Introducing complex dependence structures into supervised component–based models

Jocelyn CHAUVET

work supervised by
Catherine TROTTIER and **Xavier BRY**

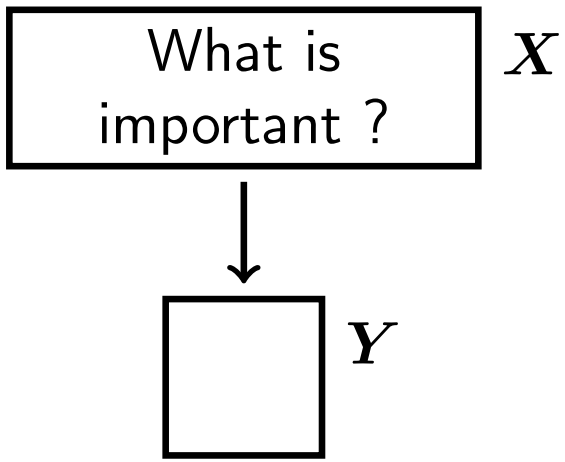
Montpellier
April 19^{th}, 2019

General context

$\hookrightarrow$ Why regularise a regression model ?

Fuzzy conceptual model, large amount of variables

- Ill–conditioned matrix (almost singular)
  $\hookrightarrow$ Instability of coefficients
- High dimensional data ($p > n$)
  $\hookrightarrow$ Multicollinearity, singularity

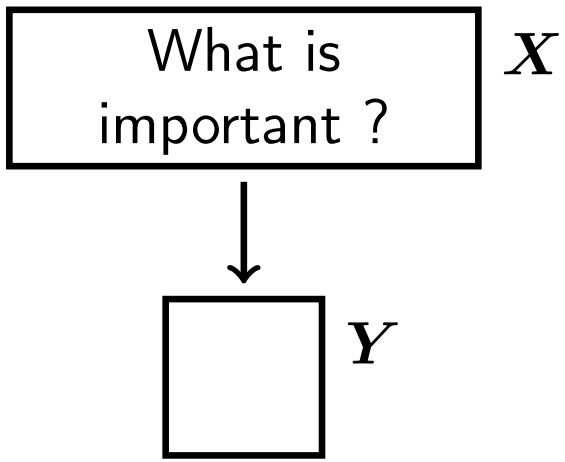


General context

$\hookrightarrow$ Why regularise a regression model ?

Fuzzy conceptual model, large amount of variables

- Ill–conditioned matrix (almost singular)
  $\hookrightarrow$ Instability of coefficients
- High dimensional data ($p > n$)
  $\hookrightarrow$ Multicollinearity, singularity

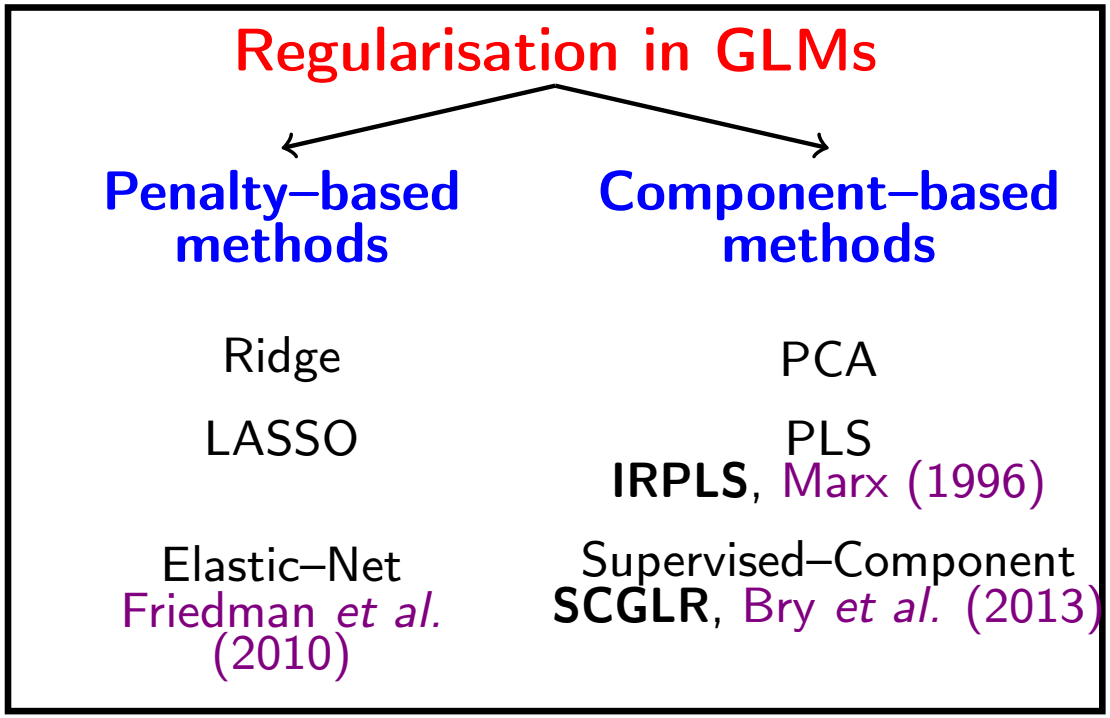


Regularisation (definition)

Introduction of additional criteria besides the Goodness–of–Fit in the estimation process in order to

- solve an ill–posed problem
- prevent overfitting

Existing methods



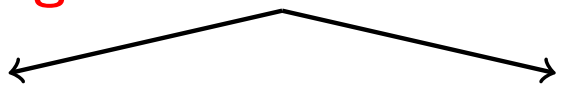
Existing methods

Regularisation in GLMMs (GLMs + random effects)

Penalty–based methods:

- LMM–Ridge — Eliot *et al.* (2011)
- GLMM–LASSO — Groll and Tutz (2014)

Regularisation in GLMs



Penalty–based methods	**Component–based methods**
Ridge	PCA
LASSO	PLS **IRPLS**, Marx (1996)
Elastic–Net Friedman *et al.* (2010)	Supervised–Component **SCGLR**, Bry *et al.* (2013)

Component–based methods?

Existing methods

Regularisation in GLMMs (GLMs + random effects)

Penalty–based methods:

- LMM–Ridge — Eliot *et al.* (2011)
- GLMM–LASSO — Groll and Tutz (2014)

Regularisation in GLMs

Penalty–based methods	**Component–based methods**
Ridge	PCA
LASSO	PLS **IRPLS**, Marx (1996)
Elastic–Net Friedman *et al.* (2010)	Supervised–Component **SCGLR**, Bry *et al.* (2013)

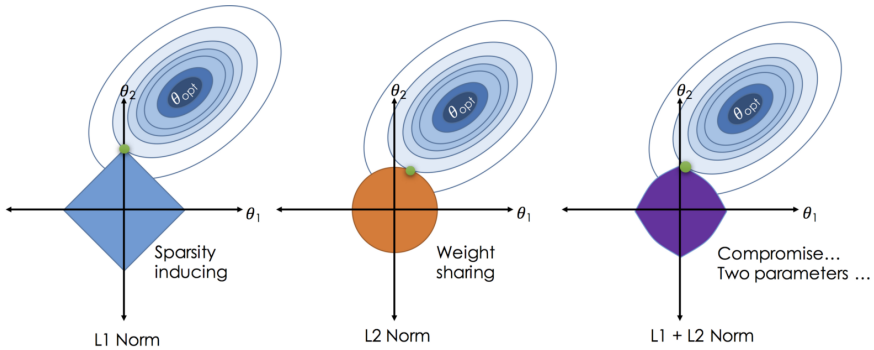
Component–based methods?

Extension of SCGLR to the GLMMs:
Mixed–SCGLR

Penalty–based regularisation

Penalised log–likelihood

$$\ell(\boldsymbol{\beta}; \boldsymbol{y}) - \lambda \, \text{pen}(\boldsymbol{\beta})$$



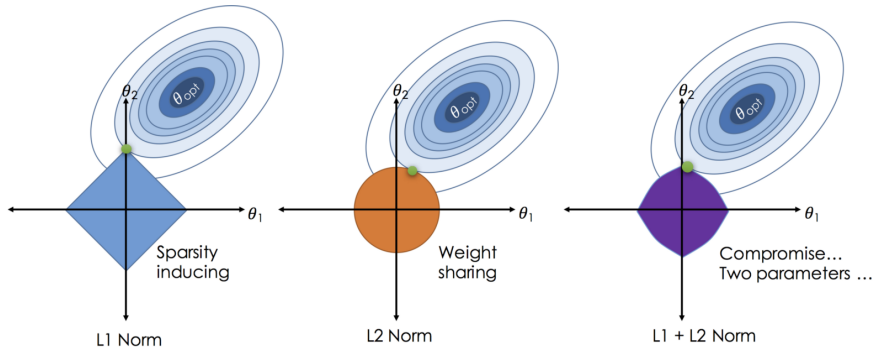
- **LASSO:** $\text{pen}(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_1$
  - ↪ Sparse solutions
  - ↪ Variable selection
- **Ridge:** $\text{pen}(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_2^2$
  - ↪ Shrinkage towards 0
  - ↪ Reduce the estimator's variance
- **Elastic–net:**
  $\text{pen}(\boldsymbol{\beta}) = (1-\alpha)\|\boldsymbol{\beta}\|_1 + \alpha\|\boldsymbol{\beta}\|_2^2$
  - ↪ Variable selection
  - ↪ Grouping effect

Penalty–based regularisation

Penalised log–likelihood

$$\ell(\boldsymbol{\beta}; \boldsymbol{y}) - \lambda \, \text{pen}(\boldsymbol{\beta})$$

- **LASSO:** $\text{pen}(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_1$
  - ↪ Sparse solutions
  - ↪ Variable selection
- **Ridge:** $\text{pen}(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_2^2$
  - ↪ Shrinkage towards 0
  - ↪ Reduce the estimator's variance
- **Elastic–net:**
  $\text{pen}(\boldsymbol{\beta}) = (1-\alpha)\|\boldsymbol{\beta}\|_1 + \alpha\|\boldsymbol{\beta}\|_2^2$
  - ↪ Variable selection
  - ↪ Grouping effect



Framework of interest

- Many highly correlated explanatory variables
- Proxies to latent phenomena to be found and interpreted

⇓

Component–based approaches

Component–based regularisation

Components = synthetic variables $\left\{ \boldsymbol{f_h} = \boldsymbol{X u_h} \,|\, h = 1, \ldots, H \right\}$

$$\boldsymbol{u_h} = \begin{cases} \arg \max\limits_{\boldsymbol{u} \in \mathbb{R}^p} \ \text{crit}(\boldsymbol{u}) \\ \text{w.r.t. } \|\boldsymbol{u}\| = 1 \text{ and } \boldsymbol{Xu} \perp \boldsymbol{Xu_1}, \ldots, \boldsymbol{Xu_{h-1}} \end{cases}$$

Component–based regularisation

Components = synthetic variables $\left\{ \boldsymbol{f_h} = \boldsymbol{X u_h} \,|\, h = 1, \ldots, H \right\}$

$$\boldsymbol{u_h} = \begin{cases} \arg \max\limits_{\boldsymbol{u} \in \mathbb{R}^p} \ \text{crit}(\boldsymbol{u}) \\ \text{w.r.t. } \|\boldsymbol{u}\| = 1 \text{ and } \boldsymbol{Xu} \perp \boldsymbol{Xu_1}, \ldots, \boldsymbol{Xu_{h-1}} \end{cases}$$

PCA

$$\text{crit}(\boldsymbol{u}) = \underbrace{\|\boldsymbol{Xu}\|_2^2}_{\substack{\text{Component Variance} \\ \textbf{(CV)}}}$$

↪ Information in $\boldsymbol{X}$ ✔

↪ Link between $\boldsymbol{X}$ and $\boldsymbol{y}$ ✗

PLS

$$\text{crit}(\boldsymbol{u}) = \underbrace{\|\boldsymbol{Xu}\|_2^2}_{\substack{\text{Comp. Var.} \\ \textbf{(CV)}}} \underbrace{\|\boldsymbol{y}\|_2^2 \cos^2\left(\boldsymbol{y}, \text{span}\left\{\boldsymbol{Xu}\right\}\right)}_{\substack{\text{Goodness–of–Fit} \\ \textbf{(GoF)}}}$$

↪ Information in $\boldsymbol{X}$ ✔

↪ Link between $\boldsymbol{X}$ and $\boldsymbol{y}$ ✔

Component–based regularisation

Components = synthetic variables $\left\{ \boldsymbol{f_h} = \boldsymbol{X u_h} \,|\, h = 1, \ldots, H \right\}$

$$\boldsymbol{u_h} = \begin{cases} \arg \max\limits_{\boldsymbol{u} \in \mathbb{R}^p} \ \text{crit}(\boldsymbol{u}) \\ \text{w.r.t. } \|\boldsymbol{u}\| = 1 \text{ and } \boldsymbol{Xu} \perp \boldsymbol{Xu_1}, \ldots, \boldsymbol{Xu_{h-1}} \end{cases}$$

PCA

$$\text{crit}(\boldsymbol{u}) = \underbrace{\|\boldsymbol{Xu}\|_2^2}_{\substack{\text{Component Variance} \\ \textbf{(CV)}}}$$

↪ Information in $\boldsymbol{X}$ ✔

↪ Link between $\boldsymbol{X}$ and $\boldsymbol{y}$ ✗

PLS

$$\text{crit}(\boldsymbol{u}) = \underbrace{\|\boldsymbol{Xu}\|_2^2}_{\substack{\text{Comp. Var.} \\ \textbf{(CV)}}} \underbrace{\|\boldsymbol{y}\|_2^2 \cos^2\left(\boldsymbol{y}, \text{span}\left\{\boldsymbol{Xu}\right\}\right)}_{\substack{\text{Goodness–of–Fit} \\ \textbf{(GoF)}}}$$

↪ Information in $\boldsymbol{X}$ ✔

↪ Link between $\boldsymbol{X}$ and $\boldsymbol{y}$ ✔

A "flexible PLS"

$$\text{crit}(\boldsymbol{u}) = \left[\text{CV}(\boldsymbol{u})\right]^{\boldsymbol{s}} \left[\text{GoF}(\boldsymbol{u})\right]^{\boldsymbol{1-s}}$$

$\boldsymbol{s} \in [\boldsymbol{0}, \boldsymbol{1}]$ a trade–off parameter

Structural relevance

↪ Introduced by Bry and Verron (2015)

Supervised Components via the Structural Relevance

$$\text{crit}(\boldsymbol{u}) = \left[\text{SR}(\boldsymbol{u})\right]^{\boldsymbol{s}} \left[\text{GoF}(\boldsymbol{u})\right]^{\boldsymbol{1-s}}$$

$$\text{SR}(\boldsymbol{u}) = \phi_{\boldsymbol{l}}(\boldsymbol{u}) = \left(\sum_{j=1}^{p} \left[\text{cor}^2 \left(\boldsymbol{X}\boldsymbol{u}, \boldsymbol{x_j} \right) \right]^{\boldsymbol{l}} \right)^{\frac{\boldsymbol{1}}{\boldsymbol{l}}}$$

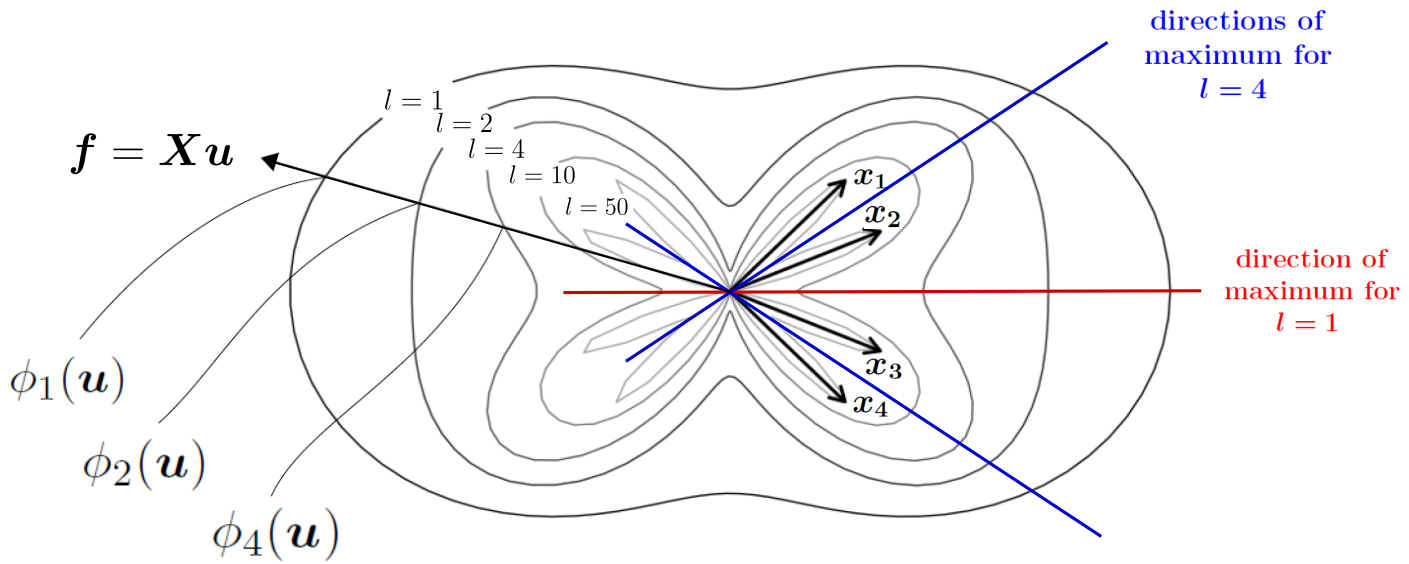
Structural relevance

↪ Introduced by Bry and Verron (2015)

Supervised Components via the Structural Relevance

$$\text{crit}(\boldsymbol{u}) = \left[\text{SR}(\boldsymbol{u})\right]^{\boldsymbol{s}} \left[\text{GoF}(\boldsymbol{u})\right]^{\boldsymbol{1-s}}$$

$$\text{SR}(\boldsymbol{u}) = \phi_{\boldsymbol{l}}(\boldsymbol{u}) = \left(\sum_{j=1}^{p} \left[\text{cor}^2 \left(\boldsymbol{X}\boldsymbol{u}, \boldsymbol{x_j} \right) \right]^{\boldsymbol{l}} \right)^{\frac{\boldsymbol{1}}{\boldsymbol{l}}}$$

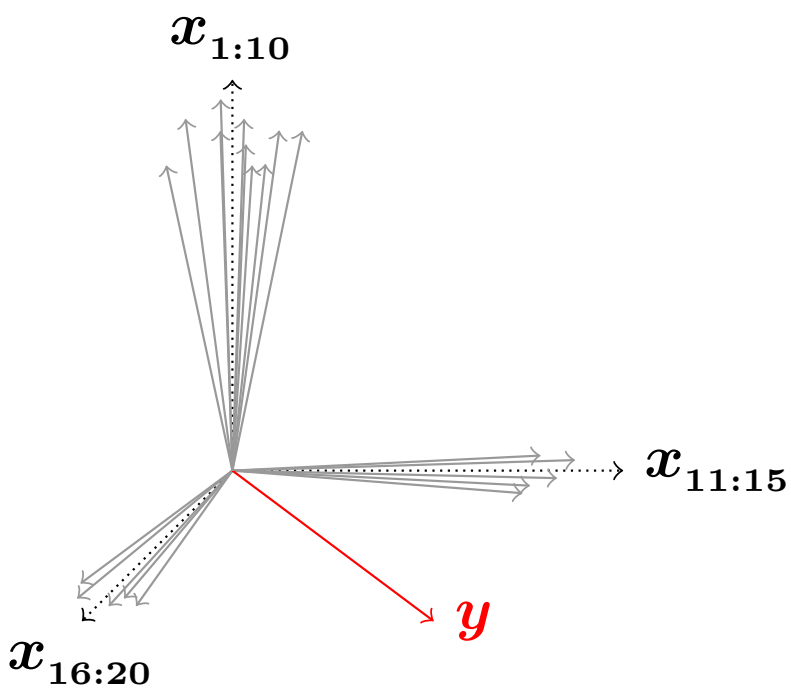


Flash method–comparison

A simple Gaussian model

- $\boldsymbol{y} \sim \mathcal{N}_n\left(\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta},\, \boldsymbol{\Sigma} = \boldsymbol{I}_n\right)$
- $\boldsymbol{X} = \Big[\underbrace{\boldsymbol{x_1} \ldots\ldots\ldots\ldots \boldsymbol{x_{10}}}_{\text{large bundle}} \; \underbrace{\boldsymbol{x_{11}} \ldots\ldots \boldsymbol{x_{15}}}_{\text{small bundle}} \; \underbrace{\boldsymbol{x_{16}} \ldots\ldots \boldsymbol{x_{20}}}_{\text{small bundle}} \Big]$

  ↪ nuisance ; ↪ predicts y ; ↪ predicts y



☛ PCR vs PLSR vs "Supervised Component Regression"

Issues

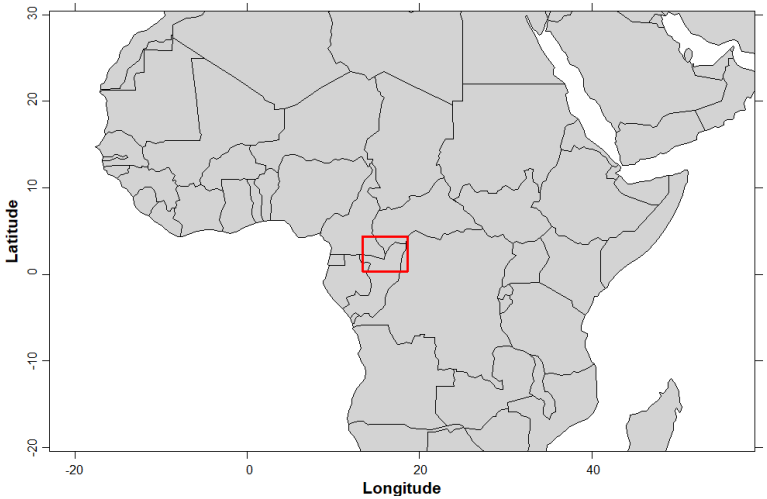
- **Non–independent observations**
  - ↪ Grouped and panel data
    - ⇒ From GLM to GLMM (use of random effects)
- **Multivariate framework**
  - ↪ Several responses of various types $\boldsymbol{Y} = \left[\boldsymbol{y_1} \,|\, \ldots \,|\, \boldsymbol{y_q} \right]$
- **Additional explanatory variables**
  - ↪ with little redundancy
  - ↪ Requiring no regularisation

1. Supervised Component–based regularisation for multivariate GLMMs
   - The Congo–Basin floristic data
   - The mixed–SCGLR method
   - Simulation study
   - Results on the floristic data
2. Introducing a time–specific random effect
3. Perspectives

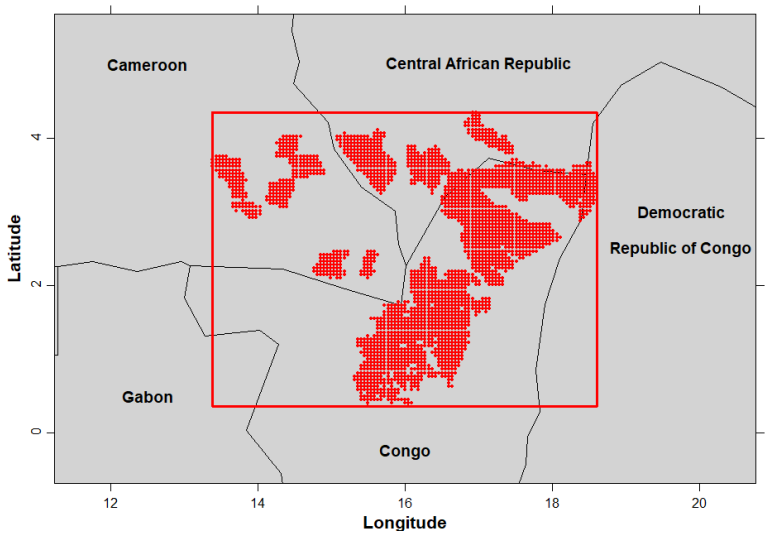
Floristic data: a multivariate GLM ?

Problem: Model and predict the abundance of tree species in the tropical moist forest of the Congo–Basin

The Congo–Basin



2615 land–plots



Responses:

$q = 8$ abundances of selected tree species

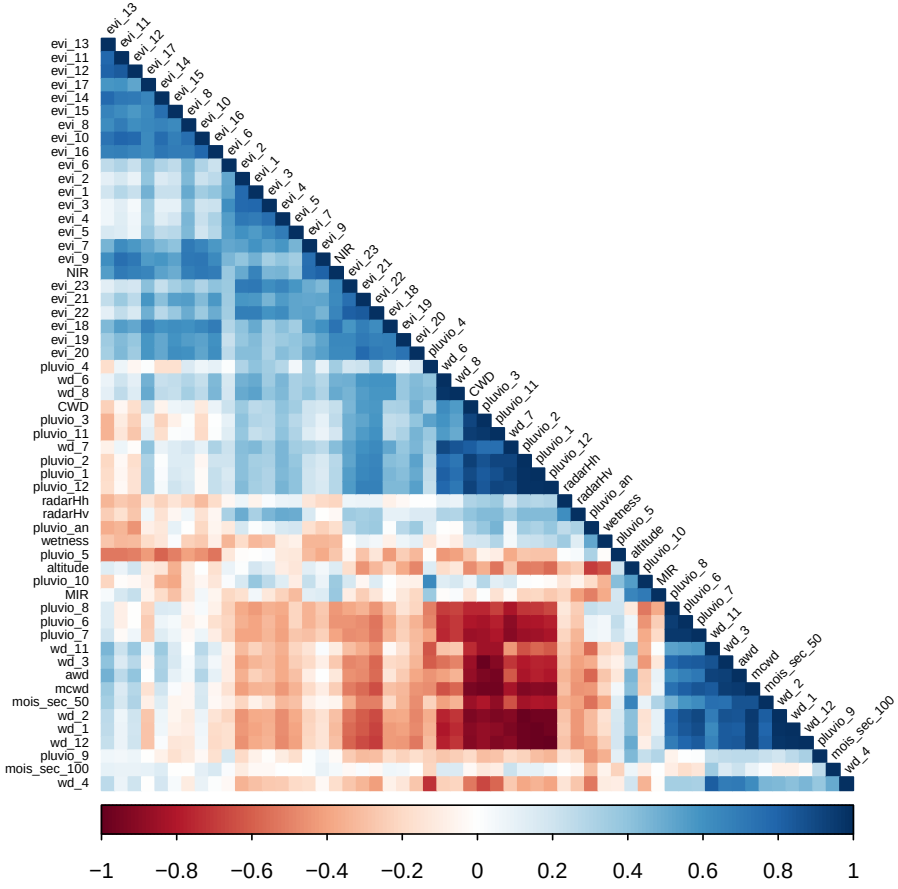
(i.e. multivariate count responses)

$\Downarrow$

Multivariate GLM

Regularisation needed

Correlation–heatmap



Explanatory variables:

- $p = 56$ **highly correlated** environmental variables
  - ↪ Information **redundancy**
  - ↪ Model **instability**
- $r = 2$ **additional covariates** (geology and anthropogenic interference)

⇓

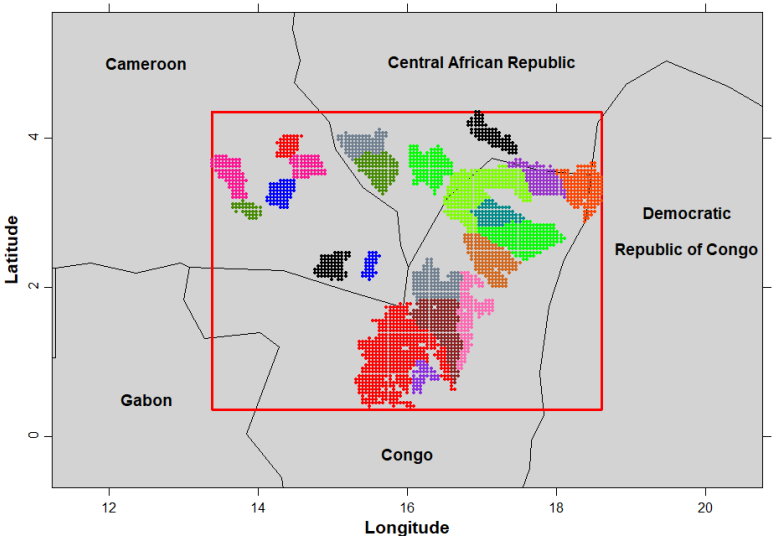
Regularisation

via **supervised components**

common to all responses

Random effects needed: GLMM

22 forest concessions



SCGLR:

- The land–plots are assumed **independent**...
- ...yet they are grouped into **22 concessions**

Mixed–SCGLR:

- Within–group dependence modelled by a **random effect**

$\Downarrow$

Multivariate GLMM

Mixed–SCGLR (1)

Notations:

- $\boldsymbol{Y}_{n\times q}$: matrix of q response vectors $\boldsymbol{y_1}, \ldots, \boldsymbol{y_q}$
- $\boldsymbol{X}_{n\times p}$: explanatory variables $\boldsymbol{x_1}, \ldots, \boldsymbol{x_p}$ (many, redundant)
- $\boldsymbol{A}_{n\times r}$: additional covariates $\boldsymbol{a_1}, \ldots, \boldsymbol{a_r}$ (few, not redundant)
- $\boldsymbol{U}_{n\times N}$: design matrix of the random effects

Mixed–SCGLR (1)

Notations:

- $\boldsymbol{Y}_{n\times q}$: matrix of q response vectors $\boldsymbol{y_1}, \ldots, \boldsymbol{y_q}$
- $\boldsymbol{X}_{n\times p}$: explanatory variables $\boldsymbol{x_1}, \ldots, \boldsymbol{x_p}$ (many, redundant)
- $\boldsymbol{A}_{n\times r}$: additional covariates $\boldsymbol{a_1}, \ldots, \boldsymbol{a_r}$ (few, not redundant)
- $\boldsymbol{U}_{n\times N}$: design matrix of the random effects

Single component multivariate GLMM

For each $k \in \{1, \ldots, q\}$,

$$g_k\left(\mathbb{E}\left(\boldsymbol{Y_k} \,|\, \boldsymbol{\xi_k}\right)\right) = \boldsymbol{\eta_k}$$

$$\boldsymbol{\eta_k} = (\boldsymbol{Xu})\gamma_k + \boldsymbol{A\delta_k} + \boldsymbol{U\xi_k}$$

$$\boldsymbol{\xi_k} \overset{\text{ind.}}{\sim} \mathcal{N}_N\left(\mathbf{0},\ \sigma_k^2 \boldsymbol{I}_N\right), \text{ with } N \text{ the number of groups}$$

Mixed–SCGLR (2)

Estimation method: Iterative procedure based on a linearisation of the model

$\hookrightarrow$ Pseudo–responses: $\boldsymbol{z_k}$

"Linearised" model

$$\boldsymbol{z_k} = \underbrace{(\boldsymbol{Xu})\gamma_k + \boldsymbol{A\delta_k} + \boldsymbol{U\xi_k}}_{\boldsymbol{\eta_k}} + \boldsymbol{e_k} \quad \text{with:} \quad \begin{cases} \mathbb{E}\left(\boldsymbol{e_k} \,|\, \boldsymbol{\xi_k}\right) = \mathbf{0} \\ \mathbb{V}\left(\boldsymbol{e_k} \,|\, \boldsymbol{\xi_k}\right) =: \boldsymbol{W_k}^{-1} \end{cases}$$

Mixed–SCGLR (2)

Estimation method: Iterative procedure based on a linearisation of the model

$\hookrightarrow$ Pseudo–responses: $\boldsymbol{z_k}$

"Linearised" model

$$\boldsymbol{z_k} = \underbrace{(\boldsymbol{Xu})\gamma_k + \boldsymbol{A\delta_k} + \boldsymbol{U\xi_k}}_{\boldsymbol{\eta_k}} + \boldsymbol{e_k} \quad \text{with:} \quad \begin{cases} \mathbb{E}\left(\boldsymbol{e_k} \,|\, \boldsymbol{\xi_k}\right) = \boldsymbol{0} \\ \mathbb{V}\left(\boldsymbol{e_k} \,|\, \boldsymbol{\xi_k}\right) =: \boldsymbol{W_k^{-1}} \end{cases}$$

Alternate procedure

(i) Given γ_k, $\boldsymbol{\delta_k}$, $\boldsymbol{\xi_k}$ and σ_k^2, we compute the component $\boldsymbol{f} = \boldsymbol{Xu}$

(ii) Given $\boldsymbol{u}$, we estimate γ_k, $\boldsymbol{\delta_k}$, $\boldsymbol{\xi_k}$ and σ_k^2

$\hookrightarrow$ Schall's algorithm, Henderson's system

Mixed–SCGLR (3)

Step (ii)

Henderson's systems

$\hookrightarrow$ Given $\boldsymbol{f} = \boldsymbol{X}\boldsymbol{u}$, for each $k \in \{1, \ldots, q\}$:

$$\begin{pmatrix} \boldsymbol{f}^\mathsf{T}\boldsymbol{W_k}\boldsymbol{f} & \boldsymbol{f}^\mathsf{T}\boldsymbol{W_k}\boldsymbol{A} & \boldsymbol{f}^\mathsf{T}\boldsymbol{W_k}\boldsymbol{U} \\ \boldsymbol{A}^\mathsf{T}\boldsymbol{W_k}\boldsymbol{f} & \boldsymbol{A}^\mathsf{T}\boldsymbol{W_k}\boldsymbol{A} & \boldsymbol{A}^\mathsf{T}\boldsymbol{W_k}\boldsymbol{U} \\ \boldsymbol{U}^\mathsf{T}\boldsymbol{W_k}\boldsymbol{f} & \boldsymbol{U}^\mathsf{T}\boldsymbol{W_k}\boldsymbol{A} & \boldsymbol{U}^\mathsf{T}\boldsymbol{W_k}\boldsymbol{U} + \boldsymbol{D_k}^{-1} \end{pmatrix} \begin{pmatrix} \gamma_k \\ \boldsymbol{\delta_k} \\ \boldsymbol{\xi_k} \end{pmatrix} = \begin{pmatrix} \boldsymbol{f}^\mathsf{T}\boldsymbol{W_k}\,\boldsymbol{z_k} \\ \boldsymbol{A}^\mathsf{T}\boldsymbol{W_k}\,\boldsymbol{z_k} \\ \boldsymbol{U}^\mathsf{T}\boldsymbol{W_k}\,\boldsymbol{z_k} \end{pmatrix}$$

Update variance components

$$\sigma_k^2 \longleftarrow \frac{\boldsymbol{\xi_k}^\mathsf{T}\,\boldsymbol{\xi_k}}{N - \frac{1}{\sigma_k^2}\,\mathsf{Trace}\left[\left(\boldsymbol{U}^\mathsf{T}\boldsymbol{W_k}\boldsymbol{U} + \boldsymbol{D_k}^{-1}\right)^{-1}\right]}$$

Mixed–SCGLR (4)

Step (i)

Goodness–of–Fit

$$\psi(\boldsymbol{u}) = \sum_{k=1}^{q} \left\| \Pi_{\text{span}\{\boldsymbol{X}\boldsymbol{u},\boldsymbol{A}\}} \, \boldsymbol{z_k} \right\|^2_{\boldsymbol{W_k}}$$

Mixed–SCGLR (4)

Step (i)

Goodness–of–Fit

$$\psi(\boldsymbol{u}) = \sum_{k=1}^{q} \left\| \Pi_{\text{span}\{\boldsymbol{X}\boldsymbol{u}, \boldsymbol{A}\}} \, \boldsymbol{z_k} \right\|^2_{\boldsymbol{W_k}}$$

Structural Relevance

$$\phi(\boldsymbol{u}) = \left(\sum_{j=1}^{p} \left[\text{cor}^2 \left(\boldsymbol{X}\boldsymbol{u}, \boldsymbol{x_j} \right) \right]^{\boldsymbol{l}} \right)^{\frac{1}{\boldsymbol{l}}}, \quad \text{with } \boldsymbol{l} \in [\mathbf{1}, +\infty)$$

Mixed–SCGLR (4)

Step (i)

Goodness–of–Fit

$$\psi(\boldsymbol{u}) = \sum_{k=1}^{q} \left\| \Pi_{\text{span}\{\boldsymbol{X}\boldsymbol{u}, \boldsymbol{A}\}} \, \boldsymbol{z_k} \right\|^2_{\boldsymbol{W_k}}$$

Structural Relevance

$$\phi(\boldsymbol{u}) = \left(\sum_{j=1}^{p} \left[\text{cor}^2 \left(\boldsymbol{X}\boldsymbol{u}, \boldsymbol{x_j} \right) \right]^{\boldsymbol{l}} \right)^{\frac{\mathbf{1}}{\boldsymbol{l}}}, \quad \text{with } \boldsymbol{l} \in [\mathbf{1}, +\infty)$$

Trade–off GoF/SR

$$\max \quad \left[\psi(\boldsymbol{u}) \right]^{\mathbf{1}-\boldsymbol{s}} \left[\phi(\boldsymbol{u}) \right]^{\boldsymbol{s}}, \quad \text{with } \boldsymbol{s} \in [\mathbf{0}, \mathbf{1}]$$

$$\text{w.r.t.} \quad \|\boldsymbol{u}\| = 1 \quad \text{(Identification constraint)}$$

Simulations (1)

Two random responses: $\boldsymbol{Y} = \left[\boldsymbol{y_1} \,|\, \boldsymbol{y_2} \right]$

Simulations (1)

Two random responses: $\boldsymbol{Y} = \left[\boldsymbol{y_1} \,|\, \boldsymbol{y_2} \right]$

Fixed effects:

- 30 explanatory variables $\mathcal{N}(0, 1)$:

$$\boldsymbol{X} = \Big[\underbrace{\boldsymbol{x_1} \ldots\ldots\ldots\ldots\ldots \boldsymbol{x_{15}}}_{\text{bundle } \boldsymbol{X_0} \text{ (large)}} \; \underbrace{\boldsymbol{x_{16}} \ldots\ldots\ldots\ldots \boldsymbol{x_{25}}}_{\text{bundle } \boldsymbol{X_1} \text{ (medium)}} \; \underbrace{\boldsymbol{x_{26}} \ldots\ldots \boldsymbol{x_{30}}}_{\text{bundle } \boldsymbol{X_2} \text{ (small)}} \Big]$$

$\hookrightarrow$ **nuisance** | $\hookrightarrow$ **predicts y_1** | $\hookrightarrow$ **predicts y_2**

- Within each bundle:

$$\text{cor}(\boldsymbol{x_j}, \boldsymbol{x_k}) = \begin{cases} \mathbf{1} & \text{if } j = k \\ \boldsymbol{\tau} & \text{if } j \neq k \end{cases} \quad \text{with } \boldsymbol{\tau} \in \{\mathbf{0.1}, \mathbf{0.3}, \mathbf{0.5}, \mathbf{0.7}, \mathbf{0.9}\}$$

Simulations (1)

Two random responses: $\boldsymbol{Y} = \left[\boldsymbol{y_1} \,|\, \boldsymbol{y_2} \right]$

Fixed effects:

- 30 explanatory variables $\mathcal{N}(0, 1)$:

$$\boldsymbol{X} = \Big[\underbrace{\boldsymbol{x_1} \ldots\ldots\ldots\ldots\ldots \boldsymbol{x_{15}}}_{\text{bundle } \boldsymbol{X_0} \text{ (large)}} \; \underbrace{\boldsymbol{x_{16}} \ldots\ldots\ldots\ldots \boldsymbol{x_{25}}}_{\text{bundle } \boldsymbol{X_1} \text{ (medium)}} \; \underbrace{\boldsymbol{x_{26}} \ldots\ldots \boldsymbol{x_{30}}}_{\text{bundle } \boldsymbol{X_2} \text{ (small)}} \Big]$$

$\hookrightarrow$ **nuisance** | $\hookrightarrow$ **predicts y_1** | $\hookrightarrow$ **predicts y_2**

- Within each bundle:

$$\text{cor}(\boldsymbol{x_j}, \boldsymbol{x_k}) = \begin{cases} \mathbf{1} & \text{if } j = k \\ \boldsymbol{\tau} & \text{if } j \neq k \end{cases} \quad \text{with } \boldsymbol{\tau} \in \{\mathbf{0.1}, \mathbf{0.3}, \mathbf{0.5}, \mathbf{0.7}, \mathbf{0.9}\}$$

Random effects:

- $\boldsymbol{N = 10}$ groups and $\boldsymbol{R = 10}$ units per group

  $\hookrightarrow$ Design matrix $\boldsymbol{U} = \boldsymbol{I}_N \otimes \mathbf{1}_R$

Simulations (2)

Model

$$\mathcal{M}: \begin{cases} \boldsymbol{y_1} = \boldsymbol{X}\boldsymbol{\beta_1} + \boldsymbol{U}\boldsymbol{\xi_1} + \boldsymbol{\varepsilon_1} \\ \boldsymbol{y_2} = \boldsymbol{X}\boldsymbol{\beta_2} + \boldsymbol{U}\boldsymbol{\xi_2} + \boldsymbol{\varepsilon_2} \end{cases} \quad \text{with}$$

$$\boldsymbol{\xi_k} \sim \mathcal{N}_N\left(\mathbf{0}, \boldsymbol{I}_N\right) \text{ and } \boldsymbol{\varepsilon_k} \sim \mathcal{N}_{NR}\left(\mathbf{0}, \boldsymbol{I}_{NR}\right)$$

100 samples of model $\mathcal{M}$ for each value of τ

Comparison with:

LMM–Ridge (2011)

Eliot et al.

↪ EM algorithm

↪ GCV at each step

GLMM–LASSO (2014)

Groll, A. and Tutz, G.

↪ Laplace approximation

↪ Coordinate Gradient Descent

Simulations (3)

τ	LMM (No reg.)	GLMM–LASSO $\lambda^{\star}_{\text{lasso}}$ (shrinkage)	LMM–Ridge $\lambda^{\star}_{\text{ridge}}$ (shrinkage)	Mixed–SCGLR ($l = 4$) $H^{\star}$ (nb comp.)	$s^{\star}$ (trade–off)
0.1		65	24	25	0.50
0.3		92	54	5	0.58
0.5		124	73	3	0.70
0.7		163	78	3	0.73
0.9		175	85	2	0.80

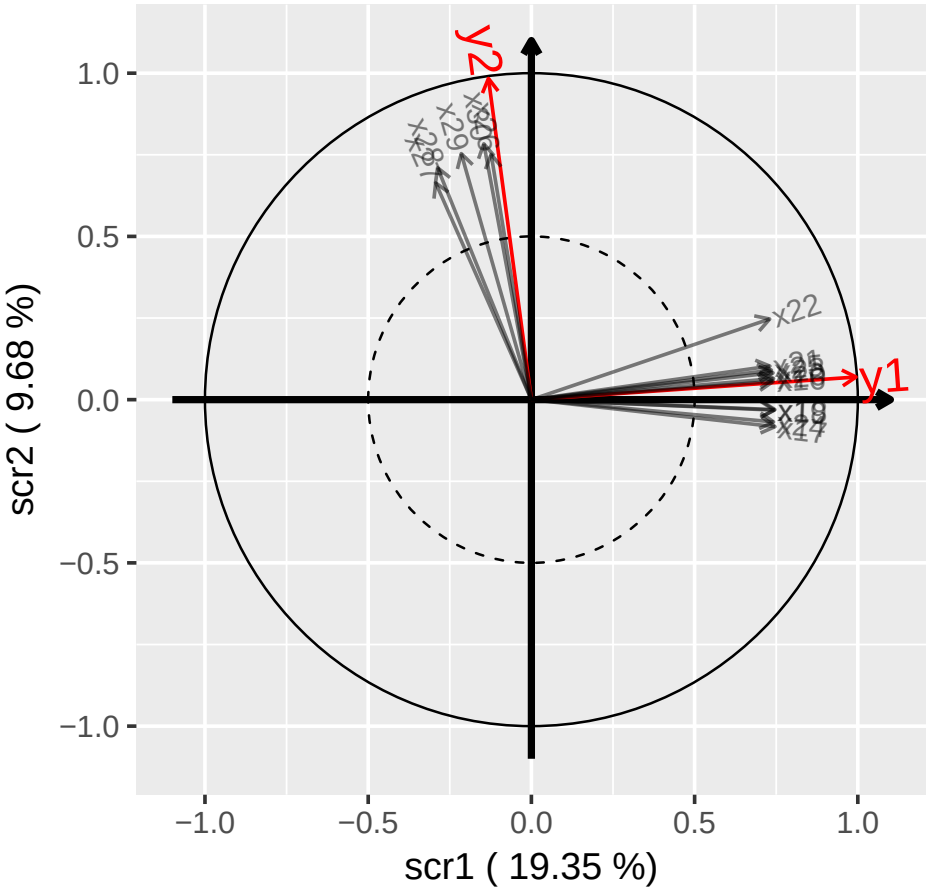
τ	Ave $\left[\max\left(\frac{\|\widehat{\boldsymbol{\beta_1}} - \boldsymbol{\beta_1}\|_2^2}{\|\boldsymbol{\beta_1}\|_2^2}, \frac{\|\widehat{\boldsymbol{\beta_2}} - \boldsymbol{\beta_2}\|_2^2}{\|\boldsymbol{\beta_2}\|_2^2}\right)\right]$			
0.1	0.12	0.05	0.08	0.12
0.3	**0.33**	0.12	0.13	0.10
0.5	**0.61**	0.20	0.16	**0.07**
0.7	**1.32**	0.25	0.20	**0.06**
0.9	**4.62**	0.26	0.31	**0.05**

Simulations (4)

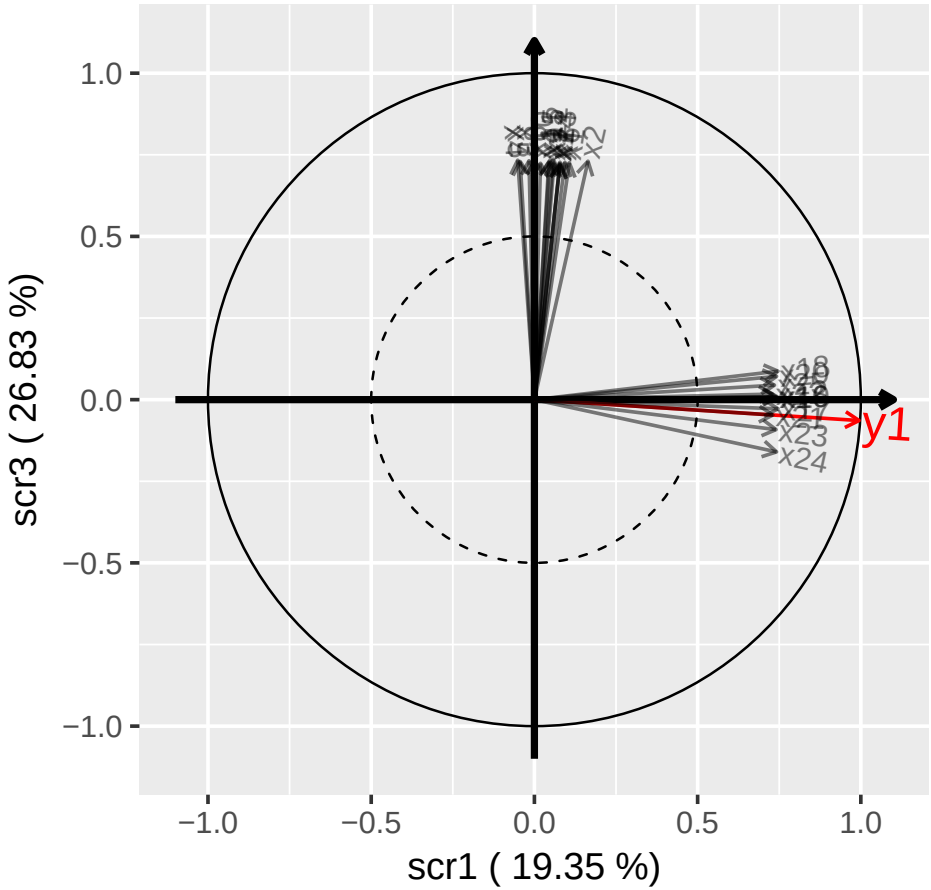
Redundancy level

$\tau = 0.5$

Component plane (1,2)



Component plane (1,3)



Simulations (5)

Model

$$\mathcal{M}: \begin{cases} \boldsymbol{y_1} \sim \mathcal{B}\text{er}\Big(\boldsymbol{p} = \text{logit}^{-1}\left[\boldsymbol{X\beta_1} + \boldsymbol{U\xi_1}\right] \Big) \\ \boldsymbol{y_2} \sim \mathcal{P}\text{oi}\Big(\boldsymbol{\lambda} = \exp\left[\boldsymbol{X\beta_2} + \boldsymbol{U\xi_2}\right] \Big), \end{cases}$$

Fixed–effect squared relative errors:

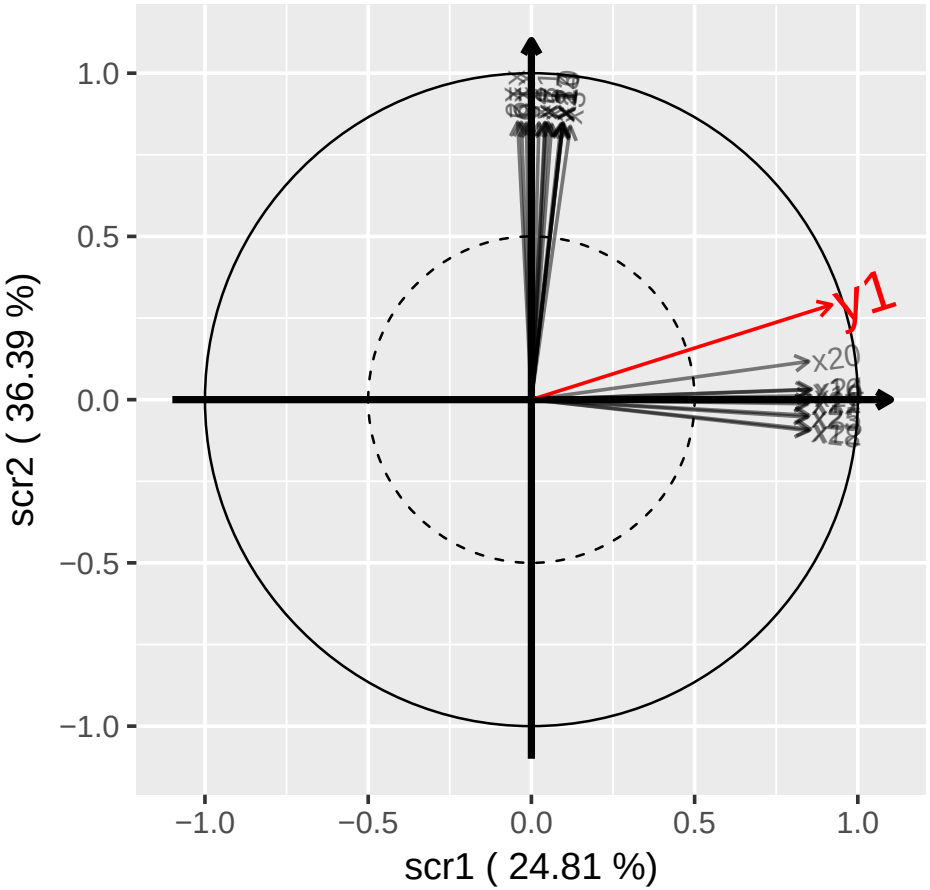
	GLMM (no reg.)		GLMM–LASSO		mixed–SCGLR	
τ	$\mathcal{B}$er	$\mathcal{P}$oi	$\mathcal{B}$er	$\mathcal{P}$oi	$\mathcal{B}$er	$\mathcal{P}$oi
0.1	316.48	0.54	8.61	0.30	14.71	0.46
0.3	398.78	0.64	9.23	0.36	7.21	0.21
0.5	576.68	0.87	14.48	0.44	2.01	0.09
0.7	886.04	1.28	17.37	0.47	1.50	0.07
0.9	2840.10	3.72	17.24	0.59	1.31	0.05

Simulations (6)

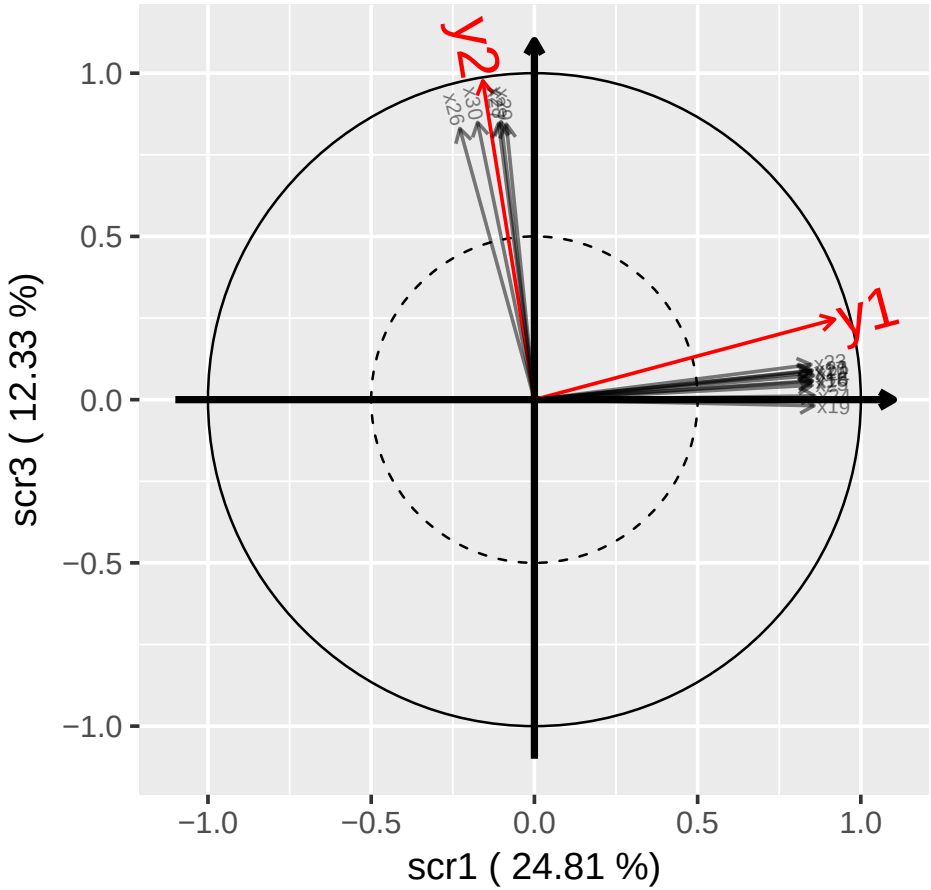
Redundancy level

$\tau = 0.9$

Component plane (1,2)



Component plane (1,3)



Results on the floristic data (1)

- $n = 2615$ land–plots, divided in $N = 22$ forest concessions (considered as groups)
- $q = 8$ abundances of tree genera (responses Y)
- $p = 56$ explanatory variables (X)
- $r = 2$ additional covariates (A)

Results on the floristic data (1)

- $n = 2615$ land–plots, divided in $N = 22$ forest concessions (considered as groups)
- $q = 8$ abundances of tree genera (responses $\boldsymbol{Y}$)
- $p = 56$ explanatory variables ($\boldsymbol{X}$)
- $r = 2$ additional covariates ($\boldsymbol{A}$)

Model

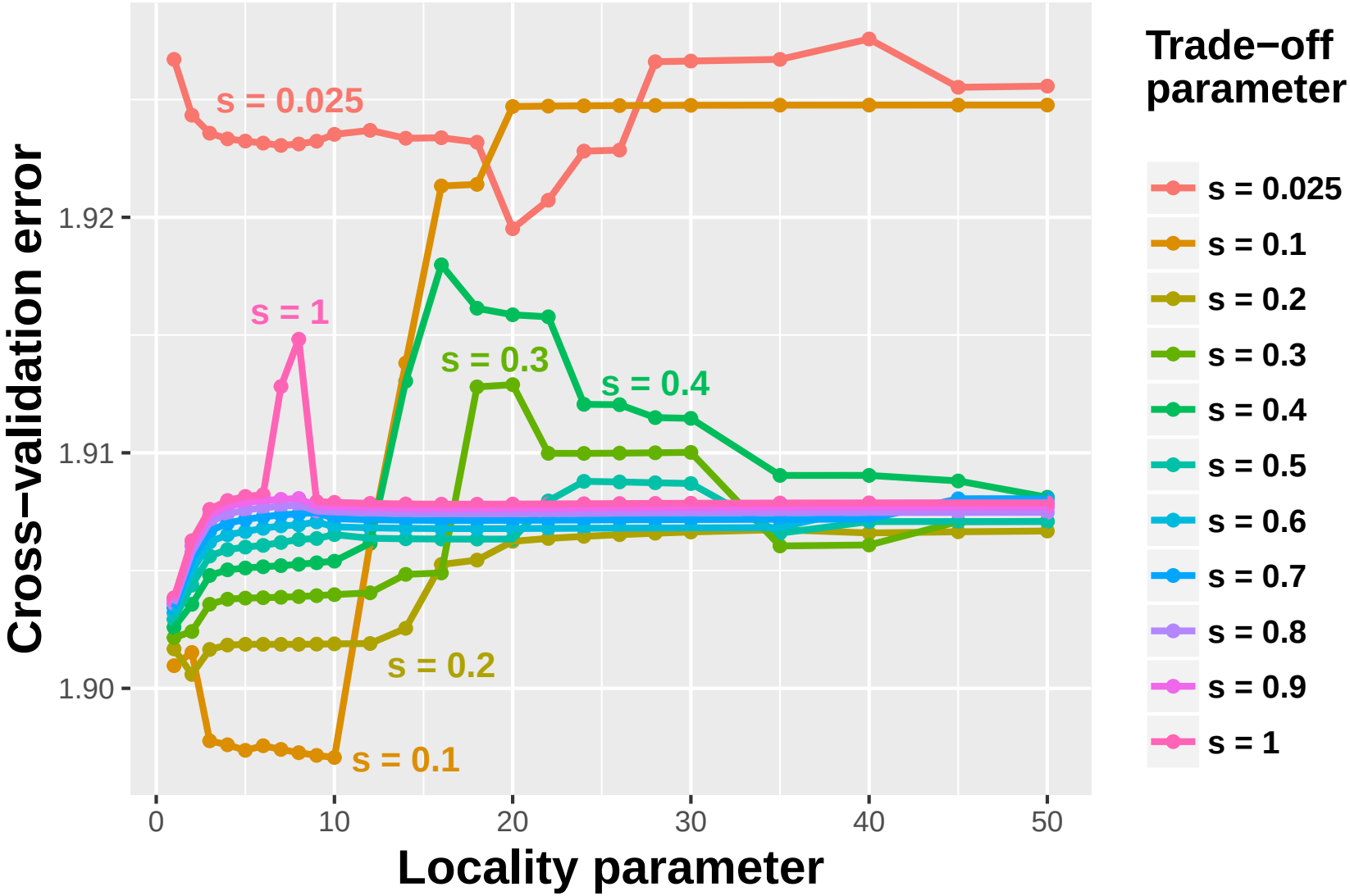
Abundance of tree species: count data

$\hookrightarrow$ **Poisson regression with log link**

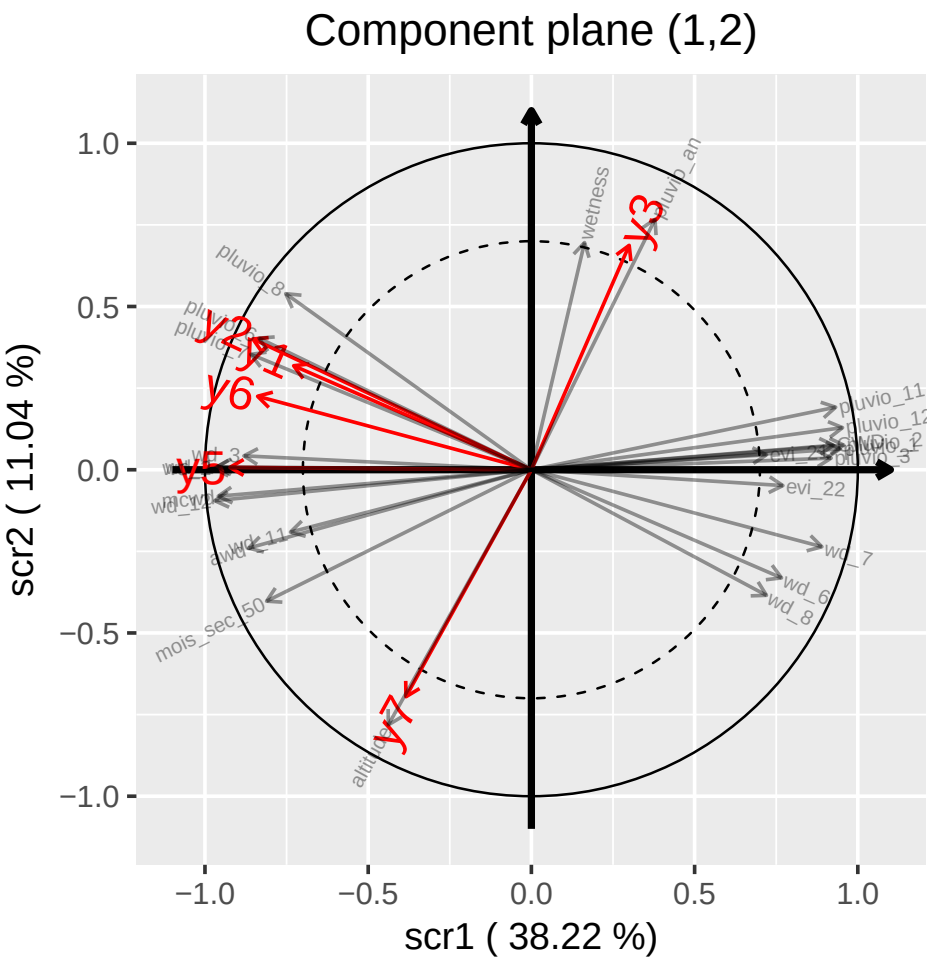
$$\boldsymbol{y_k} \sim \mathcal{P}\text{oi}\left(\boldsymbol{\lambda} = \exp\left[\boldsymbol{\eta_k}\right]\right)$$

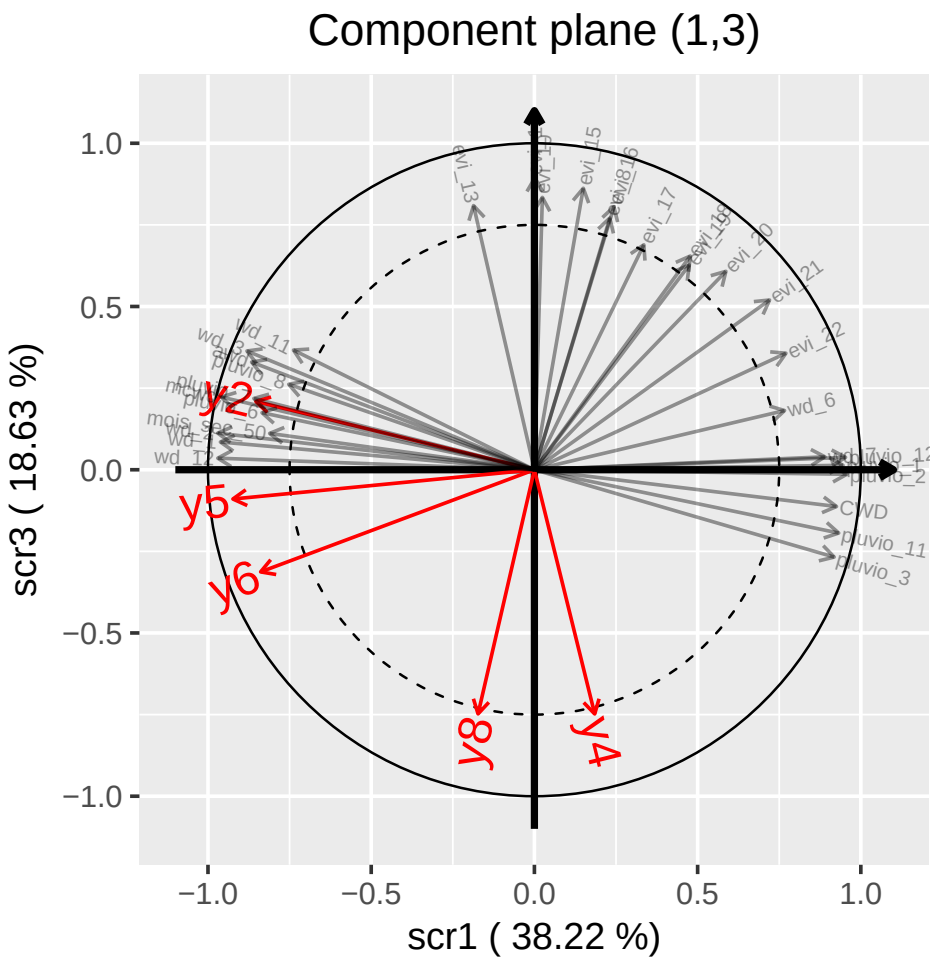
$$\boldsymbol{\eta_k} = \sum_{h=1}^{H} \left(\boldsymbol{X u_h}\right) \gamma_{k,h} + \boldsymbol{A \delta_k} + \boldsymbol{U \xi_k}$$

Results on the floristic data (2), $H^\star = 4$

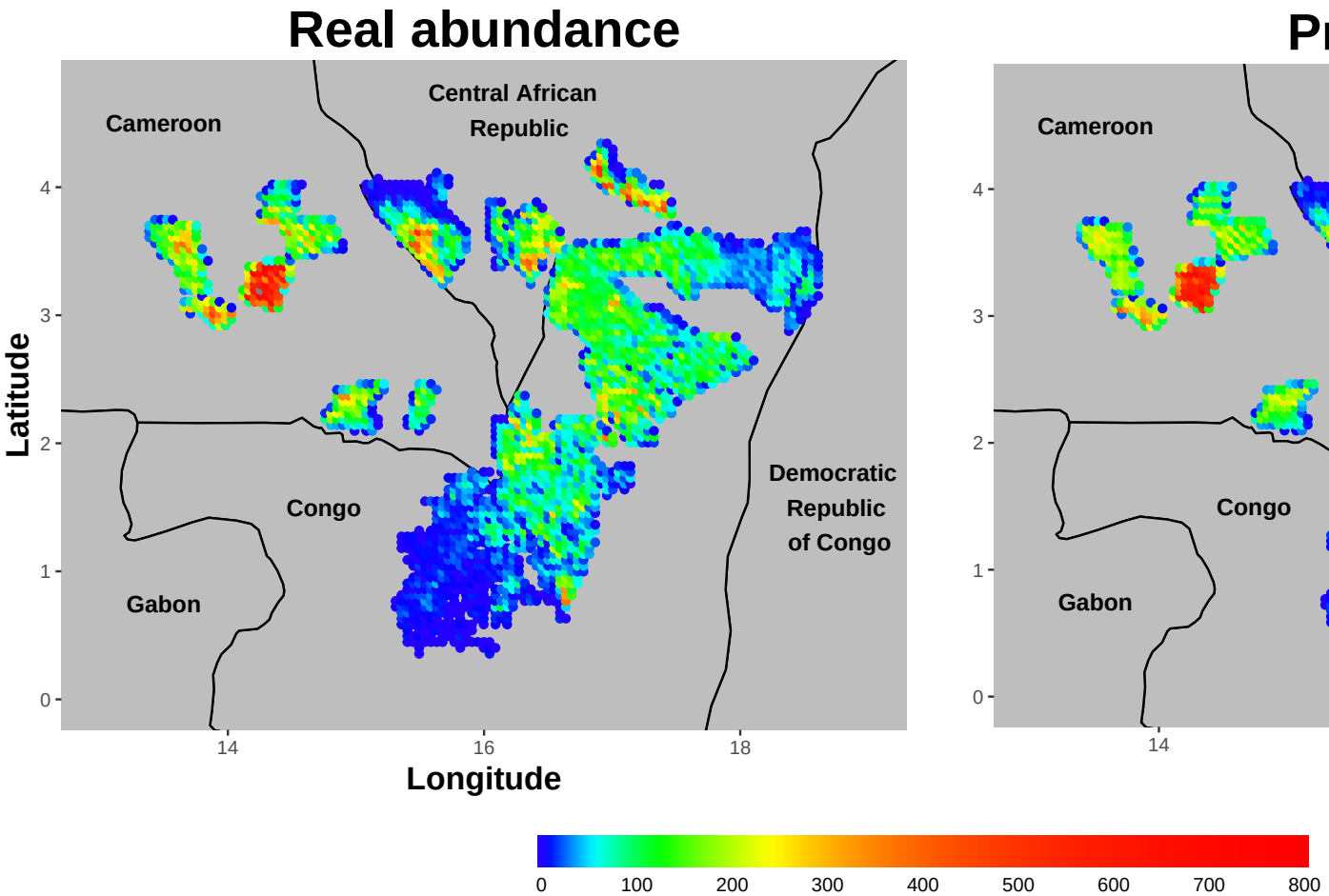


Results on the floristic data (3)





Results on the floristic data (4)



1. Supervised Component–based regularisation for multivariate GLMMs

2. Introducing a time–specific random effect
   - Model definition
   - Regularisation frameworks
     - Ridge–penalised EM
     - SC–regularised EM
   - Application to the GLMMs
   - Simulation study

3. Perspectives

Framework of interest

Balanced panel data with:

- N **individuals**...
- ...observed at the same R **time–points**

Notations:

- $\boldsymbol{y}_{NR\times 1}$: response vector
- $\boldsymbol{X}_{NR\times p}$: design matrix of the **many and redundant** explanatory variables

Framework of interest

Balanced panel data with:

- N individuals...
- ...observed at the same R time–points

Notations:

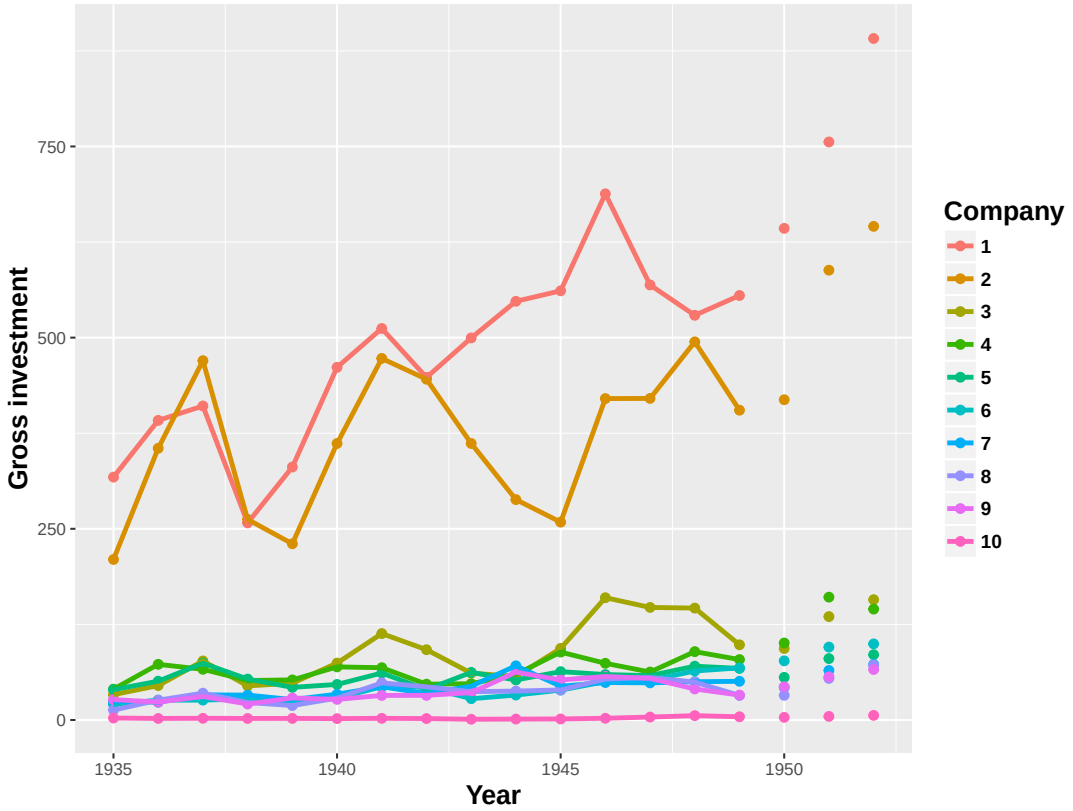
- $\boldsymbol{y}_{NR\times 1}$: response vector
- $\boldsymbol{X}_{NR\times p}$: design matrix of the **many and redundant** explanatory variables

Difficulties

- High level of correlation among the explanatory variables
  $\Rightarrow$ **Regularisation**
- Individual– and time–specific effects
  $\Rightarrow$ Take into account the **complex dependence structure**

The Grunfeld data

Aim: Predict the gross investments from 1950 onwards



Two elements to consider:

- The specific behaviour of each company
- The economic climate: latent phenomenon shared by all the companies which tends to persist over time

Two–way random effects model

In general, we consider data with

- a **within–individual dependence**
  - ↪ **random effect with independent levels**
- a **time dependence**
  - ↪ **random effect with AR(1) levels**

⇓

GLMM

(in order to deal with non–Gaussian response)
with both individual– and time–specific random effects

The GLMM framework

$$\boldsymbol{y} = (y_{11},\ y_{12},\ \ldots,\ y_{1R},\ y_{21},\ y_{22},\ \ldots,\ y_{2R},\ \ldots\ y_{N1}, y_{N2}, \ldots, y_{NR})^{\mathsf{T}}$$

$$g\big(\mathbb{E}\left(Y \,|\, \boldsymbol{\xi}\right)\big) = \boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{U_1}\boldsymbol{\xi_1} + \boldsymbol{U_2}\boldsymbol{\xi_2}$$

$$\boldsymbol{U_1} = \boldsymbol{I}_N \otimes \mathbf{1}_R \ \text{ and } \ \boldsymbol{U_2} = \mathbf{1}_N \otimes \boldsymbol{I}_R$$

The GLMM framework

$$\boldsymbol{y} = (y_{11},\ y_{12},\ \ldots,\ y_{1R},\ y_{21},\ y_{22},\ \ldots,\ y_{2R},\ \ldots\ y_{N1}, y_{N2}, \ldots, y_{NR})^{\mathsf{T}}$$

$$g\big(\mathbb{E}\left(Y \,|\, \boldsymbol{\xi}\right)\big) = \boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{U_1}\boldsymbol{\xi_1} + \boldsymbol{U_2}\boldsymbol{\xi_2}$$

$$\boldsymbol{U_1} = \boldsymbol{I}_N \otimes \boldsymbol{1}_R \ \text{ and } \ \boldsymbol{U_2} = \boldsymbol{1}_N \otimes \boldsymbol{I}_R$$

- **Individual–specific random effect:**
  $\boldsymbol{\xi_1} = (\xi_{11}, \xi_{12}, \ldots, \xi_{1N})^{\mathsf{T}} \sim \mathcal{N}_N\left(\boldsymbol{0},\, \sigma_1^2\, \boldsymbol{A_1}\right), \quad \boldsymbol{A_1} = \boldsymbol{I}_N$
- **Time–specific random effect:**
  $\boldsymbol{\xi_2} = (\xi_{21}, \xi_{22}, \ldots, \xi_{2R})^{\mathsf{T}} \sim \mathcal{N}_R\left(\boldsymbol{0},\, \sigma_2^2\, \boldsymbol{A_2}(\rho)\right)$

$$\boldsymbol{A_2}(\rho) = \left(\frac{\rho^{|i-j|}}{1-\rho^2}\right)_{1 \leqslant i,j \leqslant R}$$

- $\boldsymbol{\xi_1} \perp \boldsymbol{\xi_2}$

Estimation through penalised EM (1)

Principle

Green, P.J. (1990) *On use of the EM for penalized likelihood estimation. Journal of the Royal Statistical Society.* Series B (Methodological), 443-452.

$\boldsymbol{\theta} = \left(\boldsymbol{\beta}, \sigma_1^2, \sigma_2^2, \rho\right)$, $\boldsymbol{\xi} = (\boldsymbol{\xi_1}, \boldsymbol{\xi_2})$

$$\mathsf{E} : \mathcal{Q}_{\mathsf{pen}}\left(\boldsymbol{\theta} \,|\, \boldsymbol{\theta}^{[t]}\right) := \mathbb{E}_{\boldsymbol{\xi}|\boldsymbol{y}}\left[\ell^{\mathsf{c}}_{\mathsf{pen}}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{\xi}) \,|\, \boldsymbol{\theta}^{[t]}\right]$$

$$\mathsf{M} : \boldsymbol{\theta}^{[t+1]} \longleftarrow \arg\max_{\boldsymbol{\theta}} \mathcal{Q}_{\mathsf{pen}}\left(\boldsymbol{\theta} \,|\, \boldsymbol{\theta}^{[t]}\right)$$

Estimation through penalised EM (1)

Principle

Green, P.J. (1990) *On use of the EM for penalized likelihood estimation. Journal of the Royal Statistical Society.* Series B (Methodological), 443-452.

$\boldsymbol{\theta} = \left(\boldsymbol{\beta}, \sigma_1^2, \sigma_2^2, \rho\right)$, $\boldsymbol{\xi} = (\boldsymbol{\xi_1}, \boldsymbol{\xi_2})$

$$\mathsf{E} : \mathcal{Q}_{\mathsf{pen}}\left(\boldsymbol{\theta} \,|\, \boldsymbol{\theta}^{[t]}\right) := \mathbb{E}_{\boldsymbol{\xi}|\boldsymbol{y}}\left[\ell^{\mathsf{c}}_{\mathsf{pen}}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{\xi}) \,|\, \boldsymbol{\theta}^{[t]}\right]$$

$$\mathsf{M} : \boldsymbol{\theta}^{[t+1]} \longleftarrow \arg\max_{\boldsymbol{\theta}} \mathcal{Q}_{\mathsf{pen}}\left(\boldsymbol{\theta} \,|\, \boldsymbol{\theta}^{[t]}\right)$$

Usual penalised complete log-likelihood

$$\ell^{\mathsf{c}}_{\mathsf{pen}}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{\xi}) = \ell^{\mathsf{c}}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{\xi}) - \lambda\, \mathsf{pen}(\boldsymbol{\beta})$$

$$\mathsf{pen}(\boldsymbol{\beta}) = \begin{cases} \|\boldsymbol{\beta}\|_1 \\ \|\boldsymbol{\beta}\|_2^2 = \boldsymbol{\beta}^\intercal \boldsymbol{\beta} \\ \alpha\|\boldsymbol{\beta}\|_2^2 + (1-\alpha)\|\boldsymbol{\beta}\|_1, \quad 0 \leqslant \alpha \leqslant 1 \end{cases}$$

Estimation through penalised EM (2)

Ridge-based regularisation

$$\hookrightarrow \boldsymbol{\theta} = \left(\boldsymbol{\beta}, \sigma_1^2, \sigma_2^2, \rho\right),\ \boldsymbol{\xi} = (\boldsymbol{\xi_1}, \boldsymbol{\xi_2})$$

$$\mathsf{E}: \mathcal{Q}_{\textbf{ridge}}\left(\boldsymbol{\theta}, \lambda \,|\, \boldsymbol{\theta}^{[t]}\right) := \mathbb{E}_{\boldsymbol{\xi}|\boldsymbol{y}}\left[\ell^{\mathsf{C}}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{\xi}) - \lambda\, \boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\beta} \,|\, \boldsymbol{\theta}^{[t]}\right]$$

$$\mathsf{M}: \begin{cases} \lambda^{[t+1]} \longleftarrow \mathrm{GCV}^{[t+1]}(\lambda) \\ \boldsymbol{\theta}^{[t+1]} \longleftarrow \arg\max\limits_{\boldsymbol{\theta}} \mathcal{Q}_{\textbf{ridge}}\left(\boldsymbol{\theta}, \lambda^{[t+1]} \,|\, \boldsymbol{\theta}^{[t]}\right) \end{cases}$$

Eliot, M., Ferguson, J., Reilly, M.P. and Foulkes, A.S. (2011) *Ridge Regression for Longitudinal Biomarker Data.* The International Journal of Biostatistics, **7**, 1–11.

Estimation through SC–regularised EM

Linear predictor

$$\boldsymbol{\eta} = \begin{cases} (\boldsymbol{X}\boldsymbol{u})\gamma + \boldsymbol{U_1}\boldsymbol{\xi_1} + \boldsymbol{U_2}\boldsymbol{\xi_2} & \text{for a single component} \\ \sum_{h=1}^{H} (\boldsymbol{X}\boldsymbol{u_h})\gamma_h + \boldsymbol{U_1}\boldsymbol{\xi_1} + \boldsymbol{U_2}\boldsymbol{\xi_2} & \text{for } H \text{ components} \end{cases}$$

Estimation through SC–regularised EM

Linear predictor

$$\boldsymbol{\eta} = \begin{cases} (\boldsymbol{X}\boldsymbol{u})\gamma + \boldsymbol{U_1}\boldsymbol{\xi_1} + \boldsymbol{U_2}\boldsymbol{\xi_2} & \text{for a single component} \\ \sum_{h=1}^{H} (\boldsymbol{X}\boldsymbol{u_h})\gamma_h + \boldsymbol{U_1}\boldsymbol{\xi_1} + \boldsymbol{U_2}\boldsymbol{\xi_2} & \text{for } H \text{ components} \end{cases}$$

SC–based complete log–likelihood

$\boldsymbol{\theta} = \left(\boldsymbol{u}, \gamma, \sigma_1^2, \sigma_2^2, \rho\right)$, trade-off parameter $s \in [0, 1]$

$$\ell^{\mathsf{c}}_{\mathsf{SC}}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{\xi}) = (1 - s)\, \ell^{\mathsf{c}}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{\xi}) + s \log\left[\phi(\boldsymbol{u})\right]$$

- **Complete log–likelihood :** measures the fit of the model (based on component $\boldsymbol{f} = \boldsymbol{X}\boldsymbol{u}$) to the data
- **Structural relevance criterion :** measures the closeness of component $\boldsymbol{f}$ to the strongest structures of $\boldsymbol{X}$

Ridge– versus SC–EM

Ridge–based penalisation

$$\ell^{\mathsf{C}}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{\xi}) - \lambda \boldsymbol{\beta}^{\mathsf{T}} \boldsymbol{\beta}$$

- **Penalises** the "large" coefficients
- Sees the high correlations among the explanatory variables as **pure nuisance**
- η **hard to interpret**

Component–based regularisation

$$\ell^{\mathsf{C}}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{\xi}) + \frac{s}{1-s} \log\left[\phi(\boldsymbol{u})\right]$$

- **Gives a bonus** to the most interpretable bundles in $\boldsymbol{X}$
- **Takes advantage** of the high correlations among the explanatory variables
- η **easier to interpret** through decomposition on components

Application to the GLMMs

LINEARISATION step

- Given $\boldsymbol{\mu}_i := \mathbb{E}\left(\boldsymbol{Y}_i \,|\, \boldsymbol{\xi}\right)$, the working variable $\boldsymbol{z}_i$ writes

$$\boldsymbol{z}_i = \underbrace{g(\boldsymbol{\mu}_i)}_{\boldsymbol{\eta}_i} + (\boldsymbol{y}_i - \boldsymbol{\mu}_i) g'(\boldsymbol{\mu}_i)$$

- Linearised model:

$$\boldsymbol{\mathcal{M}}\text{:}\;\; \boldsymbol{z} = \underbrace{\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{U_1}\boldsymbol{\xi_1} + \boldsymbol{U_2}\boldsymbol{\xi_2}}_{\boldsymbol{\eta}} + \boldsymbol{e}, \quad \text{with } \mathbb{V}(\boldsymbol{e}) = \boldsymbol{W}$$

ESTIMATION step

- Instead of the classical Henderson's systems, we propose a Penalised/Regularised EM algorithm on $\boldsymbol{\mathcal{M}}$

Ridge–EM for GLMM–AR$_1$

$\boldsymbol{\theta} = \left(\boldsymbol{\beta}, \sigma_1^2, \sigma_2^2, \rho\right)$

Linearised model

$$\boldsymbol{\mathcal{M}}^{[t]}\text{: } \boldsymbol{z}^{[t]} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{U_1}\boldsymbol{\xi_1} + \boldsymbol{U_2}\boldsymbol{\xi_2} + \boldsymbol{e}^{[t]}, \quad \text{with } \mathbb{V}\left(\boldsymbol{e}^{[t]}\right) = \boldsymbol{W}^{[t]}$$

Ridge estimation

$$\textbf{E: } \mathcal{Q}_{\textbf{ridge}}\left(\boldsymbol{\theta}, \lambda \,|\, \boldsymbol{\theta}^{[t]}\right) := \mathbb{E}_{\boldsymbol{\xi}|\boldsymbol{z}^{[t]}}\left[\ell^{\mathsf{c}}\left(\boldsymbol{\theta}\,;\, \boldsymbol{z}^{[t]}, \boldsymbol{\xi}\right) - \lambda\, \boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\beta} \,|\, \boldsymbol{\theta}^{[t]}\right]$$

$$\textbf{M: } \begin{cases} \lambda^{[t+1]} \longleftarrow \mathsf{GCV}^{[t+1]}(\lambda) \\ \boldsymbol{\theta}^{[t+1]} \longleftarrow \arg\max_{\boldsymbol{\theta}} \mathcal{Q}_{\textbf{ridge}}\left(\boldsymbol{\theta}, \lambda^{[t+1]} \,|\, \boldsymbol{\theta}^{[t]}\right) \end{cases}$$

Update

Compute $\boldsymbol{\xi}^{[t+1]}$, $\boldsymbol{z}^{[t+1]}$, $\boldsymbol{W}^{[t+1]}$

SC–EM for GLMM–AR$_1$

$\boldsymbol{\theta} = \left(\boldsymbol{u}, \gamma, \sigma_1^2, \sigma_2^2, \rho\right)$

Linearised model

$$\boldsymbol{\mathcal{M}}^{[t]}\text{: } \boldsymbol{z}^{[t]} = (\boldsymbol{X}\boldsymbol{u})\gamma + \boldsymbol{U_1}\boldsymbol{\xi_1} + \boldsymbol{U_2}\boldsymbol{\xi_2} + \boldsymbol{e}^{[t]}, \quad \text{with } \mathbb{V}\left(\boldsymbol{e}^{[t]}\right) = \boldsymbol{W}^{[t]}$$

SC–estimation

$$\textbf{E: } \mathcal{Q}_{\textsf{SC}}\left(\boldsymbol{\theta} \,|\, \boldsymbol{\theta}^{[t]}\right) := \mathbb{E}_{\boldsymbol{\xi}|\boldsymbol{z}^{[t]}}\left[(1-s)\ell^{\textsf{c}}\left(\boldsymbol{\theta}\,;\, \boldsymbol{z}^{[t]}, \boldsymbol{\xi}\right) + s\log\left[\phi(\boldsymbol{u})\right] \,|\, \boldsymbol{\theta}^{[t]}\right]$$

$$\textbf{M: } \begin{cases} \sigma_1^{2\,[t+1]},\ \sigma_2^{2\,[t+1]},\ \rho^{[t+1]} \text{ computed as previously} \\ \boldsymbol{u}^{[t+1]} \longleftarrow \underset{||\boldsymbol{u}||=1}{\arg\max}\ \widetilde{\mathcal{Q}}_{\textsf{SC}}\left(\boldsymbol{u}, \gamma^{[t]} \,|\, \boldsymbol{\theta}^{[t]}\right) \\ \gamma^{[t+1]} \longleftarrow \underset{\gamma}{\arg\max}\ \widetilde{\mathcal{Q}}_{\textsf{SC}}\left(\boldsymbol{u}^{[t+1]}, \gamma \,|\, \boldsymbol{\theta}^{[t]}\right) \end{cases}$$

Update

Compute $\boldsymbol{\xi}^{[t+1]},\ \boldsymbol{z}^{[t+1]},\ \boldsymbol{W}^{[t+1]}$

Simulations (1)

Poisson regression with log link

- $\boldsymbol{y} \sim \mathcal{P}\text{oi}\Big(\boldsymbol{\lambda} = \exp\left[\boldsymbol{X\beta} + \boldsymbol{U_1\xi_1} + \boldsymbol{U_2\xi_2}\right]\Big)$
- $\boldsymbol{X} = \Big[\underbrace{\boldsymbol{x_1} \ldots\ldots\ldots\ldots \boldsymbol{x_{10}}}_{\text{large bundle}} \; \underbrace{\boldsymbol{x_{11}} \ldots\ldots \boldsymbol{x_{15}}}_{\text{small bundle}} \; \underbrace{\boldsymbol{x_{16}} \ldots\ldots \boldsymbol{x_{20}}}_{\text{small bundle}}\Big]$

  large bundle ↪ noise

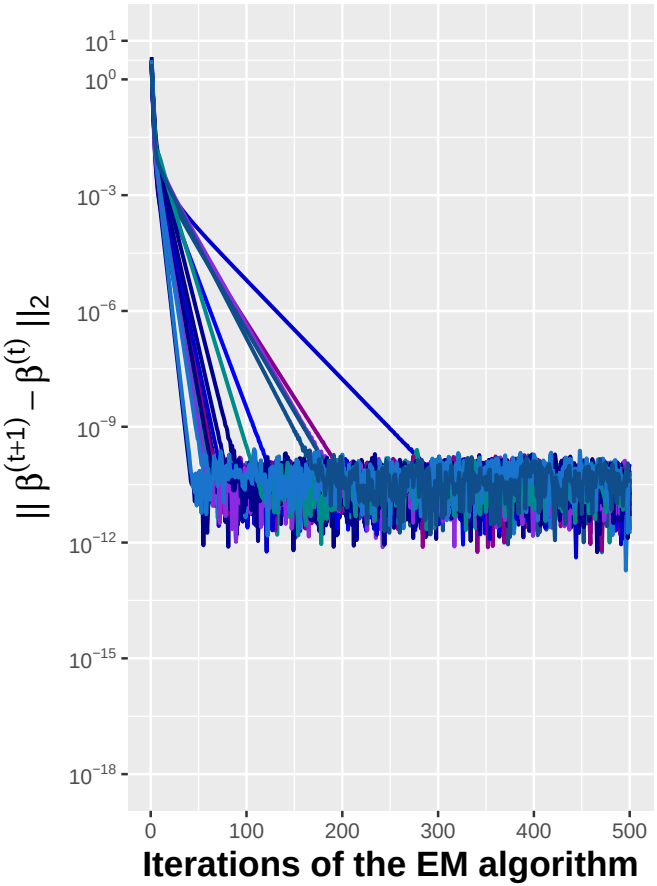
  small bundle ↪ predicts y

  small bundle ↪ predicts y

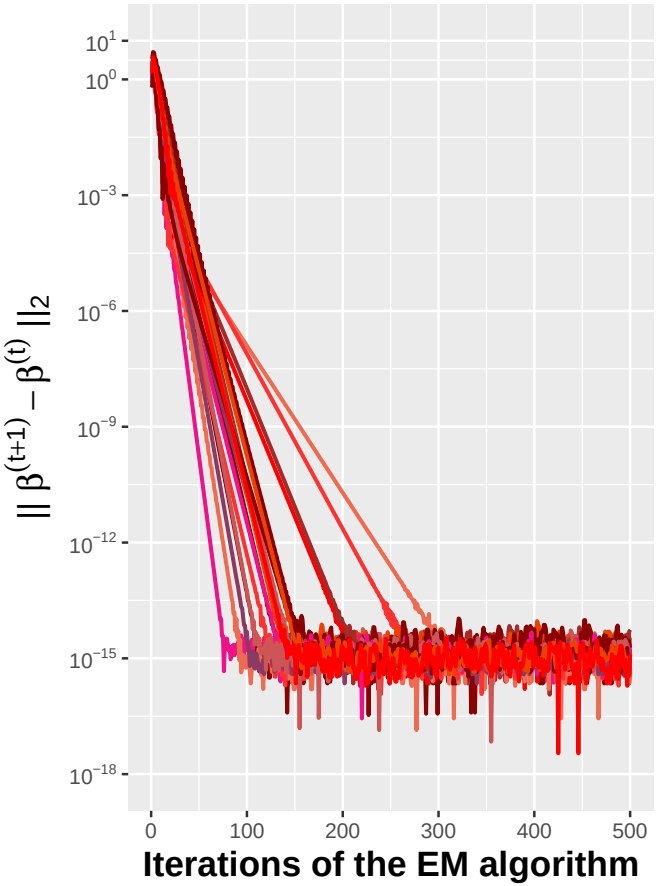
Simulations (2)

How does convergence go?

Ridge penalty

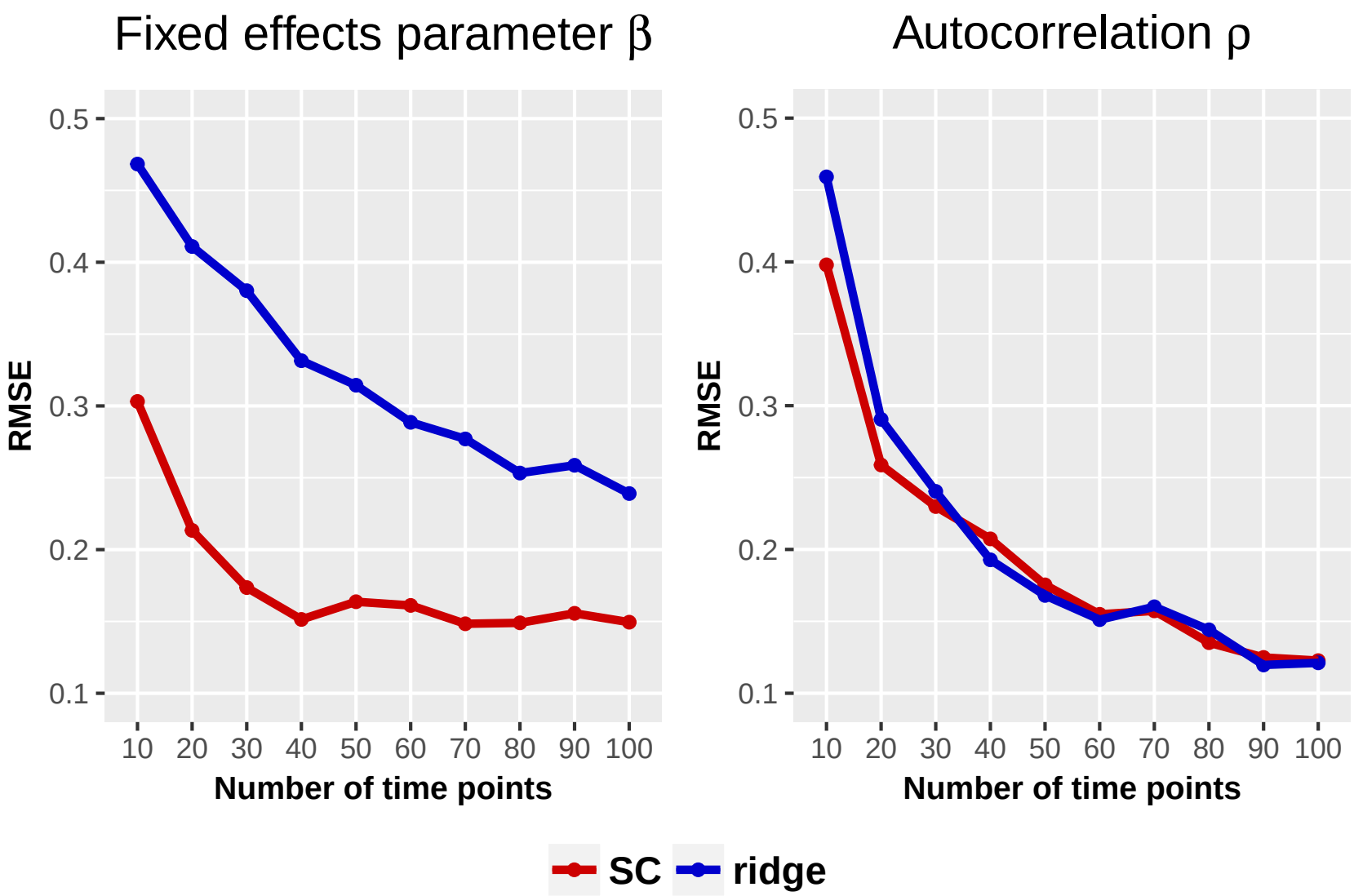


Supervised Component regularisation



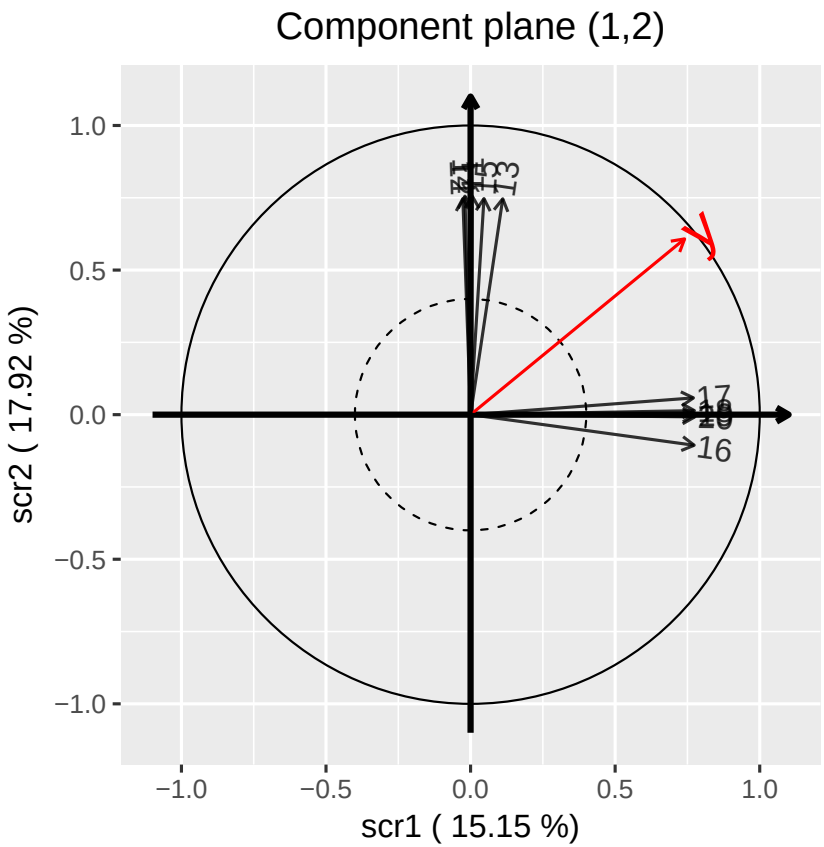
Simulations (3)

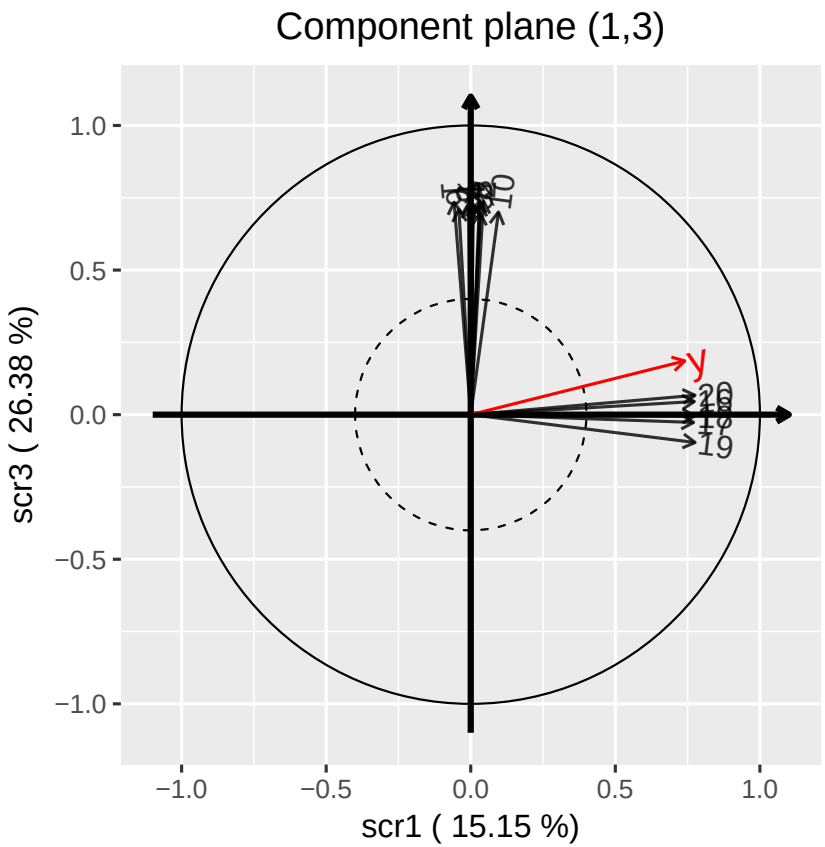
Accuracy of the estimates



Simulations (4)

Power for model interpretation





Overview on Mixed–SCGLR

Powerful trade–off between

- multivariate GLMM
- component–based regularisation

Model interpretation ↗

- Mixed–SCGLR provides graphical diagnoses (component planes)
- reveals the multidimensional explanatory and predictive structures

Estimate–accuracy ↗

Mixed–SCGLR now suitable for

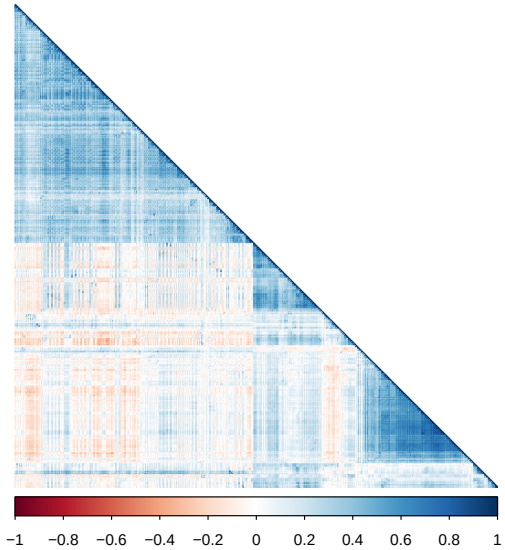
- grouped data
- panel data

Application in progress

Major depressive disorders and grey–matter volume reduction

- 15 years follow–up neuropsychiatric study including **636 participants**
- **Repeated binary response:** Depressive or not
- **528 explanatory variables:** thickness, area, volume and curvature of several brain areas
- **Additional covariates:** Gender, age...

Correlation–heatmap



Future developments (1)

Sparse Supervised Component $\left\{ \boldsymbol{f_h} = \boldsymbol{X}\boldsymbol{u_h} \,|\, h = 1, \ldots, K \right\}$

- Variable selection via supervised components
  - $\hookrightarrow$ Supervised components that are weighted combination of only a few explanatory variables
- Idea: add a sparsity constraint on the input variables
  - $\hookrightarrow$ Generic program:

$$\boldsymbol{u_h} = \begin{cases} \arg \max\limits_{\boldsymbol{u} \in \mathbb{R}^p} & s \log\left[\phi(\boldsymbol{u})\right] + (1-s) \log\left[\psi(\boldsymbol{u})\right] - \lambda \|\boldsymbol{u}\|_1 \\ \|\boldsymbol{u}\| = 1 \text{ and } & \boldsymbol{X}\boldsymbol{u} \perp \boldsymbol{X}\boldsymbol{u_1}, \ldots, \boldsymbol{X}\boldsymbol{u_{h-1}} \end{cases}$$

Future developments (2)

Spatial correlation modelling

$$\boldsymbol{\xi} \sim \mathcal{N}_q\left(\mathbf{0}, \varsigma^2 \boldsymbol{B}^{-1}\right), \text{ where}$$

- ς^2 is the unknown spatial–specific variance component, and
- $\boldsymbol{B} = \boldsymbol{I} - \rho \boldsymbol{A}$, where ρ is the unknown autoregressive spatial parameter and $\boldsymbol{A} = (a_{ij})_{1 \leqslant i,j \leqslant q}$ is the adjacency matrix defined by

$$a_{ij} = \begin{cases} 1 & \text{if } j \text{ is adjacent to } i \\ 0 & \text{otherwise.} \end{cases}$$

A general covariance structure on $\boldsymbol{Y} = \left[\boldsymbol{y_1} \,|\, \ldots \,|\, \boldsymbol{y_q}\right]$?

- e.g. modelling species interactions and competition
  - ↪ Congo–Basin floristic data: some species tend to appear together
  - ↪ Other species are antagonistic

References

Bry, X., Verron, T. (2015). *THEME: THEmatic model exploration through multiple co–structure maximization.* Journal of Chemometrics, **29**, 637-647.

Bry, X., Trottier, C., Verron, T. and Mortier, F. (2013). *Supervised component generalized linear regression using a PLS-extension of the Fisher scoring algorithm.* Journal of Multivariate Analysis, **119**, 47-60.

+ Package R : SCGLR

https://github.com/SCnext/SCGLR

Chauvet, J., Bry, X., Trottier, C. (2019). *Component–based regularisation of multivariate generalised mixed models.* Journal of Computational and Graphical Statistics, *in press*.

+ Package R : mixedSCGLR

https://github.com/SCnext/mixedSCGLR

Chauvet, J., Bry, X., Trottier, C. (2019). *Regularisation of GLMMs with an autoregressive random time–specific effect. in progress*.

Eliot, M., Ferguson, J., Reilly, M.P. and Foulkes, A.S. (2011) *Ridge Regression for Longitudinal Biomarker Data.* The International Journal of Biostatistics, **7**, 1–11.

Green, P.J. (1990) *On use of the EM for penalized likelihood estimation.* Journal of the Royal Statistical Society. Series B (Methodological), 443-452.

Marx, B. D. (1996) *Iteratively reweighted partial least squares estimation for generalized linear regression.* Technometrics, **38**, 4, 374–381.

High dimensional data (1)

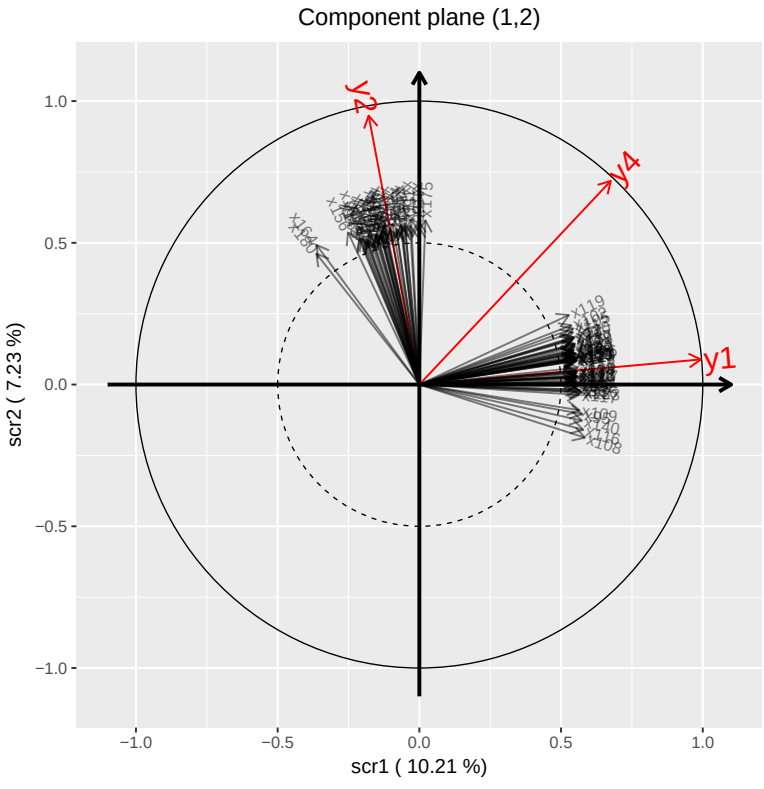
- 100 observations
- 150 explanatory variables: $\boldsymbol{X} = \left[\boldsymbol{X_0} \mid \boldsymbol{X_1} \mid \boldsymbol{X_2} \mid \boldsymbol{X_3} \right]$
  $\hookrightarrow$ 60 variables in $\boldsymbol{X_0}$, 45 variables in $\boldsymbol{X_1}$,
  30 variables in $\boldsymbol{X_2}$, 15 variables in $\boldsymbol{X_3}$
- Responses

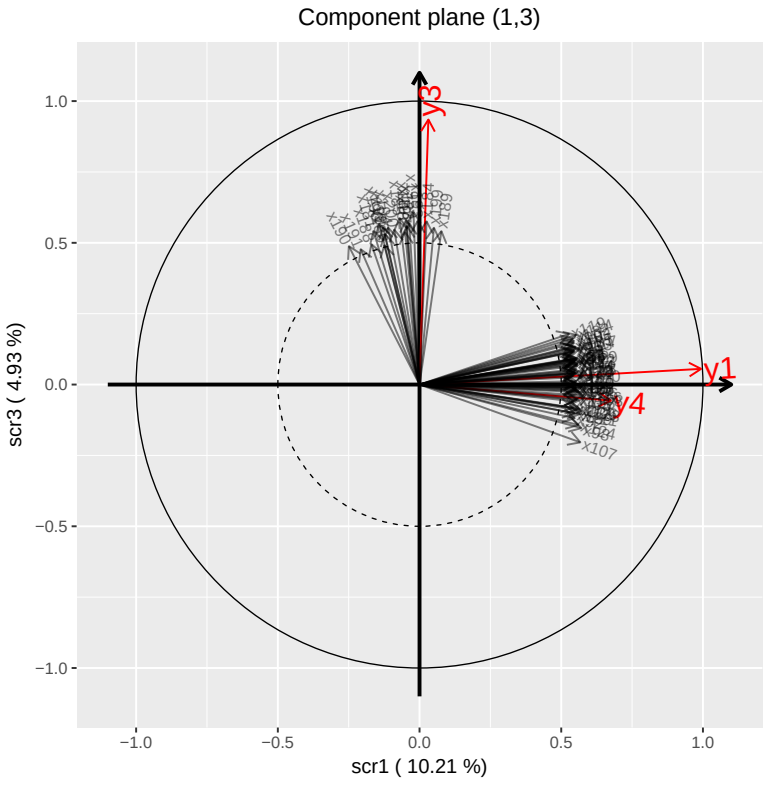
$$\begin{cases} \boldsymbol{y_1} \sim \mathcal{N}_n\Big(\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta_1} + \boldsymbol{U}\boldsymbol{\xi_1},\, \boldsymbol{\Sigma} = \boldsymbol{I}_n\Big) \\ \boldsymbol{y_2} \sim \mathcal{B}\Big(\boldsymbol{p} = \text{logit}^{-1}\Big[\boldsymbol{X}\boldsymbol{\beta_2} + \boldsymbol{U}\boldsymbol{\xi_2}\Big]\Big) \\ \boldsymbol{y_3} \sim \mathcal{B}\text{in}\Big(\textbf{trials} = 30\,\mathbf{1}_n,\, \boldsymbol{p} = \text{logit}^{-1}\Big[\boldsymbol{X}\boldsymbol{\beta_3} + \boldsymbol{U}\boldsymbol{\xi_3}\Big]\Big) \\ \boldsymbol{y_4} \sim \mathcal{P}\Big(\boldsymbol{\lambda} = \exp\Big[\boldsymbol{X}\boldsymbol{\beta_4} + \boldsymbol{U}\boldsymbol{\xi_4}\Big]\Big). \end{cases}$$

- $\boldsymbol{X_0}$ = nuisance bundle
  $\boldsymbol{y_1}$ predicted only by $\boldsymbol{X_1}$, $\boldsymbol{y_2}$ only by $\boldsymbol{X_2}$, $\boldsymbol{y_3}$ only by $\boldsymbol{X_3}$,
  $\boldsymbol{y_4}$ by both $\boldsymbol{X_2}$ and $\boldsymbol{X_3}$.

High dimensional data (2)

Results





High dimensional data (3)

First idea: replace X with the matrix C of its principal components associated with non–negligible eigenvalues

- $\boldsymbol{C = XV}$, with $\boldsymbol{V}$ the matrix of unit–eigenvectors
- Modified GoF and SR criteria

$$\widetilde{\psi}(\boldsymbol{u}) = \sum_{k=1}^{q} \left\| \Pi_{\text{span}\{\boldsymbol{Cu},\boldsymbol{A}\}} \, \boldsymbol{z_k} \right\|^2_{\boldsymbol{W_k}}$$

$$\widetilde{\phi}(\boldsymbol{u}) = \left(\sum_{j=1}^{p} \Big[\text{cor}^2 \left(\boldsymbol{Cu}, \boldsymbol{x_j} \right) \Big]^l \right)^{\frac{1}{l}}$$

- Maximisation program

$$\begin{cases} \max & s \log \left[\widetilde{\phi}(\boldsymbol{u}) \right] + (1-s) \log \left[\widetilde{\psi}(\boldsymbol{u}) \right] \\ \text{subject to} & \boldsymbol{u}^\intercal \boldsymbol{C}^\intercal \boldsymbol{P} \boldsymbol{C} \boldsymbol{u} = 1, \; \boldsymbol{P} = n^{-1} \boldsymbol{I}_n \end{cases}$$

High dimensional data (4)

Better idea:

- Preserve the GoF and SR criteria

$$\psi(\boldsymbol{u}) = \sum_{k=1}^{q} \left\| \Pi_{\text{span}\{\boldsymbol{X}\boldsymbol{u}, \boldsymbol{A}\}} \, \boldsymbol{z_k} \right\|_{\boldsymbol{W_k}}^2$$

$$\phi(\boldsymbol{u}) = \left(\sum_{j=1}^{p} \Big[\, \text{cor}^2 \left(\boldsymbol{X}\boldsymbol{u}, \boldsymbol{x_j} \right) \Big]^l \right)^{\frac{1}{l}}$$

- Modify the norm–constraint

$$\begin{cases} \text{max} & s \log \left[\phi(\boldsymbol{u}) \right] + (1-s) \log \left[\psi(\boldsymbol{u}) \right] \\ \text{subject to} & \boldsymbol{u}^\intercal \left[\tau \boldsymbol{I} + (1-\tau) \, \boldsymbol{X}^\intercal \boldsymbol{P} \boldsymbol{X} \right] \boldsymbol{u} = 1, \; \boldsymbol{P} = n^{-1} \boldsymbol{I}_n \end{cases}$$

The PING algorithm

- **Generic program:**

$$\begin{cases} \max & \mathcal{J}(\boldsymbol{v}) \\ \text{subject to} & \boldsymbol{v}^{\mathsf{T}}\boldsymbol{v} = 1 \text{ and } \boldsymbol{\Delta}^{\mathsf{T}}\boldsymbol{v} = \mathbf{0}. \end{cases}$$

- **Direction of ascent:**

$$\boldsymbol{v}^{[t+1]} = \frac{\Pi_{\text{span}\{\boldsymbol{\Delta}\}^{\perp}} \Gamma\left(\boldsymbol{v}^{[t]}\right)}{\left\| \Pi_{\text{span}\{\boldsymbol{\Delta}\}^{\perp}} \Gamma\left(\boldsymbol{v}^{[t]}\right) \right\|}, \text{ with } \Gamma(\boldsymbol{v}) = \boldsymbol{\nabla}\mathcal{J}(\boldsymbol{v})$$

The PING algorithm

The Projected Iterated Normed Gradient Algorithm

while *convergence of* $\boldsymbol{v}$ *non reached* **do**

$$\boldsymbol{m} \leftarrow \frac{\Pi_{\text{span}\{\boldsymbol{\Delta}\}^{\perp}}\ \Gamma\left(\boldsymbol{v}^{[t]}\right)}{\left\|\Pi_{\text{span}\{\boldsymbol{\Delta}\}^{\perp}}\ \Gamma\left(\boldsymbol{v}^{[t]}\right)\right\|}$$

while $\mathcal{J}(\boldsymbol{m}) < \mathcal{J}\left(\boldsymbol{v}^{[t]}\right)$ **do**

$$\boldsymbol{m} \leftarrow \frac{\boldsymbol{v}^{[t]} + \boldsymbol{m}}{\left\|\boldsymbol{v}^{[t]} + \boldsymbol{m}\right\|}$$

end

$$\boldsymbol{v}^{[t+1]} \leftarrow \boldsymbol{m}$$

$$t \leftarrow t + 1$$

end