Some contributions on variable selection in nonlinear mixed-effects models.

Maud Delattre
Université Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France.
https://madelattre.github.io/

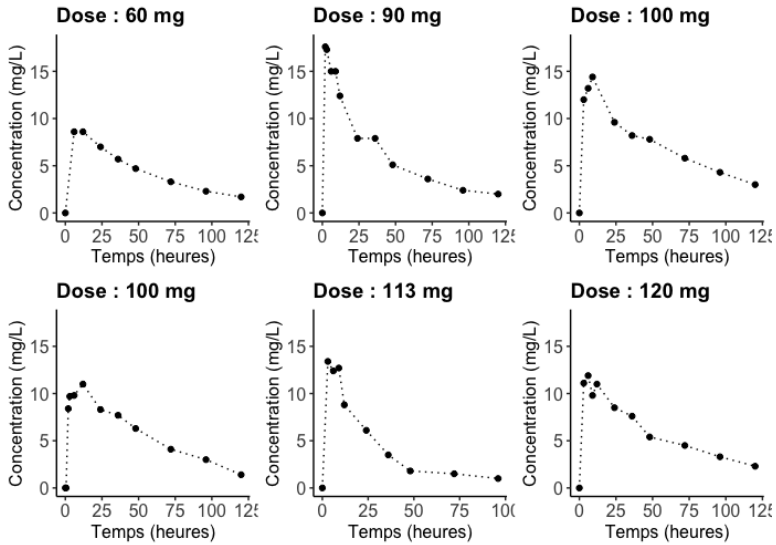
1. Introduction and model formulation

2. Covariate selection
   - On the use of BIC in NLMEM
   - High-dimensional variable selection in NLMEM

1. Introduction and model formulation

2. Covariate selection
   - On the use of BIC in NLMEM
   - High-dimensional variable selection in NLMEM

Motivating example : pharmacokinetic (PK) studies

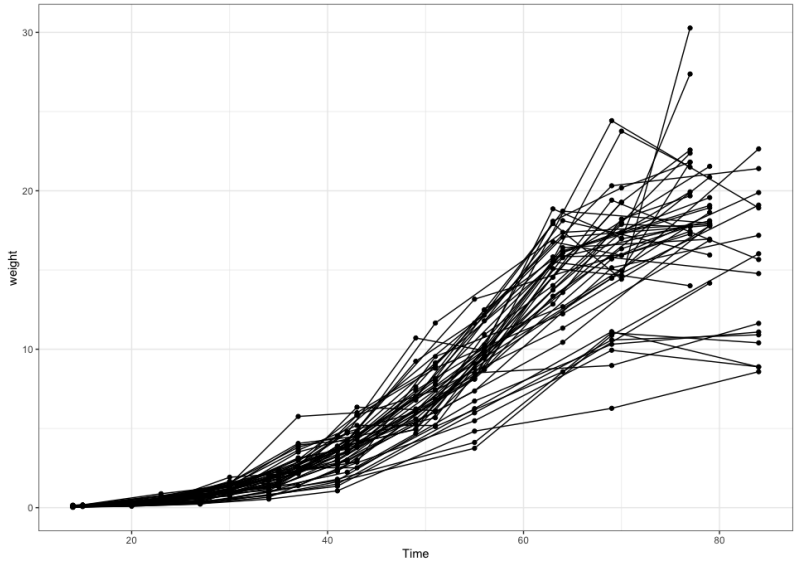


Clinical issues : define the optimal modalities of drug administration

Statistical issues :

- describe the time course of drug absorption, distribution, metabolism, and excretion in the body
- estimate key PK parameters (such as absorption or elimination rates)
- evaluate the influence of covariates (such as age or weight)

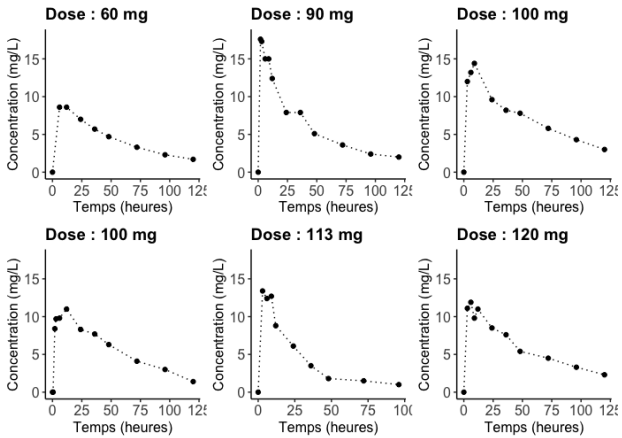
Other example : growth of soybean plants (`nlme`)

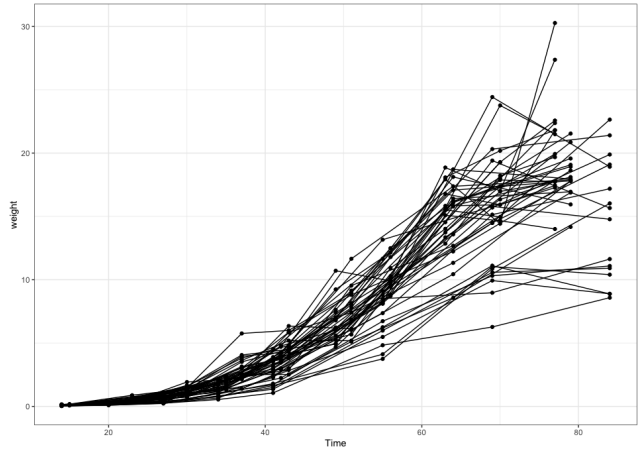


Biological issues : optimize crops

Statistical issues :

- describe the growth curves and estimate meaningful quantities (growth rate, growth plateau, etc.)
- characterize the part due to the genotype and the part due to the year of culture in the observed differences
- be able to make predictions





Common features that require special consideration in the statistical models

- repeated measurements from several subjects
- similarly-shaped individual profiles
- peak/maximum value, rise, decay vary from one individual to another

1) Description of the intra-individual variability

$$y_{ij} \underset{ind.}{\sim} f(t_{ij}, \phi_i) + g(t_{ij}, \phi_i) e_{ij} \ , \ e_{ij} \underset{i.i.d.}{\sim} \mathcal{N}(0,1) \ , \ j = 1 : n_i \ , \ i = 1 : N$$

where

- y_{ij} : jth observation for individual i
- f, g : (nonlinear, often mechanistic) functions governing within-individual behavior/residual variance
- $\phi_i \in \mathbb{R}^d$: parameters for individual i

2) Description of the inter-individual variability

$$\phi_i = \mu + \beta C_i + \eta_i, \ \eta_i \underset{i.i.d.}{\sim} \mathcal{N}(0, \Omega)$$

- $C_i \in \mathbb{R}^q$: vector of covariates for individual i
- $\mu \in \mathbb{R}^d$: intercept vector
- β : $d \times q$ matrix of fixed-effects
- $\theta = (\mu, \beta, \Omega)$

PK example (Data)

1) Observations :

$$y_{ij} = f_{ij} + f_{ij} e_{ij} \ , \ f_{ij} = f(t_{ij}, D_i, \phi_i) \ , \ e_{ij} \underset{i.i.d.}{\sim} \mathcal{N}(0,1)$$

$$f(t, D, k_a, V, Cl) = \frac{D\, k_a}{V\, k_a - Cl} \left(e^{-(Cl/V)\, t} - e^{-k_a\, t} \right)$$

2) Individual parameters :

$$\begin{aligned} \log k_{a,i} &= \mu_{ka} \ , \\ \log V_i &= \mu_V + \beta_V c_i + \eta_{V,i} \ , \ \eta_{V,i} \underset{i.i.d.}{\sim} \mathcal{N}(0, \omega_V^2) \\ \log Cl_i &= \mu_{Cl} + \eta_{Cl,i} \ , \ \eta_{Cl,i} \underset{i.i.d.}{\sim} \mathcal{N}(0, \omega_{Cl}^2) \end{aligned}$$

Typical questions

1) Estimation of the population parameter θ
   - Computation ?
   - Theoretical properties ?
2) Computation of the Fisher information matrix, computation of the likelihood, ...
3) Model building, ...

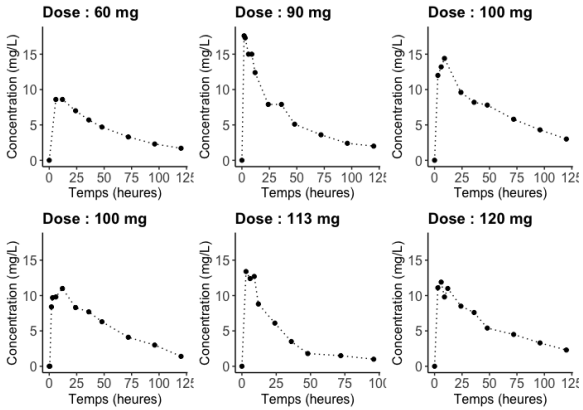
Difficult questions because of non linearity + latent variable structure of the model

1. Introduction and model formulation

2. Covariate selection
   - On the use of BIC in NLMEM
   - High-dimensional variable selection in NLMEM

1. Joint work with Marie-Anne Poursat and Marc Lavielle published in *A note on BIC in mixed effects models.*. (2014), Electronic Journal of Statistics 8(1) p. 456-475
2. Joint work with Marie-Anne Poursat published in *An iterative algorithm for joint covariate and random effect selection in mixed effects models.* (2020), The International Journal of Biostatistics 16(2), 20190082

Purpose : Characterizing inter-individual variability



$$y_{ij} \underset{ind.}{\sim} f(t_{ij}, \phi_i) + g(t_{ij}, \phi_i) e_{ij} \ , \ e_{ij} \underset{i.i.d.}{\sim} \mathcal{N}(0,1)$$
$$\phi_i = \mu + \beta C_i + \eta_i, \eta_i \underset{i.i.d.}{\sim} \mathcal{N}(0, \Omega)$$

Questions

- which covariates for which components of ϕ_i ?
  $\hookrightarrow$ support of β (i.e. non nul coefficients in β) ?

$$S_\beta = \{(k, k') \in \{1, \dots, d\} \times \{1, \dots q\} : \beta_{k,k'} \neq 0\}$$

- on which parameters to include random effects, with which covariance structure (i.e. non nul elements in Ω) ?
  $\hookrightarrow$ Related work : tests on variance components in NLMEM, see (Baey et al., 2019)

- State of the art
  - Many methods for linear mixed-effects models
    - LASSO (Bondell H. D. et al., 2010), (Schelldorfer J. et al., 2011)
    - ...

    ...but no contribution in nonlinear models
  - In practice : BIC (Bayesian information criterion) is widely used (Schwarz G., 1978)

$$-2\mathcal{LL} + \dim(\theta) \times \log(\text{sample size})$$

$\rightarrow$ Sample size definition is unclear : $= N$ or $n_{tot} = \sum_{i=1}^{N} n_i$?

no consensus, both definitions are implemented in softwares

- Purpose : propose a good definition of the BIC by revisiting the theoretical foundations of the criterion

Theoretical fundations of the criterion (Schwarz G., 1978)

- Model collection : $\mathfrak{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_K\}$
- Bayesian model choice :

$$\begin{aligned} \hat{\mathcal{M}} &= \arg\max_{\mathcal{M}\in\mathfrak{M}} p(\mathcal{M}_k|\mathrm{Y}) \\ \text{where } p(\mathcal{M}|\mathrm{Y}) &\propto \pi(\mathcal{M})p(\mathrm{Y}|\mathcal{M}) \\ \text{and } p(\mathrm{Y}|\mathcal{M}) &= \int p(\mathrm{Y}|\theta,\mathcal{M})p(\theta|\mathcal{M})d\theta. \end{aligned}$$

- Laplace approximation :

$$\log p(\mathrm{Y}|\mathcal{M}) \approx \log p(\mathrm{Y}|\hat{\theta},\mathcal{M}) - \frac{1}{2}\log\det\left(H_{\hat{\theta}}\right), H_{\hat{\theta}} = -\frac{\partial^2 \log p(\mathrm{Y}|\theta,\mathcal{M})}{\partial\theta\partial\theta'}|_{\theta=\hat{\theta}}$$

- In models based on n independent observations, $H_{\hat{\theta}} \approx n\mathcal{I}_{\hat{\theta}}$ when $n \to \infty$ and

$$\frac{1}{2}\log\det\left(H_{\hat{\theta}}\right) \approx \frac{\dim(\theta)}{2}\log n + \frac{1}{2}\log\det\left(\mathcal{I}_{\hat{\theta}}\right),$$

- Parameter partitioning :

$$\begin{pmatrix} \phi_{F,i} \\ \phi_{R,i} \end{pmatrix} = \begin{pmatrix} \mu_F \\ \mu_R \end{pmatrix} + \begin{pmatrix} \beta_F \\ \beta_R \end{pmatrix} C_i + \begin{pmatrix} 0 \\ \eta_{R,i} \end{pmatrix} \ , \ \eta_{R,i} \underset{i.i.d.}{\sim} \mathcal{N}(0, \Omega_R)$$

- Asymptotic framework : $N \to \infty$, $\min_i n_i \to \infty$ [Nie, 2007]

$$\begin{pmatrix} \sqrt{n_{tot}} & 0 \\ 0 & \sqrt{N} \end{pmatrix} \begin{pmatrix} \widehat{\beta_F} - \beta_F^\star \\ \widehat{\beta_R} - \beta_R^\star \end{pmatrix} \longrightarrow \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \mathcal{I}_1 & 0 \\ 0 & \mathcal{I}_2 \end{pmatrix} \right)$$

  - $H_{\hat{\theta}} \approx$ block diagonal
  - $\log\det\left(H_{\hat{\theta}}\right) \approx \log\det(n_{\text{tot}}\mathcal{I}_1) + \log\det(N\,\mathcal{I}_2)$

Result

Select model (i.e. covariate structure) minimizing :

$$BIC_h = -2\mathcal{LL} + \dim(\beta_R)\log N + \dim(\beta_F)\log n_{\text{tot}}$$

Now implemented in the `saemix` *R package*

$$BIC_h = -2\mathcal{LL} + \dim(\beta_R)\log N + \dim(\beta_F)\log n_{\text{tot}}$$

Remarks

- Covariate selection depends on Ω.
- Hybrid version of the two simplest (empirical) BIC definitions :
  - If $\dim(\beta_R) = 0$,
$$BIC_{n_{\text{tot}}} = -2\mathcal{LL} + \dim(\beta)\log n_{\text{tot}}.$$
  - If $\dim(\beta_F) = 0$,
$$BIC_N = -2\mathcal{LL} + \dim(\beta)\log N.$$

Numerical experiments

- BIC_h performs better than BIC_N and $BIC_{n_{\text{tot}}}$
- Results evoke theoretical consistency for BIC_h (theoretical study under progress)

Algorithm for joint selection

- Simultaneous selection of covariates and random effects

$$\phi_i = \mu + \beta C_i + \eta_i, \eta_i \underset{i.i.d.}{\sim} \mathcal{N}(0, \Omega)$$

$\rightarrow$ Exhaustive search impossible

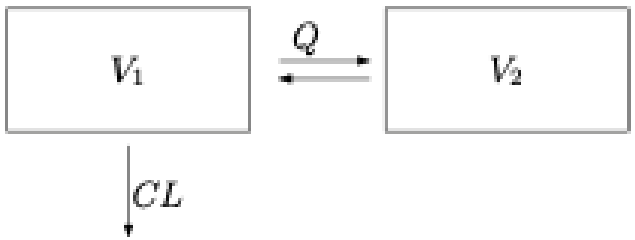
Ex. $q = 6$ covariates, $d = 3$ individual parameters $\Rightarrow \approx 2$ millions possible combinations

$\rightarrow$ Incorrect specification of $\beta \Leftrightarrow$ Incorrect specification of Ω

- Iterative algorithm, iteration t
  1. Selection of the random effect structure (BIC_Ω, $pen = \cdots \log N$)
  2. Selection of the covariate structure (stepwise, BIC_β, $pen = \cdots \log N + \cdots \log n_{tot}$)

- Numerical experiments :
  - Performance equivalent to exhaustive search or competitive methods
  - Substantial reduction of the calculation time

- Biological context : PK of an antibiotic administered by infusion (Burdet C. et al., 2015)
- Model :



$$\frac{dA_1(t,\phi)}{dt} = \frac{Q}{V_2}A_2(t,\phi) - \frac{Q+CL}{V_1}A_1(t,\phi) + \frac{dose}{T_{inf}}\frac{1}{V_1}\mathbb{1}_{0\leq t-t_D\leq T_{inf}}$$

$$\frac{dA_2(\phi)}{dt} = \frac{Q}{V_1}A_1(t,\phi) - \frac{Q}{V_2}A_2(t,\phi)$$

$$Y_{ij} = A_1(t_{ij},\phi_i) + (a + bA_1(t_{ij},\phi_i))\epsilon_{ij} \ , \ \epsilon_{ij} \underset{i.i.d.}{\sim} \mathcal{N}(0,1)$$

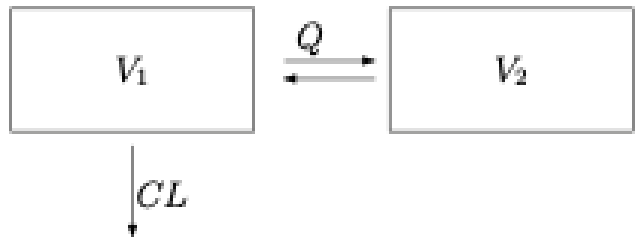
$$\phi_i = (\log CL_i, \log V_{1,i}, \log Q_i, \log V_{2,i})^\top = \mu + \beta C_i + \eta_i \ , \ \eta_i \underset{i.i.d.}{\sim} \mathcal{N}(0,\Omega)$$

- Observations (y_{ij}) : blood concentration of the antibiotic
  - $N = 53$ patients, $n_{tot} = 247$
  - $q = 12$ covariates

Pharmacokinetics example

- Model for the individual parameters :

$$\phi_i = (\log CL_i, \log V_{1,i}, \log Q_i, \log V_{2,i})^\top$$
$$= \mu + \beta C_i + \eta_i \ , \ \eta_i \underset{i.i.d.}{\sim} \mathcal{N}(0, \Omega)$$



- Result :

covariates — random effects

ϕ	$CICr$	age	$weight$	sex	P/F	η
CL	$\star\bullet$	$\star$				$\star\bullet$
V_1			$\star\bullet$		$\star\bullet$	$\star\bullet$
Q				$\star$		$\bullet$
V_2	$\star$					$\star\bullet$

	BIC_N
$\star$	1492.624
$\bullet$	1643.676

($\star$) our method

($\bullet$) (Burdet C. et al., 2015)

$\rightarrow$ best compromise between covariates and random effects

$\rightarrow$ best choice of dose according to the sex and age of the patients ?

1 Introduction and model formulation

2 Covariate selection
- On the use of BIC in NLMEM
- High-dimensional variable selection in NLMEM

First part of M. Naveau's PhD

Preprint available at `https://hal.inrae.fr/hal-03685060`

M. Naveau et al. (2022) *Bayesian high-dimensional covariate selection in non-linear mixed-effects models using the SAEM algorithm.*

Model

1) Description of the intra-individual variability

$$y_{ij} \underset{ind.}{\sim} f(t_{ij}, \phi_i) + g(t_{ij}, \phi_i) e_{ij} \ , \ e_{ij} \underset{i.i.d.}{\sim} \mathcal{N}(0,1) \ , \ j = 1 : n_i \ , \ i = 1 : N$$

2) Description of the inter-individual variability

$$\phi_i = \mu + \beta C_i + \eta_i, \ \eta_i \underset{i.i.d.}{\sim} \mathcal{N}(0, \Omega)$$

Issue : covariate selection in large dimension framework, *ie* $q >> N$
$\hookrightarrow$ BIC can not be used anymore

Consider $d = \dim(\phi_i) = 1$ (but extension to $d > 1$ should be straightforward)

Frequentist framework :

- Linear mixed-effects models : theoretical consistency results on Lasso-type methods (Schelldorfer J. et al., 2011).
- Nonlinear mixed-effects models : computational aspects only (Bertrand J. and Balding D. J., 2013), (Ollier, 2022).

Bayesian framework :

- Linear (fixed-effects) models : use of sparsity priors with both computational and theoretical contributions (see Tadesse and Vannucci (2021)).
- Nonlinear mixed-effects models : implementation of MCMC methods (Lee, 2022).

Our approach : Combining *spike-and-slab* priors for variable selection and the *MCMC-SAEM* algorithm.

Spike-and-slab prior (George and McCulloch, 1997)

For each $1 \leq \ell \leq q$: inclusion variable

$$\delta_\ell = \begin{cases} 1 & \text{if covariate } \ell \text{ is to be included in the model,} \\ 0 & \text{otherwise.} \end{cases}$$

and distribution

$$\pi(\beta|\delta) = \mathcal{N}_p(0, \text{diag}((1-\delta_\ell)\nu_0 + \delta_\ell \nu_1)) \text{ , } 0 \leq \nu_0 < \nu_1 \text{ fixed.}$$

i.e. the β_ℓ's are independent and :

- $\beta_\ell|(\delta_\ell = 0) \sim \mathcal{N}(0, \nu_0)$: "spike" distribution, ν_0 small
- $\beta_\ell|(\delta_\ell = 1) \sim \mathcal{N}(0, \nu_1)$: "slab" distribution, ν_1 large

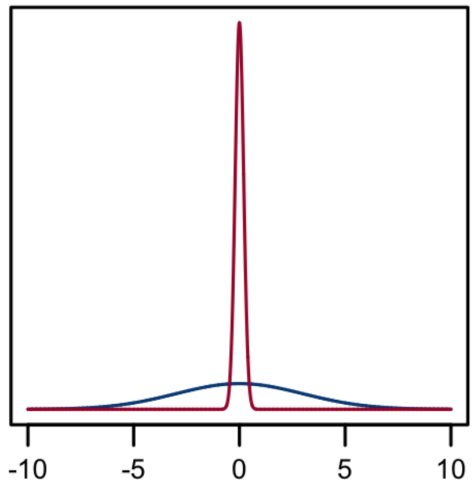
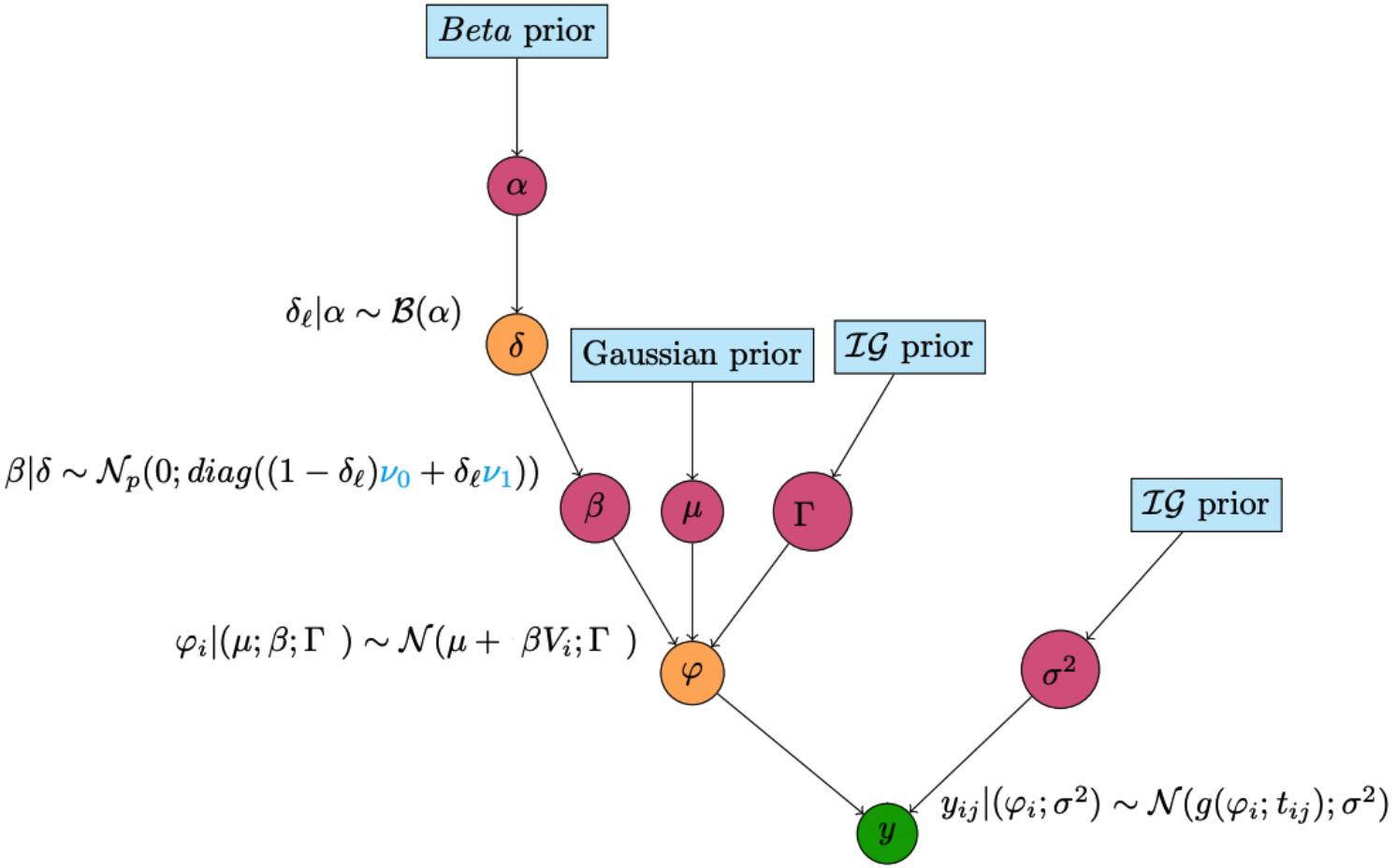


Figure – Prior spike-and-slab. Source : Deshpande et al. (2019)

Hierarchical bayesian model



Proposed method : explore different levels of sparsity for β by trying different values for ν_0

1. For $\nu_0 \in \Delta$,
   - Compute the maximum *a posteriori* estimator of Θ with an MCMC-SAEM algorithm :

$$\widehat{\Theta}_{\nu_0}^{MAP} = \underset{\Theta \in \Lambda}{\text{argmax}}\ \pi(\Theta|y)$$

   - Select covariates (Ročková and George, 2014) :

$$\widehat{S}_{\nu_0} = \left\{ \ell \in \{1, \ldots, p\} \;\middle|\; |(\widehat{\beta}_{\nu_0}^{MAP})_\ell| \geq s_\beta(\nu_0, \nu_1, \widehat{\alpha}_{\nu_0}^{MAP}) \right\}$$

2. Select the "best" model among $(\widehat{S}_{\nu_0})_{\nu_0 \in \Delta}$ with eBIC (Chen and Chen, 2008) :

$$\hat{\nu}_0 = \underset{\nu_0 \in \Delta}{\text{argmin}} \left\{ eBIC(\widehat{S}_{\nu_0}) \right\}$$

Proposed method : explore different levels of sparsity for β by trying different values for ν_0

1. For $\nu_0 \in \Delta$,
   - Compute the maximum *a posteriori* estimator of Θ with an MCMC-SAEM algorithm :

$$\widehat{\Theta}^{MAP}_{\nu_0} = \underset{\Theta \in \Lambda}{\text{argmax}}\ \pi(\Theta|y)$$

   - Select covariates (Ročková and George, 2014) :

$$\widehat{S}_{\nu_0} = \left\{ \ell \in \{1, \ldots, p\} \;\middle|\; |(\widehat{\beta}^{MAP}_{\nu_0})_\ell| \geq s_\beta(\nu_0, \nu_1, \widehat{\alpha}^{MAP}_{\nu_0}) \right\}$$

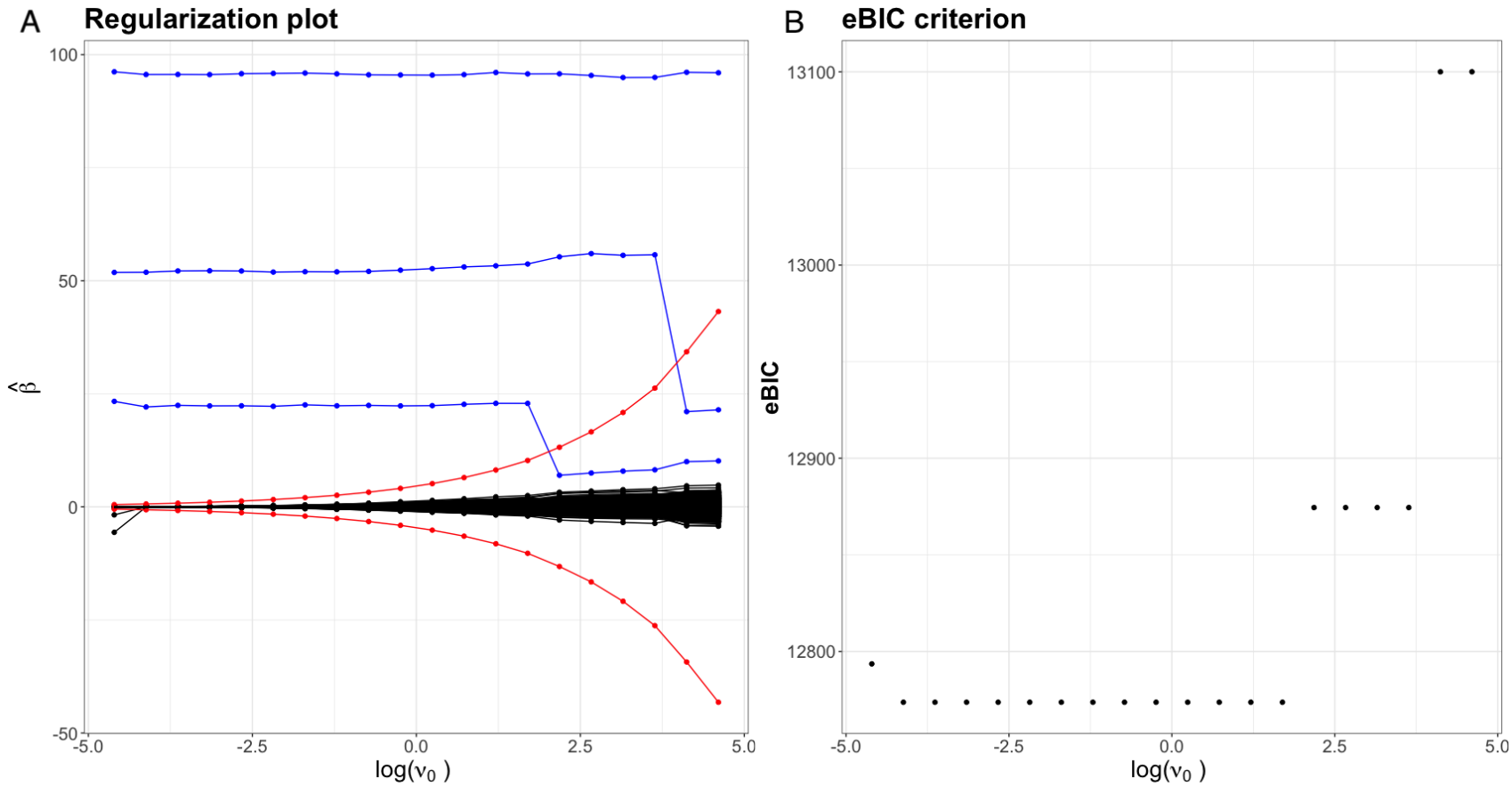
   Threshold ? Estimate the *a posteriori* inclusion probabilities of each covariate :

$$\hat{\delta} = \underset{\delta}{\text{argmax}}\ P(\delta|\hat{\Theta}^{MAP}_{\nu_0}) \text{ s.t. } \hat{\delta}_\ell = 1 \Longleftrightarrow \mathbb{P}(\delta_\ell = 1|\hat{\Theta}^{MAP}_{\nu_0}) \geq 0.5$$

2. Select the "best" model among $(\widehat{S}_{\nu_0})_{\nu_0 \in \Delta}$ with eBIC (Chen and Chen, 2008) :

$$\hat{\nu}_0 = \underset{\nu_0 \in \Delta}{\text{argmin}} \left\{ eBIC(\widehat{S}_{\nu_0}) \right\}$$

Regularization plot



Numerical experiments

Model

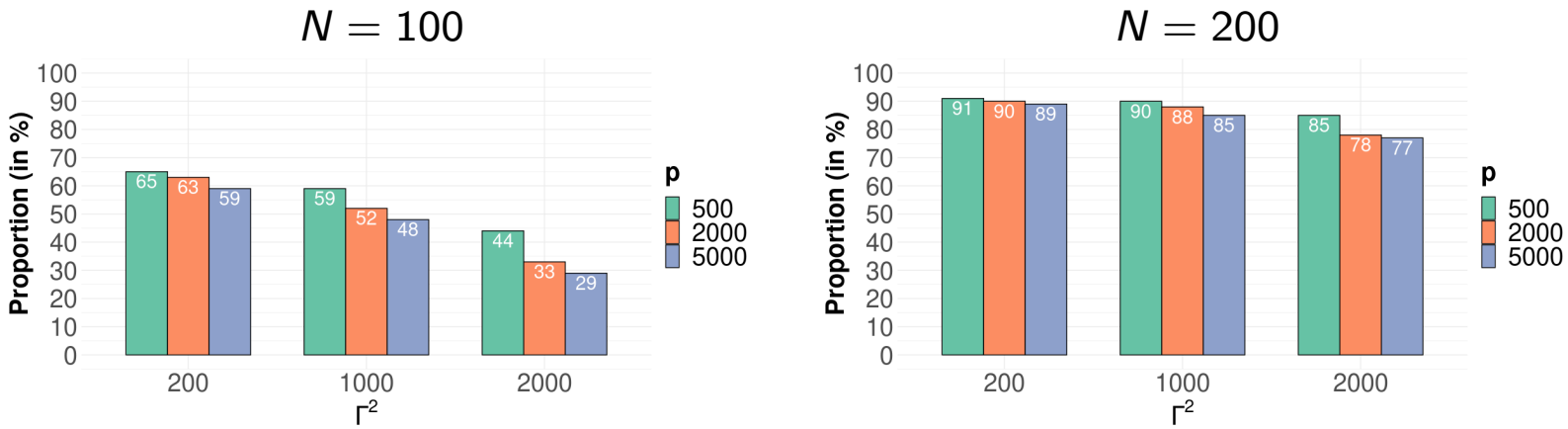
- $y_{ij} = f(\varphi_i, \psi, t_{ij}) + \varepsilon_{ij},\ \varepsilon_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$ where

$$f(\varphi_i, \psi, t_{ij}) = \frac{\psi_1}{1 + \exp\left(-\frac{t_{ij} - \varphi_i}{\psi_2}\right)}$$

  $\psi = (\psi_1, \psi_2)$ fixed effects.

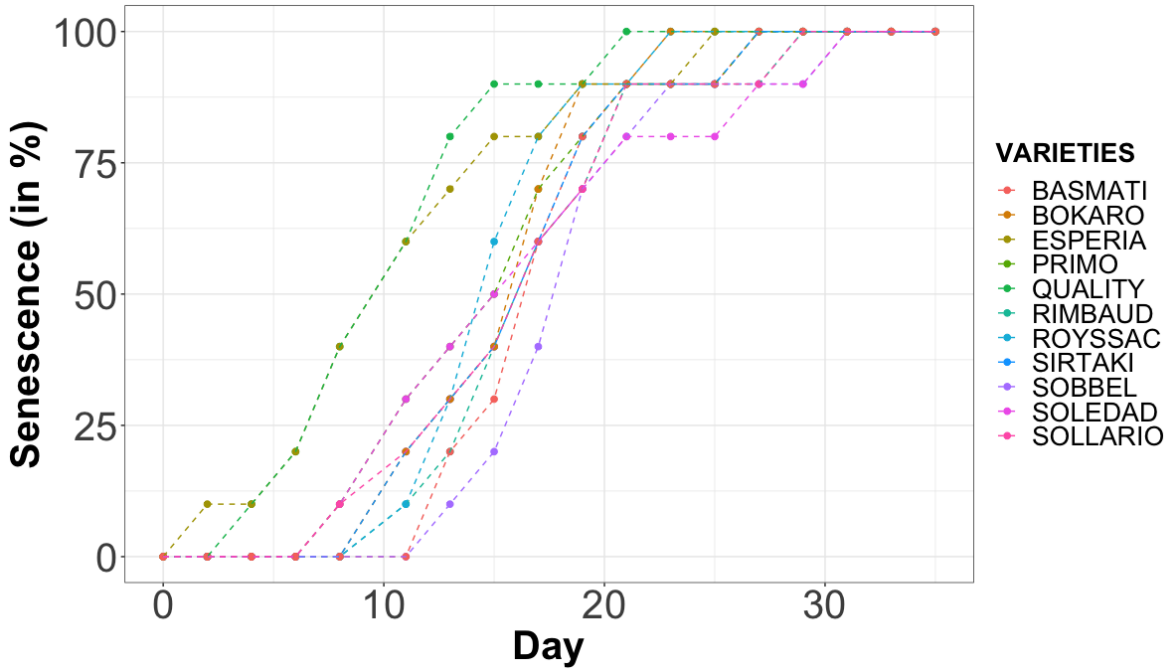
- $\varphi_i = \mu + {}^t\beta V_i + \xi_i,\ \xi_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \Gamma^2)$

Results



Application : plant senescence genetic marker identification

- Observations : proportion of senesced surface of the canopy (y_{ij}) for $N = 220$ wheat varieties at $n = 18$ time points



- Goal : identify genetic markers associated to senescence
  - 26 189 SNPs in total
  - chromosome by chromosome analysis $\Rightarrow q \equiv 2000 >> N$ for each chromosome

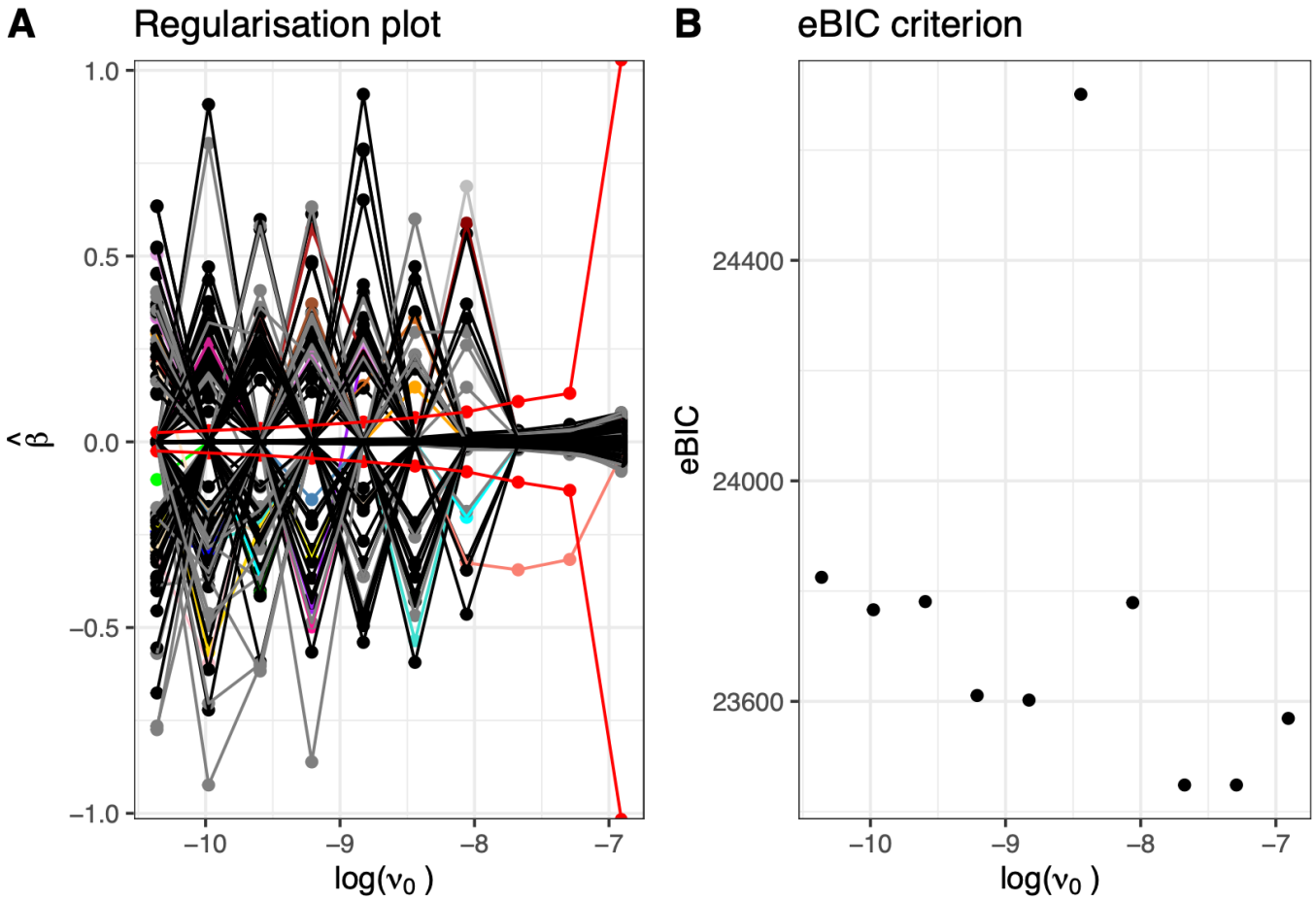
Model

$$\begin{cases} y_{ij} = \dfrac{100}{1 + \exp\left(-\dfrac{t_{ij} - \varphi_i}{\psi_i}\right)} + \varepsilon_{ij}, & \varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2), \\ \varphi_i = \mu + \lambda v_i + \beta C_i + \xi_i, & \xi_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Gamma^2), \\ \psi_i = \eta + \omega_i, & \omega_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Omega^2), \end{cases}$$

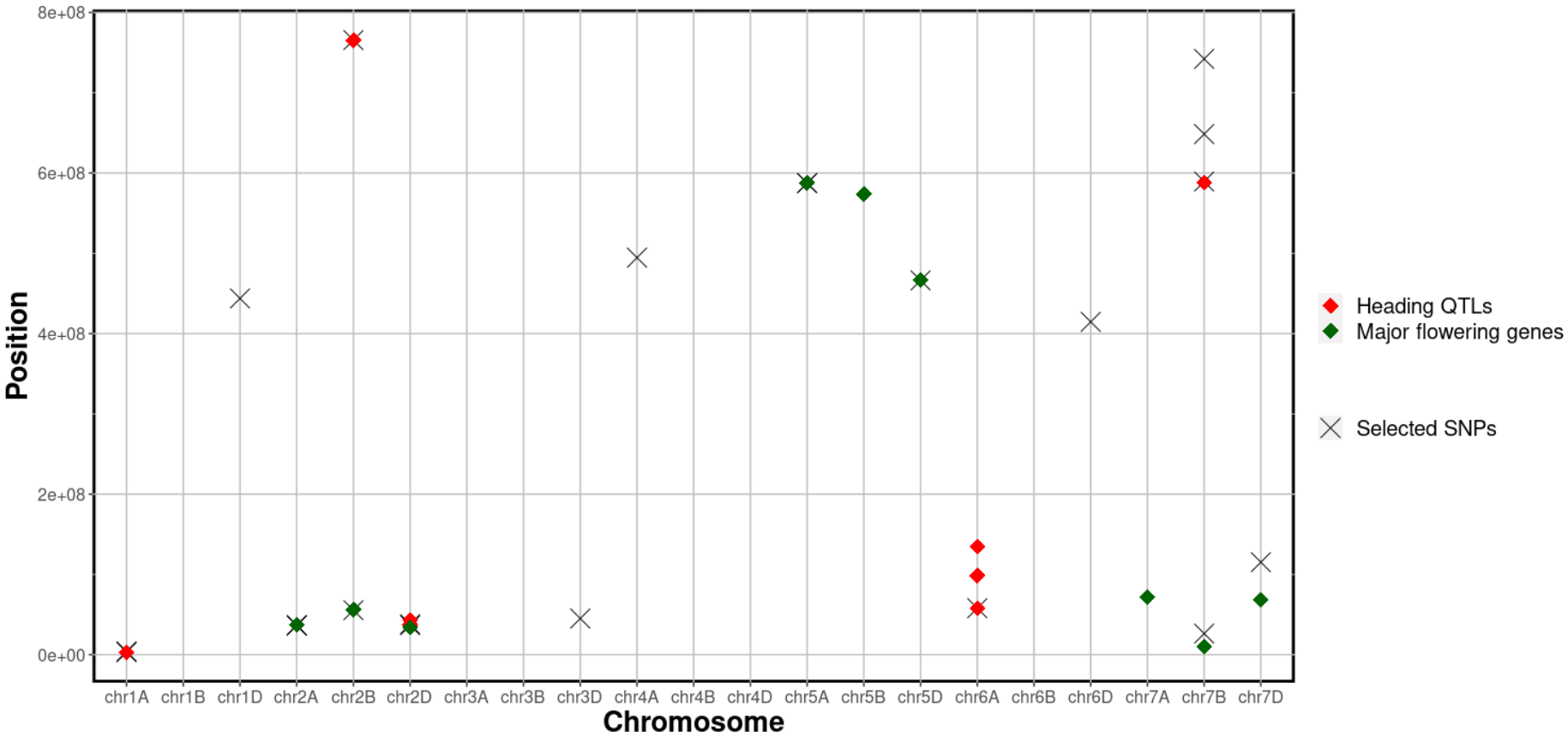
where

- C_i : q-vector of genetic markers for variety i, $1 \leq i \leq N$
- v_i : sub-populations covariates, not subject to selection

Regularization plot for Chromosome 6A



Results


8e+08
6e+08
4e+08
2e+08
0e+00
Position
chr1A chr1B chr1D chr2A chr2B chr2D chr3A chr3B chr3D chr4A chr4B chr4D chr5A chr5B chr5D chr6A chr6B chr6D chr7A chr7B chr7D
Chromosome
Heading QTLs
Major flowering genes
Selected SNPs

References I

Baey, C., Cournède, P.-H., and Kuhn, E. (2019). Asymptotic distribution of likelihood ratio test statistics for variance components in nonlinear mixed effects models. Computational Statistics and Data Analysis, 135 :107–122.

Bertrand J. and Balding D. J. (2013). Multiple single nucleotide polymorphism analysis using penalized regression in nonlinear mixed-effect pharmacokinetic models. Pharmacogenetics and Genomics, 23 :167–174.

Bondell H. D., Krishna A., and Ghosh S. K. (2010). Joint Variable Selection for Fixed and Random Effects in Linear Mixed-Effects Models. Biometrics, 66(4) :1069–1077.

Burdet C., Pajot O., Couffignal C., Armand-Lefevre L., Foucrier A., LaouÃ©nan C., Wolff M., Massias M., and MentrÃ© F. (2015). Population pharmacokinetics of single-dose amikacin in critically ill patients with suspected ventilator-associated pneumonia. European Journal of Clinical Pharmacology, 71 :75–83.

Chen, J. and Chen, Z. (2008). Extended bayesian information criteria for model selection with large model spaces. Biometrika, 95(3) :759–771.

Deshpande, S. K., Ročková, V., and George, E. I. (2019). Simultaneous variable and covariance selection with the multivariate spike-and-slab lasso. Journal of Computational and Graphical Statistics, 28(4) :921–931.

George, E. I. and McCulloch, R. E. (1997). Approaches for bayesian variable selection. Statistica sinica, pages 339–373.

Lavielle M. (2014). Mixed Effects Models for the Population Approach. Models, Tasks, Methods and Tools. Chapman and Hall/CRC Biostatistics Series. Chapman and Hall/CRC.

Lee, S. Y. (2022). Bayesian Nonlinear Models for Repeated Measurement Data : An Overview, Implementation, and Applications. Publisher : Preprints.

Ollier, E. (2022). Fast selection of nonlinear mixed effect models using penalized likelihood. Computational Statistics & Data Analysis, 167.

Pinheiro J. C. and Bates D. M. (2000). Mixed-Effects Models in S and S-PLUS. Statistics and Computing. Springer-Verlag New York Inc., 1st ed. edition.

References II

Ročková, V. and George, E. I. (2014). Emvs : The em approach to bayesian variable selection. Journal of the American Statistical Association, 109(506) :828–846.

Schelldorfer J., Buhlmann P., and Van De Geer S. (2011). Estimation for High-Dimensional Linear Mixed-Effects Models Using L1-Penalization. Scandinavian Journal of Statistics, 38 :197–214.

Schwarz G. (1978). Estimating the dimension of a model. The Annals of Statistics, 6(2) :461–464.

Tadesse, M. G. and Vannucci, M. (2021). Handbook of bayesian variable selection.