

# Supervised Component-based regularisation of multivariate generalised linear mixed models

Jocelyn CHAUVENT

with

Catherine TROTTIER  
Xavier BRY

Sherbrooke

June 19<sup>th</sup>, 2019

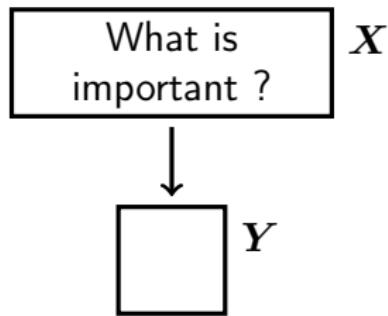


# General context

## → Why regularise a regression model ?

Fuzzy conceptual model, large amount of variables

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

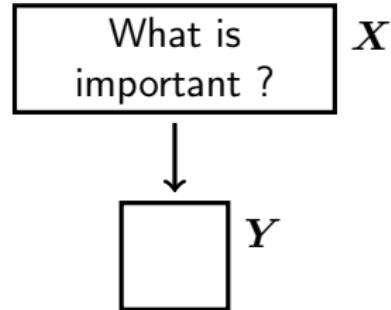


# General context

## → Why regularise a regression model ?

Fuzzy conceptual model, large amount of variables

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

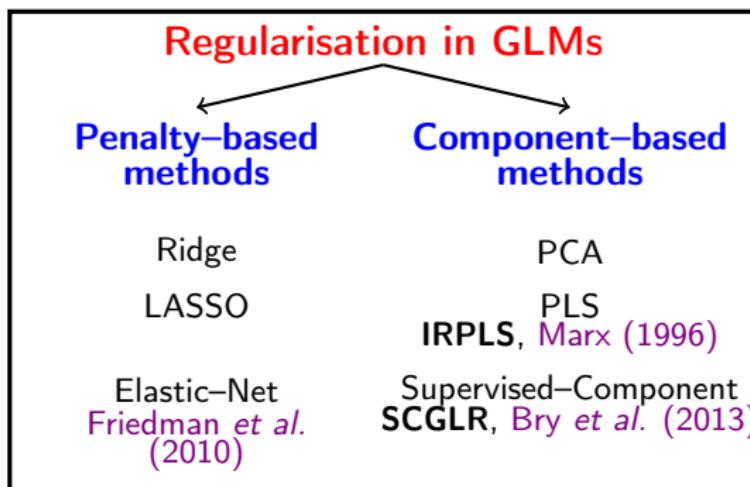


## Regularisation (definition)

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

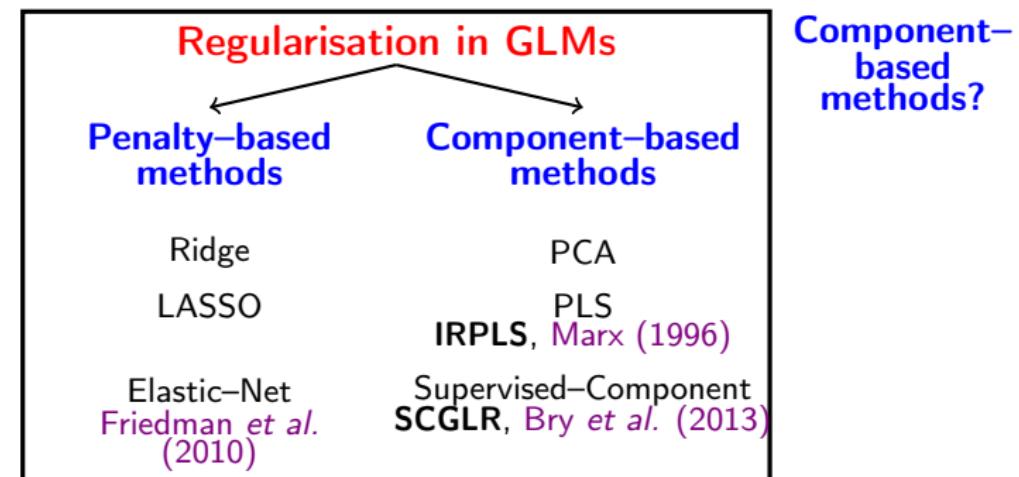


## Regularisation in GLMMs (GLMs + random effects)

Penalty-based  
methods:

LMM–Ridge  
Eliot *et al.* (2011)

GLMM–LASSO  
Groll and Tutz (2014)



# 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

Ridge

LASSO

Elastic–Net  
Friedman *et al.*  
(2010)

Component–based  
methods

PCA

PLS  
IRPLS, Marx (1996)

Supervised–Component  
SCGLR, Bry *et al.* (2013)

Component–  
based  
methods?

Extension of  
SCGLR  
to the GLMMs:  
Mixed–SCGLR

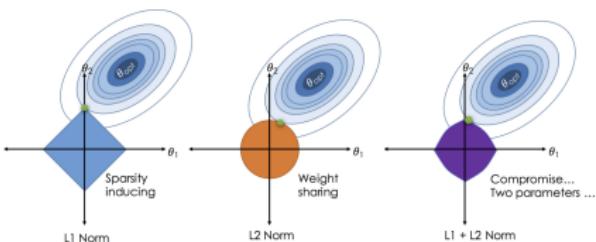


# Penalty-based regularisation

Penalised log-likelihood

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

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

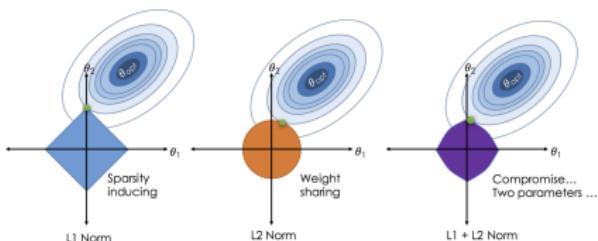


# Penalty-based regularisation

## Penalised log-likelihood

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

- ▶ **LASSO:**  $\text{pen}(\beta) = \|\beta\|_1$ 
  - ↪ Sparse solutions
  - ↪ Variable selection
- ▶ **Ridge:**  $\text{pen}(\beta) = \|\beta\|_2^2$ 
  - ↪ Shrinkage towards 0
  - ↪ Reduce the estimator's variance
- ▶ **Elastic-net:**  
 $\text{pen}(\beta) = (1 - \alpha)\|\beta\|_1 + \alpha\|\beta\|_2^2$ 
  - ↪ 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  $\{f_h = Xu_h \mid h = 1, \dots, H\}$

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

# Component-based regularisation

Components = synthetic variables  $\{f_h = Xu_h \mid h = 1, \dots, H\}$

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

## PCA

$$\text{crit}(u) = \underbrace{\|Xu\|_2^2}_{\text{Component Variance (cv)}}$$

- ↪ Information in  $X$  ✓
- ↪ Link between  $X$  and  $y$  ✗

## PLS

$$\text{crit}(u) = \underbrace{\|Xu\|_2^2}_{\text{Comp. Var. (cv)}} \underbrace{\|y\|_2^2 \cos^2(y, \text{span}\{Xu\})}_{\text{Goodness-of-Fit (GoF)}}$$

- ↪ Information in  $X$  ✓
- ↪ Link between  $X$  and  $y$  ✓

# Component-based regularisation

Components = synthetic variables  $\{f_h = Xu_h \mid h = 1, \dots, H\}$

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

## PCA

$$\text{crit}(u) = \underbrace{\|Xu\|_2^2}_{\text{Component Variance (cv)}}$$

- ↪ Information in  $X$  ✓
- ↪ Link between  $X$  and  $y$  ✗

## PLS

$$\text{crit}(u) = \underbrace{\|Xu\|_2^2}_{\text{Comp. Var. (cv)}} \underbrace{\|y\|_2^2 \cos^2(y, \text{span}\{Xu\})}_{\text{Goodness-of-Fit (GoF)}}$$

- ↪ Information in  $X$  ✓
- ↪ Link between  $X$  and  $y$  ✓

## A "flexible PLS"

$$\text{crit}(u) = [\text{CV}(u)]^s [\text{GoF}(u)]^{1-s}$$
$$s \in [0, 1] \text{ a trade-off parameter}$$

# Structural relevance

↪ Introduced by Bry and Verron (2015)

## Supervised Components via the Structural Relevance

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

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

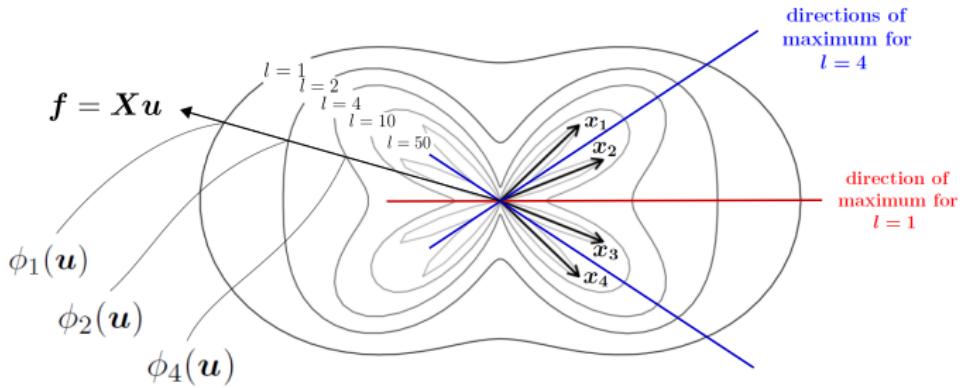
# Structural relevance

↪ Introduced by Bry and Verron (2015)

## Supervised Components via the Structural Relevance

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

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



# Flash method–comparison

## A simple Gaussian model

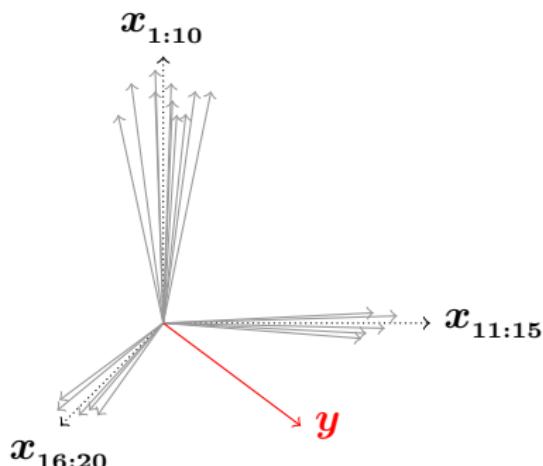
►  $y \sim \mathcal{N}_n(\mu = X\beta, \Sigma = I_n)$

►  $X = \begin{bmatrix} x_1 \dots x_{10} & x_{11} \dots x_{15} & x_{16} \dots x_{20} \end{bmatrix}$

large bundle  
↪ **nuisance**

small bundle  
↪ **predicts y**

small bundle  
↪ **predicts y**



👉 PCR vs PLSR vs "Supervised Component Regression"

- ▶ Non-independent observations
  - ↪ Grouped and panel data
  - ⇒ From GLM to GLMM (use of random effects)
- ▶ Multivariate framework
  - ↪ Several responses of various types  $\mathbf{Y} = [\mathbf{y}_1 | \dots | \mathbf{y}_q]$
- ▶ Additional explanatory variables
  - ↪ with little redundancy
  - ↪ Requiring no regularisation

## 1 The Congo–Basin floristic data

2 The mixed–SCGLR method

3 Simulation study

4 Results on the floristic data

5 Conclusion and perspectives

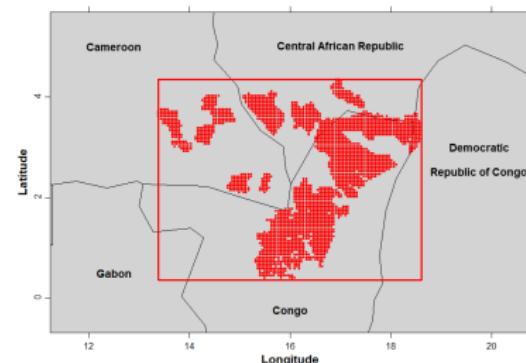
# A multivariate GLM ?

**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:

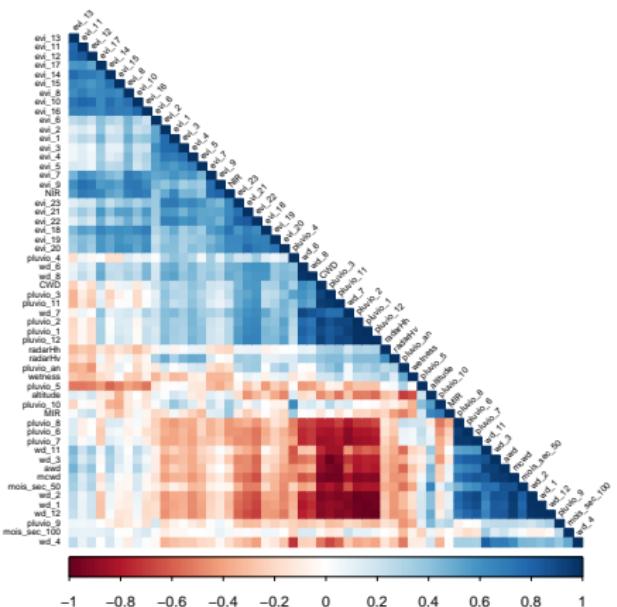
$q = 8$  abundances of selected tree species  
(i.e. multivariate count responses)



Multivariate GLM

# Regularisation needed

## Correlation-heatmap



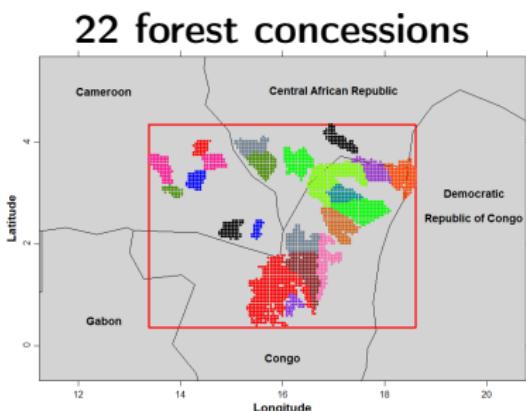
## Explanatory variables:

- ▶  $p = 56$  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



## SCGLR:

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

## Mixed-SCGLR:

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



**Multivariate GLMM**

- 1 The Congo–Basin floristic data
- 2 The mixed–SCGLR method
- 3 Simulation study
- 4 Results on the floristic data
- 5 Conclusion and perspectives

# Model

## Notations:

- ▶  $Y_{n \times q}$ : matrix of  $q$  response vectors  $y_1, \dots, y_q$
- ▶  $X_{n \times p}$ : explanatory variables  $x_1, \dots, x_p$  (many, redundant)
- ▶  $A_{n \times r}$ : additional covariates  $a_1, \dots, a_r$  (few, not redundant)
- ▶  $U_{n \times N}$ : design matrix of the random effects

# Model

## Notations:

- ▶  $Y_{n \times q}$ : matrix of  $q$  response vectors  $y_1, \dots, y_q$
- ▶  $X_{n \times p}$ : explanatory variables  $x_1, \dots, x_p$  (many, redundant)
- ▶  $A_{n \times r}$ : additional covariates  $a_1, \dots, a_r$  (few, not redundant)
- ▶  $U_{n \times N}$ : design matrix of the random effects

## Single component multivariate GLMM

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

$$g_k(\mathbb{E}(Y_k | \xi_k)) = \eta_k$$

$$\eta_k = (\mathbf{X}u)\gamma_k + A\delta_k + U\xi_k$$

$\xi_k \stackrel{\text{ind.}}{\sim} \mathcal{N}_N(\mathbf{0}, \sigma_k^2 I_N)$ , with  $N$  the number of groups

# Estimation

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

↪ Pseudo-responses:  $z_k$

## "Linearised" model

$$z_k = \underbrace{(\mathbf{X}u)\gamma_k + A\delta_k + U\xi_k}_{\eta_k} + e_k \quad \text{with: } \begin{cases} \mathbb{E}(e_k | \xi_k) = \mathbf{0} \\ \mathbb{V}(e_k | \xi_k) =: W_k \end{cases}$$

# Estimation

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

↪ Pseudo-responses:  $z_k$

## "Linearised" model

$$z_k = \underbrace{(\mathbf{X}\mathbf{u})\gamma_k + A\delta_k + U\xi_k}_{\eta_k} + e_k \quad \text{with: } \begin{cases} \mathbb{E}(e_k | \xi_k) = \mathbf{0} \\ \mathbb{V}(e_k | \xi_k) =: W_k \end{cases}$$

## Alternate procedure

- (i) Given  $\gamma_k$ ,  $\delta_k$ ,  $\xi_k$  and  $\sigma_k^2$ , we compute the component  $f = \mathbf{X}\mathbf{u}$
- (ii) Given  $\mathbf{u}$ , we estimate  $\gamma_k$ ,  $\delta_k$ ,  $\xi_k$  and  $\sigma_k^2$ 
  - ↪ Schall's algorithm, Henderson's system

# Updating parameters

## Step (ii)

### Henderson's systems

Given  $f = Xu$ , for each  $k \in \{1, \dots, q\}$ :

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

### Update variance components

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

# Computing the component

## Step (i)

### Goodness-of-Fit (GoF)

$$\psi(\mathbf{u}) = \sum_{k=1}^q \left\| \Pi_{\text{span}\{\mathbf{X}\mathbf{u}, \mathbf{A}\}} \mathbf{z}_k \right\|_{\mathbf{W}_k^{-1}}^2$$

# Computing the component

## Step (i)

### Goodness-of-Fit (GoF)

$$\psi(\mathbf{u}) = \sum_{k=1}^q \left\| \Pi_{\text{span}\{\mathbf{X}\mathbf{u}, \mathbf{A}\}} \mathbf{z}_k \right\|_{\mathbf{W}_k^{-1}}^2$$

### Structural Relevance (SR)

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

# Computing the component

## Step (i)

### Goodness-of-Fit (GoF)

$$\psi(\mathbf{u}) = \sum_{k=1}^q \left\| \Pi_{\text{span}\{\mathbf{X}\mathbf{u}, \mathbf{A}\}} \mathbf{z}_k \right\|_{\mathbf{W}_k^{-1}}^2$$

### Structural Relevance (SR)

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

### Trade-off GoF/SR

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

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

- 1 The Congo–Basin floristic data
- 2 The mixed–SCGLR method
- 3 Simulation study
- 4 Results on the floristic data
- 5 Conclusion and perspectives

# Design

Two random responses:  $\mathbf{Y} = [y_1 \mid y_2]$

# Design

Two random responses:  $\mathbf{Y} = [y_1 \mid y_2]$

Fixed effects:

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

$$\mathbf{X} = \left[ \underbrace{\mathbf{x}_1 \dots \mathbf{x}_{15}}_{\text{bundle } \mathbf{X}_0 \text{ (large)}} \quad \underbrace{\mathbf{x}_{16} \dots \mathbf{x}_{25}}_{\text{bundle } \mathbf{X}_1 \text{ (medium)}} \quad \underbrace{\mathbf{x}_{26} \dots \mathbf{x}_{30}}_{\text{bundle } \mathbf{X}_2 \text{ (small)}} \right]$$

$\hookrightarrow \text{nuisance}$        $\hookrightarrow \text{predicts } y_1$        $\hookrightarrow \text{predicts } y_2$

- ▶ Within each bundle:

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

# Design

Two random responses:  $\mathbf{Y} = [y_1 \mid y_2]$

Fixed effects:

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

$$\mathbf{X} = \left[ \underbrace{\mathbf{x}_1 \dots \mathbf{x}_{15}}_{\text{bundle } \mathbf{X}_0 \text{ (large)}} \quad \underbrace{\mathbf{x}_{16} \dots \mathbf{x}_{25}}_{\text{bundle } \mathbf{X}_1 \text{ (medium)}} \quad \underbrace{\mathbf{x}_{26} \dots \mathbf{x}_{30}}_{\text{bundle } \mathbf{X}_2 \text{ (small)}} \right]$$

↳ **nuisance**      ↳ **predicts  $y_1$**       ↳ **predicts  $y_2$**

- ▶ Within each bundle:

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

Random effects:

- ▶  $N = 10$  groups and  $R = 10$  units per group
  - ↪ Design matrix  $\mathbf{U} = \mathbf{I}_N \otimes \mathbf{1}_R$

# Gaussian responses

## Model

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

$$\xi_k \sim \mathcal{N}_N(\mathbf{0}, I_N) \quad \text{and} \quad \varepsilon_k \sim \mathcal{N}_{NR}(\mathbf{0}, I_{NR})$$

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

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

# Behavioural comparison

$\tau$	LMM (No reg.)	GLMM-LASSO $\lambda_{\text{lasso}}^*$ (shrinkage)	LMM-Ridge $\lambda_{\text{ridge}}^*$ (shrinkage)	Mixed-SCGLR ( $l = 4$ ) $H^*$ (nb comp.)	$s^*$ (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

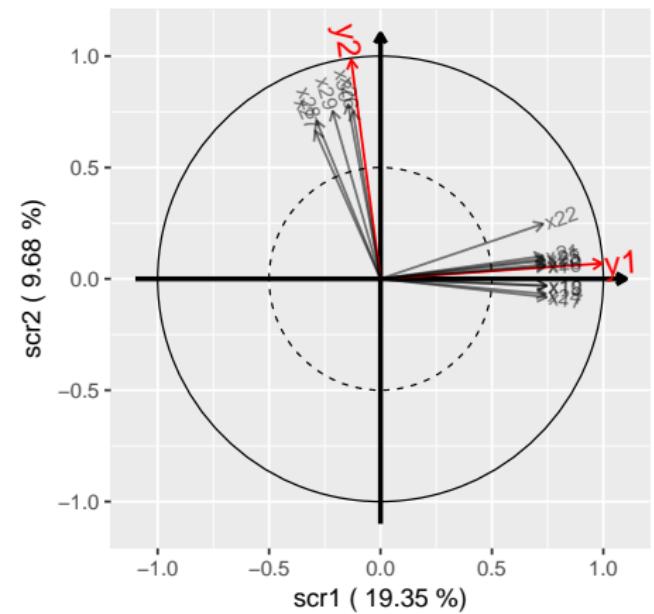
$\tau$	Ave $\left[ \max \left( \frac{\ \widehat{\beta_1} - \beta_1\ _2^2}{\ \beta_1\ _2^2}, \frac{\ \widehat{\beta_2} - \beta_2\ _2^2}{\ \beta_2\ _2^2} \right) \right]$
0.1	0.12
0.3	0.33
0.5	0.61
0.7	1.32
0.9	4.62

# Model interpretation

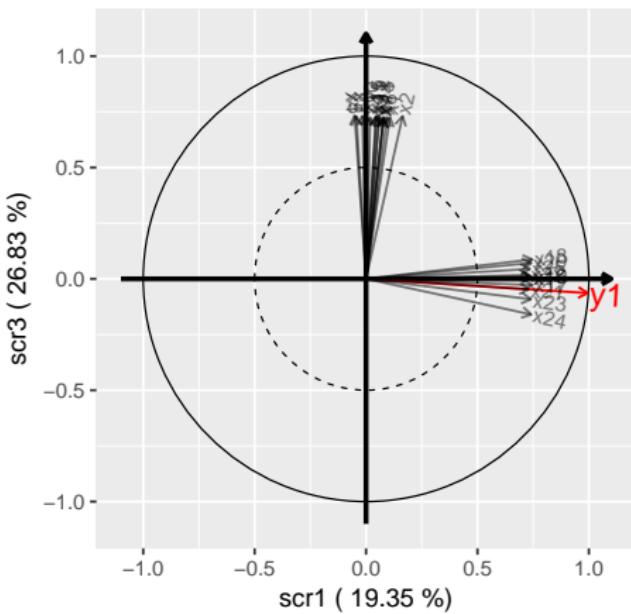
Redundancy level

$$\tau = 0.5$$

Component plane (1,2)



Component plane (1,3)



# Bernoulli and Poisson responses

## Model

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

Fixed-effect squared relative errors:

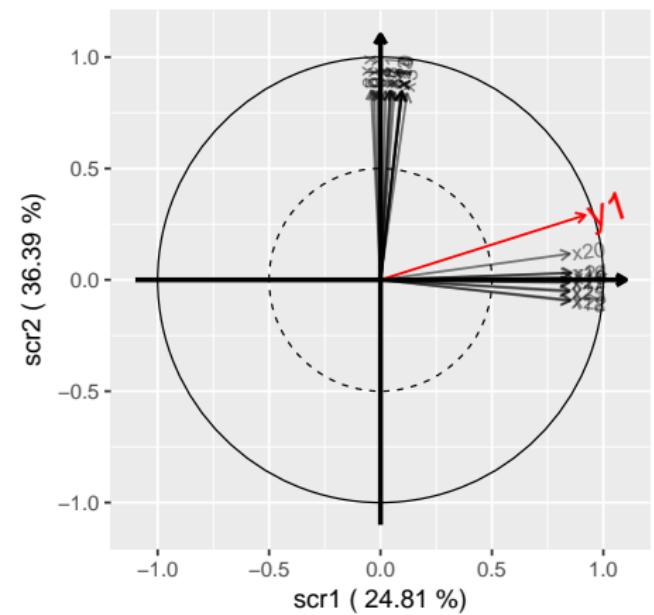
$\tau$	GLMM (no reg.)		GLMM-LASSO		mixed-SCGLR	
	Ber	Poi	Ber	Poi	Ber	Poi
0.1	316.48	0.54	8.61	0.30	14.71	0.46
0.3	398.78	0.64	9.23	0.36	7.21	0.21
0.5	576.68	0.87	14.48	0.44	2.01	0.09
0.7	886.04	1.28	17.37	0.47	1.50	0.07
0.9	2840.10	3.72	17.24	0.59	1.31	0.05

# Model interpretation

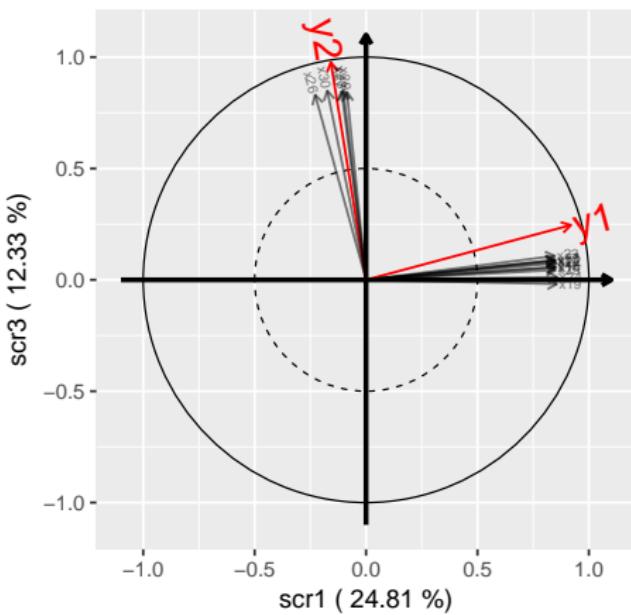
Redundancy level

$$\tau = 0.9$$

Component plane (1,2)



Component plane (1,3)



- 1 The Congo–Basin floristic data
- 2 The mixed–SCGLR method
- 3 Simulation study
- 4 Results on the floristic data
- 5 Conclusion and perspectives

# Poisson model

- ▶  $n = 2615$  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 ( $A$ )

# Poisson model

- ▶  $n = 2615$  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 ( $A$ )

## Model

Abundance of tree species: count data

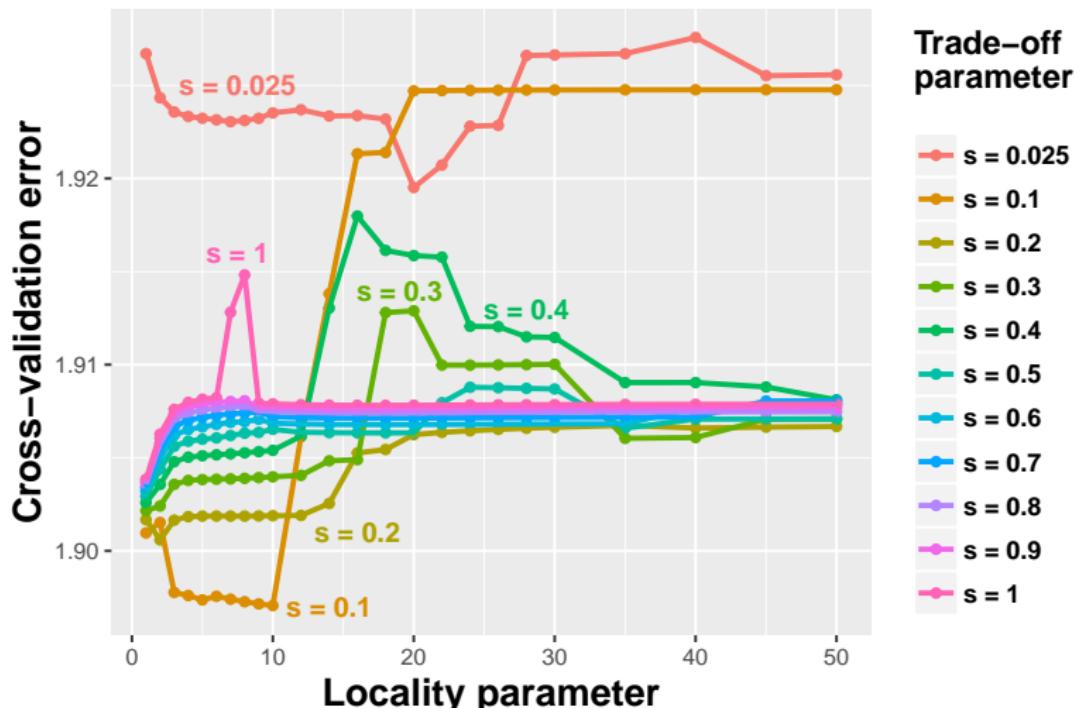
↪ Poisson regression with log link

$$y_k \sim \text{Poi}(\lambda = \exp[\eta_k])$$

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

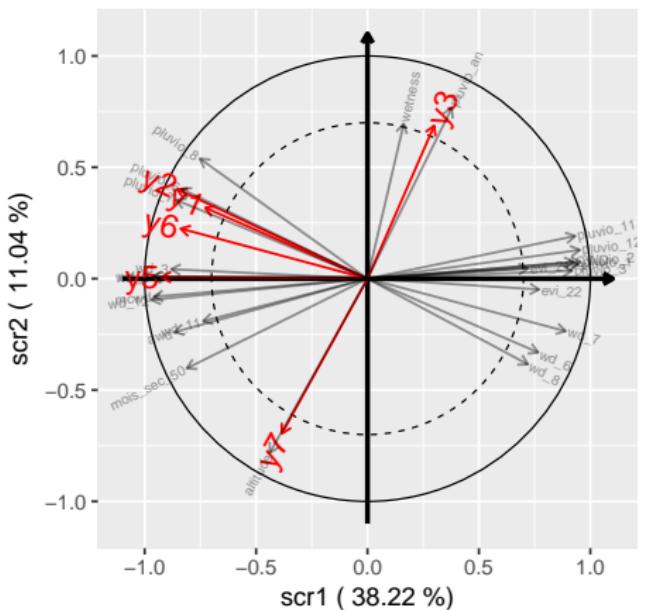
# Parameter calibration

## Behaviour of the cross-validation error ( $H^* = 4$ )

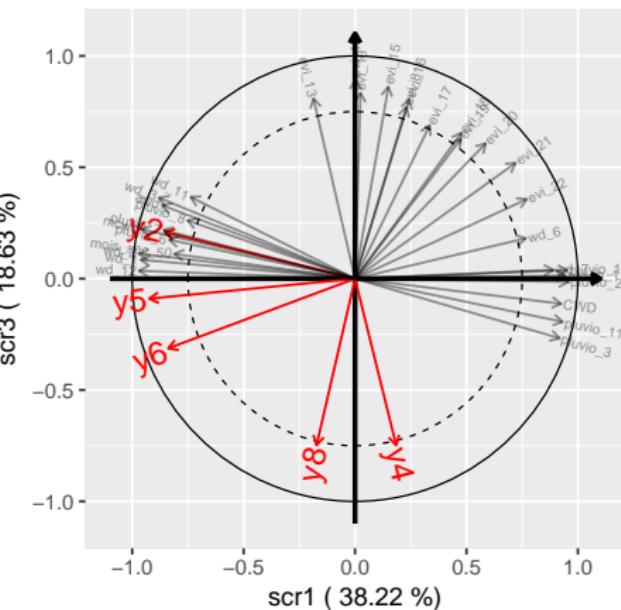


# Model interpretation

## Component plane (1,2)

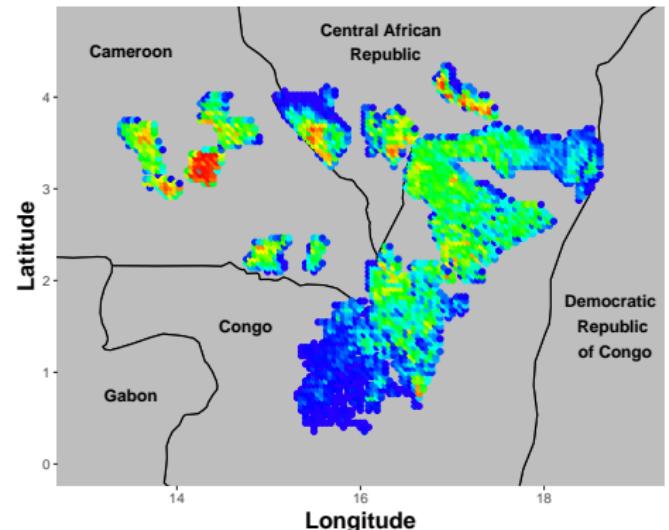


### Component plane (1,3)

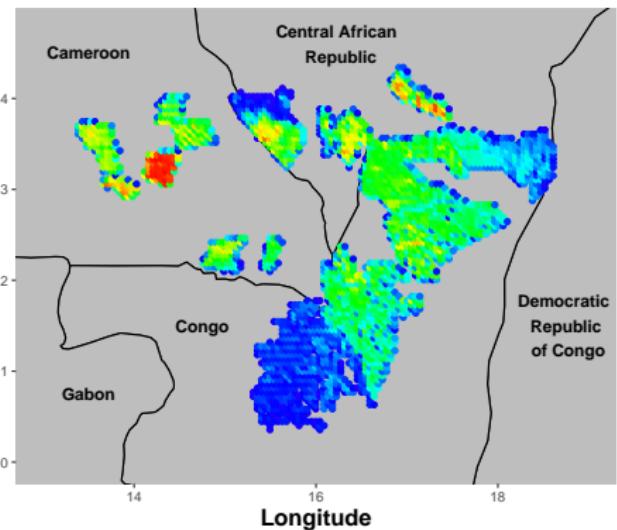


# Abundance map

## Real abundance



## Prediction



- 1 The Congo–Basin floristic data
- 2 The mixed–SCGLR method
- 3 Simulation study
- 4 Results on the floristic data
- 5 Conclusion and perspectives

# 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

# The $p > n$ case — First strategy

Replace  $X$  with the matrix  $\mathbf{C}$  of its principal components associated with non-negligible eigenvalues

- ▶  $\mathbf{C} = \mathbf{X}\mathbf{V}$ , with  $\mathbf{V}$  the matrix of unit-eigenvectors
- ▶ Modified GoF and SR criteria

$$\tilde{\psi}(\mathbf{u}) = \sum_{k=1}^q \left\| \Pi_{\text{span}\{\mathbf{C}\mathbf{u}, \mathbf{A}\}} \mathbf{z}_k \right\|_{\mathbf{W}_k^{-1}}^2$$

$$\tilde{\phi}(\mathbf{u}) = \left( \sum_{j=1}^p \left[ \text{cor}^2 (\mathbf{C}\mathbf{u}, \mathbf{x}_j) \right]^l \right)^{\frac{1}{l}}$$

- ▶ Maximisation program

$$\begin{cases} \max & s \log [\tilde{\phi}(\mathbf{u})] + (1-s) \log [\tilde{\psi}(\mathbf{u})] \\ \text{subject to} & \mathbf{u}^\top [\mathbf{C}^\top \mathbf{P} \mathbf{C}] \mathbf{u} = 1, \quad \mathbf{P} = n^{-1} \mathbf{I}_n \end{cases}$$

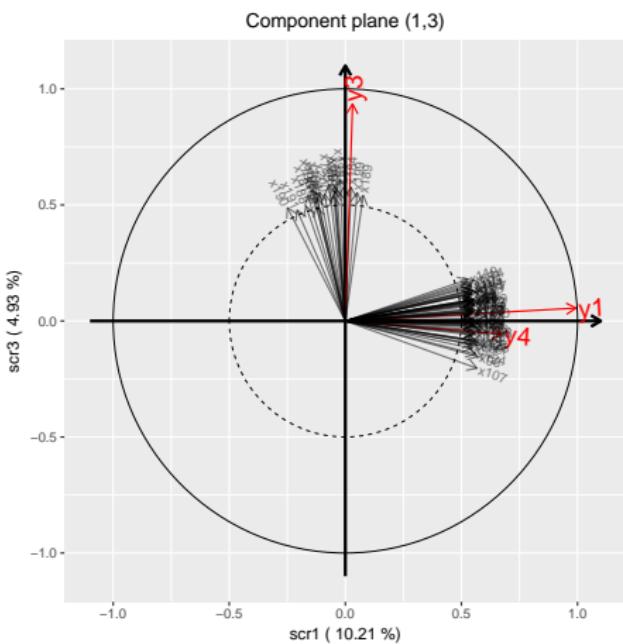
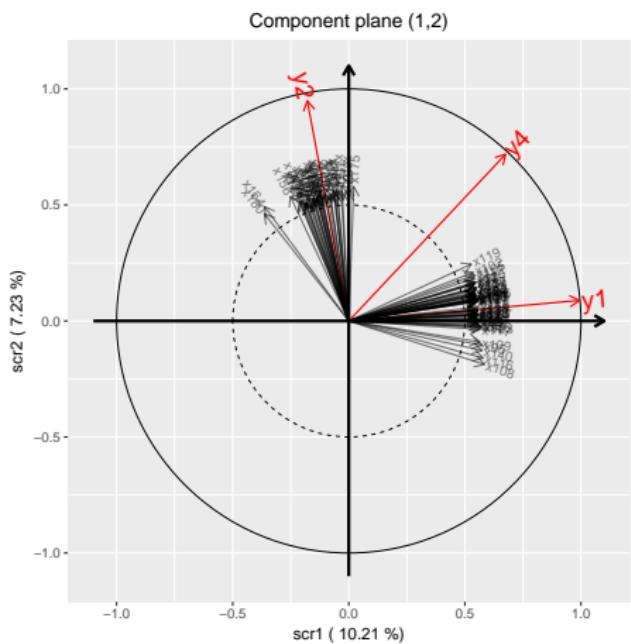
# The $p > n$ case — Design example

- $n = 100$  observations
- $p = 150$  explanatory variables:  $X = [X_0 \mid X_1 \mid X_2 \mid X_3]$
- Responses:

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

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

# The $p > n$ case — Model interpretation



# The $p > n$ case — Alternative strategy

- Preserve the GoF and SR criteria

$$\psi(\mathbf{u}) = \sum_{k=1}^q \left\| \Pi_{\text{span}\{\mathbf{X}\mathbf{u}, \mathbf{A}\}} \mathbf{z}_k \right\|_{\mathbf{W}_k^{-1}}^2$$

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

- Modify the norm-constraint

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

# Sparse Supervised Component

**Synthetic variables**  $\{f_h = \mathbf{X}\mathbf{u}_h \mid h = 1, \dots, K\}$

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

$$\mathbf{u}_h = \begin{cases} \arg \max_{\mathbf{u} \in \mathbb{R}^p} s \log [\phi(\mathbf{u})] + (1-s) \log [\psi(\mathbf{u})] - \lambda \|\mathbf{u}\|_1 \\ \|\mathbf{u}\| = 1 \text{ and } \mathbf{X}\mathbf{u} \perp \mathbf{X}\mathbf{u}_1, \dots, \mathbf{X}\mathbf{u}_{h-1} \end{cases}$$

# 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.