\left\{-\gamma(r^\beta - 1)/\beta\right\}, & \beta > 0, \\ 1 - r^{-\gamma}, & \beta = 0, \end{cases} \qquad r \geq 1.$$

for $\beta \geq 0, \gamma > 0$.

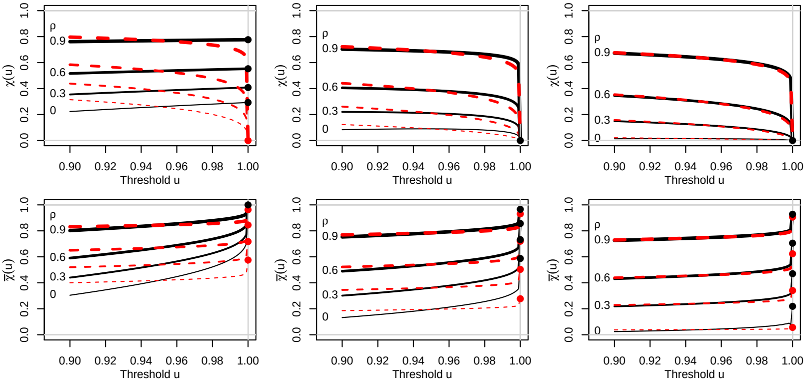
- This distribution forms a continuous parametric family in β.
- AI/AD is determined by the value of β:

$$\beta > 0 \Rightarrow \text{AI}$$

$$\beta = 0 \Rightarrow \text{AD}.$$

- The Dirac mass at 1, and thus the standard Gaussian process, is obtained as $\beta \to \infty$ or as $\gamma \to \infty$.

A flexible model for extremal dependence



Coefficients $\chi(u)$ (top) and $\overline{\chi}(u)$ (bottom), $u \in [0.9, 1]$, for our model (solid black) and for the Gaussian copula matching at $u = 0.95$ (dashed red). Parameter configurations are $\beta = 0, \gamma = 1$ (left), $\beta = 1, \gamma = 1$ (middle), $\beta = 5, \gamma = 1$ (right). Thin to thick curves correspond to increasing $\rho = 0, 0.3, 0.6, 0.9$ for our model.

Inference for threshold exceedances

Fit our copula model to extremes

- We want to use our copula model for threshold exceedances.
- Multivariate threshold exceedances: there is no unique definition. Here we define exceedances of the threshold $\boldsymbol{v} \in \mathbb{R}^D$ as observations $\boldsymbol{x}_i \in \mathbb{R}^D$ for which at least one component x_{ij} exceed v_j.
- We assume that in the joint tail region corresponding to large observed values, the multivariate distribution of our data is well described by a continuous joint distribution H with margins $H_1, \ldots, H_D$ and copula C stemming from a Gaussian scale mixture.
- We will use a two-step approach to deal with the marginals and dependence separately:
  (1) transform marginals to Unif$(0, 1)$, and
  (2) fit our copula model using a censored likelihood approach.

Step 1: marginal transformations

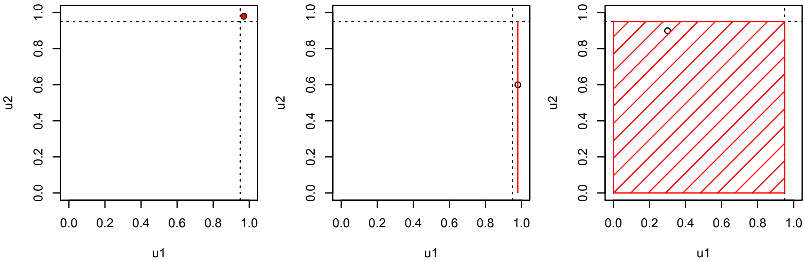
- Let $\boldsymbol{y}_i, \ldots, \boldsymbol{y}_n \in \mathbb{R}^D$ denote our observations.
- We estimate marginal distributions $H_1, \ldots, H_D$ using empirical distribution functions. Defining $\widehat{H}_k(y) = (n+1)^{-1} \sum_{i=1}^{n} I(y_{ki} \leq y)$ we transform the data to a pseudo-uniform scale as

$$u_{ki} = \widehat{H}_k(y_{ki}) = \frac{\text{rank}(y_{ki})}{n+1}, \qquad k = 1, \ldots, D,\ i = 1, \ldots, n,$$

  where $\text{rank}(y_{ki})$ is the rank of y_{ki} among $y_{k1}, \ldots, y_{kn}$.
- Since $\widehat{H}_k$ is a consistent estimator of H_k as $n \to \infty$, $\{u_{ki}\}$ form an approximate $\text{Unif}(0,1)$ sample for large n.

Step 2: censored likelihood for the copula

- We fit our copula model $C(\cdot;\boldsymbol{\psi})$ to the sample $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$.
- We don't want non-extreme values to influence the fit.
- Let $\boldsymbol{v} = (v_1, \ldots, v_D)$ denote a high threshold (typically $v_i = .95$) and $\boldsymbol{u}_i^\star = \max(\boldsymbol{u}_i, v_i)$ the censored observations.
- We then use the likelihood for the censored data $\boldsymbol{u}_i^\star$. Three distinct scenarios can occur:



Likelihood contributions depend on the number of components of $\boldsymbol{u}_i$ exceeding $\boldsymbol{v}$.

Properties of the pseudo-MLE

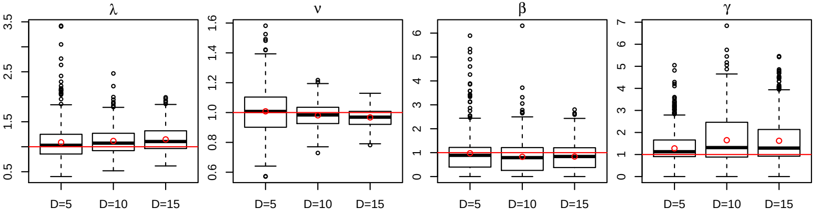
- The censored log likelihood is the sum of all individual contributions:

$$L(\boldsymbol{\psi}) = \sum_{i=1}^{n} L(\boldsymbol{u}_i^\star; \boldsymbol{\psi}).$$

- $\widehat{\boldsymbol{\psi}} = \operatorname{argmax} L(\boldsymbol{\psi})$ is a full likelihood estimator for the censored observations $\boldsymbol{u}_i^\star$: if the marginal estimation performed in Step 1 is perfect $\widehat{\boldsymbol{\psi}}$ obeys classical likelihood theory (under std. reg. cond.).
- Two subtilities with our model and the two-step approach:
  (1) The case $\beta = 0$ is nonstandard (boundary of parameter space).
  (2) The nonparametric rank transformation results in a slight bias for finite n (asymptotically unbiased). Genest et al. (1995) show that under mild conditions, the pseudo-MLE has similar asymptotic properties to the MLE, although with a slight loss in efficiency.

Simulation study I

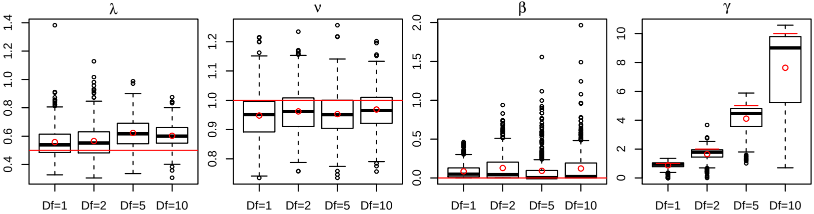
- We test our two-step pseudo-likelihood estimation procedure: (1) generate data from the RW model, (2) use ranks to transform to approx Unif$(0, 1)$, and (3) fit the copula model.



Boxplots of estimated parameters for our model with correlation function $\rho(\boldsymbol{s}_1, \boldsymbol{s}_2) = \exp\{-(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|/\lambda)^\nu\}$ and parameters $\lambda = \nu = \beta = \gamma = 1$. Simulations are based on $n = 1000$ independent replicates observed at $D = 5, 10, 15$ uniform locations in $[0, 1]^2$. Estimation uses the threshold $\boldsymbol{v} = (.95, \ldots, .95)$.

Simulation study II

- Misspecified model: generate data from a Student t process.
- Conclusion: Our model provides a good approximation to the tail.

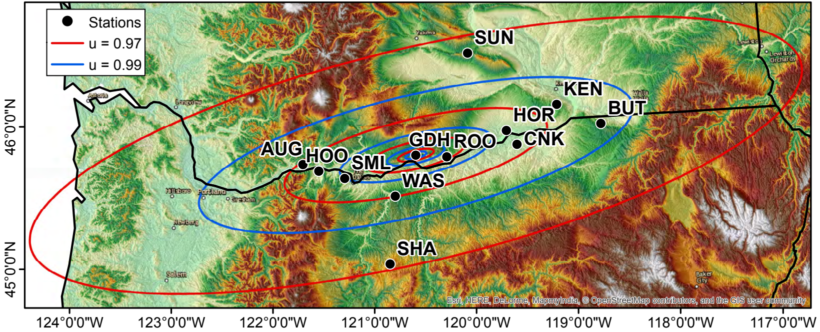


Boxplots of estimated parameters for our model when data are generated from a Student t process with correlation function $\rho(\boldsymbol{s}_1, \boldsymbol{s}_2) = \exp\{-(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|/\lambda)^{\nu}\}$ and parameters $\lambda = 0.5$, $\nu = 1$, and $\mathrm{Df} = 1, 2, 5, 10$. Simulations are based on $n = 1000$ independent replicates observed at $D = 15$ uniform locations in $[0, 1]^2$. Estimation uses the threshold $\boldsymbol{v} = (.95, \ldots, .95)$.

Application to hourly wind speed in the Pacific Northwest

Data

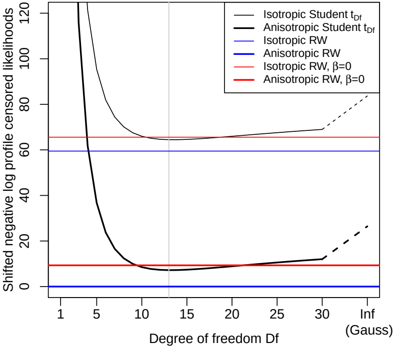
- Fit hourly wind speed extremes recorded at 12 sites during 2012–2014.
- Temporal nonstationarity: we focus on winter months (DJF).
- About 6504 hourly observations at each site ($\approx 8\%$ of values missing).
- Anisotropy: wind patterns are mainly characterized by easterly and westerly winds.

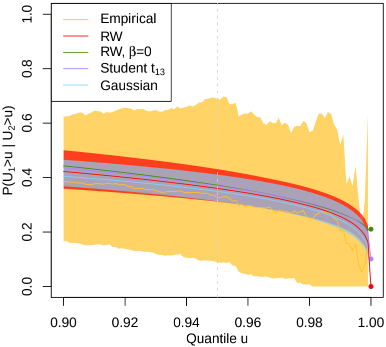


Locations of the 12 sites.

Results

- We ignore temporal dependence in the estimation but account for it when calculating standard errors (block bootstrap).
- We compare different copula models. Isotropic and anisotropic. Gaussian, t, and our new model.





Left: Log-likelihood differences for isotropic and anisotropic models; baseline is our Gaussian scale mixture. Right: Fitted conditional exceedance probabilities.

Conclusion

Conclusion

- Summary:
  - Sub-asymptotic models for extremes.
  - We have a flexible copula model that can link AI and AD.
  - The censored likelihood approach is appropriate for extremes.
- Extensions/limitations:
  - Computation is slow.
  - Bayesian?

Reference: Huser, Opitz & Thibaud (2016) Bridging Asymptotic Independence and Dependence in Spatial Extremes Using Gaussian Scale Mixtures. arXiv:1610.04536.

Conclusion

- Summary:
  - Sub-asymptotic models for extremes.
  - We have a flexible copula model that can link AI and AD.
  - The censored likelihood approach is appropriate for extremes.
- Extensions/limitations:
  - Computation is slow.
  - Bayesian?

Reference: Huser, Opitz & Thibaud (2016) Bridging Asymptotic Independence and Dependence in Spatial Extremes Using Gaussian Scale Mixtures. arXiv:1610.04536.

Thanks for your attention!