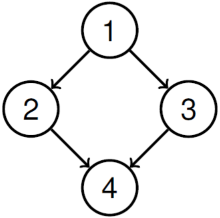
# Max-linear models on graphs

Claudia Klüppelberg
(joint with Nadine Gissibl and Moritz Otto)

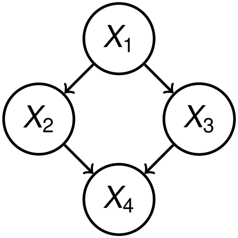
Technical University of Munich



## Graphical models [Lauritzen (1996)]

- $\mathbb{D} = (V, E)$: directed acyclic graph (DAG)
- $\mathbf{X} = (X_1, \ldots, X_d)$: joint probability distribution
- Markov relative to $\mathbb{D}$

**Example.** $V = \{1, 2, 3, 4\}$, $E = \{(1, 2), (1, 3), (2, 4), (3, 4)\}$




(local) Markov property:

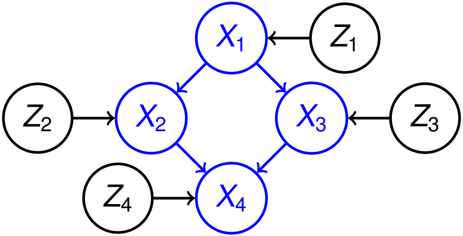
$$X_v \perp\!\!\!\perp \mathbf{X}_{\text{nd}(v)\setminus\text{pa}(v)} \mid \mathbf{X}_{\text{pa}(v)}$$

$$X_4 \perp\!\!\!\perp X_1 \mid X_2, X_3$$

## Structural equation models [Pearl (2009)]

For $i = 1, \ldots, d$:

- $f_i$ measurable functions
- $Z_i$ independent noise variables
- Define $X_i := f_i(\mathbf{X}_{\mathrm{pa}(i)}, Z_i)$

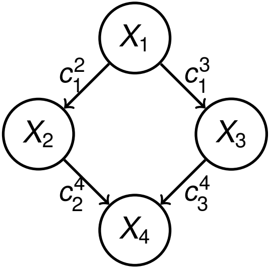



**Examples:** in the literature mainly discrete models and Gaussian models with $X_i = f_i(\mathbf{X}_{\mathrm{pa}(i)}, Z_i) = \sum_{k \in \mathrm{pa}(i)} c_k^i X_k + c_i^i Z_i$.

## Max-linear structural equation models (ML-SEM)

For $Z_1, \dots, Z_d > 0$ independent, continuous with support $\mathbb{R}^+$ and $c_k^i \in (0,1]$, we define the **max-linear structural equation model**

$$X_i := \bigvee_{k \in \mathrm{pa}(i)} c_k^i X_k \vee c_i^i Z_i \quad i = 1, \dots, d$$




$$X_1 = c_1^1 Z_1$$

$$X_2 = c_1^2 X_1 \vee c_2^2 Z_2 = c_1^1 c_1^2 Z_1 \vee c_2^2 Z_2$$

$$X_3 = c_1^3 X_1 \vee c_3^3 Z_3 = c_1^1 c_1^3 Z_1 \vee c_3^3 Z_3$$

$$X_4 = c_2^4 X_2 \vee c_3^4 X_3 \vee c_4^4 Z_4$$

$$= c_2^4 (c_1^1 c_1^2 Z_1 \vee c_2^2 Z_2) \vee c_3^4 (c_1^1 c_1^3 Z_1 \vee c_3^3 Z_3) \vee c_4^4 Z_4$$

$$= (c_1^1 c_1^2 c_2^4 \vee c_1^1 c_1^3 c_3^4) Z_1 \vee c_2^2 c_2^4 Z_2 \vee c_3^3 c_3^4 Z_3 \vee c_4^4 Z_4$$

## Max-linearity of X by path analysis

Let $\mathbf{X} = (X_1, \dots, X_d)$ be generated by a max-linear SEM with coefficients $c_k^i \in (0,1]$ and DAG $\mathbb{D} = (V, E)$.
For a path $p = [j = k_0 \to k_1 \to \cdots \to k_n = i]$ define the coefficients

$$d_{ji}(p) := c_{k_0}^{k_1} c_{k_1}^{k_2} \cdots c_{k_{n-1}}^{k_n}$$

and for all $i = 1, \dots, d$,

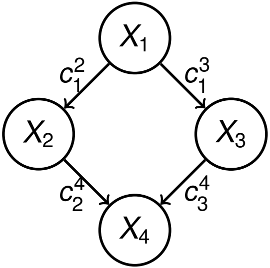
$$b_{ji} := \bigvee_{p \in P_{ji}} d_{ji}(p) \quad \forall j \in \mathrm{an}(i), \quad b_{ii} = c_i^i \quad \text{and all other} \quad b_{ji} = 0,$$

We call the specific path/paths giving $b_{ji}$ **max-weighted paths**.
**Theorem.** $\mathbf{X}$ is a **max-linear model:** For $i = 1, \dots, d$,

$$X_i = \bigvee_{j=1}^{d} b_{ji} Z_j = \bigvee_{j \in \mathrm{An}(i)} b_{ji} Z_j \qquad (\mathrm{An}(i) = \mathrm{an}(i) \cup \{i\}).$$

## A SEM as max-linear model on a DAG

The **ML coefficient matrix** $B$ is a weighted **reachability matrix**.




For our example we find:

$$B = \begin{bmatrix} c_1^1 & c_1^2 & c_1^3 & c_1^2 c_2^4 \vee c_1^3 c_3^4 \\ 0 & c_2^2 & 0 & c_2^4 \\ 0 & 0 & c_3^3 & c_3^4 \\ 0 & 0 & 0 & c_4^4 \end{bmatrix}$$

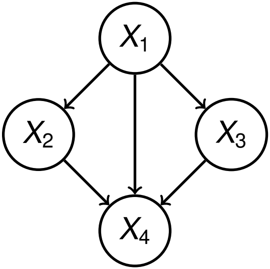
Reachability matrix: $R = \operatorname{sgn}(B)$

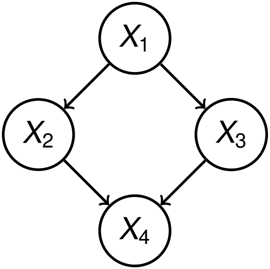
## Transitive reduction

A DAG $\mathbb{D}^{\text{tr}} = (V, E^{\text{tr}})$ is called **transitive reduction** of $\mathbb{D}$, if

(a) for all $i, j \in V$ the DAG $\mathbb{D}^{\text{tr}}$ has a path from $j$ to $i$ if and only if $\mathbb{D}$ has a path from $j$ to $i$, and

(b) there is no DAG with less edges satisfying condition (a).



## Theorem

Let $(\mathbb{D}, \mathbf{X})$ be a ML model on a DAG with coeff. matrix $B = (b_{ij})_{d \times d}$. Let further $\mathbb{D}^{\text{tr}} = (V, E^{\text{tr}})$ be the transitive reduction of $\mathbb{D}$. Define

$$B^{=} := \left\{ (k,i) \in V \times V : k \in \text{pa}(i) \backslash \text{pa}^{\text{tr}}(i) \text{ and } b_{ki} = \bigvee_{l \in \text{de}(k) \cap \text{pa}(i)} \frac{b_{kl} b_{li}}{b_{ll}} \right\}$$

and for $E^B := E \setminus B^{=}$ the DAG $\mathbb{D}^B := (V, E^B)$.

Then $(\mathbb{D}^B, \mathbf{X})$ is a **minimal ML model on a DAG**.

**Remark:** $\mathbb{D}^B$ is minimal causal w.r.t. $\mathbf{X}$.

## Max-weighted ML model on a DAG

A ML model on a DAG $(\mathbb{D}, \mathbf{X})$ is called **max-weighted**, if all paths are max-weighted: for all paths $p = [j = k_0 \to k_1 \to \ldots \to k_n = i]$ we have

$$b_{ji} = c_{k_0}^{k_0} c_{k_0}^{k_1} \cdots c_{k_{n-2}}^{k_{n-1}} c_{k_{n-1}}^{k_n} = d_{ji}(p).$$

**Proposition.** (1) Let $(\mathbb{D}, \mathbf{X})$ be a max-weighted ML model on a DAG. Then $\mathbb{D}^B = \mathbb{D}^{\text{tr}}$.
(2) A ML model $(\mathbb{D}, \mathbf{X})$ on a directed tree is max-weighted.
(3) For every DAG we can construct a max-weighted ML model by choosing $c_k^i = n_k / n_i$, $c_i^i = 1/n_i$ for $n_i := |\text{An}(i)|$ for $k \in \text{pa}(i)$.

Asadi, P., Davison, A.C. and Engelke, S. (2015) Extremes on river networks.

2 P. ASADI, A. C. DAVISON, AND S. ENGELKE

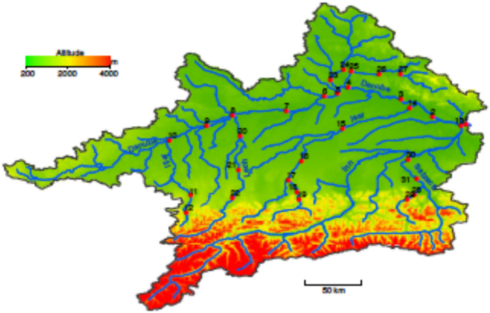
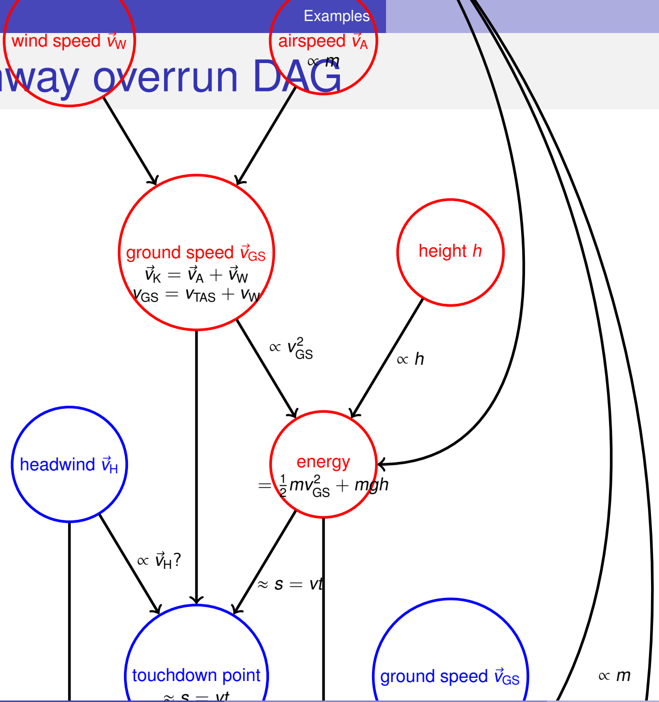
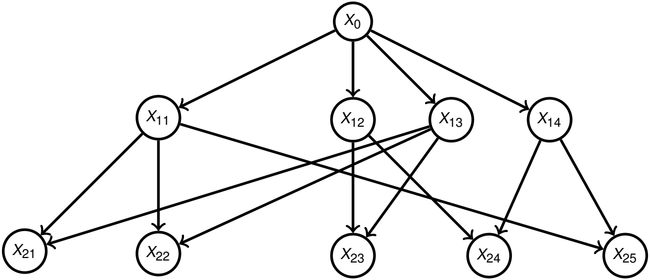



FIGURE 1. Topographic map of the upper Danube basin, showing sites of 31 gauging stations (red blobs) along the Danube and its tributaries. Water flows broadly from left to right.

# Runway overrun DAG



Einmahl, Kiriliouk and Segers (2016) A continuous updating weighted least squares estimator of tail dependence in high dimensions.




$X_0$ (EURO STOXX 50),
$X_{11}, X_{12}, X_{13}, X_{14}$ (chemical industry, insurance, DAX, CAC40),
$X_{21}, X_{22}, X_{23}, X_{24}, X_{25}$ (Bayer, BASF, Allianz, Axa, Airliquide)

## The multivariate distribution function of a ML model on a DAG

Let $Z_1, \dots, Z_d \in \mathrm{MDA}(\Phi_\alpha)$ with $\Phi_\alpha(x) = e^{-x^{-\alpha}}, x > 0$.
Then $\mathbf{X} = (X_1, \dots, X_d) \in \mathrm{MDA}(G)$, where for $\mathbf{x} = (x_1, \dots, x_d) > \mathbf{0}$

$$G(\mathbf{x}) = \exp\Big\{ - \sum_{k=1}^{d} \bigvee_{k \in \mathrm{An}(i)} b_{ki}^{\alpha} x_i^{-\alpha} \Big\}$$

In particular,

$$\begin{aligned} G_{X_i}(x) &= \exp\Big\{ - x^{-\alpha} \sum_{k \in \mathrm{An}(i)} b_{ki}^{\alpha} \Big\} \\ G_{X_i,X_j}(x_i, x_j) &= \exp\Big\{ - \sum_{k \in \mathrm{An}(i) \cap \mathrm{An}(j)} \Big(\frac{b_{ki}}{x_i}\Big)^{\alpha} \wedge \Big(\frac{b_{kj}}{x_j}\Big)^{\alpha} \Big\} \end{aligned}$$

## Tail dependence coefficient

For notational simplicity assume from now on

$$\sum_{k \in \mathrm{An}(i)} b_{ki}^{\alpha} = 1 \quad \text{for} \quad i \in V$$

Then $G$ has standard marginal distributions $\Phi_\alpha$.
For $i, j \in V$ the **tail dependence coefficient** between $X_k$ and $X_l$

$$\chi(i,j) \;=\; \lim_{u \to \infty} P(X_i > u \mid X_j > u) = \sum_{k \in \mathrm{An}(i) \cap \mathrm{An}(j)} b_{ki}^{\alpha} \wedge b_{kj}^{\alpha}$$

We also assume from now on

$$\alpha = 1 \quad \text{such that} \quad \chi(i,j) = \sum_{k \in \mathrm{An}(i) \cap \mathrm{An}(j)} b_{ki} \wedge b_{kj}, \quad i, j \in V.$$

**Goal:** Estimate a max-weighted ML model $(\mathbb{D}, \mathbf{X})$ from $\chi$.

## Max-weighted ML model on a DAG

**Proposition.** Let $(\mathbb{D}, \mathbf{X})$ be a max-weighted ML model on a DAG.

1. For $j \in \mathrm{An}(i)$ we have $\chi(j,i) = \frac{b_{ji}}{b_{ii}}$.
2. For $j \in \mathrm{An}(i)$ with path $[j = k_0 \to k_1 \to \cdots \to k_n = i]$ we have $\chi(j,i) = \chi(k_0,k_1)\cdots\chi(k_{n-1},k_n)$.

□

**Corollary.** Let $V_0$ denote the set of initial nodes.

1. Then $k \in \mathrm{An}(i)$ if and only if $\chi(k,i) > 0$ and for all $j \in \mathrm{An}(i) \cap \mathrm{An}(k) \cap V_0$ we have $\chi(j,i) = \chi(j,k)\chi(k,i)$.
2. There is a path $[j = k_0 \to \cdots \to k_n]$ if and only if $\chi(k_m, k_{m+1}) > 0$ for $m = 0, \ldots, n-1$ and for all $j \in \mathrm{An}(i) \cap V_0$ we have $\chi(j,i) = \chi(k_0,k_1)\cdots\chi(k_{n-1},k_n)$.

□

**Theorem.** The following are equivalent

1. $\chi(i,j) = 0$
2. $X_i$ and $X_j$ are independent
3. $\mathrm{An}(i) \cap \mathrm{An}(j) = \emptyset$

We call $W \subseteq V$ a **$\chi$-clique** of $\mathbb{D}$ if $\chi(i,j) = 0$ for all $i, j \in W$, $i \neq j$.

**Lemma.** Let $V_0$ denote the set of initial nodes of $\mathbb{D}$.

1. For all $i, j \in V_0$ we have $\chi(i,j) = 0$; i.e. $V_0$ is a $\chi$-clique.
2. Let $W \subseteq V$ such that $\chi(i,j) = 0$ for all $i, j \in W$.
   Then $|W| \leq |V_0|$; i.e. $V_0$ is a maximal $\chi$-clique.
   Hence, $|\mathrm{An}(i) \cap V_0| = 1$ for all $i \in W$.

**Theorem.** The matrix $B$ is identifiable from $\chi$ and $V_0$.

# Identify $(\mathbb{D}, \mathbf{X})$ from data

(I) Find $\mathbb{D}$ from the given (or estimated) tail dependence matrix $\chi$.
(1) Calculate all maximum cliques.
– There is only one maximum clique $\Rightarrow$ this is $V_0$.
– There are various maximum cliques
$\Rightarrow$ there may be several DAGs with different $V_0$.
(2) Construct a reachability matrix $R$ (hence $\mathbb{D}$) from $\chi$ and every maximum clique $V_0$.
Use: for $k, i \in V$, $k \in \mathrm{An}(i)$ if and only if $\chi(k, i) > 0$ and $\chi(j, i) = \chi(j, k)\chi(k, i)$ for all $j \in V_0$ with $\chi(j, i)\chi(j, k) > 0$.
Thus we find for every node its ancestors, giving a possible $R$.

(II) Find $B$.
For all $j \in V_0$ we know $b_{jj} = 1$.
For $j \in \mathrm{An}(i)$ we have $b_{ji} = \chi(i, j)$.
Since $\sum_{l \in \mathrm{An}(i)} b_{li} = 1$,

$$b_{ii} = 1 - \sum_{j \in \mathrm{an}(i)} b_{ji} = 1 - \sum_{j \in \mathrm{an}(i)} b_{jj} \chi(i, j).$$

# References


- Gissibl, N. and Klüppelberg, C. (2015) Max-linear models on directed acyclic graphs. Submitted. `http://arxiv.org/abs/1512.07522`
- Gissibl, N., Klüppelberg, C. and Otto, M. (2016) Tail dependence of max-linear models on directed acyclic graphs. In preparation.
- Lauritzen, S.L. (1996). *Graphical Models.* Oxford University Press.
- Pearl, J. (2009) *Causality: Models, Reasoning, and Inference (2nd Ed.)* Cambridge University Press.