HAL
open science

Continuous shrinkage priors for fixed and random e↩ects selection in linear mixed models: application to genetic mapping

Benjamin Heuclin, Marie Denis, Catherine Trottier, Sébastien Tisné, Frédéric Mortier

To cite this version:

Benjamin Heuclin, Marie Denis, Catherine Trottier, Sébastien Tisné, Frédéric Mortier. Continuous shrinkage priors for fixed and random e↩ects selection in linear mixed models: application to genetic mapping. 2023. hal-04238536

HAL Id: hal-04238536

https://hal.inrae.fr/hal-04238536

Preprint submitted on 12 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Continuous shrinkage priors for fixed and random effects selection in linear mixed models: application to genetic mapping

Benjamin Heuclin[1,2,3], Marie Denis[2,3], Catherine Trottier[1,6], Sébastien Tisné[2,3] and Frédéric Mortier[4,5,*]

[1]IMAG, Univ Montpellier, CNRS, Montpellier, France,
[2]CIRAD, UMR AGAP Institut, F-34398 Montpellier, France,
[3]UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, F-34398 Montpellier, France,
[4]Forêts et Sociétés, Cirad, F-34398 Montpellier, France,
[5]Forêts et Sociétés, Univ Montpellier, Cirad, Montpellier, France,
[6]AMIS, Univ Paul-Valéry Montpellier 3, Montpellier, France.
*email: frederic.mortier@cirad.fr

Summary: The identification of random factors to include in a linear mixed model is crucial for modeling dependence structures while avoiding over-fitting. Random effects selection can be achieved by shrinking non-relevant variance parameters towards zero. We propose extending the horseshoe prior for variance components selection in a folded version. Motivated by two applications, the folded-horseshoe prior is evaluated either in a genetic breeding or in a functional mapping context. In the latter, we use a polar parametrization of the correlation matrix of random effects, using sinusoidal priors for angular parameters. Finally, we design efficient MCMC algorithms taking advantage of Kronecker product properties. From a statistical point of view, we show that the folded-horseshoe prior outperforms the folded-Cauchy when the number of parameters is close to the sample size. For variance component selection, it performs as well as the folded-spike-and-slab but it is computationally more efficient. We also show the impact of erroneous dependence structures assumptions on the selection and the estimation of variance components. From a genetic point of view, the numerical results highlight the efficiency of the folded-horseshoe prior. In particular, this prior selects molecular markers already identified in these data but also new markers. Finally, we discuss how and why linear mixed models are an interesting alternative to usual functional mapping approaches.

Key words: Angular parametrization; Fixed and random effects selection; Horseshoe prior; Linear mixed model; Quantitative genetics.

1. Introduction

Linear mixed models (LMM) are flexible tools for modeling data from a wide range of data types in various applied fields including ecology and evolution (Bolker et al., 2009; Ives and Helmus, 2011), quantitative genetics (Lynch and Walsh, 1998) or medical research (Brown and Prescott, 2014). An important practical point when using linear mixed models is the choice of random effect components. Choosing which random grouping factors to include is vital to model appropriate dependence structures within data. This issue compounds when the possible number of random effects is large, which also leads to identifiability problems and estimation instability. With new technologies, high-resolution satellite images, high-throughput genotyping/phenotyping techniques, such contexts are now very common. For instance, in quantitative genetics, linear mixed models are commonly used for IBD-QTL mapping (George et al., 2000; van Eeuwijk et al., 2010; Tisné et al., 2015) or gene-set analyses (GSA) (Fridley and Biernacka, 2011), allowing to decompose the global genetic effect as a sum of local/specific effects related to each position/block considered on the genome. In such studies, the genetic effects are associated to random effects that can vary from tens to a few hundreds.

This work has been motivated by two original applications. The first one takes place in the IBD-QTL mapping context and aims at identifying QTLs related to the *oil palm* production (*Elaeis guineensis*, Jacq). In this application, 135 random effects associated to 135 genetic positions are considered. In the functional mapping framework, the second application aims at modeling the effects of 38 genetic markers assumed varying with time, and at selecting those involved in the dynamics of shoot growth of *Arabidopsis thaliana* (L. Heynh) (Marchadier et al., 2019). While providing an alternative modeling approach to the usual functional mapping methods based on non-parametric strategies (Ma et al., 2002), our work also allows to estimate potential dependencies between random effects.

The usual solution for dealing with identifiability and/or inference instability problems is to reduce the number of variables, using model choice procedures based on information criteria (Müller, Scealy, and Welsh, 2013). An alternative strategy relies on regularization approaches (Bickel et al.,

2006). In linear models (LM), regularization procedures have been widely studied and a large set of penalty functions has been proposed (see for example Tibshirani (1996); Desboulets (2018)). Most regularization methods, initially developed in a frequentist context, have been proposed in the Bayesian framework (Kyung et al., 2010). Prior distributions act as penalty terms in the frequentist approach. A set of priors has been extensively developed, among others: spike-and-slab (George and McCulloch, 1993), Bayesian Lasso (Park and Casella, 2008), Elastic-Net (Kyung et al., 2010), normal-gamma (Griffin et al., 2010) or horseshoe (Carvalho et al., 2009) priors.

In the LMM context, literature is less developed. The first approaches to simultaneously select fixed effects and variance components used model choice criteria (Rao and Wu, 1989; Vaida and Blanchard, 2005; Müller et al., 2013; Delattre and Poursat, 2020). As in the LM context, penalized likelihood approaches have also been developed as alternatives, especially when the number of predictors increases. For instance, Bondell, Krishna, and Ghosh (2010) and Ibrahim et al. (2011) combined penalized likelihood techniques, using adaptive lasso or smoothly clipped absolute deviation (SCAD) penalties, with the Cholesky decomposition of the random effects covariance matrix or its modified version. In the high dimensional context, Fan and Li (2012) proposed a two steps approach, while Li et al. (2018) proposed a doubly regularized estimation and selection of fixed and random effects in longitudinal data. The reparametrization of the LMM by the use of a modified Cholesky decomposition of the random effects covariance matrix has initially been put forward in a Bayesian framework by Chen and Dunson (2003). Such a technique allows to consider the standard deviations of random effects as regression parameters. In their approach, a spike-and-slab prior is used for the fixed effects and the standard deviations, with a truncated Gaussian distribution for the slab part associated to the standard deviations. Frühwirth-Schnatter and Tüchler (2008) propose a related approach, modeling directly the Cholesky decomposition elements using a spike-and-slab prior where the slab distribution is an unconstrained Gaussian distribution. As for the fixed effect selection, alternative approaches to spike and-slab priors have also been developed and compared for the selection of variance components. In particular, continuous shrinkage priors, or their mixture versions, such as Student and Cauchy distributions, Bayesian Lasso and normal-gamma priors have been studied in the context of random intercept models (Frühwirth-Schnatter and Wagner, 2011). Finally, we note that the question of shrinking variance parameters towards zero does not raise only in the LMM context. Such objectives have been studied in different statistical contexts. In structured additive regression models (Fahrmeir, Kneib, and Lang, 2004) for instance, groups of fixed effects are selected using a spike-and-slab prior on specific group variance components based on a mixture of inverse gamma distributions (Scheipl, Fahrmeir, and Kneib, 2012). In time varying parameters and state-space models, Bitto and Frühwirth-Schnatter (2019) and Cadonna, Frühwirth-Schnatter, and Knaus (2020) propose the use of double or triple gamma priors extending Normal-Gamma (Griffin et al., 2010) or more generally the scaled-Beta distribution class (Pérez et al., 2017).

In this paper, we propose to combine the horseshoe prior with its folded version to simultaneously select fixed and random effects. We study the performances of the proposed prior through the two applications and discuss results in comparison with the two commonly used alternative priors: the folded Cauchy prior and the folded spike-and-slab prior. In the second application, to model dependency structures between random slopes effects, we apply polar parametrization (Pinheiro and Bates, 1996) using sinusoidal prior on angles (Pourahmadi and Wang, 2015) to ensure symmetry and positive-definiteness of the unknown correlation matrix. The paper is organized as follows. Section 2 presents the general model, priors formulation, and the specific context associated with each of the two applications along with their dedicated models. In section 3, we present the computational aspects of the Bayesian inference in order to optimize the MCMC algorithms. In Section 4, we firstly discuss results obtained by the three priors from a statistical point of view, and we then interpret results from a biological point of view.

2. Model specification and priors formulation

2.1 *General considerations*

LMMs can be expressed in the following general form:

$$y = \boldsymbol{X}\beta + \boldsymbol{Z}\tilde{u} + \varepsilon \tag{1}$$

where y is a n-response vector, $\boldsymbol{X}$ a $n \times (p+1)$-matrix of p covariates with a first unitary column for the intercept, $\boldsymbol{Z}$ a $n \times s$ known sparse random effects design matrix associated to $\tilde{u}$ a s-vector of random effects assumed to be distributed as a multivariate Gaussian distribution with null expectation and covariance matrix denoted by $\boldsymbol{\Omega}$. In the following, $\tilde{u}$ and ε are assumed independent. Such a formulation encompasses a broad set of LMMs. Each model leads to consider different design matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$, and variance matrices $\boldsymbol{\Omega}$.

For the variance component model, considering q independent random effects $\tilde{u}_l$, with c_l levels each, following a Gaussian distribution centered on zero with covariance matrix Ω_l ($l = 1, \ldots, q$), then $\boldsymbol{Z} = \bigoplus_{l=1}^{q} Z_l$ with Z_l the lth random effect design matrix, $\tilde{u} = (\tilde{u}_1', \ldots, \tilde{u}_q')'$, $\boldsymbol{\Omega}$ is a block diagonal matrix where each block is the $c_l \times c_l$ matrix Ω_l ($\boldsymbol{\Omega} = bdiag(\Omega_1, \ldots, \Omega_q)$) and $s = \sum_{l=1}^{q} c_l$. Here $\bigoplus$ denotes the column concatenation operator. $\boldsymbol{Z}\tilde{u}$ can be decomposed as a sum of random effects $\boldsymbol{Z}\tilde{u} = Z_1\tilde{u}_1 + \cdots + Z_q\tilde{u}_q$. Two special cases of the variance component model may be considered. The first one is the usual variance component model where $\Omega_l = \lambda_l^2 I_{c_l}$ with I_{c_l} the identity matrix of size c_l. In this case, levels of each random effect are assumed independent. The second one is the animal model where $c_l = n$, $Z_l = I_n$ and $\Omega_l = \lambda_l^2 A_l$, where A_l is a known $n \times n$-IBD matrix.

In the random intercept and slope (RIS) context with one grouping factor with c levels and p covariates, $\boldsymbol{Z} = J \bullet \boldsymbol{X}$ where $\bullet$ is the face-splitting product (row-by-row Kronecker product, see web appendix B for more details), J corresponds to the $n \times c$-0/1-design matrix associated to the random effect and $\boldsymbol{\Omega}$ is a block diagonal matrix such as $\boldsymbol{\Omega} = I_c \otimes \Omega$, where

Ω is a $(p+1)\times(p+1)$ unknown correlation matrix related to dependencies between random intercept and slopes. It is straightforward to extend to q random effects with c_l levels each.

When p and s are large, such models (see equation 1) must be regularized. The already proposed approaches are mainly based on the Cholesky decomposition of $\boldsymbol{\Omega}$ (Chen and Dunson, 2003; Bondell et al., 2010). In this paper, we propose to use this decomposition: $\boldsymbol{\Omega} = \boldsymbol{\Lambda R \Lambda}$, where $\boldsymbol{\Lambda}$ is a diagonal matrix and $\boldsymbol{R}$ the associated correlation matrix. A general LMM model (see equation 1) can then be reformulated as:

$$\begin{aligned} y &= \boldsymbol{X}\beta + \boldsymbol{Z\Lambda}u + \varepsilon \\ &= \left[\boldsymbol{X}, \left(u' \otimes \boldsymbol{Z}\right) P\right] \begin{bmatrix} \beta \\ \lambda \end{bmatrix} + \varepsilon \end{aligned} \tag{2}$$

where λ is the unique diagonal elements vector of $\boldsymbol{\Lambda}$ and P is the matrix that transforms λ to $vec(\boldsymbol{\Lambda})$ (Ibrahim et al., 2011). Calculation details are presented in web appendix B. Now, u is the vector of Gaussian random effects $\mathcal{N}_s\left(0, \boldsymbol{R}\right)$. Finally, when $\boldsymbol{\Omega}$ is supposed to be unknown (RIS models), we propose to write the correlation matrix $\boldsymbol{R}$ using the polar parametrization (Pinheiro and Bates, 1996; Pourahmadi and Wang, 2015). Such an approach ensures that the correlation matrix R, sampled though the posterior distribution, is a valid symmetric and positive-definite matrix with 1's on the diagonal. More details are given in section 2.3.2 bellow. Finally, ε is a multivariate Gaussian residual vector assumed to be independent of u.

2.2 *Priors formulation*

To achieve the selection of fixed effects (β) and scale parameters (λ), we consider in this work local-global priors (Polson and Scott, 2012; Piironen and Vehtari, 2017). Such priors, initially used for the selection of fixed effects only, consist in a scale mixture of Gaussian distribution on parameters a_j subjects to selection:

$$a_j|\tau^2, \omega_j^2, \sigma^2 \sim \mathcal{N}\left(0, \sigma^2\tau^2\omega_j^2\right), \quad j = 1, \ldots, q, \tag{3}$$

where σ^2 is the residual variance, τ^2 is the global parameter while ω_j^2 are the local ones. τ^2 allows to shrink all coefficients towards zero while local parameters ω_j^2 highlight non-null parameters. Such a prior encompasses a large set of well known priors. Assuming $\omega_j^2 \equiv 1$ leads to global priors such as the well known normal-inverse-gamma (NIG) prior (Gelman et al., 2004), while assuming $\tau^2 \equiv 1$ leads to local priors such as the Laplace prior (Park and Casella, 2008), the student prior (Gelman, 2006) or the normal-gamma prior (Griffin et al., 2010) (see table 1 in web appendix A). Among local-global priors, the horseshoe prior assumes that local, as well as global parameters, are distributed from folded-Cauchy distributions (Carvalho, Polson, and Scott, 2009). The horseshoe prior has demonstrated high performances for the selection of fixed effects, comparable to the spike-and-slab prior (van Erp, Oberski, and Mulder, 2019). In this article, we propose to investigate horseshoe priors to simultaneously select fixed effects and standard deviations. We note that since λ_j is positive, the Gaussian distribution is replaced by a folded-Gaussian distribution $\mathcal{N}^+\left(0, \sigma^2\tau^2\omega_j^2\right)$. The horseshoe prior is systematically applied to fixed effect. For standard deviations, we evaluate the folded horseshoe prior (fHS) addressing two specific questions: (i) what are the performances of such priors to select random effects, and (ii) which impacts on the fixed and random effect estimations. These results are discussed relatively to the use of alternative priors: folded Cauchy (fC) and folded zero-inflated (spike-and-slab, fSS).

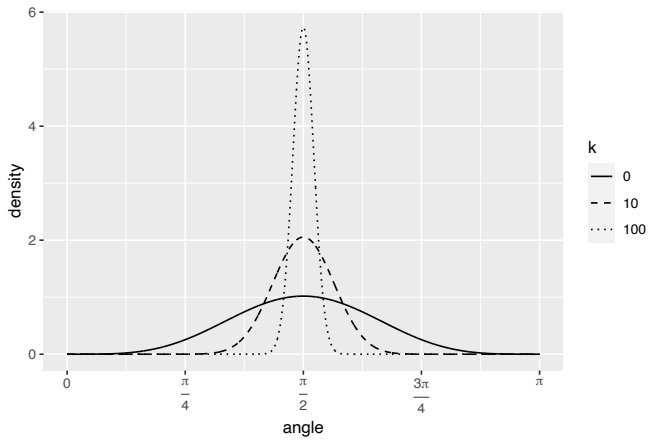

Figure 1: *Prior density distribution of the angle $\theta_{i,j}$ (see equation 5) according to different values of k: 0, 10 of 100.*

In the specific RIS context, random intercept and slopes are commonly assumed to be non-independent through an unknown correlation matrix R such that $\boldsymbol{R} = I_c \otimes R$. Different priors have been proposed (Lewandowski et al., 2009; Pourahmadi and Wang, 2015). Here, we adopt the polar parametrization introduced by Pinheiro and Bates (1996). It consists in the use of a hyperspherical parametrization of the Cholesky factors of the correlation matrix and an appropriate distribution on related angles. Pinheiro and Bates (1996) demonstrate that any correlation matrix R can be factorized as BB', with $B_{1,1} = 1,\ B_{i,1} = cos(\theta_{i,1}),\ i = 2, \ldots, q$ and

$$B_{j,i} = \begin{cases} \prod_{m=1}^{j-1} \sin(\theta_{i,m}) & \text{for } i = j, \\ \cos(\theta_{i,j}) \prod_{l=1}^{j-1} \sin(\theta_{i,l}) & \text{for } 2 \leqslant j \leqslant i-1, \end{cases} \tag{4}$$

Pourahmadi and Wang (2015) proposed the following θ's sinusoidal distribution :

$$\theta_{i,j} \propto sin(\theta)^{2k+(p+1)-j} \mathbb{1}_{0<\theta<\pi},\ i = j+1, \ldots, p, \tag{5}$$

where k is a non-negative constant. This distribution ensures that angles are centered on $\pi/2$ or equivalently that the distribution of R is centered on the identity matrix. Moreover and interestingly, parameter k can be interpreted as a shrinkage parameter (Ghosh, Mallick, and Pourahmadi, 2020). For instance, if $k = 0$ then R is distributed as a uniform distribution on the set of all $(p+1)\times(p+1)$-positive-definite correlation matrices, while if k tends to infinity, the distribution of R tends to a point mass on the unit diagonal $(p+1)\times(p+1)$-matrix (see Figure 1). In this work, k is chosen based on cross-validation procedure.

2.3 *Specific applied contexts*

2.3.1 *The oil palm dataset.* We analyse this data set within the animal model framework and give the results in section 4.1. Indeed, this first application aims at identifying the genomic positions involved in the variability of *oil palm* production traits. A total of 144 palm trees belonging to the breeding program of PalmElit, a Cirad subsidiary and leading *oil palm* breeding company (`www.palmelit.com`), were analyzed. Palm trees were genotyped with 226 molecular markers and 1,007 IBD matrices were estimated on a grid of 3 centimorgan (cM) along the genome (Tisné et al., 2015). Each genetic position l is associated to a random effect u_l with a variance equal to $\lambda^2_{u_l}$ and a correlation matrix A_l equal to the identity-by-descent (IBD) matrix (see equation 6). Then, the identification of the genomic positions is equivalent to the variance components selection. Due to the genetic characteristics of the population, *i.e.* a moderate number of individuals and generations, a subset of 135 genetic positions, spaced 10 cM apart, was considered to avoid a redundant information between consecutive genetic positions. In the next section, we will present the results for the bunch number trait.

As previously explained in subsection 2.1, the animal model can be formulated as follows:

$$y = \mu + u_1\lambda_1 + \cdots + u_q\lambda_q + \varepsilon, \tag{6}$$

where μ is an intercept and u_l is now assumed to follow a Gaussian distribution $\mathcal{N}_n(0, A_l)$, $l = 1, \ldots, q$ and λ_l is the standard deviation associated to u_l. Finally, the matrix version of the animal model is given by:

$$y = \mu + U\lambda + \varepsilon, \tag{7}$$

where U is a $n \times q$-matrix of the concatenation of the random effects $U = \bigoplus_{l=1}^{q} u_l$.
In a fully Bayesian framework, the intercept μ is supposed to be proportional to one and the residual variance σ^2 is supposed to follow an inverse-gamma distribution $\mathcal{IG}(s_{\sigma^2}, r_{\sigma^2})$ (shape and rate parametrization). The Bayesian hierarchical model is presented in web appendix C.

2.3.2 *The arabidopsis thaliana dataset.* We analyse this data set within the RIS model framework and give the results in section 4.2. Indeed, in this second application, we are interested by disentangling the evolution over time of the complex genetic architecture of shoot growth of *Arabidopsis thaliana* (L. Heynh). Data consists of leaf compactness phenotypic trait measured over $T = 21$ time points on $n = 358$ individuals. We use genetic covariates X containing $p = 38$ markers (Marchadier et al., 2019; Heuclin et al., 2020). In the RIS model framework, we consider time as the grouping factor. This model is an alternative approach to the usual non-parametric functional mapping (Ma et al., 2002). It can be expressed as follows:

$$y_{i,t} = x_i\beta + x_i\tilde{u}_t + \alpha_i + \varepsilon_{i,t} \tag{8}$$

where $y_{i,t}$ is the observation of individual i at time t ($i = 1, \ldots, n$ and $t = t_1, \ldots, t_T$). x_i is a $(p+1)$-row vector of p genetic markers (constant over time) associated to the ith individual. The first element is fixed to one and is related to the intercept. β is a $(p+1)$-vector of fixed effects, $\tilde{u}_t$ a $(p+1)$-vector of random intercept and slopes effects assumed to follow a Gaussian distribution $\mathcal{N}_{p+1}(0, \Lambda R \Lambda)$, where Λ is an unknown $(p+1) \times (p+1)$-diagonal matrix of standard deviation and R is an unknown $(p+1) \times (p+1)$-correlation matrix. In this application, R is assumed to be block diagonal where each block is related to one chromosome (Ghosh et al., 2020). The random intercept is also assumed independent from the random slopes. α_i is a Gaussian individual random effect ($\mathcal{N}\left(0, \sigma^2_\alpha\right)$) not subject to selection. $\varepsilon_{i,t}$ corresponds to the residual part such that $\varepsilon_i = (\varepsilon_{i,t_1}, \ldots, \varepsilon_{i,t_T})$ is distributed from a multivariate Gaussian distribution $\mathcal{N}_{t_T}(0, \sigma^2_\varepsilon\Gamma)$ where Γ is a $t_T \times t_T$-correlation matrix of a first-order autoregressive structure with unknown parameter ρ.

Let $y = (y'_{t_1}, \ldots, y'_{t_T})'$ be the concatenation of all measurements over time for all individuals where $y_t = (y_{1,t}, \ldots, y_{n,t})'$. Since the genetic information varies between individuals but is constant over time, $\boldsymbol{X}$ can be simplified such that $\boldsymbol{X} = (\mathbb{1}_{t_T} \otimes X)$ where X is the $n \times (p+1)$-matrix containing the p genetic markers (and the intercept) of all individuals. Matrix J is here equal to $I_{t_T} \otimes \mathbb{1}_n$. The random effects design matrix $\boldsymbol{Z}$ can also be simplified as:

$$\boldsymbol{Z} = I_{t_T} \otimes X. \tag{9}$$

Calculation details are presented in web appendix B. Finally, $\boldsymbol{\Lambda}$ is decomposed as $I_{t_T} \otimes \Lambda$ and $\boldsymbol{R} = I_{t_T} \otimes R$. P is the matrix that transforms λ to $\mathbb{V}ec(\Lambda)$ (or equivalently $\Lambda = diag(\lambda)$ and $\mathbb{V}ec(\Lambda) = P\lambda$). Then, as proposed in section 2.1, this model can be expressed as:

$$y = [\mathbb{1}_{t_T} \otimes X, (U' \otimes X)P] \begin{bmatrix} \beta \\ \lambda \end{bmatrix} + D\alpha + \varepsilon. \tag{10}$$

Calculation details are presented in web appendix B. U is a $(p+1) \times t_T$-matrix of the collection of the t_T reparametrized vectors of random intercept and slopes associated to each time $U = \bigoplus_k^{t_T} u_k$ (U follows a matrix Gaussian distribution $\mathcal{MN}_{(p+1)\times t_T}(0, R, I_{t_T})$). $D = \mathbb{1}_{t_T} \otimes I_n$ is the design matrix associated to the individual random effects. $\varepsilon = (\varepsilon'_1, \ldots, \varepsilon'_{t_T})'$ is the concatenation of all residuals over time and for all individuals, where $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{n,t})'$ is a n-vector of residuals associated to all individuals at time t. ε is supposed to follow a Gaussian distribution centered on zero with covariance $\sigma^2\boldsymbol{\Gamma}$ where $\boldsymbol{\Gamma} = \Gamma \otimes I_n$. While the introduction of time random effects allows to capture dependencies between observations within the same time measurement and to model dynamics of genetic effects through the dependence structure. Moreover, introducing a random individual effect combined with a specific residual correlation structure allows to take into account dependencies between measurements over time.
Finally, in a fully bayesian framework, the variance associated to the individual random effect σ^2_α is supposed to follows an inverse-gamma distribution $\mathcal{IG}(s_\alpha, r_\alpha)$, the residual variance σ^2 is supposed to follows an inverse-gamma distribution $\mathcal{IG}(s_{\sigma^2}, r_{\sigma^2})$ and the autoregressive parameter ρ is supposed to follows a uniform distribution $\mathcal{U}(-1, 1)$. The Bayesian hierarchical model is presented in web appendix C.

3. Computational aspects of the Bayesian inference

Both applications raise computational challenges mainly due to the number of parameters, dependency structures but also the number of latent variables. In the animal model,

the number of parameters is equal to 137 (the intercept, 135 standard deviations and the residual variance) and 19440 unobserved latent variables should be updated (number of elements of U). In the RIS model, the number of parameters is equal to 214 (39 fixed effects, 39 standard deviations, one individual and one residual variances, one autoregressive parameter and 132 angles) and 1176 unobserved latent variables should be updated (number of elements of U). While the animal model looks simpler (with a simple additive form) compared to the RIS model (involving complex unknown dependency structures), both complexities are high and the difference between them are not clear. MCMC algorithms have to be appropriately designed for optimization purposes. These optimizations are achieved by reparametrizing standard deviations and by proposing an efficient sampling scheme to avoid inversion of dense posterior covariance matrices.

The first difficulty relies on sampling the standard deviations λ_j according to their full conditional distributions. These distributions are proportional to a non-central multivariate folded-Gaussian distribution. Such a distribution does not have a closed form and cannot easily be sampled. To overcome this challenge, following Gelman's work, we propose to reparametrize λ_l as $sign(\xi_l)\xi_l$ where ξ_l is a parameter which can be positive or negative. It follows that ξ_l is distributed from a Gaussian distribution (and not from a folded-Gaussian): $\xi_l \sim \mathcal{N}(0, \sigma^2\tau^2\omega_l^2)$. Thus, to sample a standard deviation λ_l from its full conditional distribution $p(\lambda_l|.)$, we can more simply (i) sample ξ_l from its full conditional distribution $p(\xi_l|.)$ which is a Gaussian distribution and then (ii) compute $\lambda_l = sign(\xi_l)\xi_l$. Demonstrations are provided for both models in web appendix C.
High dimensionality causes a second issue to arise. Indeed, at each iteration of the MCMC algorithms, the random effects sampling step involves either the inversion of q $n \times n$-dense covariance matrices for the animal model (complexity $O(qn^3)$) or one $t_T(p+1) \times t_T(p+1)$-matrix for the RIS model (complexity in $O((t_T(p+1))^3)$). However, these covariance matrices have the form $\Sigma_u = (aA + bI)^{-1}$ (after a reparametrization under the RIS model), which is the inverse of the addition of a dense matrix A and a unit identity matrix (a and b are scalars, a, b and A depend on the specific context). This form is very convenient because SVD of the dense matrix A can be used to compute the Cholesky decomposition of Σ_u efficiently. Thus, to sample a random effect u from its full conditional distribution of the form $p(u|.) \sim \mathcal{N}(\Sigma_u h, \Sigma_u = (aA + bI)^{-1})$ (where h is a vector), we can (i) compute $A = WDW$, the SVD of A, where W is an orthogonal matrix of singular vectors and D is a diagonal matrix of singular values, (ii) compute $L = W(aD+bI)^{-1/2}$, the Cholesky decomposition of Σ_u, (iii) sample z from a standard Gaussian distribution and then (iv) compute $u = L(z + L'h)$. For the animal model context, dense matrices A are known IBD matrices and SVD can be computed only once at the beginning of the algorithm. Thus, the complexity of the sample scheme is in $O(qn^2)$. For the RIS model context, matrix A is unknown. However, using specific reformulations of matrices $\boldsymbol{B}$, $\boldsymbol{\Lambda}$ and $\boldsymbol{\Omega}$ as Kronecker products, matrix A can be reformulated as a Kronecker product of two matrices and SVD of A can be computed using SVDs of both matrices. Thus, the complexity of the sample scheme is $O((p+1)^3)$ if $p+1 > t_T$, $O(t_T^3)$ otherwise. Such algebraic simplifications considerably accelerate MCMC algorithms.
The third challenge, specific to the RIS context, is related to the sampling of fixed effects, β, and of random individual effects, from their full conditional Gaussian distributions. Again, algebraic simplifications based on reformulations of $\boldsymbol{X}$, D and $\boldsymbol{\Omega}$ matrices as Kronecker products allow the simplification of posterior covariance matrices and highly increase the speed of MCMC algorithms.

All these manipulations allow to deal with full conditional posterior distributions and to propose an efficient Gibbs sampler algorithm for the animal model (see web appendix C) or a faster Metropolis-within Gibbs algorithm in the RIS context. A Metropolis-Hasting step is proposed to update angle parameters associated to the correlation matrix between random intercept and random slopes (see web appendix C). All results presented in the next section, are based on 3 MCMC chains initialized at random starting values, each with 50,000 iterations, a burn-in of 10,000 iterations and a thinning of ten. All output statistics are based on the pooled 120,000 posterior samples. The Gelman and Rubin's Potential Scale Reduction Factors (PSRF) statistics (Gelman et al., 1992) is used to evaluate chains convergence. For standard deviation parameters, estimation is based on the posterior median.

4. Results

In the next subsections, we show that the fHS prior distribution is efficient to infer and select fixed effects and variance component parameters. As expected, when the number of parameters is large compared to the number of observations (first application), the fC prior does not shrink enough parameters towards zero, leading to clear over-fitting. We highlight that fHS and fSS priors perform similarly to select variance components as it has been shown in the multivariate linear context. In the second application, where the number of parameters is low compared to the number of observations, we show that the three priors perform well and no criteria, based on cross-validation procedure, allows to favour one more than the other.

4.1 *The oil palm dataset (animal model)*

Statistical results. Considering the algorithm does not converge, we adopt a fC prior for standard derivations as an alternative to the commonly used inverse-Gamma prior for variance parameters. The fC prior is not dedicated to selection but should allow for better model regularization than the inverse-Gamma. However, results show that even this prior does not shrink enough towards zero leading to a systematic bias in the estimations (see figure 2), with posterior medians varying around 0.17. The fC prior leads to over-fitting, which is particularly noticeable when analyzing the residual variance: it is estimated around zero (see figure 3) and it has a notable impact on the converge of the Gibbs

sampling algorithm by leading to a PSRF greater than 2 for a few continuous parameters. Comparatively, the fHS prior exhibits a very different behavior. It shrinks towards zero most standard deviations and let some of them be far from zero. Thus, it enables the selection of random effects and improves the MCMC convergence (PSRFs are always close to one for all continuous parameters). In this application, we propose the selection of variance components representing at least 0.05 percent of the total phenotypic variance (0.0023 or equivalently a threshold of 0.048 on standard deviations). This threshold leads to select 10 random effects (see figure 2 and table 1). The use of the fSS prior, with marginal inclusion posterior probability threshold equal to 0.1, leads to the selection of 7 standard deviations (see figure 2 and table 1). Six markers are commonly selected by fHS or fSS priors. The selection of variance components is comparable. Such similarities have already been observed in the selection of the fixed effects. Interestingly, the four markers selected using the fHS prior that are not selected using the fSS prior, have also been reported to impact phenotypic variability in different studies.

Thus, the fHS prior seems to efficiently shrink towards zero the non-relevant random effects while properly estimating relevant parameters. Moreover, it presents better computational performances than the fSS prior. Indeed, computational time for the fHS prior is twice faster than the fSS prior (40 and 80 minutes respectively for 50,000 iterations). Then, the fHS prior should clearly be promoted in a high dimensional quantitative genetic context.

Biological interpretation. We turn to biological interpretations by focusing on the results obtained by the fHS prior. Comparing with the Tisné et al. (2015) study that analyzed the same data using maximum likelihood ratio tests combined with a forward approach, all but one position identified in the former study were found. Surprisingly, the common positions were all identified at the 0.1% threshold selection, but none for the 0.05% selection. This could be due to the genetic design of the population studied derived from a breeding pedigree with unequal contributions of contrasted genetic groups: among the 144 palm trees, 73% were from La Mé (LM) genetic background, 15% from Yangambi (YBI) and 3% from their combination. Several other studies analyzed both genetic backgrounds with different genetic designs and common genetic markers. Billotte et al. (2010), with 25% LM and 25% YBI, found four common positions including two at the 0.05% threshold, Ukoskit et al. (2014), with 50% YBI, four common positions including two at the 0.05% threshold and Seng et al. (2016), three common positions including two at the 0.05% threshold. The ability of selecting positions corresponding to YBI QTL that were segregating in a minor fraction of the population indicates that the method evaluated in this study performs well even with unbalanced genetic designs and rare allele segregations. This result highlights that a multivariate approach increases the power of detection of subtle effects.

4.2 *The arabidopsis thaliana dataset (RIS model)*

In this second application, markers are labelled by their chromosome numbers and their positions (within the whole dataset of 538 markers) separated by a dash, such that marker 1-2 corresponds to the second position on the first chromosome. This notation was used by Heuclin et al. (2020) and will be used for comparison purposes. We compare our results with those of Heuclin et al. (2020), which used a non parametric functional mapping method, but also with the approach of Marchadier et al. (2019), which is based on a stepwise strategy. For the three approaches, PSRF statistics of all continuous parameters are lower than 1.1 indicating chains' convergence.

Selection of fixed effects.
Fixed effects are considered selected if zero does not belong to their credible intervals, leading to three selected markers (2-32, 2-62 and 5-104). Whatever the prior for variance components, the selection of fixed effects using HS priors performs well and provides the same results (see table 2).

Selection of variance components and impact of the correlation matrix between random effects.
Using fC or fHS priors, a threshold representing 0.1 percent of the total phenotypic variance (0.0068, or equivalently a threshold of 0.083 on standard deviations) is used. For the fSS prior, a threshold of 0.5 is considered. While the time random intercept is systematically included in the model, the number of selected random slopes varies according to priors but also to the correlation matrix prior we use between random effects. In particular, selection appears sensitive to the assumption we make on the angle shrinkage global parameter k which plays an important role on the correlation prior distribution (see equation 5). For example, when k tends to infinity (identity case) the number of selected random slopes is equal to 13, 10 and 10 for fC, fHS and fSS priors respectively. When k is fixed to one, numbers increase to 24, 18 and 11 respectively. To choose the most appropriate k value for each prior, a 10 cross-validation scheme is performed. The log pointwise predictive density (*lppd*, Gelman, Hwang, and Vehtari (2014)), related to $k = 1, 3, 5, 7, 10$ and for independence assumption, are reported on table 3. Small differences can be observed. For parsimony reasons, random effects are assumed independent. In this example, the fC prior leads to select only few more markers (13) than the fHS or fSS (10) (see figure 4 and table 2). Differences between priors are less pronounced than in the animal context where the use of the fC prior leads to an estimation of the residual variance close to zero and then to over-fitting problems. Here, residual variance is slightly lower using the fC prior (2), than using fHS or fSS priors (2.5). These differences cannot be used to evidence one prior rather than another. To decide if a prior can be promoted, we compare *lppd* between models (see table 3). But results are very close and no clear conclusion can be drawn from these results.

Selection of variance components and impact of the residual correlation.
The selection of fixed effects is not impacted by the residual correlation matrix, on the contrary this dependency structure impacts the selection of variance components. Such conclusions have already been observed in the functional mapping context (Ma et al., 2002; Li and Wu, 2010; Heuclin et al., 2020). When we compare selection of random effects taking into account an AR(1) residual correlation structure or assuming independence between residuals, the number of selected markers differs. It considerably increases with the

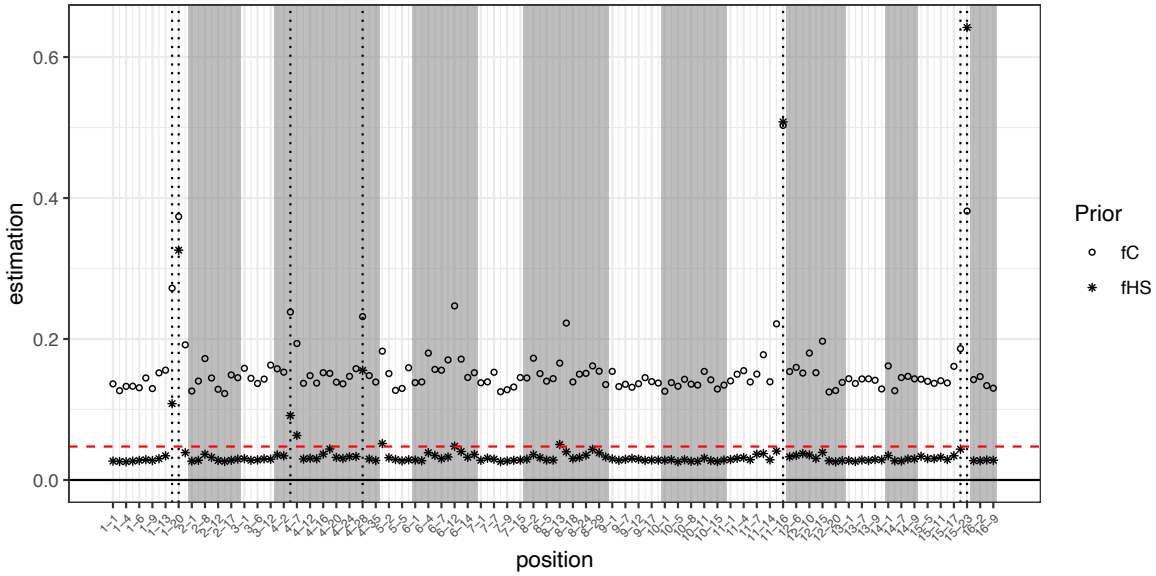

Figure 2: *Posterior median of standard deviation parameters λ_l for folded horseshoe (fHS) and folded Cauchy (fC) priors on the oil palm trees dataset. Vertical dotted lines correspond to the selected positions using the fSS prior with posterior marginal probability of inclusion upper than a threshold of 0.1. The horizontal red dashed line corresponds to a threshold of 0.048 which is the root of 0.05% of the response variance. The alternated white and grey areas delimit the 16 chromosomes.*

Table 1: *Selected standard deviation parameters λ_l for the oil palm trees dataset using folded horseshoe (fHS) and folded spike-and-slab (fSS) priors.*

Chromosome	1	4	5	6	8	11	15	nb
HS	1-17, 1-20	4-5, 4-7, 4-28	5-1	6-12	8-13	11-16	15-23	10
SS	1-17, 1-20	4-5, 4-28				11-16	15-21, 15-23	7

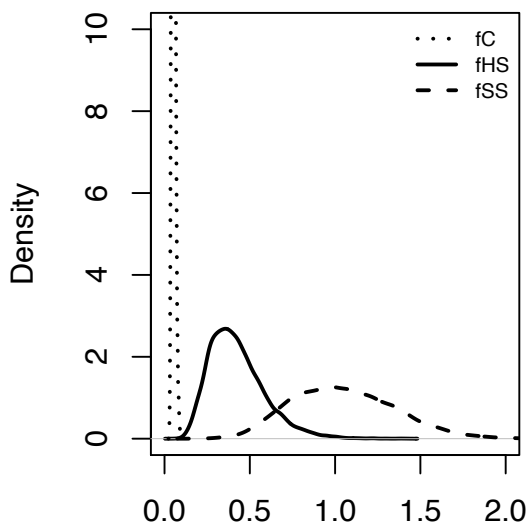

Figure 3: *Posterior density of the residual variance parameter for folded Cauchy (fC), folded horseshoe (fHS) and folded spike-and-slab (fSS) priors on the oil palm trees dataset.*

independence assumption, leading to potential over-fitting problems. Indeed, the 10 cross-validation *lppd*, considering an AR(1) residual correlation structure and the fC prior, is equal to -924, while considering an independent residual structure, it is equal to -1148. Here, results are clear and the residual correlation structure has to be included in the model.

Comparison with previous studies.
The initial study identified eight markers using the last time measurement and a forward likelihood ratio test approach (Marchadier et al., 2019). In a recent work, Heuclin et al. (2020) reanalyses this data proposing a non-functional mapping technique combined with group spike-and-slab and taking into account the full phenotypic profile over time. They identified the same eight markers but also highlighted five more effects. In our current analysis, all positions already identified by the previous approaches are selected except two on chromosome three (3-1 and 3-25), compared to Heuclin et al.'s approach. And we select two extra positions (2-47 and 5-46). Moreover, in our approach, decomposing effects as fixed and random allows to more precisely dissociate the type of effects (null, constant or varying effects). For instance, the position 5-104 selected as random effect and varying over time in Marchadier et al. or Heuclin et al. is mostly identified as fixed effect (see table 2). Finally, comparing the *lppd* statistics

Table 2: *Selection of fixed effects β and scale parameters λ on arabidopsis thaliana dataset using HS-fC, HS-fHS and HS-fSS approaches. Alternative methods proposed by Marchadier et al. (2019) and Heuclin et al. (2020) are also indicated.*

	Chromosome	**1**	**2**	**3**	**4**	**5**	**nb**
Marchadier et al. (2019)		1-20	2-62	3-3, 3-91	4-45	5-76, 5-104	8
Heuclin et al. (2020)		1-1, 1-20, 1-79, 1-97, 1-110	2-62	3-1, 3-25, 3-91	4-45	5-33, 5-76, 5-104	13
Fixed effects	HS-fC		2-32, 2-62			5-104	3
	HS-fHS		2-32, 2-62			5-104	3
	HS-hSS		2-32, 2-62			5-104	3
Scale parameters	HS-fC	1-1, 1-20, 1-79, 1-97, 1-110	2-47, 2-62	3-91	4-45	5-33, 5-46, 5-76, 5-104	13
	HS-fHS	1-1, 1-20, 1-79, 1-97, 1-110		3-91	4-45	5-33, 5-46, 5-76	10
	HS-fSS	1-1, 1-20, 1-79, 1-97, 1-110		3-91	4-45	5-33, 5-46, 5-76	10

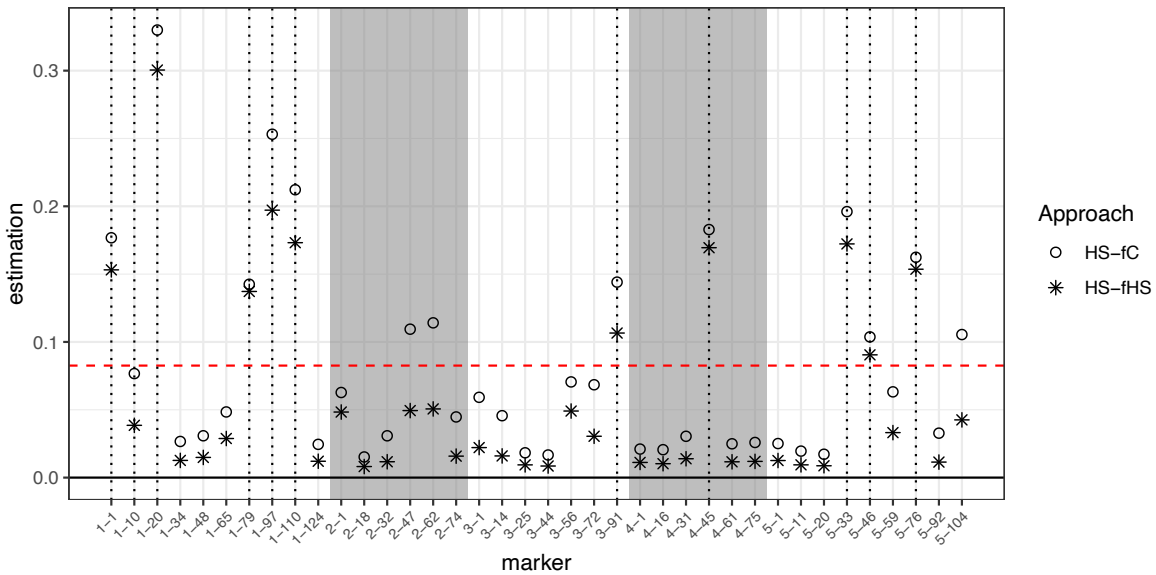

Figure 4: *Posterior median of standard deviation parameters λ_l on arabidopsis thaliana dataset. Bullets black and blue correspond to the HS-fHS and HS-fC approaches. Vertical red dotted lines correspond to the selected positions using the HS-fSS approach with posterior marginal probability of inclusion upper than a threshold of 50%. The horizontal red dashed line corresponds to a threshold of 0.083 which is the root of 0.1% of the response variance. The alternated white and gray areas delimit the 5 chromosomes.*

Table 3: *Log pointwise predictive density, considering either an unknown RIS correlation matrix with different fixed shrinkage parameters k, or an identity matrix.*

	$k=1$	$k=3$	$k=5$	$k=7$	$k=10$	Identity
HS-fC	-925	-923	-923	-923	-923	-925
HS-fHS	-923	-924	-925	-924	-925	-926
HS-fSS	-925	-925	-925	-925	-926	-927

using 10 cross-validation, favours the RIS model (-925 for the HS-fC approach with $R = I_{39}$) to the VCM model (-931).

5. Conclusion

In this paper, we show that the folded horseshoe prior should be promoted as a prior distribution for regularization in linear mixed models. Based on two real applications, we demonstrate that the folded horseshoe prior seems insensitive to high dimensional problems and leads to unbiased estimation even in low dimension. In the first example, where the number of parameters is close to the number of observations, the folded horseshoe prior shows advantages compared to the folded Cauchy and to the folded spike-and-slab priors. In particular, where the folded Cauchy prior does not allow to shrink parameters towards zero inducing a clear overfitting, the folded horseshoe prior performs well. Compared to the folded spike-and-slab prior, the folded horseshoe prior

presents similar effectiveness in terms of selection but a much greater computational efficiency. In the second application, where the number of observations is much greater than the number of parameters, no prior seems to take advantage. Such results observed in multivariate linear regression (van Erp et al., 2019) can then be extended to the linear mixed model framework. However, the folded horseshoe prior does not lead to biased estimations or under or over-fitting of models compared to the two other priors. Be that as it may, we recommend to use local-global priors.

We also propose a polar reparametrization of the model random effect correlation matrix. This approach has received little attention in the past few decades. While Pourahmadi and Wang were the first to develop a prior to generate high-dimensional random correlation matrix, Ghosh et al. were the first to infer, in a Bayesian framework, a correlation matrix in a longitudinal context. In this article, we show how this approach can be used to infer RIS correlation matrix. We also show that assuming independence or not can impact variance components selection. However, the number of parameters (angles) is equal to the number of elements of the sub-diagonal correlation matrix. Appropriate priors for the selection of angles such as considered by Ghosh et al. (2020) should be studied in combination with standard deviations shrinkage priors.

From a biological point of view, in the palm oil context, the folded horseshoe prior allows to identify positions which were segregated in a minor fraction of the population due to the unbalanced genetic design, while the frequentist stepwise selection approach considered by Tisné et al. (2015) does not. In the *Arabidobsis* context, as already noticed by Heuclin et al. (2020), we show that a longitudinal approach allows a better detection of relevant markers compared to an approach that analyzes a single time point as proposed by Marchadier et al. (2019). Both applications highlight that multivariate approaches increase the statistical power.

Acknowledgments:

F. Mortier and C. Trottier were supported by the GAMBAS project funded by the French National Research Agency (ANR-18-CE02-0025). M. Denis was fully supported by the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 840383. We thank all people from Cirad/PalmElit (France) who planned this trial. We acknowledge P.T. Socfin Indonesia (Indonesia) for planting, observing and collecting data, and authorizing use of the phenotypic data for this study.

References

Bickel, P. J., Li, B., Tsybakov, A. B., van de Geer, S. A., Yu, B., Valdés, T., Rivero, C., Fan, J., and van der Vaart, A. (2006). Regularization in statistics. *Test* **15,** 271–344.

Billotte, N., Jourjon, M.-F., Marseillac, N., Berger, A., Flori, A., Asmady, H., Adon, B., Singh, R., Nouy, B., Potier, F., et al. (2010). Qtl detection by multi-parent linkage mapping in oil palm (elaeis guineensis jacq.). *Theoretical and Applied Genetics* **120,** 1673–1687.

Bitto, A. and Frühwirth-Schnatter, S. (2019). Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics* **210,** 75–97.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H., and White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution* **24,** 127–135.

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,** 1069–1077.

Brown, H. and Prescott, R. (2014). *Applied Mixed Models in Medicine, Third Edition.* John Wiley Sons, Ltd.

Cadonna, A., Frühwirth-Schnatter, S., and Knaus, P. (2020). Triple the gamma—a unifying shrinkage prior for variance and variable selection in sparse state space and tvp models. *Econometrics* **8,**.

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2009). Handling sparsity via the horseshoe. In *Artificial Intelligence and Statistics*, pages 73–80.

Chen, Z. and Dunson, D. B. (2003). Random effects selection in linear mixed models. *Biometrics* **59,** 1069–1077.

Delattre, M. and Poursat, M.-A. (2020). An iterative algorithm for joint covariate and random effect selection in mixed effects models. *The International Journal of Biostatistics* **1,**.

Desboulets, L. D. D. (2018). A review on variable selection in regression analysis. *Econometrics* **6,** 45.

Fahrmeir, L., Kneib, T., and Lang, S. (2004). Penalized structured additive regression for space-time data: a bayesian perspective. *Statistica Sinica* **14,** 731–761.

Fan, J. and Li, R. (2012). Variable selection in linear mixed effects models. *Annals of Statistics* **40,** 2043–2068.

Fridley, B. L. and Biernacka, J. M. (2011). Gene set analysis of snp data: benefits, challenges, and future directions. *European Journal of Human Genetics* **19,** 837–843.

Frühwirth-Schnatter, S. and Tüchler, R. (2008). Bayesian parsimonious covariance estimation for hierarchical linear mixed models. *Statistics and Computing* **18,** 1–13.

Frühwirth-Schnatter, S. and Wagner, H. (2011). Bayesian variable selection for random intercept modeling of gaussian and non-gaussian data.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian Analysis* **1,** 515–534.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis.* Chapman and Hall/CRC.

Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for bayesian models. *Statistics and computing* **24,** 997–1016.

Gelman, A., Rubin, D. B., et al. (1992). Inference from iterative simulation using multiple sequences. *Statistical science* **7,** 457–472.

George, A. W., Visscher, P. M., and Haley, C. S. (2000). Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach. *Genetics* **156,** 2081–2092.

George, E. I. and McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical*

Association **88,** 881–889.

Ghosh, R. P., Mallick, B., and Pourahmadi, M. (2020). Bayesian estimation of correlation matrices of longitudinal data. *Bayesian Analysis* .

Griffin, J. E., Brown, P. J., et al. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian analysis* **5,** 171–188.

Heuclin, B., Mortier, F., Trottier, C., and Denis, M. (2020). Bayesian varying coefficient model with selection: An application to functional mapping. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* .

Ibrahim, J. G., Zhu, H. T., Garcia, R. I., and Guo, R. (2011). Fixed and random effects selection in mixed effects models. *Biometrics* **67,** 495–503.

Ives, A. R. and Helmus, M. H. (2011). Generalized linear mixed models for phylogenetic analyses of community structure. *Ecological Monograph* **81,** 511–525.

Kyung, M., Gill, J., Ghosh, M., Casella, G., et al. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian Analysis* **5,** 369–411.

Lewandowski, D., Dorota Kurowick, D., and Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis* **100,** 1989–2001.

Li, Y., Wang, S., Song, P. X., Wang, N., Zhou, L., and Zhu, J. (2018). Doubly regularized estimation and selection in linear mixed-effects models for high-dimensional longitudinal data. *Stat Interface.* **11,** 721–737.

Li, Y. and Wu, R. (2010). Functional mapping of growth and development. *Biological Reviews* **85,** 207–216.

Loudet, O. (2018). Raw phenotypic data obtained on the arabidopsis rils with the phenoscope robots (marchadier, hanemian, tisnÃ© et al., 2019).

Lynch, M. and Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits.* Sinauer.

Ma, C. X., Casella, G., and Wu, R. (2002). Functional Mapping of Quantitative Trait Loci Underlying the Character Process: A Theoretical Framework. *Genetics* page 12.

Marchadier, E., Hanemian, M., Tisne, S., Bach, L., Bazakos, C., Gilbault, E., Haddadi, P., Virlouvet, L., and Loudet, O. (2019). The complex genetic architecture of shoot growth natural variation in Arabidopsis thaliana. *Plos Genetics* **15,**.

Müller, S., Scealy, J. L., and Welsh, A. H. (2013). Model selection in linear mixed models. *Statistical Science* **28,** 135–167.

Park, T. and Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association* **103,** 681–686.

Pérez, M.-E., Pericchi, L. R., and Ramírez, I. C. (2017). The scaled beta2 distribution as a robust prior for scales. *Bayesian Analysis* **12,** 615–637.

Piironen, J. and Vehtari, A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics* **11,** 5018–5051.

Pinheiro, J. C. and Bates, D. M. (1996). Unconstrained parametrizations for variance-covariance matrices. *Statistics and computing* **6,** 289–296.

Polson, N. G. and Scott, J. G. (2012). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Bayesian Analysis* **7,** 887–902.

Pourahmadi, M. and Wang, X. (2015). Distribution of random correlation matrices: Hyperspherical parameterization of the cholesky factor. *Statistics & Probability Letters* **106,** 5–12.

Rao, R. and Wu, Y. (1989). A strongly consistent procedure for model selection in a regression problem. *Biometrika* **76,** 369–374.

Scheipl, F., Fahrmeir, L., and Kneib, T. (2012). Spike-and-slab priors for function selection in structured additive regression models. *Journal of the American Statistical Association* **107,** 1518–1532.

Seng, T.-Y., Ritter, E., Saad, S. H. M., Leao, L.-J., Singh, R. S. H., Zaman, F. Q., Tan, S.-G., Alwee, S. S. R. S., and Rao, V. (2016). Qtls for oil yield components in an elite oil palm (elaeis guineensis) cross. *Euphytica* **212,** 399–425.

Team, R. C. et al. (2013). R: A language and environment for statistical computing.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* **58,** 267–288.

Tisné, S., Denis, M., Cros, D., Pomiès, V., Riou, V., Syahputra, I., Omoré, A., Durand-Gasselin, T., Bouvet, J.-M., and Cochard, B. (2015). Mixed model approach for ibd-based qtl mapping in a complex oil palm pedigree. *BMC genomics* **16,** 1–12.

Ukoskit, K., Chanroj, V., Bhusudsawang, G., Pipatchartlearnwong, K., Tangphatsornruang, S., and Tragoonrung, S. (2014). Oil palm (elaeis guineensis jacq.) linkage map, and quantitative trait locus analysis for sex ratio and related traits. *Molecular breeding* **33,** 415–424.

Vaida, F. and Blanchard, S. (2005). Conditional akaike information for mixed-effects models. *Biometrika* **92,** 351–370.

van Eeuwijk, F. A., Boer, M., Totir, L. R., Bink, M., Wright, D., Winkler, C. R., Podlich, D., Boldman, K., Baumgarten, A., Smalley, M., et al. (2010). Mixed model approaches for the identification of qtls within a maize hybrid breeding program. *Theoretical and Applied Genetics* **120,** 429–440.

van Erp, S., Oberski, D. L., and Mulder, J. (2019). Shrinkage priors for bayesian penalized regression. *Journal of Mathematical Psychology* **89,** 31–50.

Supporting Information

Web appendix A, B and C, referenced in Section 1, 2 and 3, are available with this paper at the Biometrics website on Wiley Online Library. Algorithms for animal and RIS models are available in the R language (Team et al., 2013) on GitHub `https://github.com/Heuclin/variance_component_selection`. The *oil palm* dataset is available on request. For the *arabidopsis thaliana* dataset, the complete phenotypic dataset is freely available at: `https://data.inra.fr/dataset.xhtml?persistentId=doi:10.15454/OCOP9B` (Loudet, 2018). The genotypic dataset is freely available at: `http://publiclines.versailles.inra.fr/page/8`.