Modelling and inference for the body and tail regions of multivariate data

Lídia André

Jenny Wadsworth

16th April 2025

Outline

Introduction to the problem

1. Develop dependence models able to represent the non-extremal (i.e., the body) and extremal (i.e., the tail) regions accurately
   - Bypasses the need to define an extremal region
   - Important to model regions where only a subset of variables is extreme
2. Inference of complex (computationally expensive) dependence copula models able to
   - characterise the body and tail regions
   - interpolate between asymptotic dependence (AD) and independence (AI) at an interior point

Univariate framework

There have been proposed parametric, semi-parametric and non-parametric models

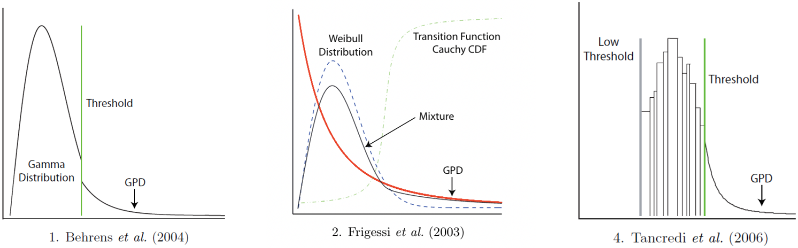


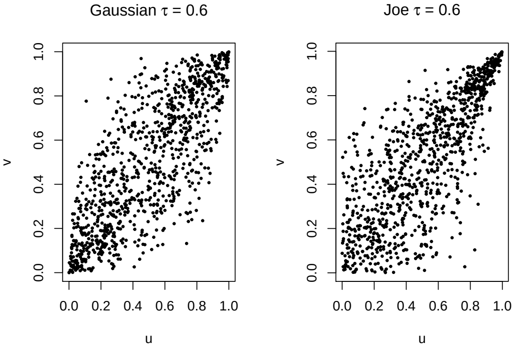
Figure 1: Taken from Scarrott and MacDonald (2012)

Copulas

In a multivariate setting we are also concerned about the dependence between variables.

A copula C is a joint distribution of a random vector $(X_1, \ldots, X_d)$

$$F(x_1, \ldots, x_d) = C\left(F_{X_1}(x_1), \ldots, F_{X_d}(x_d)\right), \quad d \geq 2$$



Extremal dependence properties

It is important to know if extreme values of the variables are likely to occur together (**asymptotic dependence**) or not (**asymptotic independence**)

$$\chi = \lim_{r\to 1} P\left[F_Y(y) > r \mid F_X(x) > r\right],$$

$$P\left[F_Y(y) > r \mid F_X(x) > r\right] \sim \mathcal{L}(1-r)(1-r)^{\frac{1}{\eta}-1} \qquad \text{as } r \to 1$$

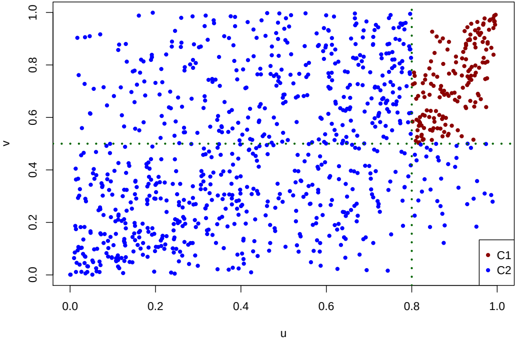
- Asymptotic Dependence (AD): $\chi > 0$ and $\eta = 1$
- Asymptotic Independence (AI): $\chi = 0$ and $\eta \neq 1$

Joint modelling of the bulk and tail of bivariate data

Joint work with Jenny Wadsworth and Adrian O'Hagan

Multivariate framework: Aulbach et al. (2012)

- Fit one copula to the body and another to the upper tail
- Sometimes does not offer a smooth transition between the two copulas
- Requires the choice of thresholds



Weighted copula model (WCM)

For $(u^*, v^*) \in [0,1]^2$, we define the density c^* as

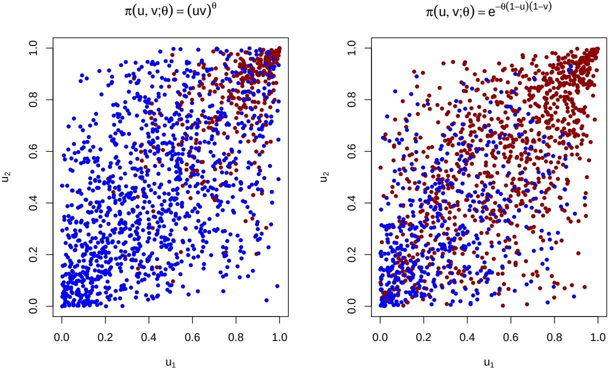
$$c^*(u^*, v^*; \boldsymbol{\gamma}) = \frac{\pi(u^*, v^*; \theta) c_t(u^*, v^*; \boldsymbol{\alpha}) + [1 - \pi(u^*, v^*; \theta)] c_b(u^*, v^*; \boldsymbol{\beta})}{K(\boldsymbol{\gamma})}$$

- $c_t, \ c_b$: copula densities tailored to the tail and body, respectively
- $\pi(u^*, v^*; \theta)$: dynamic weighting function, defined in $[0,1]^2$ and increasing in u^* and v^*
- $\boldsymbol{\gamma} = (\theta, \boldsymbol{\alpha}, \boldsymbol{\beta})$: vector of model parameters
- $K(\boldsymbol{\gamma})$: normalising constant [1]

[1] For more details see André et al. (2024)

Weighted copula model (WCM)

- Does not require a choice of threshold
- Offers a smooth transition between the body and tail copulas
- However, it is hard to perform inference on it



Inference

- Fit the copula of the density c^*

$$c(u,v;\boldsymbol{\gamma}) = \frac{c^*\left(F_{U^*}^{-1}(u), F_{V^*}^{-1}(v); \boldsymbol{\gamma}\right)}{f_{U^*}\left(F_{U^*}^{-1}(u)\right) f_{V^*}\left(F_{V^*}^{-1}(v)\right)}$$

with

$$F_{U^*}(u^*) = P[U^* \leq u^*] = \int_0^{u^*}\int_0^1 c^*(u,v)\mathrm{d}v\mathrm{d}u$$

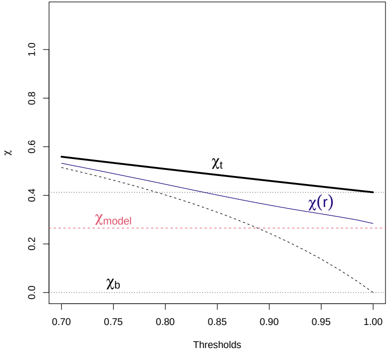
$$f_{U^*}(u^*) = \int_0^1 c^*(u^*,v)\ \mathrm{d}v, \qquad v \in (0,1)$$

Extremal dependence properties

Depending on the weighting function used, c_b has an influence in χ in some cases:

- If $\pi(u^*, v^*; \theta) = (u^* v^*)^\theta$ and c_t is AD, χ is dominated by χ_t with an influence of χ_b
- If $\pi(u^*, v^*; \theta) = (u^* v^*)^\theta$ and c_t is AI, χ is that from c_t
- If $\pi(u^*, v^*; \theta) = \exp\{-\theta(1 - u^*)(1 - v^*)\}$, χ is that from c_t (independently of the nature of c_t)

Extremal dependence properties



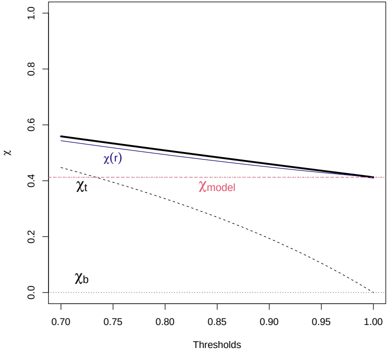


Figure 2: Weighting functions: $\pi(u^*, v^*; \theta) = (u^* v^*)^\theta$ (left) and $\pi(u^*, v^*; \theta) = \exp\{-\theta(1 - u^*)(1 - v^*)\}$ (right) with $\gamma = (1.5, 2, 3.488889)$

Extremal dependence properties

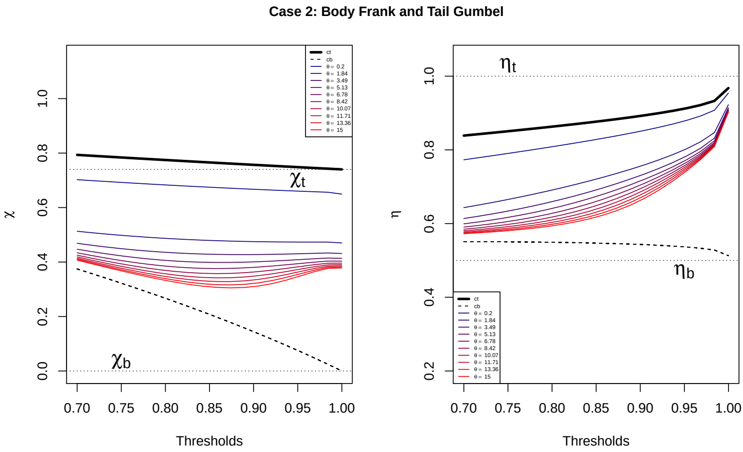


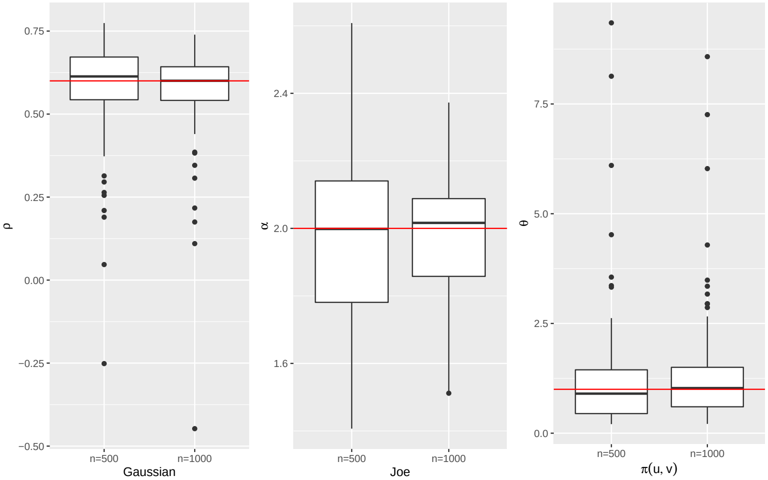
Figure 3: Weighting function: $\pi(u^*, v^*; \theta) = (u^* v^*)^{\theta}$.

Parameter estimation

Simulation setup:

- c_t : Gaussian copula with $\rho = 0.6$
- c_b : Joe copula with $\alpha = 2$
- $\pi(u^*, v^*; \theta) = (u^* v^*)^\theta$ with $\theta = 1$
- $n = 500$ and $n = 1000$
- 100 repetitions

Parameter estimation

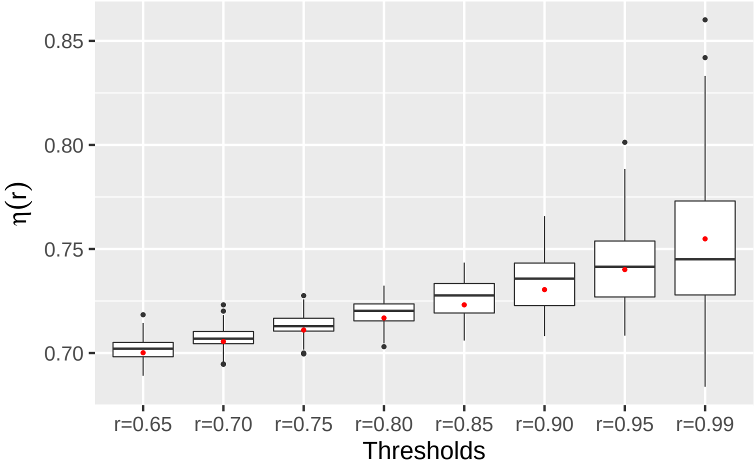


Model misspecification

Simulation setup:

- True data from a Gaussian copula with $\rho = 0.65$
- Models considered:
  1. c_t : Joe copula; c_b : Frank copula
  2. c_t : Hüsler-Reiss copula; c_b : Clayton copula
  3. c_t : **Inverted Gumbel copula;** c_b : **Student t copula** $\rightarrow$ best average AIC
  4. c_t : Coles-Tawn copula; c_b : Galambos copula
- $\pi(u^*, v^*; \theta) = (u^* v^*)^\theta$
- $n = 1000$
- Each model was fitted 50 times

Model misspecification



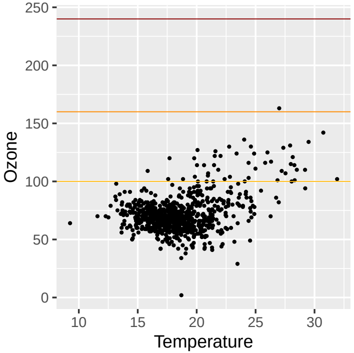
Case study: Ozone and Temperature

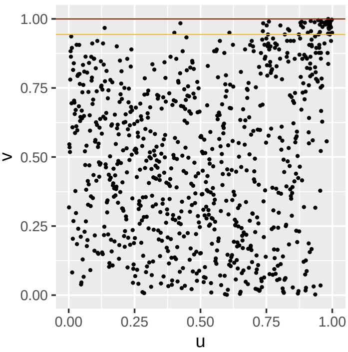
- Temperature may influence the levels of Ozone concentration in the air
- The legal thresholds for O_3 levels in the UK might then be found in the body and not just in the tails of the data
- Summers between 2011 and 2019 of Blackpool, UK

Table 1: UK legal thresholds

Levels	Low	Moderate	High	Very High
$O_3\,(\mu g/m^3)$	$[0, 100]$	$[101, 160]$	$[161, 240]$	> 240

Case study: Ozone and Temperature

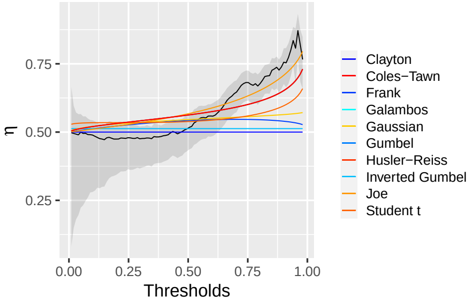




Case study: Ozone and Temperature

Table 2: Single copula fit

Copula	AIC
Clayton	2.0
Gaussian	-28.6
Frank	-15.8
Joe	**-143.6**
Gumbel	-97.4
Student t	-52.8
Inverted Gumbel	0.1
Hüsler-Reiss	-99.1
Coles-Tawn	-99.0
Galambos	-95.9

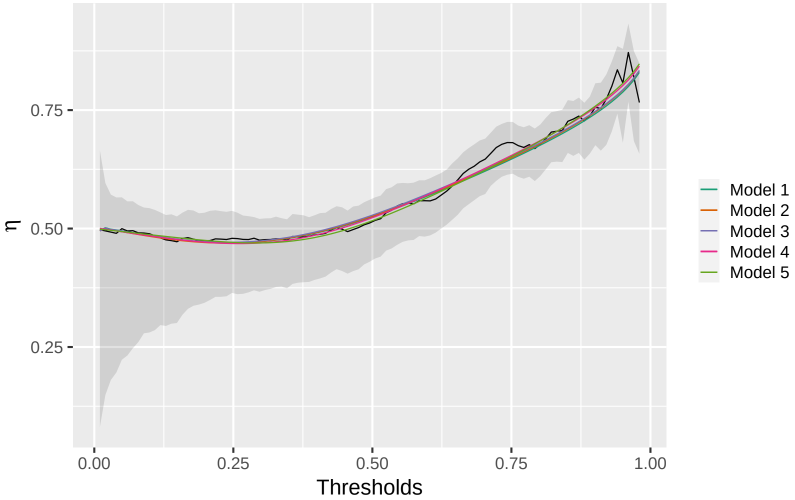


Case study: Ozone and Temperature

Table 3: WCM with $\pi(u^*, v^*; \theta) = \exp\{-\theta(1-u^*)(1-v^*)\}$

Model		Parameters			AIC
c_b	c_t	$\hat{\beta}$	$\hat{\boldsymbol{\alpha}}$	$\hat{\theta}$	
Gaussian	Hüsler-Reiss	-0.74	1.33	3.32	**-240.1**
Gaussian	Galambos	-0.72	0.90	3.55	-237.2
Gaussian	Coles-Tawn	-0.74	0.85, 0.79	3.25	-234.8
Frank	Coles-Tawn	-4.51	0.87, 1.02	4.33	-235.7
Frank	Joe	-6.49	1.72	2.45	-232.9

Case study: Ozone and Temperature



Case study: Ozone and Temperature

Table 4: Other diagnostics

Models	Kendall's τ	$P[T \geq 24, O_3 \geq 100]$	$P[O_3 \geq 100 \mid 22 \leq T \leq 23]$
Empirical	0.0812	0.0302	0.1330
(95% CI)	(0.0173 , 0.1867)	(0.0147 , 0.0544)	(0.0227 , 0.1944)
Model 1	0.0690	0.0246	0.1441
Model 2	0.0663	0.0250	0.1412
Model 3	0.0770	0.0251	0.1429
Model 4	0.0779	0.0262	0.1392
Model 5	0.0718	0.0267	0.1366

Conclusion: Advantages

- No need to define an extremal region
- Smooth transition between the two regions
- Weighting function determines the influence of each copula in $[0,1]^2$
- Better representation of both regions than just considering a simple single copula model
- Flexible under misspecified scenarios
- Suitable for AI and AD

Conclusion: Limitations

- Choice of copula families a priori
- No general derivations of the extremal dependence properties
- Computationally expensive to fit — infeasible beyond $d = 2$
- Defined only for stationary settings

Gaussian mixture copulas for flexible dependence modelling in the body and tails of joint distributions

Joint work with Jon Tawn

Gaussian mixture copula (GMC): Model and inference

- Similarly to the WCM, we fit the copula of $\boldsymbol{Y} := (Y_1, \ldots, Y_d)$ where

$$\boldsymbol{Y} = \boldsymbol{Z}_j \sim \mathrm{MVN}(\boldsymbol{\mu}_j, \Sigma_j)$$

with probability p_j for $j = 1 \ldots, k$.

  - $\boldsymbol{Z}_j = (Z_j^1, Z_j^2, \ldots, Z_j^d)$
  - $\boldsymbol{\mu}_j = (\mu_j^1, \ldots, \mu_j^d)$
  - $\Sigma_j = \begin{pmatrix} \sigma_{1j}^2 & \ldots & \rho_j^{1,d}\sigma_{1j}\sigma_{dj} \\ \rho_j^{1,2}\sigma_{1j}\sigma_{2j} & \ldots & \rho_j^{2,d}\sigma_{2j}\sigma_{dj} \\ \vdots & \ddots & \vdots \\ \rho_j^{1,d}\sigma_{1j}\sigma_{dj} & \ldots & \sigma_{dj}^2 \end{pmatrix}$

Gaussian mixture copula (GMC)

- Does not require a choice of threshold
- No need to define a priori which copulas/distributions to include
- Scales relatively well to dimensions beyond $d = 2$

Model fit and diagnostics

- We have AI and $\chi = 0$ due to the model specification
- Assess the performance of the GCM in data with different dependence structures
  1. Underlying data is AD
  2. Underlying data is generated from a WCM specification

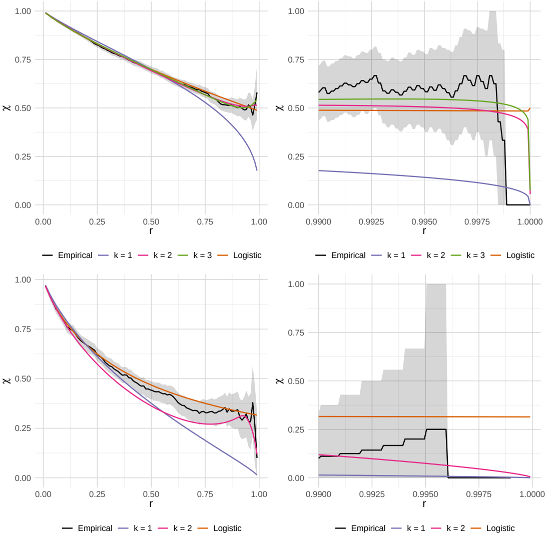
Asymptotically dependent data

- Data from a Logistic copula with $\alpha_L = 0.6$ and $d = \{2, 5\}$
- GMC specifications
  - When $d = 2$, $n = 5000$ and $k = 1, 2, 3$ mixture components
  - When $d = 5$, $n = 1000$ and $k = 1, 2$ mixture components

Table 5: Change in AIC values obtained for the Gaussian mixture copula for $k > 1$ relative to when $k = 1$ for $d = \{2, 5\}$.

Dimension	$\text{AIC}_{k_1 - k_2}$	$\text{AIC}_{k_1 - k_3}$
$d = 2$	-219.28	-226.18
$d = 5$	-148.69	—

Asymptotically dependent data



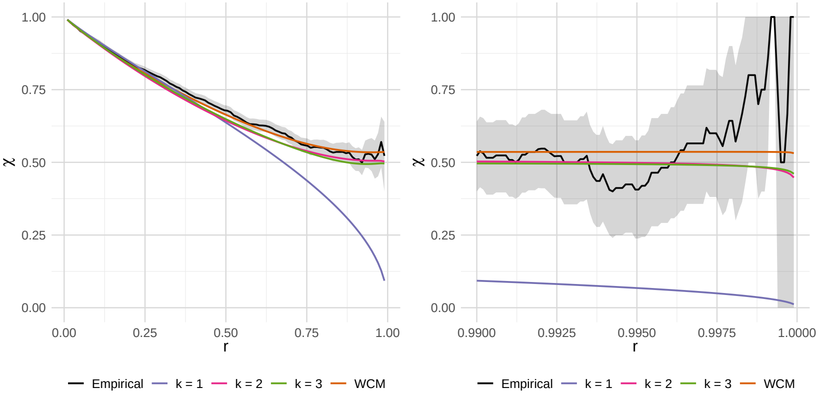
Data from WCM

- $d = 2$
- $n = 5000$ observations generated from a WCM with
    - c_t : Logistic copula with $\alpha_L = 0.3$
    - c_b : Frank copula with $\alpha_F = 2$
    - Weighting function: $\pi(u^*, v^*; \theta) = (u^* v^*)^\theta$ with $\theta = 1.5$
- GMC with $k = 1, 2, 3$ mixture components

Table 6: Change in AIC values obtained for the Gaussian mixture copula for $k > 1$ relative to when $k = 1$ for $d = 2$.

Dimension	$\text{AIC}_{k_1 - k_2}$	$\text{AIC}_{k_1 - k_3}$
$d = 2$	-974.27	-1017.67

Data from WCM



Case study: air pollution data

- Air pollution data set analysed by Heffernan and Tawn (2004)
- Daily maxima of the hourly means of ground level measurements of O_3, NO_2, NO, SO_2 and PM_{10} recorded at Leeds, UK, from 1994 to 1998.
- Winter data
- Pairwise ($d = 2$) and trivariate ($d = 3$) analyses

Case study: air pollution data

Pairwise analysis

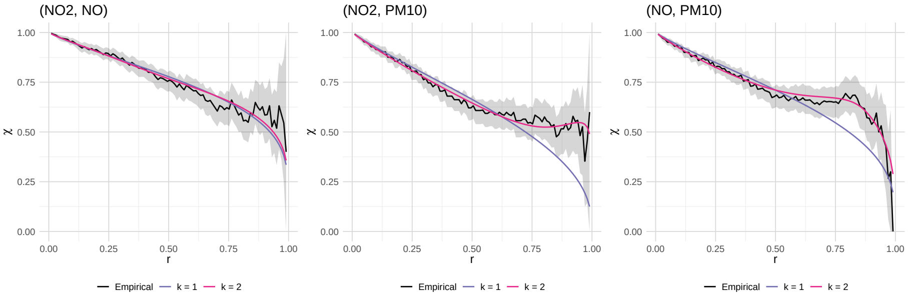
- Pairs (NO_2, NO), (NO_2, PM_{10}) and (NO, PM_{10})
- GMC with $k = 1, 2$ mixture components.

Table 7: Change in AIC values obtained for the Gaussian mixture copula for $k = 2$ relative to when $k = 1$. The estimated mixing probabilities $(\hat{p}_1, \hat{p}_2)$ are reported for the $k = 2$ model.

Pair	$\text{AIC}_{k_1 - k_2}$	$(\hat{p}_1, \hat{p}_2)$
(NO_2, NO)	4.01	(0.37, 0.63)
(NO_2, PM_{10})	-34.40	(0.91, 0.09)
(NO, PM_{10})	-50.87	(0.78, 0.22)

Case study: air pollution data

Pairwise analysis



Case study: air pollution data

Trivariate analysis

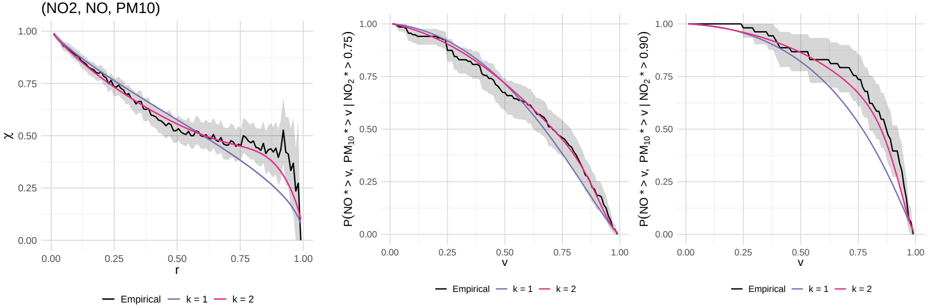
- Triple (NO_2, NO, PM_{10})
- GMC with $k = 1, 2$ mixture components.

Table 8: Change in AIC values obtained for the Gaussian mixture copula for $k = 2$ relative to when $k = 1$. The estimated mixing probabilities $(\hat{p}_1, \hat{p}_2)$ are reported for the $k = 2$ model.

Triple	$\text{AIC}_{k_1-k_2}$	$(\hat{p}_1, \hat{p}_2)$
(NO_2, NO, PM_{10})	-61.22	(0.73, 0.21)

Case study: air pollution data

Trivariate analysis



Conclusion: Advantages

- No need to define an extremal region
- No need to choose copula families/distributions a priori
- Scales relatively well to $d > 2$
- Suitable for data which are non-exchangeable or exhibit a more complex dependence structure
- Suitable for AI and AD (sub-asymptotically)

Conclusion: Limitations

- Choice of number of mixture components
- Exhibits AI in the limit due to the model specification
- No general derivations of the extremal dependence properties
- Computationally expensive to fit — highly parameterised
- Defined only for stationary settings

Neural Bayes inference for complex bivariate extremal dependence models

Joint work with Jenny Wadsworth and Raphaël Huser

Likelihood inference

- Requires the knowledge of a likelihood function
- Might be computationally costly when there is
  - inversion of functions;
  - numerical integration;
- Examples:
  - WCM (André et al., 2024)
  - Models that are available to **interpolate** between AD and AI (e.g. Wadsworth et al., 2017)

Point estimation

- General setting:
  - Replicate data: $\boldsymbol{Z} := (\boldsymbol{Z}_1', \dots, \boldsymbol{Z}_n')' \in \mathcal{S}^n$ where $\boldsymbol{Z}_i \sim f(\boldsymbol{z}_i \mid \boldsymbol{\theta})$
  - Sampling space: $\mathcal{S} = \mathbb{R}^d$
  - Parameter space: $\Theta = \mathbb{R}^p$
- Point estimators: $\hat{\boldsymbol{\theta}} : \mathcal{S}^n \to \Theta$
- Bayes estimators: minimise a weighted average of the risk at $\boldsymbol{\theta}$ (Bayes risk)

$$r_\Omega(\hat{\boldsymbol{\theta}}(\cdot)) = \int_\Theta \int_{\mathcal{S}^n} L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(\boldsymbol{z})) f(\boldsymbol{z} \mid \boldsymbol{\theta}) \mathrm{d}\boldsymbol{z} \mathrm{d}\Omega(\boldsymbol{\theta})$$

  - $\Omega(\cdot)$: prior measure for $\boldsymbol{\theta}$
  - $L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(\boldsymbol{z}))$: absolute error loss

Neural Bayes estimators: Sainsbury-Dale et al. (2024)

- Bayes estimator that is approximated using a **neural network** as function approximator
- Neural point estimator: $\hat{\boldsymbol{\theta}}(\boldsymbol{Z} \mid \boldsymbol{\gamma})$
  - $\boldsymbol{\gamma}$: parameters of the neural network
- Neural Bayes estimator (NBE): $\hat{\boldsymbol{\theta}}(\boldsymbol{Z} \mid \boldsymbol{\gamma}^*)$

$$\boldsymbol{\gamma}^* = \underset{\boldsymbol{\gamma}}{\arg\min}\, r_\Omega(\hat{\boldsymbol{\theta}}(\cdot; \boldsymbol{\gamma}))$$

- NBEs just need to be trained **once**!
  - subsequent estimates are obtained in (milli)seconds

Neural Network architecture

- For any permutation $\tilde{\boldsymbol{Z}}$ of the independent replicates in $\boldsymbol{Z}$:

$$\hat{\boldsymbol{\theta}}(\boldsymbol{Z};\gamma) = \hat{\boldsymbol{\theta}}(\tilde{\boldsymbol{Z}};\gamma)$$

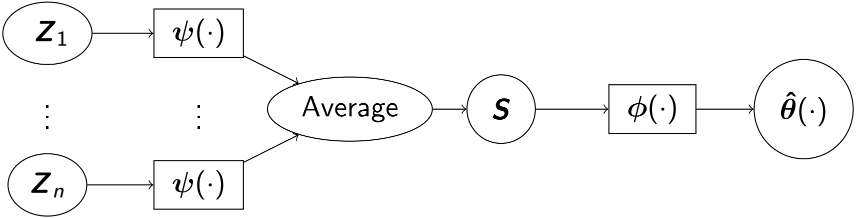
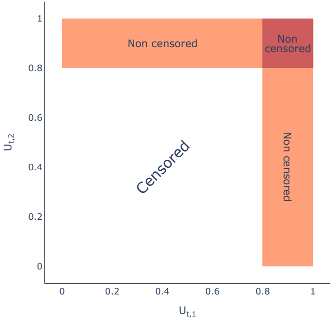


Figure 4: Schematic of the DeepSets architecture (Zaheer et al., 2017). $\boldsymbol{\psi} : \mathbb{R}^d \to \mathbb{R}^q$ and $\boldsymbol{\phi} : \mathbb{R}^q \to \mathbb{R}^p$ are neural networks, and $\boldsymbol{S}$ are summary statistics.

NBEs for censored data: Richards et al. (2024)

Censor non-extreme values to prevent them affecting the extremal dependence estimation

- $\boldsymbol{Z}^* = ((\boldsymbol{Z}_1^*)', \ldots, (\boldsymbol{Z}_n^*)')'$
- Censored values set to $c \in \mathbb{R}$ outside the support
- $\boldsymbol{I}_i$: indicator vectors
  - if 1 then the observations are censored



NBEs for censored data: Richards et al. (2024)

- NBEs are trained using an augmented data set $\boldsymbol{A} = ((\boldsymbol{Z}^*)', \boldsymbol{I}')$
- Censoring level τ is treated as **variable**

$$\hat{\theta}(\boldsymbol{A}; \tau, \gamma) = \phi\left(\boldsymbol{S}(\boldsymbol{A}; \gamma_{\boldsymbol{\psi}}, \tau); \gamma_{\boldsymbol{\phi}}\right)$$

with $\boldsymbol{S}(\boldsymbol{A}; \gamma_{\boldsymbol{\psi}}, \tau) = \left(\boldsymbol{S}(\boldsymbol{A}; \gamma_{\boldsymbol{\psi}})', \tau\right)'$ and $\boldsymbol{S}(\boldsymbol{A}; \gamma_{\boldsymbol{\psi}})$ defined as before

Parameter estimation: WCM

- c_t : Logistic copula with $\alpha_L \in (0, 1)$
- c_b : Gaussian copula with $\rho \in (-1, 1)$
- Weighting function: $\pi(u^*, v^*; \gamma) = (u^* v^*)^\gamma$ with $\gamma > 0$

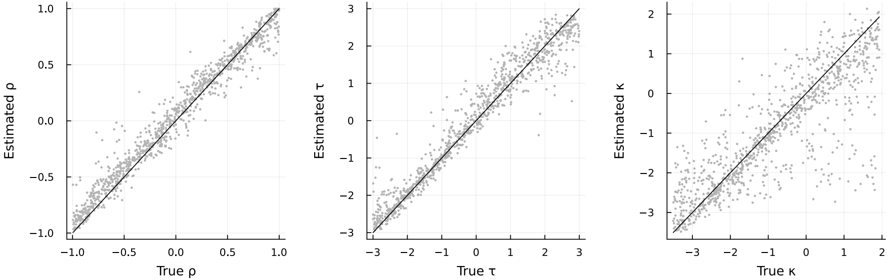
Reparameterisation: $\tau_L = \text{logit}(\alpha_L)$ and $\kappa = \log(\gamma)$

Parameter estimation: Priors

- $\tau_L \sim \mathrm{Unif}(-3, 3)$, which results in $\alpha_L \in (0.05, 0.95)$
- $\rho \sim \mathrm{Unif}(-1, 1)$
- $\kappa \sim \mathrm{Unif}(-3.51, 1.95)$, which results in $\gamma \in (-0.03, 7.03)$
- $N \sim \mathrm{Unif}(\{100, 101, \ldots, 1500\})$

Sample size N is treated as a random variable.

Assessment of NBEs



Assessment of NBEs: Uncertainty quantification

- Non-parametric bootstrap procedure:
  - $B = 400$ bootstrap samples
  - $\boldsymbol{\theta}$ is re-estimated
  - 95% confidence intervals are obtained

Assessment of NBEs: Uncertainty quantification

Table 9: Coverage probability and average length of the 95% uncertainty intervals for the parameters obtained via a non-parametric bootstrap procedure averaged over 1000 models fitted using a NBE (rounded to 2 decimal places).

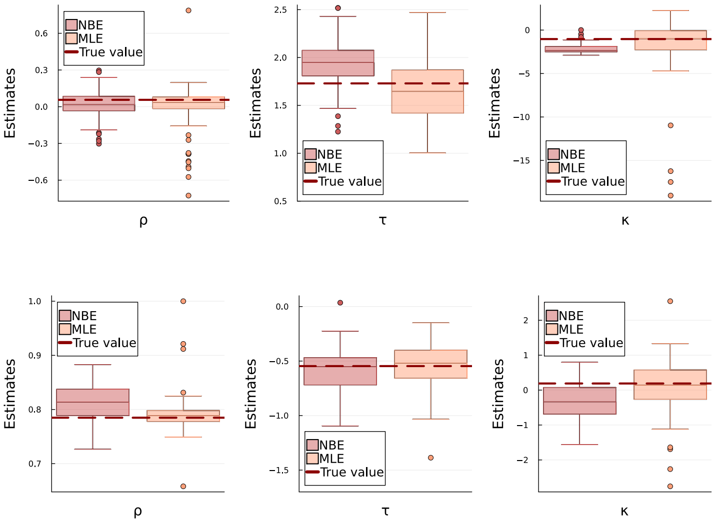
Parameter	Coverage	Length
ρ	0.71	0.26
τ_L	0.75	0.76
κ	0.69	1.33

Assessment of NBEs: Extremal dependence structure

Table 10: Coverage probability and average length of the 95% uncertainty intervals for $\chi(u)$ at levels $u = \{0.50, 0.80, 0.95\}$ obtained via a non-parametric bootstrap procedure averaged over 1000 models fitted using a NBE (rounded to 2 decimal places).

$\chi(u)$	Coverage	Length
$\chi(0.50)$	0.77	0.05
$\chi(0.80)$	0.79	0.08
$\chi(0.95)$	0.78	0.09

Comparison with MLE



Comparison with MLE

- Once trained, getting an estimate through this NBE takes on average 0.653 seconds.
- An estimate through MLE takes on average 3h and 12 minutes.
- This is a $17,663$ fold speed-up

Model selection: neural Bayes classifier (NBC)

- Information criteria like AIC/BIC cannot be used
- **Solution:** Treat model type as a random variable M
- $K \geq 2$ candidate models and M takes values in $\{1, \dots, K\}$
- M is inferred jointly with $\boldsymbol{\theta}$ (based on $\boldsymbol{Z}$): $(\boldsymbol{\theta}', M)' \mid \boldsymbol{Z}$
- Can be decomposed as the product of $\boldsymbol{\theta} \mid (\boldsymbol{Z}', M)'$ and $M \mid \boldsymbol{Z}$
- $\boldsymbol{\theta} \mid (\boldsymbol{Z}', M)'$ is split into m problems: $\boldsymbol{\theta}_m \mid (\boldsymbol{Z}', M = m)' -$ trained with NBEs

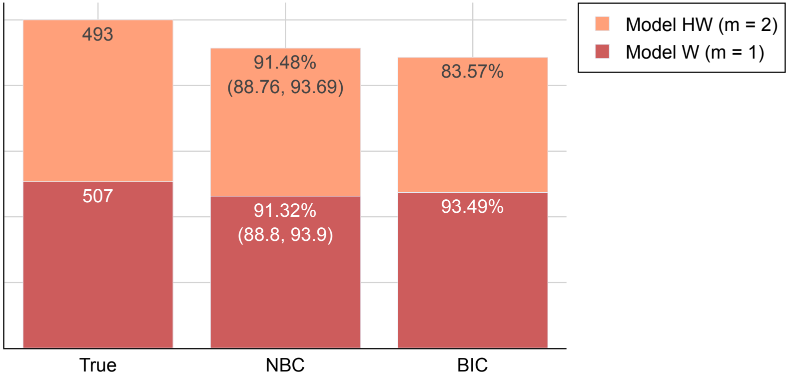
Model selection: neural Bayes classifier (NBC)

- Construct a neural network that approximates $M \mid \boldsymbol{Z} = \boldsymbol{z}$ for any data input $\boldsymbol{Z} = z$
- Neural Bayes classifier (NBC): $\hat{\boldsymbol{p}}(\boldsymbol{Z}; \gamma)$

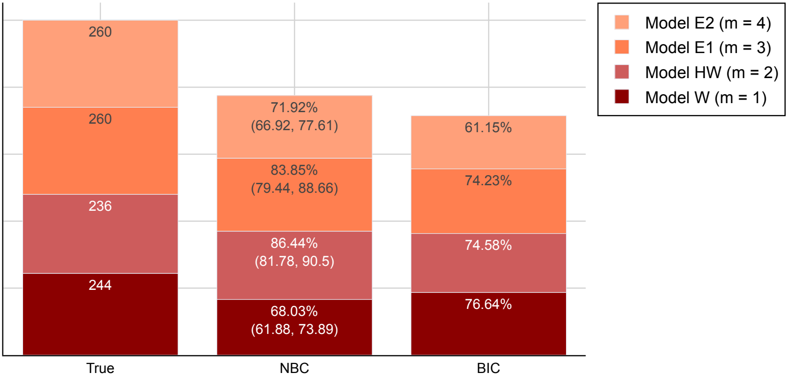
$$\boldsymbol{\gamma}^* = \underset{\boldsymbol{\gamma}}{\arg\min} \; - \sum_{m=1}^{K} p_m \int_{\Omega_m} \int_{\mathcal{S}_m^n} \log\left(\hat{p}_m(\boldsymbol{z}; \boldsymbol{\gamma})\right) f_m\left(\boldsymbol{z} \mid \boldsymbol{\theta}_m\right) \mathrm{d}\boldsymbol{z} \mathrm{d}\Omega_m(\boldsymbol{\theta}_m)$$

  - $p_m = \Pr(M = m) = 1/K$, and $\sum_{m=1}^{K} p_m = 1$
  - $\hat{p}_m(\boldsymbol{z}; \boldsymbol{\gamma})$: approximate posterior probability of model m
- Identical to a classification problem
- Loss function: categorical cross-entropy

$K = 2$ candidate models



$K = 4$ candidate models



Misspecified scenarios

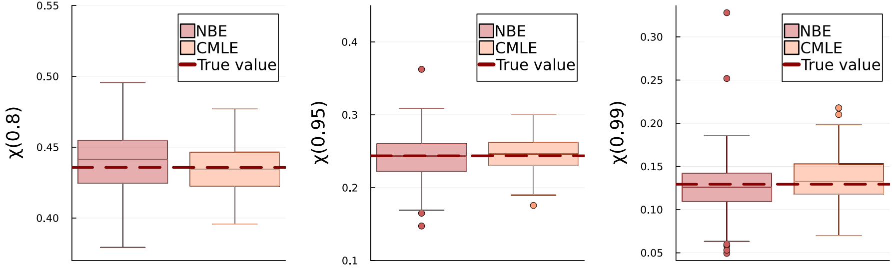
- Data from a Gaussian copula with $\rho = 0.5$ (AI) and $\tau = 0.65$
- 100 samples each with $n = 1000$

Table 11: Proportion of times each model was selected through the NBC and through BIC (left), and proportion of AD and AI samples identified by the NBE and CMLE (right). All the values are rounded up to 2 decimal places.

Model	NBC	BIC
Model W	0.02	0.30
Model HW	**0.88**	**0.69**
Model E1	0.02	0.00
Model E2	0.08	0.01

Method	AD	AI
NBE	0.02	0.98
CMLE	0.03	0.97

Misspecified scenarios



Case study: changes in geomagnetic field fluctuations

- Space weather events cause large fluctuations in the geomagnetic field - geomagnetically induced currents (GICs)
- GICs can cause: disruptions on power grids, railway systems, etc
- **Interest:** assess whether a large magnitude of GICs occurring in one location has an effect on another location
- Pairwise χ of the rate of change of the horizontal component of geomagnetic field $\mathrm{d}B_H/\mathrm{t}$ as a measure of magnitude of GICs

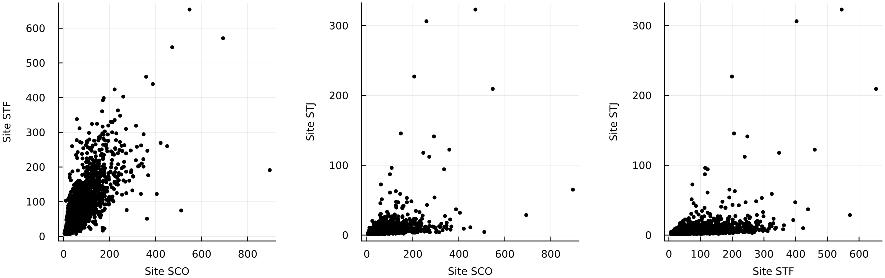
Case study: changes in geomagnetic field fluctuations

- $n = 1500$ and $\tau \in \{0.60, 0.65, \dots, 0.95\}-$ results for $\tau = 0.85$
- Pairs: (SCO, STF), (SCO, STJ) and (STF, STJ)

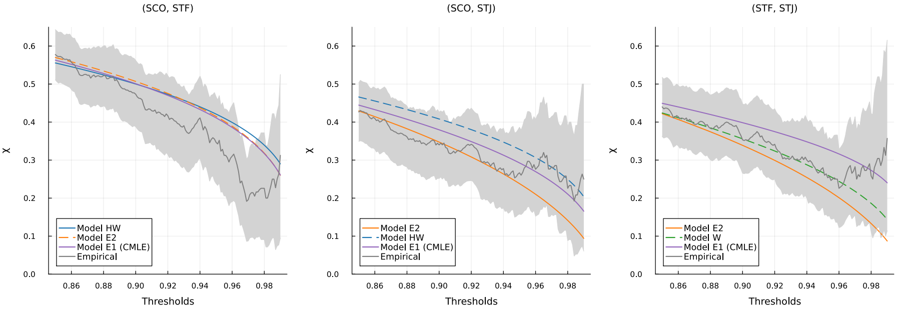
Table 12: International Association of Geomagnetism and Aeronomy (IAGA) code, and location of the observatory for the three locations considered.

IAGA code	Country	Latitude	Longitude
SCO	Greenland	70.48	-21.97
STF	Greenland	67.02	-50.72
STJ	Canada	47.60	-52.68

Case study: changes in geomagnetic field fluctuations



Case study: changes in geomagnetic field fluctuations



Conclusion: Advantages

- Robust and amortised statistical toolbox
- Fast inference method
- Well calibrated extremal dependence properties
- Sensitivity analysis for censoring level

Conclusion: Limitations

- Biased results
- Poor coverage of bootstrap-based uncertainty results
- Subjectivity in the neural network architecture
- Need to choose prior distributions

Thank you all for listening ☺

Questions?

lidia.andre@unamur.be

WCM **GMC** **NBE(C)**

References I

André, L., Wadsworth, J., and O'Hagan, A. (2024). Joint modelling of the body and tail of bivariate data. *Computational Statistics and Data Analysis*, 189:107841.

Aulbach, S., Bayer, V., and Falk, M. (2012). A multivariate piecing-together approach with an application to operational loss data. *Bernoulli*, 18(2):455–475.

Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B*, 66:497–546.

Richards, J., Sainsbury-Dale, M., Zammit-Mangion, A., and Huser, R. (2024). Neural Bayes estimators for censored inference with peaks-over-threshold models. *Journal of Machine Learning Research*, (to appear).

References II

Sainsbury-Dale, M., Zammit-Mangion, A., and Huser, R. (2024). Likelihood-free parameter estimation with neural Bayes estimators. *The American Statistician*, 78(1):1–14.

Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT-Statistical Journal*, 10(1):33–60.

Wadsworth, J. L., Tawn, J. A., Davison, A. C., and Elton, D. M. (2017). Modelling across extremal dependence classes. *Journal of the Royal Statistical Society: Series B*, 79:149–175.

Zaheer, M., Kottur, S., Ravanbhakhsh, S., Póczos, B., Salakhutdinov, R., and Smola, A. J. (2017). Deep Sets. *Advances in Neural Information Processing Systems*, 30.