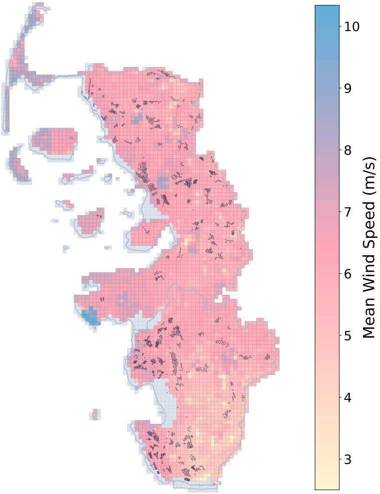
Linear factor models for tail dependence in high dimensions with applications to wind turbine cut-in risk

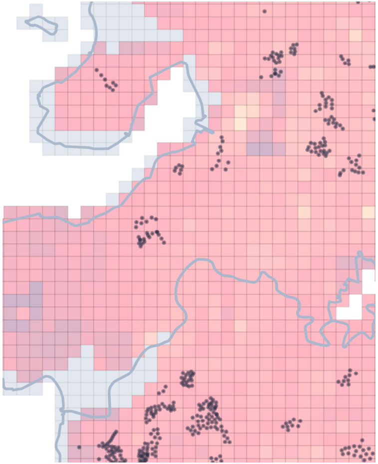
Alexis Boulin - Ruhr University Bochum
Axel Bücher - Ruhr University Bochum

EXALT workshop 2026
May 27, 2026

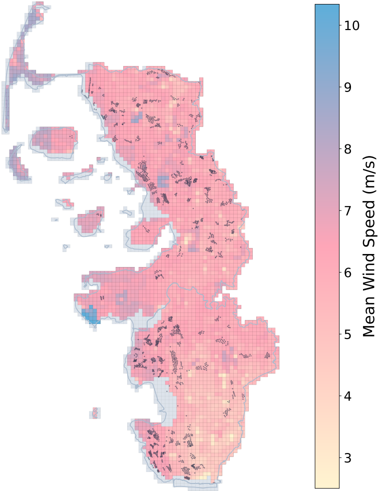
Flaute: low-wind events for wind energy production I

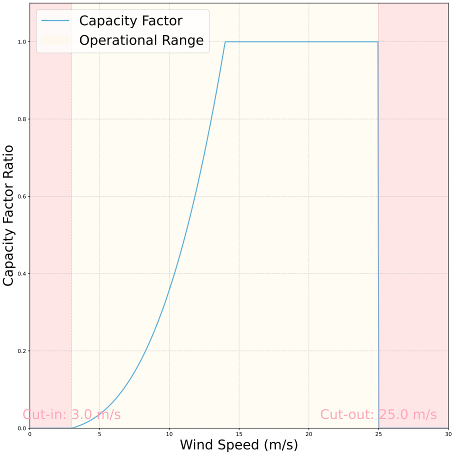


Flaute: low-wind events for wind energy production I



Flaute: low-wind events for wind energy production I





Formal description:

- In 2025: 1917 wind turbines located in $d = 563$ grid cells.
- C_j total installed capacity (2025) in grid cell j.
- W_j maximal daily wind speed at turbine height in grid cell j.
- Observations for $n = 2708$ winter days over the past 30 years, collected from HOSTRADA

Formal description:

- In 2025: 1917 wind turbines located in $d = 563$ grid cells.
- C_j total installed capacity (2025) in grid cell j.
- W_j maximal daily wind speed at turbine height in grid cell j.
- Observations for $n = 2708$ winter days over the past 30 years, collected from HOSTRADA

Target quantity:

$$p_{0.8}(3) = \mathbb{P}\Big(\exists J \subset [d] \text{ with installed capacity in } J \text{ exceeding } 80\% \text{ of the total capacity: } W_j \leq 3 \text{ for all } j \in J\Big).$$

Outline

Tail dependence

- Margin-free measures of tail dependence

The linear factor model with heavy tailed factors

- A (semi-)parametric tail dependence model
- The pure variable assumption and identfiability
- Estimation

Application: the probability of low-wind events

- Estimation of $p_{0.8}(3)$ and beyond.

Tail dependence

Simple bivariate measures of tail dependence

Throughout: $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ is a d-variate observable rv with continuous marginal cdfs $F_1, \ldots, F_d$.

The **tail correlation** between X_j and X_ℓ is

$$\chi(j,\ell) = \lim_{t \downarrow 0} t^{-1} \mathbb{P}\Big(F_j(X_j) > 1 - t, F_\ell(X_\ell) > 1 - t\Big), \qquad j, \ell \in [d],$$

provided the limit exists. $\mathcal{X} = (\chi(j,\ell))_{j,\ell \in [d]}$ is the **tail correlation matrix** .

Throughout: $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ is a d-variate observable rv with continuous marginal cdfs $F_1, \ldots, F_d$.

The **tail correlation** between X_j and X_ℓ is

$$\chi(j,\ell) = \lim_{t \downarrow 0} t^{-1} \mathbb{P}\Big(F_j(X_j) > 1 - t, F_\ell(X_\ell) > 1 - t \Big), \qquad j, \ell \in [d],$$

provided the limit exists. $\mathcal{X} = (\chi(j,\ell))_{j,\ell \in [d]}$ is the **tail correlation matrix** .

Margin-free dependence measure:

- The coefficient is invariant under increasing transformations of the margins.

Strength of tail dependence:

- We have $\chi(j,\ell) \in [0,1]$, and X_j, X_ℓ are called tail independent iff $\chi(j,\ell) = 0$.

The stable tail dependence function $\boldsymbol{X}$:

$$L(\boldsymbol{x}) = \lim_{t \downarrow 0} t^{-1} \mathbb{P}\Big(\exists j \in [d] : F_j(X_j) > 1 - t x_j\Big), \qquad \boldsymbol{x} \in [0, \infty)^d$$

provided the limit exists. Then $\chi(j, \ell) = 2 - L(\boldsymbol{e}_j + \boldsymbol{e}_\ell)$.

The stable tail dependence function $\boldsymbol{X}$:

$$L(\boldsymbol{x}) = \lim_{t \downarrow 0} t^{-1} \mathbb{P}\Big(\exists j \in [d] : F_j(X_j) > 1 - t x_j\Big), \qquad \boldsymbol{x} \in [0,\infty)^d$$

provided the limit exists. Then $\chi(j,\ell) = 2 - L(\boldsymbol{e}_j + \boldsymbol{e}_\ell)$.

The tail copula of $\boldsymbol{X}$:

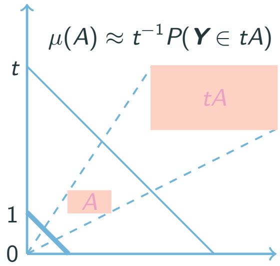
$$R(\boldsymbol{x}) = \lim_{t \downarrow 0} t^{-1} \mathbb{P}\Big(\forall j \in [d] : F_j(X_j) > 1 - t x_j\Big),$$

provided the limit exists, such that $\chi(j,\ell) = R(\infty,\dots,\infty,1,\infty,\dots\infty,1,\infty,\dots,\infty)$. It relates to L via inclusion-exclusion.

Exponent measure: the STDF of $\boldsymbol{X}$ exists if and only if $\boldsymbol{Y}$ with coordinates $Y_j = 1/(1 - F_j(X_j))$ is regularly varying on $[0, \infty)^d \setminus \{\mathbf{0}\}$, i.e.,

$$t\mathbb{P}(\boldsymbol{Y} \in t\,\cdot\,) \to \mu(\cdot), \qquad t \to \infty,$$

for non-degenerate measure μ called the **exponent measure**.



Exponent measure: the STDF of $\boldsymbol{X}$ exists if and only if $\boldsymbol{Y}$ with coordinates $Y_j = 1/(1 - F_j(X_j))$ is regularly varying on $[0, \infty)^d \setminus \{\mathbf{0}\}$, i.e.,

$$t\mathbb{P}(\boldsymbol{Y} \in t\,\cdot\,) \to \mu(\cdot), \qquad t \to \infty,$$

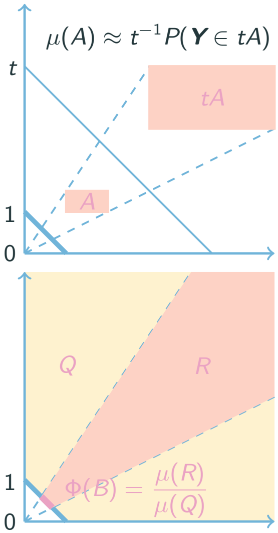
for non-degenerate measure μ called the **exponent measure**.

Spectral measure: the angular component of μ:

$$\Phi(B) := \frac{\mu(\{\boldsymbol{y} : \|\boldsymbol{y}\| > 1, \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|} \in B\})}{\mu(\{\boldsymbol{y} : \|\boldsymbol{y}\| > 1\})}.$$

Φ is in one-to-one correspondence to L:

$$L(\boldsymbol{x}) = d \int_{\mathbb{S}_+^{d-1}} \max_{j \in [d]} (\lambda_j x_j) \Phi(d\boldsymbol{\lambda}),$$



Linear factor models in extremes

Proposition.

Let

- $K \in [d]$,
- $\boldsymbol{A} \in [0,1]^{d\times K}$ a matrix with non-zero row and column sums,
- $\boldsymbol{Z} = (Z_1, \dots, Z_K)^\top$ has identically distributed, pairwise asymptotically independent coordinates that are regularly varying of order α.

Then

$$\boldsymbol{X}' := \boldsymbol{A}\boldsymbol{Z}$$

has stable tail dependence function

$$L_{K,\bar{\boldsymbol{A}}}(\boldsymbol{x}) = \sum_{a\in[K]} \bigvee_{j\in[d]} \bar{A}_{ja} x_j, \qquad \boldsymbol{x} \in [0,\infty)^d,$$

where $\bar{\boldsymbol{A}} = (\bar{A}_{ja})_{ja} \in [0,1]^{d\times K}$ has entries $\bar{A}_{ja} = \frac{A_{ja}^\alpha}{\sum_{b\in[K]} A_{jb}^\alpha}$, which has row sums 1.

Linear Factor Tail Dependence Model: The observable random vector $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ has continuous marginal cdfs and its stable tail dependence function L satisfies

$$L(\boldsymbol{x}) = L_{K,\bar{\boldsymbol{A}}}(\boldsymbol{x}) = \sum_{a \in [K]} \bigvee_{j \in [d]} \bar{\boldsymbol{A}}_{ja} x_j, \qquad \boldsymbol{x} \in [0, \infty)^d,$$

for some unknown parameter $\theta = (K, \bar{\boldsymbol{A}})$ with $\bar{\boldsymbol{A}}$ having row sums 1 and non-zero column sums.

Linear Factor Tail Dependence Model: The observable random vector $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ has continuous marginal cdfs and its stable tail dependence function L satisfies

$$L(\boldsymbol{x}) = L_{K,\bar{\boldsymbol{A}}}(\boldsymbol{x}) = \sum_{a \in [K]} \bigvee_{j \in [d]} \bar{\boldsymbol{A}}_{ja} x_j, \qquad \boldsymbol{x} \in [0, \infty)^d,$$

for some unknown parameter $\theta = (K, \bar{\boldsymbol{A}})$ with $\bar{\boldsymbol{A}}$ having row sums 1 and non-zero column sums.

The matrix $\bar{\boldsymbol{A}}$ can be interpreted as a *factor loading matrix for tail dependence*: the same L-function also arises in a linear factor model with $\alpha = 1$ and $\boldsymbol{A} = \bar{\boldsymbol{A}}$ (as in **factor copula models**; see Krupskii and Joe, 2014, or Oh and Patton, 2017).

Estimation strategies:

- If K is known: moment-based estimators matching $L_{K,\bar{\boldsymbol{A}}}$ to the empirical stdf $\hat{L}_n$ (Einmahl, Krajina and Segers, 2012).
- When $K = d$, $\bar{\boldsymbol{A}}$ can be estimated with the empirical Tail Pairwise Dependence Matrix through a completely positive decomposition (Kiriliouk and Zhou, 2022).
- If $K = d$ and $\bar{\boldsymbol{A}}$ is of a specific form that arises in structural equations models on DAGs: recursive estimation after graph recovery (Klüppelberg and Krali, 2021, ...).

Fitting the Linear Factor Tail Dependence Model

Estimation strategies:

- If K is known: moment-based estimators matching $L_{K,\bar{\boldsymbol{A}}}$ to the empirical stdf $\hat{L}_n$ (Einmahl, Krajina and Segers, 2012).
- When $K = d$, $\bar{\boldsymbol{A}}$ can be estimated with the empirical Tail Pairwise Dependence Matrix through a completely positive decomposition (Kiriliouk and Zhou, 2022).
- If $K = d$ and $\bar{\boldsymbol{A}}$ is of a specific form that arises in structural equations models on DAGs: recursive estimation after graph recovery (Klüppelberg and Krali, 2021, ...).
- If $K \leq d$ is unknown and $\bar{\boldsymbol{A}}$ satisfies the **pure variable assumption** (Bing et al., 2020)

$$\forall \text{ factor indexes } a \in [K], \exists \text{ variable index } j \in [d] : \quad \bar{\boldsymbol{A}}_{j\cdot} = \boldsymbol{e}_a :$$

estimation based on the pairwise empirical tail correlation matrix (Bücher and Boulin, 2025).

Estimation strategies:

- If K is known: moment-based estimators matching $L_{K,\bar{\boldsymbol{A}}}$ to the empirical stdf $\hat{L}_n$ (Einmahl, Krajina and Segers, 2012).
- When $K = d$, $\bar{\boldsymbol{A}}$ can be estimated with the empirical Tail Pairwise Dependence Matrix through a completely positive decomposition (Kiriliouk and Zhou, 2022).
- If $K = d$ and $\bar{\boldsymbol{A}}$ is of a specific form that arises in structural equations models on DAGs: recursive estimation after graph recovery (Klüppelberg and Krali, 2021, ...).
- If $K \leq d$ is unknown and $\bar{\boldsymbol{A}}$ satisfies the **pure variable assumption** (Bing et al., 2020)

$$\forall \text{ factor indexes } a \in [K], \exists \text{ variable index } j \in [d] : \quad \bar{\boldsymbol{A}}_{j\cdot} = \boldsymbol{e}_a :$$

  estimation based on the pairwise empirical tail correlation matrix (Bücher and Boulin, 2025).

However:

- The asymptotic independence of the factors makes the model too restrictive for many applications.

Proposition (B. and Bücher, 2026). Let

- $K \in [d]$,
- $\bar{\boldsymbol{A}} \in [0,1]^{d \times K}$ a full column rank matrix with row sums 1,
- $\boldsymbol{Z} = KR\boldsymbol{\Lambda}$, where R is standard Pareto and where $\boldsymbol{\Lambda} \sim \psi$ is independent of R, with ψ a probability measure on $\mathbb{S}_+^{K-1} = \{\boldsymbol{x} \in [0,1]^K : \|\boldsymbol{x}\|_1 = 1\}$ whose margins have expectation $1/K$.

Then

$$\boldsymbol{X}' := \bar{\boldsymbol{A}}\boldsymbol{Z}$$

has stable tail dependence function

$$L_{K,\bar{\boldsymbol{A}},\psi}(\boldsymbol{x}) = K \int_{\mathbb{S}_+^{K-1}} \bigvee_{j \in [d]} \left(x_j \sum_{a \in [K]} \bar{\boldsymbol{A}}_{ja} z_a \right) \psi(dz), \qquad \boldsymbol{x} \in [0,\infty)^d.$$

Proposition (B. and Bücher, 2026). Let

- $K \in [d]$,
- $\bar{\boldsymbol{A}} \in [0,1]^{d \times K}$ a full column rank matrix with row sums 1,
- $\boldsymbol{Z} = KR\boldsymbol{\Lambda}$, where R is standard Pareto and where $\boldsymbol{\Lambda} \sim \psi$ is independent of R, with ψ a probability measure on $\mathbb{S}_+^{K-1} = \{\boldsymbol{x} \in [0,1]^K : \|\boldsymbol{x}\|_1 = 1\}$ whose margins have expectation $1/K$.

Then

$$\boldsymbol{X}' := \bar{\boldsymbol{A}}\boldsymbol{Z}$$

has stable tail dependence function

$$L_{K,\bar{\boldsymbol{A}},\psi}(\boldsymbol{x}) = K \int_{\mathbb{S}_+^{K-1}} \bigvee_{j \in [d]} \Big(x_j \sum_{a \in [K]} \bar{\boldsymbol{A}}_{ja} z_a \Big) \psi(dz), \qquad \boldsymbol{x} \in [0,\infty)^d.$$

The ‘traditional’ linear model is included, by taking $\psi = K^{-1} \sum_{a \in [K]} \delta_{\boldsymbol{e}_a}$ (the spectral dependence measure corresponding to K asymptotically independent variables).

Generalized Linear Factor Tail Dependence Model: The observable random vector $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ has continuous marginal cdfs and its stable tail dependence function L satisfies

$$L(\boldsymbol{x}) = L_{K,\bar{\boldsymbol{A}},\psi}(\boldsymbol{x}) = K \int_{\mathbb{S}_+^{K-1}} \bigvee_{j \in [d]} \left(x_j \sum_{a \in [K]} \bar{\boldsymbol{A}}_{ja} z_a \right) \psi(dz), \qquad \boldsymbol{x} \in [0, \infty)^d$$

for some unknown parameter $\theta = (K, \bar{\boldsymbol{A}}, \psi) \in \Theta$, where

$$\Theta = \Big\{ (K, \bar{\boldsymbol{A}}, \psi) \mid K \in [d], \bar{\boldsymbol{A}} \in [0, \infty)^{d \times K} \text{ has row sums 1 and full column rank,} \\ \psi \text{ a probability measure on } \mathbb{S}_+^{K-1} \text{ with margins having expectation } 1/K \Big\}.$$

Generalizing the classical linear factor model II

Generalized Linear Factor Tail Dependence Model: The observable random vector $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ has continuous marginal cdfs and its stable tail dependence function L satisfies

$$L(\boldsymbol{x}) = L_{K,\bar{\boldsymbol{A}},\psi}(\boldsymbol{x}) = K \int_{\mathbb{S}_+^{K-1}} \bigvee_{j \in [d]} \left(x_j \sum_{a \in [K]} \bar{\boldsymbol{A}}_{ja} z_a \right) \psi(dz), \qquad \boldsymbol{x} \in [0,\infty)^d$$

for some unknown parameter $\theta = (K, \bar{\boldsymbol{A}}, \psi) \in \Theta$, where

$$\Theta = \Big\{ (K, \bar{\boldsymbol{A}}, \psi) \mid K \in [d], \bar{\boldsymbol{A}} \in [0,\infty)^{d \times K} \text{ has row sums 1 and full column rank}, \; \psi \text{ a probability measure on } \mathbb{S}_+^{K-1} \text{ with margins having expectation } 1/K \Big\}.$$

The model is fully flexible: any stable tail dependence function L can be written as $L = L_{d,I_d,\psi}$ for some suitable ψ.

Theorem (B. and Bücher, 2026). Suppose that

1. $\bar{\boldsymbol{A}}$ satisfies the pure variable assumption,
2. ψ satisfies an additional mild assumption preventing two factors from being perfectly tail dependent.

Then $(K, \bar{\boldsymbol{A}}, \psi)$ can be uniquely recovered from L. More formally: for a suitable set Θ_L, the mapping

$$\Phi_L : \Theta_L \to \{\text{all stdfs}\}, \quad (K, \bar{\boldsymbol{A}}, \psi) \to L_{K, \bar{\boldsymbol{A}}, \psi}$$

is injective up to label permutations.

Identifiability from L: proof idea I

Some definitions:

- Pure variables: $I_a = \{j \in [d] : \bar{\boldsymbol{A}}_{j\cdot} = \boldsymbol{e}_a\}$ the pure variables for the ath factor, and $I = \bigcup_{a\in[K]} I_a$.
- Impure variables: $J = [d] \setminus I$.
- The tail pairwise dependence matrix: with $\Phi_{K,\bar{\boldsymbol{A}},\psi}$ the spectral measure associated with $L_{K,\bar{\boldsymbol{A}},\psi}$:

$$\Sigma = \int \boldsymbol{x}\boldsymbol{x}^\top \, \Phi_{K,\bar{\boldsymbol{A}},\psi}(d\boldsymbol{x})$$

- Rowwise Max: $m_j = \max_{\ell\in[d]} \Sigma_{j\ell}$. Rowwise Argmax: $S_j = \arg\max_{\ell\in[d]} \Sigma_{j\ell}$.

Identifiability from L: proof idea I

Some definitions:

- Pure variables: $I_a = \{j \in [d] : \bar{\boldsymbol{A}}_{j\cdot} = \boldsymbol{e}_a\}$ the pure variables for the ath factor, and $I = \bigcup_{a \in [K]} I_a$.
- Impure variables: $J = [d] \setminus I$.
- The tail pairwise dependence matrix: with $\Phi_{K,\bar{\boldsymbol{A}},\psi}$ the spectral measure associated with $L_{K,\bar{\boldsymbol{A}},\psi}$:

$$\Sigma = \int \boldsymbol{x}\boldsymbol{x}^\top \, \Phi_{K,\bar{\boldsymbol{A}},\psi}(d\boldsymbol{x})$$

- Rowwise Max: $m_j = \max_{\ell \in [d]} \Sigma_{j\ell}$. Rowwise Argmax: $S_j = \arg\max_{\ell \in [d]} \Sigma_{j\ell}$.

Step 1: Identification of pure variables:

- $I = \{j \in [d] \mid \forall \ell \in S_j : m_j = m_\ell\}$.
- K is the number of equivalence classes under the relation

$$j \sim \ell \quad \Longleftrightarrow \quad \ell \in S_j$$

- The K equivalence classes correspond to $I_1, \ldots, I_K$.

Identifiability from L: proof idea II

Step 2: Identification of $\bar{\boldsymbol{A}}$:

- Rows of $\bar{\boldsymbol{A}}$ corresponding to $j \in I$: unit vectors.
- Rows of $\bar{\boldsymbol{A}}$ corresponding to $j \in J$: Define $\boldsymbol{C} \in \mathbb{R}^{K \times K}$ by $C_{ab} = \Sigma_{j_a, j_b}$ with $j_a \in I_a$ and $j_b \in I_b$. The jth row of $\bar{\boldsymbol{A}}$, say $\boldsymbol{\beta}_j \in \mathbb{R}^K$, satisfies

$$\boldsymbol{C}\boldsymbol{\beta}_j = (\Sigma_{j_1, j}, \ldots, \Sigma_{j_K, j})^\top$$

Step 2: Identification of $\bar{\boldsymbol{A}}$:

- Rows of $\bar{\boldsymbol{A}}$ corresponding to $j \in I$: unit vectors.
- Rows of $\bar{\boldsymbol{A}}$ corresponding to $j \in J$: Define $\boldsymbol{C} \in \mathbb{R}^{K \times K}$ by $C_{ab} = \Sigma_{j_a, j_b}$ with $j_a \in I_a$ and $j_b \in I_b$. The jth row of $\bar{\boldsymbol{A}}$, say $\boldsymbol{\beta}_j \in \mathbb{R}^K$, satisfies

$$\boldsymbol{C}\boldsymbol{\beta}_j = (\Sigma_{j_1,j}, \ldots, \Sigma_{j_K,j})^\top$$

Step 3: Identification of ψ:

- Define pure variables: $Z_a := |I_a|^{-1} \sum_{j \in I_a} Y_j$.
- Then $\boldsymbol{Z} = (Z_1, \ldots, Z_K)$ is regularly varying, with spectral measure ψ.

Starting point: the TPDM associated with $L_{K,\bar{\boldsymbol{A}},\psi}$ can be written as a limit of a conditional expectation:

$$\Sigma = \int \boldsymbol{x}\boldsymbol{x}^\top \, \Phi_{K,\bar{\boldsymbol{A}},\psi}(d\boldsymbol{x}) = \lim_{x\to\infty} \mathbb{E}\left[\frac{\boldsymbol{Y}\boldsymbol{Y}^\top}{\|\boldsymbol{Y}\|^2} \mid \|\boldsymbol{Y}\| > x \right],$$

where $\boldsymbol{Y}$ has coordinates $Y_j = 1/(1 - F_j(X_j))$.

Starting point: the TPDM associated with $L_{K,\bar{\boldsymbol{A}},\psi}$ can be written as a limit of a conditional expectation:

$$\Sigma = \int \boldsymbol{x}\boldsymbol{x}^\top \, \Phi_{K,\bar{\boldsymbol{A}},\psi}(d\boldsymbol{x}) = \lim_{x\to\infty} \mathbb{E}\left[\frac{\boldsymbol{Y}\boldsymbol{Y}^\top}{\|\boldsymbol{Y}\|^2} \mid \|\boldsymbol{Y}\| > x\right],$$

where $\boldsymbol{Y}$ has coordinates $Y_j = 1/(1 - F_j(X_j))$.

Empirical counterpart: the empirical TPDM $\hat{\Sigma} = (\hat{\Sigma}_{j\ell})_{j,\ell}$ has entries

$$\hat{\Sigma}_{j\ell} = \frac{1}{k} \sum_{i=1}^{n} \frac{\hat{\boldsymbol{Y}}_i \hat{\boldsymbol{Y}}_i^\top}{\hat{S}_i} \mathbf{1}(\hat{S}_i > \hat{S}_{n-k:n}),$$

where k is a threshold parameter; e.g., $k = \lfloor 0.05 \cdot n \rfloor$, and $\hat{S}_i = \|\hat{\boldsymbol{Y}}_i\|$, and $\hat{Y}_{ij} = 1/(1 - \hat{F}_{nj}(X_{ij}))$.

Estimation: the PureVar- and LassoSimplexProjector-Algorithm II

- **Estimation of** K **and** I**:** Apply the construction from the identifiability proof with $\hat{\Sigma}$, where an index is declared a row-wise argmax if it is at distant at most 2κ from the respective maximum, some small hyperparameter $\kappa > 0$ **(PureVar-Algorithm)**.

Estimation: the PureVar- and LassoSimplexProjector-Algorithm II

- **Estimation of K and I:** Apply the construction from the identifiability proof with $\hat{\Sigma}$, where an index is declared a row-wise argmax if it is at distant at most 2κ from the respective maximum, some small hyperparameter $\kappa > 0$ **(PureVar-Algorithm)**.
- **Estimation of $\bar{A}$:** solve the linear regression problem using $\hat{\Sigma}$ instead of Σ with the LASSO (with penalty $\lambda \geq 0$), and project the resulting vectors to the unit simplex **(LassoSimplexProjector)**.

Estimation: the PureVar- and LassoSimplexProjector-Algorithm II

- **Estimation of K and I:** Apply the construction from the identifiability proof with $\hat{\Sigma}$, where an index is declared a row-wise argmax if it is at distant at most 2κ from the respective maximum, some small hyperparameter $\kappa > 0$ **(PureVar-Algorithm)**.
- **Estimation of $\bar{A}$:** solve the linear regression problem using $\hat{\Sigma}$ instead of Σ with the LASSO (with penalty $\lambda \geq 0$), and project the resulting vectors to the unit simplex **(LassoSimplexProjector)**.
- **Estimation of ψ:** Define estimated pure variables $\hat{Z}_{ia} = |\hat{I}_a|^{-1} \sum_{j \in \hat{I}_a} \hat{Y}_{ij}$ and estimate ψ by any method of choice, e.g., empirical spectral measure.

Application:
on probabilities of low-wind events

Target parameter and statistical model

Recall: W_j the daily maximal wind speed and C_j the installed capacity at grid cell $j = 1, \dots, d$.

Target parameter:

$$p_{0.8}(3) = \mathbb{P}\Big(\exists J \subset [d] \text{ with installed capacity in } J \text{ exceeding } 80\% \text{ of the total capacity : } W_j \leq 3 \text{ for all } j \in J\Big)$$

Target parameter and statistical model

Recall: W_j the daily maximal wind speed and C_j the installed capacity at grid cell $j = 1, \ldots, d$.

Target parameter:

$$p_{0.8}(3) = \mathbb{P}\Big(\exists J \subset [d] \text{ with installed capacity in } J \text{ exceeding } 80\% \text{ of the total capacity} : W_j \leq 3 \text{ for all } j \in J\Big)$$

$$= \mathbb{P}\Big(\exists J \subset [d] \text{ with installed capacity in } J \text{ exceeding } 80\% \text{ of the total capacity} : X_j \geq 1/3 \text{ for all } j \in J\Big), \qquad [\text{where } X_j = 1/W_j].$$

Target parameter and statistical model

Recall: W_j the daily maximal wind speed and C_j the installed capacity at grid cell $j = 1, \dots, d$.

Target parameter:

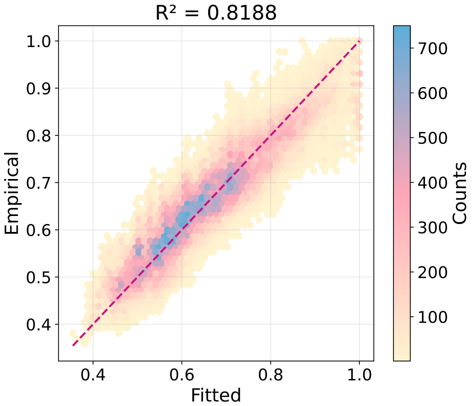
$$p_{0.8}(3) = \mathbb{P}\Big(\exists J \subset [d] \text{ with installed capacity in } J \text{ exceeding } 80\% \text{ of the total capacity} : W_j \leq 3 \text{ for all } j \in J\Big)$$

$$= \mathbb{P}\Big(\exists J \subset [d] \text{ with installed capacity in } J \text{ exceeding } 80\% \text{ of the total capacity} : X_j \geq 1/3 \text{ for all } j \in J\Big), \qquad [\text{where } X_j = 1/W_j].$$

Statistical model:

- The margins of $\boldsymbol{X}$ are in the MDA of the GEV, and the STDF satisfies $L = L_{K,\bar{\boldsymbol{A}},\psi}$.
- $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ are iid copies of $\boldsymbol{X}$

- **Threshold choice:** $k = 135$ (5% of the total sample size)
- **Hyperparameter choice:** run the algorithms on a grid of hyperparameters, and choose the one for which empirical tail correlations are as close as possible to model-implied tail correlations.
- **Estimated parameters:** $\hat{K} = 20$, with 2,251 non-zero entries in the estimated loading matrix (20% of the entries).



Implied tail correlations from the fitted model against empirical tail correlations.

Connecting the target parameter to the model

Lemma. Let $\mathcal{J}$ be a collection of subsets of $[d]$, for instance $\mathcal{J} = \mathcal{J}_{0.8}$ = all sets J such that the installed capacity in J exceeds 80% of the total capacity. Then

$$R^{\cup}_{\mathcal{J}}(\boldsymbol{x}) := \lim_{t\to\infty} t \cdot \mathbb{P}\Big(\exists J \in \mathcal{J} : F_j(X_j) > 1 - \frac{x_j}{t}\, \forall j \in J\Big) = K \int_{\mathbb{S}_+^{K-1}} \max_{J\in\mathcal{J}} \min_{j\in J} \Big(\sum_{a\in[K]} \bar{\boldsymbol{A}}_{ja} z_a x_j\Big)\, \psi(d\boldsymbol{z}).$$

Lemma. Let $\mathcal{J}$ be a collection of subsets of $[d]$, for instance $\mathcal{J} = \mathcal{J}_{0.8}$ = all sets J such that the installed capacity in J exceeds 80% of the total capacity. Then

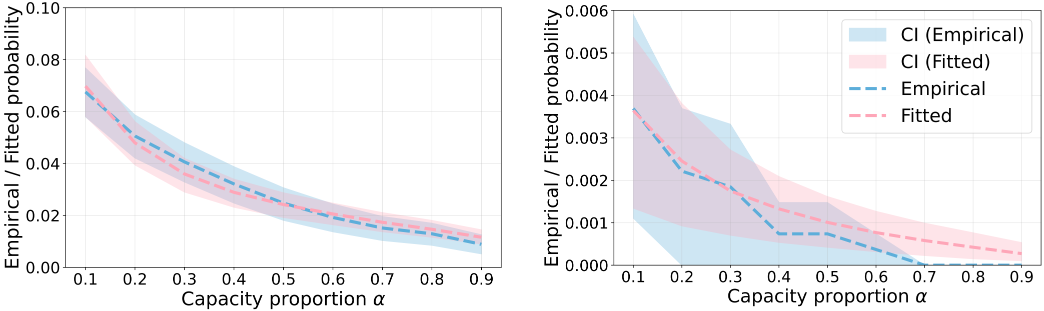
$$R^{\cup}_{\mathcal{J}}(\boldsymbol{x}) := \lim_{t\to\infty} t \cdot \mathbb{P}\Big(\exists J \in \mathcal{J} : F_j(X_j) > 1 - \frac{x_j}{t}\ \forall j \in J\Big) = K \int_{\mathbb{S}_+^{K-1}} \max_{J\in\mathcal{J}} \min_{j\in J} \Big(\sum_{a\in[K]} \bar{\boldsymbol{A}}_{ja} z_a x_j \Big)\, \psi(d\boldsymbol{z}).$$

Some straightforward calculations then yields

$$p_{0.8}(3) = p_{\mathcal{J}_{0.8}}(3) \approx K \int_{\mathbb{S}_+^{K-1}} \max_{J\in\mathcal{J}} \min_{j\in J} \Big(\sum_{a\in[K]} \bar{\boldsymbol{A}}_{ja} z_a q_j(1/3) \Big)\, \psi(d\boldsymbol{z})$$

where $q_j(x_j) = \mathbb{P}(X_j > x_j)$ [a small exceedance probability]. The maxmin can typically be evaluated efficiently in $O(d \log d)$ computations.

Estimated target parameters



Estimated probabilities $p_\alpha(w)$ for fixed $w = 5$ (left) and $w = 3$ (right) using the plug-in estimator

$$\hat{p}_\alpha(w) = \frac{\hat{K}}{k} \sum_{i \in [n]} \max_{J \in \mathcal{J}_\alpha} \min_{j \in J} \Bigg(\sum_{a \in [\hat{K}]} \frac{\hat{\boldsymbol{A}}_{ja} \hat{Z}_{ia} \hat{q}_j(1/3)}{\hat{S}_i} \Bigg) \mathbf{1}(\hat{S}_i > \hat{S}_{n-k:n}), \qquad \hat{S}_i = \|\hat{\boldsymbol{Z}}_i\|_1.$$

Conclusion

Summary

- A generalized latent linear factor model for tail dependence that allows for dependent factors.
- Identifiability results under a pure variable assumption.
- Algorithms for estimation of model parameters and derived quantities.

Summary

- A generalized latent linear factor model for tail dependence that allows for dependent factors.
- Identifiability results under a pure variable assumption.
- Algorithms for estimation of model parameters and derived quantities.

Outlook

- Statistical guarantees, parametric sub-models for ψ.

Thank you!

References:

- ▷ Alexis Boulin and Axel Bücher (2025). Structured linear factor models for tail dependence. *ArXiv:2507.16340*.
- ▷ Alexis Boulin and Axel Bücher (2026). Dimension Reduction in Multivariate Extremes via Latent Linear Factor Models. *ArXiv:2602.23143*.