Bayesian group fused priors for deciphering environmental regulation of abscission

Benjamin Heuclin, Frederic Mortier, Sébastien Tisné, Julien Gibaud, Catherine Trottier, Marie Denis

Bayesian group fused priors for deciphering environmental regulation of abscission

B. Heuclin[1,2], F. Mortier[4,5], S. Tisné[2,3], J. Gibaud[1], C. Trottier[1,6], and M. Denis[2,3]

[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] Univ Paul-Valéry Montpellier 3, Montpellier, France.

Corresponding author's e-mail address: marie.denis@cirad.fr

Abstract

The process of abscission - the shedding of parts of an organism - is closely linked to a series of physiological events whose optimal execution is key to the survival of the species. Environmental variations impact species development, and in particular abscission processes at multiple developmental stages. Identifying environmental factors and the times at which they modulate the abscission process is crucial, particularly in the context of climate change. Considering environmental variables as time series, i.e. groups of variables correlated over time, poses statistical problems for selecting the relevant groups, the environmental variables, and the variables correlated within them, the time periods. In this paper, we address these objectives by introducing and discussing a set of Bayesian fused and fusion priors via a general parameterization. This paper highlights a trade-off between the priors used on differences and coefficients. In particular, we show the effectiveness of horseshoe-type priors on differences as well as on coefficients with appropriate parameterizations in terms of selection, estimation and algorithmic stability whatever the number of groups and their size. The study was motivated by fruit abscission in oil palms, which impacts the bunch harvesting timing. Disturbance of abscission can therefore affect oil yield and quality, and consequently have an impact on economic income. This application, which is based on an experimental design in Benin Republic, illustrates the performance of the proposed prior in selecting both environmental variables and the developmental stages involved in the timing of bunch harvesting.

Keywords: Bayesian variable selection, Fusion and Fused priors, Horseshoe prior, Struc-

tured variables.

1 Introduction

Understanding the impact of environmental variables on development and adaptation processes is crucial when addressing climate change concerns. Abscission, which consists of the shedding of various parts of organisms, for example leaves in the fall or flowers after fertilization, is one of the most important adaptation process. This biological mechanism is highly sensitive to climate conditions and to climate variations both over the growing seasons and between years. A good example of the abscission process is leaf senescence and fall in deciduous trees, which was shown to be delayed in the northern hemisphere in response to increasing temperatures between 1931 and 2010 (Gill et al., 2015). Environmental stress can severely impact abscission processes due to complex regulations involving exogenous and endogenous signals (Sawicki et al., 2015). For instance, drought stress has been shown to induce activation and premature flower abscission in lupine (Wilmowicz et al., 2021) and in tomato plants (Reichardt et al., 2020) with a negative impact on crop productivity.

While in many contexts, it is clear that environmental variables can cause organ losses, it is not yet clear which one, among exogenous variables (e.g. climate, soil) or endogenous variables (e.g. development stage, carbon status) are responsible for the responses observed and at which stages of the organ development or of the abscission process the regulation occurs. The exact time at which abscission occurs is critical for the oil palm, because fruit bunches are harvested when the first fruits detach and fall to the ground. Thus, premature abscission of fruits will reduces the yield of oil palm if optimal maturity is not reached, while excessive abscission requires additional work collecting detached fruits on the ground. A recent study showed that environmental variables, such as temperature or solar radiation, affect the reproductive development of the oil palm tree by modulating the timing of fruit drop (Tisné et al., 2020).

In this paper, we identify the relevant environmental variables experienced by the fruit bunch and the periods that have an effect on the phenotypic variations of fruit abscission in the oil palm. Considering environmental variables as time series, i.e. groups of temporally correlated variables, poses at least two statistical challenges related to model regularization and variable selection. The first challenge is to select the groups, i.e. the environmental variables. The second is to select the correlated variables within groups, i.e. the time periods. Such time structure within groups leads to high correlation between consecutive measures of a given environmental variable. These dependencies have to be taken into account to avoid ill-conditioned problems and over-fitting, but also to better reflect reality and detect successive meaningful time periods. In the following, groups refer to an environmental variable measured over time. However, alternative structures could be also considered, such as those induced by biological pathways.

In recent decades, considerable attention has been paid to the selection of variables and groups. The methods developed up to now are mainly related to penalized likelihood techniques in a frequentist context, or to the use of appropriate priors reflecting desired penalties in a Bayesian context. Among others, the Least Absolute Shrinkage and Selection Operator (Lasso, Tibshirani, 1996), the Smoothly Clipped Absolute Deviation (Fan and Li, 2001) penalty or the Elastic-Net (Zou and Hastie, 2005) are most widely used. Note that Elastic-Net is appropriate when the variables are correlated. This approach is based on the combination of ℓ_1- and ℓ_2-norms on the penalization term, combining shrinkage properties from Lasso and regularization capacities from Ridge regression (Hoerl and Kennard, 1970). In a Bayesian multiple linear regression context, the set of priors for variable selection has also been extensively developed. Priors can be categorized in two classes: spike-and-slab priors (George and McCulloch, 1993) and continuous shrinkage priors (Polson and Scott, 2011). The latter class was originally developed to obtain the Bayesian version of penalized likelihood methods. Among those we are familiar with, we cite the Bayesian Lasso prior (Park and Casella, 2008), the Elastic-Net prior (Kyung et al., 2010), the normal-Gamma prior (Griffin et al., 2010) and the horseshoe (HS) prior (Carvalho et al., 2010; Piironen et al., 2017). However, these methods were not designed to account for potential group structures within covariates. To address this limitation, Lasso extensions to group selection were developed in frequentist (Yuan and Lin, 2006) or Bayesian (Kyung et al., 2010; Liquet et al., 2017) contexts. In order to select groups as well as variables within groups, Xu et al. (2015) proposed the sparse group Lasso prior. This approach mimics the frequentist sparse group Lasso penalty introduced by Simon et al. (2013). Xu et al. (2016) extended this prior by considering a horseshoe prior and a scale mixture of independent Gaussian distributions with three levels of variance parameters: one global and common to all coefficients, one specific to each group, and one for each coefficient.

However, the above methods do not allow serial correlations between successive time occurrences within groups to be taken into account. These dependencies may lead to identifiability problems that affect the estimation task which aims to assign similar effects to two adjacent variables. In a linear regression context, to allow the integration of this information and to constrain estimation, Land and Friedman (1997) and Tibshirani et al. (2005) introduced the fusion and fused Lasso. The fusion Lasso penalizes the ℓ_1-norm of successive differences in parameters, and the fused Lasso combines the fusion Lasso with the usual Lasso penalization on each coefficient. Kyung et al. (2010) proposed a Bayesian fused Lasso with a Bayesian Lasso prior on differences and also on each coefficient. However, several authors pointed out that the Bayesian Lasso, which uses Laplace distribution, does not sufficiently shrink (Carvalho et al., 2010; Polson and Scott, 2011) leading to biased and smooth estimations without possible abrupt changes (Faulkner and Minin, 2018). To allow more flexibility and sparser estimations, other continuous shrinkage priors, with stronger mass on zero and heavier tails, have been investigated on differences. For instance, Rue and Held (2005) and Song and Cheng (2020) used a Student distribution on the differences,

Shimamura et al. (2019) considered normal-Exponential-Gamma (NEG) distribution, while Kakikawa et al. (2023) placed an HS prior on differences. Note that all Bayesian fused priors place a Laplace distribution on regression coefficients. Some alternative priors, such as the HS fusion priors, assume only distributions on differences (Faulkner and Minin, 2018). These methodologies show prediction accuracy but are also able to estimate smooth functions with potentially abrupt changes. However, they were only designed for one-group. A direct extension of Bayesian fused Lasso to multi-group context was proposed by Alaíz et al. (2013). Zhang et al. (2014) used this multi-group version in a group spike-and-slab prior. Although promising, these methods may suffer from the low shrinkage properties of the Bayesian Lasso, leading to poor estimations when the number of groups is large and their size is small.

In this paper, we use a thorough simulation study to investigate the trade-off between shrinkage priors on coefficients and on their differences. Our results provide evidence in favor of using distributions with heavier tails than the usual Laplace distribution. In particular, we promote the use of horseshoe priors with random local parameters for both components but with a fixed global parameter for coefficients while remaining random for differences. Consequently, two extensions to the multi-group context are developed, assuming that the global shrinkage parameter is either specific to each group, or common to all groups. In this work, we also extend to the multi-group case, the fused priors proposed by Kakikawa et al. (2023) or the fusion one developed by Faulkner and Minin (2018) in the one-group context. Prior performances are compared through intensive simulations using different settings: the number of groups, their size and the residual variance.

The rest of the paper is organized as follows. Section 2 is dedicated to the construction of Bayesian fused priors in a linear regression framework, considering both the one-group and multi-group cases. Using simulated data, section 3 compares and evaluates the efficiency of the proposed priors according to the different settings. Section 4 identifies the environmental variables and time periods that affect the fruit abscission process in oil palm.

2 Model

2.1 Notation and model

Let $\mathbf{y} = (y_1, \ldots, y_n)'$ be a n-continuous response vector and $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_G]$ an $(n \times GT)$-matrix concatenating G matrices. For all $g = 1, \ldots, G$, $\mathbf{X}_g = \left[\mathbf{x}'_{g1}, \ldots, \mathbf{x}'_{gT}\right]$ is an $(n \times T)$-matrix describing a variable measured at T regularly spaced times with $\mathbf{x}_{gt} = (x_{1gt}, \ldots, x_{ngt})$ for $t = 1, \ldots, T$

In this paper, $\mathbf{y}$ corresponds to the abscission time measured on $n = 1,173$ bunches collected from $l = 140$ oil palm trees. Each $\mathbf{X}_g$ describes an environmental variable. As many bunches originate from the same palm tree, we used a linear mixed model such that:

$$\mathbf{y} = \boldsymbol{\mu}\mathbb{1} + \sum_{\boldsymbol{g=1}}^{\boldsymbol{G}} \mathbf{X}_{\boldsymbol{g}}\boldsymbol{\beta}_{\boldsymbol{g}} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \tag{1}$$

where μ denotes the intercept, $\mathbb{1}$ a n-vector of 1, $\boldsymbol{\beta}_g = (\beta_{g1}, \ldots, \beta_{gt}, \ldots, \beta_{gT})'$ a T-vector of regression coefficients associated with environmental variable g (i.e. group g), $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_l)'$ a l-vector of random effects assumed Gaussian distributed with zero expectation and variance equal to σ_α^2, and $\mathbf{Z}$ the known associated design matrix of dimension $(n \times l)$. This random effect enables to account for the dependence between observations made on the same oil palm tree. Finally, $\boldsymbol{\varepsilon}$ is a n-vector of residuals assumed to follow a Gaussian distribution $\mathcal{N}_n(0, \sigma^2 \boldsymbol{I}_n)$ independent from $\boldsymbol{\alpha}$.

2.2 Prior construction

In a Bayesian framework, the usual approach accounting for structure within features while imposing sparsity on coefficients, relies on the use of fused-type priors. Such priors consist in placing continuous shrinkage priors on the regression coefficients and their successive differences. The most commonly used is the Bayesian fused Lasso (Kyung et al., 2010) that assumes Laplace distribution, also called normal-Exponential (NE) distribution, on both components. In the one group case ($\boldsymbol{\beta} \equiv \boldsymbol{\beta}_1$), the associated conditional prior distribution for $\boldsymbol{\beta}$ is of the following form:

$$\pi(\boldsymbol{\beta}|\sigma^2, \lambda, \upsilon) \propto \prod_{t=1}^{T} \text{Laplace}(\beta_t|0, \upsilon, \sigma^2) \prod_{t=2}^{T} \text{Laplace}(\beta_t - \beta_{t-1}|0, \lambda, \sigma^2), \tag{2}$$

where υ and λ are positive hyperparameters. However, the Laplace prior suffers from posterior inconsistency notably due to its exponentially light tails (Castillo et al., 2015). To overcome these drawbacks, several authors have proposed using continuous shrinkage priors on differences with heavy tails such as Student (Song and Cheng, 2020), normal-Exponential-Gamma (NEG, Shimamura et al. (2019)) or horseshoe (HS, Kakikawa et al. (2023)) distributions combined with a Laplace distribution on coefficients. All continuous shrinkage priors can be reformulated as a scale mixture of Gaussian distributions (Andrews and Mallows, 1974), meaning that for fused-type prior, conditioned on the scale parameters, the regression coefficients and their differences follow normal distributions. A general formulation of the conditional prior distribution of $\boldsymbol{\beta}$ is then given by:

$$\pi(\boldsymbol{\beta}|\sigma^2, \upsilon^2, \boldsymbol{\gamma}^2, \lambda^2, \boldsymbol{\omega}^2) \propto \prod_{t=1}^{T} \mathcal{N}(\beta_t|0, \upsilon^2\gamma_t^2\sigma^2) \prod_{t=2}^{T} \mathcal{N}(\beta_t - \beta_{t-1}|0, \lambda^2\omega_t^2\sigma^2). \tag{3}$$

This formulation refers to a global-local parametrization with local shrinkage parameters γ_t and ω_t and global shrinkage parameters, υ and λ, respectively associated with coefficients and their differences. The global parameters shrink all coefficients towards zero while local parameters allow true non-zero effects to escape from overall shrinkage. The usual continuous shrinkage distributions can then be recovered by placing specific prior distributions on the shrinkage parameters (see Table 1).

The fused-type priors reported in the literature (see Table 1) focuses on the distribution assumption for differences but never on the prior distributions for coefficients. The Bayesian

Prior names	**Difference prior**	**Coefficient prior**	**References**
$\text{NE}_{\lambda}^{\omega_t}$ – $\text{NE}_{\upsilon}^{\gamma_t}$	$\lambda^2 \sim \mathcal{IG}(a,b)$ $\omega_t^2 \sim \mathcal{E}xp(1/2)$	$\upsilon^2 \sim \mathcal{IG}(s,r)$ $\gamma_t^2 \sim \mathcal{E}xp(1/2)$	(Kyung et al., 2010)
$\text{NEG}_{1}^{\omega_t}$ – $\text{NE}_{\upsilon}^{\gamma_t}$	$\lambda = 1$ $\omega_t^2 \vert \psi_t \sim \mathcal{E}xp(\psi_t)$ $\psi_t \sim \mathcal{G}(a,b)$	$\upsilon^2 \sim \mathcal{IG}(s,r)$ $\gamma_t^2 \sim \mathcal{E}xp(1/2)$	(Shimamura et al., 2019)
$\text{HS}_{\lambda}^{\omega_t}$ – $\text{NE}_{\upsilon}^{\gamma_t}$	$\lambda \sim \mathcal{C}^+(0,1)$ $\omega_t \sim \mathcal{C}^+(0,1)$	$\upsilon^2 \sim \mathcal{IG}(s,r)$ $\gamma_t^2 \sim \mathcal{E}xp(1/2)$	(Kakikawa et al., 2023)
$\text{HS}_{\lambda}^{\omega_t}$ – $\text{HS}_{\upsilon}^{\gamma_t}$	$\lambda \sim \mathcal{C}^+(0,1)$ $\omega_t \sim \mathcal{C}^+(0,1)$	$\upsilon \sim \mathcal{C}^+(0,1)$ $\gamma_t \sim \mathcal{C}^+(0,1)$	
$\text{HS}_{\lambda}^{\omega_t}$ – $\text{HS}_{1}^{\gamma_t}$	$\lambda \sim \mathcal{C}^+(0,1)$ $\omega_t \sim \mathcal{C}^+(0,1)$	$\upsilon = 1$ $\gamma_t \sim \mathcal{C}^+(0,1)$	

Table 1: Fused priors in the one-group context ($G = 1$). $\mathcal{E}xp$, $\mathcal{C}^+$, and $\mathcal{IG}(a,b)$ denote the exponential, the half-Cauchy, and the inverse-Gamma distributions. a, b, s and r are additional hyperparameters either to be set or inferred.

Laplace is systematically advocated. However, as mentioned above, Laplace distribution involves posterior inconsistency (Castillo et al., 2015). In this paper, we propose priors with HS distribution on differences, as proposed by (Kakikawa et al., 2023) but also on coefficients, which remain steady when dimension or complexity increase. We explored two parametrizations for the global shrinkage parameter υ either by half-Cauchy $\mathcal{C}^+(0,1)$, or fixing it (see the last two lines in Table 1). According to the results of the simulation, the prior that assumes a global shrinkage parameter set to 1 appeared to be more consistent and was consequently chosen. Given that the density of the half-Cauchy distribution $\mathcal{C}^+(0,1)$ defined on $\mathbb{R}^+$ is equal to $p(x) = \frac{2}{\pi(1+x^2)}$, this prior, denoted by $\text{HS}_{\lambda}^{\omega_t}$ – $\text{HS}_{1}^{\gamma_t}$, is defined as follows:

$$p(\boldsymbol{\beta}|\sigma^2) \propto \prod_{t=1}^{T} \int \frac{1}{\sqrt{2\pi\gamma_t^2\sigma^2}} \exp\left(-\frac{\beta_t^2}{2\gamma_t^2\sigma^2}\right) \frac{2}{\pi(1+\gamma_t^2)} d\gamma_t^2$$
$$\times \int \left[\prod_{t=2}^{T} \int \frac{1}{\sqrt{2\pi\lambda^2\omega_t^2\sigma^2}} \exp\left(-\frac{(\beta_t - \beta_{t-1})^2}{2\lambda^2\omega_t^2\sigma^2}\right) \frac{2}{\pi(1+\omega_t^2)} d\omega_t^2 \right] \frac{2}{\pi(1+\lambda^2)} d\lambda^2. \tag{4}$$

Equation (4) can be reformulated in a multivariate form as follows:

$$\begin{aligned} p(\boldsymbol{\beta}|\sigma^2) \;\propto\; & \int \cdots \int (\sigma^2)^{-\frac{2T-1}{2}} (\lambda^2)^{-\frac{T-1}{2}} \\ & \times \prod_{t=1}^{T} (\gamma_t^2)^{-\frac{1}{2}} \prod_{t=2}^{T} (\omega_t^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2} \boldsymbol{\beta}' \left(\boldsymbol{\Upsilon}^{-1} + \boldsymbol{D}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{D}/\lambda^2 \right) \boldsymbol{\beta} \right) \\ & \times p(\lambda) \prod_{t=1}^{T} p(\gamma_t) \prod_{t=2}^{T} p(\omega_t) d\lambda \prod_{t=1}^{T} d\gamma_t \prod_{t=2}^{T} d\omega_t. \end{aligned} \tag{5}$$

where $\boldsymbol{D}$ is the known $T \times (T-1)$-matrix associated with the finite difference operator of order 1, and $\boldsymbol{\Upsilon} = diag(\gamma_1^2, \ldots, \gamma_T^2)$ and $\boldsymbol{\Omega} = diag(\omega_1^2, \ldots, \omega_{T-1}^2)$ the $(T-1) \times (T-1)$-diagonal matrices of local parameters. $p(.)$ is the density of the half-Cauchy distribution $\mathcal{C}^+(0,1)$. Now it is straightforward to express the distribution of coefficients $\boldsymbol{\beta}$ conditionally to the shrinkage parameters:

$$\boldsymbol{\beta}|\sigma^2, \lambda, \boldsymbol{\gamma}, \boldsymbol{\omega} \;\sim\; \mathcal{N}_T\left(0, \sigma^2\left(\boldsymbol{\Upsilon}^{-1} + \boldsymbol{D}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{D}/\lambda^2\right)^{-1}\right) \tag{6}$$

This formulation is appropriate for MCMC implementation.

A direct and natural extension of the proposed prior (see Eq. 4) to the multi-group context consists in assuming that the shrinkage parameters that control sparsity are group specific. Formally, the density function can be expressed as follows:

$$p(\boldsymbol{\beta}|\sigma^2) \propto \prod_{g=1}^{G}\prod_{t=1}^{T} \int \frac{1}{\sqrt{2\pi\gamma_{gt}^2\sigma^2}} \exp\left(-\frac{\beta_{gt}^2}{2\gamma_{gt}^2\sigma^2}\right) \frac{2}{\pi(1+\gamma_{gt}^2)} d\gamma_{gt}^2$$
$$\times \int \left[\prod_{t=2}^{T} \int \frac{1}{\sqrt{2\pi\lambda_g^2\omega_{gt}^2\sigma^2}} \exp\left(-\frac{(\beta_{gt}-\beta_{gt-1})^2}{2\lambda_g^2\omega_{gt}^2\sigma^2}\right) \frac{2}{\pi(1+\omega_{gt}^2)} d\omega_t^2\right] \frac{2}{\pi(1+\lambda_g^2)} d\lambda_g^2 \tag{7}$$

This prior is hereafter denoted $\text{HS}_{\lambda_g}^{\omega_{gt}}$ – $\text{HS}_1^{\gamma_{gt}}$. However, this prior (see Eq. 7) relies on a large set of global shrinkage parameters ($\lambda_g,\ g = 1, \ldots, G$). The inference for such parameters is known to be complex and can lead to poor results mostly in terms of selection (Piironen et al., 2017). In the multi-group context, the number of groups as well as their size will reinforce such difficulties. We therefore suggest an alternative parametrization assuming a single global parameter ($\lambda_g = \lambda$) to control shrinkage over all groups while keeping local parameters ω_{gt} specific to groups. In the following, this prior is denoted by $\text{HS}_{\lambda}^{\omega_{gt}}$ – $\text{HS}_1^{\gamma_{gt}}$ and its density function is equal to:

$$p(\boldsymbol{\beta}|\sigma^2) \propto \prod_{g=1}^{G}\prod_{t=1}^{T} \int \frac{1}{\sqrt{2\pi\gamma_{gt}^2\sigma^2}} \exp\left(-\frac{\beta_{gt}^2}{2\gamma_{gt}^2\sigma^2}\right) \frac{2}{\pi(1+\gamma_{gt}^2)} d\gamma_{gt}^2$$
$$\times \int \left[\prod_{t=2}^{T} \int \frac{1}{\sqrt{2\pi\lambda^2\omega_{gt}^2\sigma^2}} \exp\left(-\frac{(\beta_{gt}-\beta_{gt-1})^2}{2\lambda^2\omega_{gt}^2\sigma^2}\right) \frac{2}{\pi(1+\omega_{gt}^2)} d\omega_t^2\right] \frac{2}{\pi(1+\lambda^2)} d\lambda^2 \tag{8}$$

A similar strategy can be applied to extend Kyung et al. (2010)'s and Kakikawa et al. (2023)'s priors to the multi-group context. In the following, the fused priors proposed by Kyung et al. (2010) are denoted by: $\text{NE}_{\lambda_g}^{\omega_{gt}}$ – $\text{NE}_{\upsilon_g}^{\gamma_{gt}}$, and $\text{NE}_{\lambda}^{\omega_{gt}}$ – $\text{NE}_{\upsilon_g}^{\gamma_{gt}}$ and Kakikawa et al. (2023)'s fused priors are denoted by $\text{HS}_{\lambda_g}^{\omega_{gt}}$ – $\text{NE}_{\upsilon_g}^{\gamma_{gt}}$, and $\text{HS}_{\lambda}^{\omega_{gt}}$ – $\text{NE}_{\upsilon_g}^{\gamma_{gt}}$. Priors for $\boldsymbol{\beta}$ coefficients as well as the associated shrinkage hyper-parameter distributions are summarized in Table 2.

Prior names	Difference prior	Coefficient prior
$\text{HS}_{\lambda_g}^{\omega_{gt}}$ – $\text{HS}_1^{\gamma_{gt}}$	$\lambda_g \sim \mathcal{C}^+(0,1)$ $\omega_{gt} \sim \mathcal{C}^+(0,1)$	$\upsilon_g = \upsilon = 1$ $\gamma_{gt} \sim \mathcal{C}^+(0,1)$
$\text{HS}_{\lambda}^{\omega_{gt}}$ – $\text{HS}_1^{\gamma_{gt}}$	$\lambda_g = \lambda$ $\lambda \sim \mathcal{C}^+(0,1)$ $\omega_{gt} \sim \mathcal{C}^+(0,1)$	$\upsilon_g = \upsilon = 1$ $\gamma_{gt} \sim \mathcal{C}^+(0,1)$
$\text{NE}_{\lambda_g}^{\omega_{gt}}$ – $\text{NE}_{\upsilon_g}^{\gamma_{gt}}$	$\lambda_g^2 \sim \mathcal{IG}(a,b)$ $\omega_{gt}^2 \sim \mathcal{E}xp(1/2)$	$\upsilon_g^2 \sim \mathcal{IG}(s,r)$ $\gamma_{gt}^2 \sim \mathcal{E}xp(1/2)$
$\text{NE}_{\lambda}^{\omega_{gt}}$ – $\text{NE}_{\upsilon_g}^{\gamma_{gt}}$	$\lambda_g^2 = \lambda^2$ $\lambda^2 \sim \mathcal{IG}(a,b)$ $\omega_{gt}^2 \sim \mathcal{E}xp(1/2)$	$\upsilon_g^2 \sim \mathcal{IG}(s,r)$ $\gamma_{gt}^2 \sim \mathcal{E}xp(1/2)$
$\text{HS}_{\lambda_g}^{\omega_{gt}}$ – $\text{NE}_{\upsilon_g}^{\gamma_{gt}}$	$\lambda_g \sim \mathcal{C}^+(0,1)$ $\omega_{gt} \sim \mathcal{C}^+(0,1)$	$\upsilon_g^2 \sim \mathcal{IG}(s,r)$ $\gamma_{gt}^2 \sim \mathcal{E}xp(1/2)$
$\text{HS}_{\lambda}^{\omega_{gt}}$ – $\text{NE}_{\upsilon_g}^{\gamma_{gt}}$	$\lambda_g = \lambda$ $\lambda \sim \mathcal{C}^+(0,1)$ $\omega_{gt} \sim \mathcal{C}^+(0,1)$	$\upsilon_g^2 \sim \mathcal{IG}(s,r)$ $\gamma_{gt}^2 \sim \mathcal{E}xp(1/2)$
$\text{HS}_{\lambda}^{\omega_{gt}}$	$\lambda_g = \lambda$ $\lambda \sim \mathcal{C}^+(0,1)$ $\omega_{gt} \sim \mathcal{C}^+(0,1)$	

Table 2: Fused priors in the multi-group context. $\mathcal{E}xp$, $\mathcal{C}^+$, and $\mathcal{IG}(a,b)$ denote the exponential, the half-Cauchy, and the inverse-Gamma distributions. a, b, s and r are additional hyperparameters either to be set or inferred.

Conditionally on shrinkage parameters, $\boldsymbol{\beta}_g$ are distributed according to a multivariate Gaussian distribution:

$$\boldsymbol{\beta}_g|\boldsymbol{\Upsilon}_g, \lambda^2, \boldsymbol{\Omega}_g, \sigma^2 \sim \mathcal{N}_T\left(0, \sigma^2\left(\boldsymbol{\Upsilon}_g^{-1} + \frac{1}{\lambda^2}\boldsymbol{D}_g^\top \boldsymbol{\Omega}_g^{-1}\boldsymbol{D}_g\right)^{-1}\right) \tag{9}$$

where $\boldsymbol{\Upsilon}_g = \text{diag}(\gamma_{g1}^2, \ldots, \gamma_{gT}^2)$, $\boldsymbol{\Omega}_g = \text{diag}(\omega_{g1}^2, \ldots, \omega_{gT-1}^2)$ and $\boldsymbol{D}_g$ is the known $T \times (T-1)$-matrix associated with the finite difference operator of order 1.

2.3 MCMC implementation

Bayesian inference is achieved using a Markov chain Monte Carlo (MCMC) algorithm. According to Makalic and Schmidt (2015), half-Cauchy distribution can be expressed as a scale mixture of inverse-Gamma distributions

$$x \sim \mathcal{C}^+(0,1) \quad \Leftrightarrow \quad x^2|\xi \sim IG(1/2, 1/\xi), \quad \xi \sim IG(1/2, 1).$$

Thus, all full conditional distributions have closed form. An efficient Gibbs sampling algorithm can consequently be implemented. Full conditional distributions are detailed in appendices A.1, A.2 and A.3 for $\text{HS}_{\lambda}^{\omega_t}$ – $\text{HS}_1^{\gamma_t}$, $\text{HS}_{\lambda_g}^{\omega_{gt}}$ – $\text{HS}_1^{\gamma_{gt}}$ and $\text{HS}_{\lambda}^{\omega_{gt}}$ – $\text{HS}_1^{\gamma_{gt}}$ respectively. Details for $\text{NE}_{\lambda}^{\omega_t}$ – $\text{NE}_{\upsilon}^{\gamma_t}$ and $\text{HS}_{\lambda}^{\omega_t}$ – $\text{NE}_{\upsilon}^{\gamma_t}$ are provided in related papers (Kyung

et al., 2010; Kakikawa et al., 2023). Computer codes are freely available on the following GitHub page: `https://github.com/Heuclin/GroupFusedHorseshoe`. Simulations (for reproducibility) and MCMC functions were implemented in R (R Core Team, 2023).

3 Simulation study

This section provides evidence for the improved performances of the proposed priors compared to the existing ones. First, we discuss the advantages of adding heavy tail distributions on coefficients compared to the usual Laplace priors used in the one-group context. Then we show the efficiency of the new priors in terms of shrinkage properties, parameter estimation and algorithmic stability in different settings: the number of groups, their size and the residual variance. The simulations are performed as follows. The number of individuals, n, was set to 150. The number of covariates, p, was set to 1500. We assumed that p was divided into $G = 1, 10, 30$ or 100 groups. Covariates within each group were sampled from a $\frac{p}{G}$-multivariate Gaussian distribution with zero mean and a covariance matrix defined as a first-order autoregressive (AR1) structure with the correlation parameter set at 0.95. Functional effects were defined as the combination of different functions: a continuous smooth function, as proposed by Faulkner and Minin (2018), and a piece-wise function of varying size (Tibshirani et al., 2014):

$$\beta_t = \begin{cases} \sin(4t/T - 2) + 2e^{-30(4t/T-2)^2} & t < T \\ 0.5 & t \in [T+1, 2T] \\ -0.5 & t \in [2T+1, (2+1/2)T] \\ 0.5 & t \in [3T+1, (3+1/3)T] \\ -0.5 & t \in [4T+1, (4+1/4)T] \\ 0 & \text{otherwise} \end{cases}$$

where $T = \min\left(\frac{p}{\max(10,G)}, 60\right)$. Finally, the residual variance, σ^2, was set at 1 or 16. 100 replicated datasets were generated for each combination of parameters G and σ^2. Performance was evaluated using the mean squared error calculated either on true zeros (MSE_z) or on non-zero coefficients (MSE_{nz}) as measures of the estimation accuracy. We also used the Matthew Correlation Coefficient (MCC, Matthews (1975)), the True Positive Rate on all non-zero coefficients (TPR) and the True Positive Rate on groups (TPR_{gr}) as measures of the selection quality. For all methods, a variable is deemed relevant if the 95% credible interval of its regression coefficient does not contain zero, and a group is selected if it contains at least one significant variable. By denoting $\mathcal{C}_z = \{t : \ \beta_t = 0\}$ and $\mathcal{C}_{nz} = \{t : \ \beta_t \neq 0\}$, the sets of indices of true zero and non-zero coefficients respectively, the performances are

calculated through:

$$MSE_z = \frac{1}{|\mathcal{C}_z|}\sum_{t\in\mathcal{C}_z}(\beta_t - \hat{\beta}_t)^2;\ MSE_{nz} = \frac{1}{|\mathcal{C}_{nz}|}\sum_{t\in\mathcal{C}_{nz}}(\beta_t - \hat{\beta}_t)^2;$$
$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}};$$
$$TPR = \frac{TP}{|\mathcal{C}_{nz}|};\ \text{and}\ TPR_{gr} = \frac{TP_{gr}}{G}$$

where $\hat{\beta}_t$ is the posterior mean of the regression coefficient for the variable X_t, and TP, TN, FP, FN correspond to True (T) and False (F) negative (N) and positive (P) coefficients, respectively. TP_{gr} is the true positive groups. Each criterion was averaged over the 100 replicates. For each repetition, the MCMC algorithm was run for $15,000$ iterations, the first $3,000$ used as burn-in and a thinning of 2.

As shown in Table 3, in the one-group context ($G = 1$), the performances of $\text{HS}_\lambda^{\omega_t}$ – $\text{HS}_1^{\gamma_t}$ and $\text{HS}_\lambda^{\omega_t}$ – $\text{NE}_\upsilon^{\gamma_t}$ were close regardless of the value of the residual variance. Both clearly outperformed the standard $\text{NE}_\lambda^{\omega_t}$ – $\text{NE}_\upsilon^{\gamma_t}$ and $\text{NEG}_1^{\omega_t}$ – $\text{NE}_\upsilon^{\gamma_t}$ priors, which demonstrated comparable performance. These results showed that using priors with heavy tails, compared to Laplace distribution on differences, considerably improves performances. Similar results were observed by Kakikawa et al. (2023). However, assuming a horseshoe prior when the global shrinkage parameter on coefficients is random and distributed as a half-Cauchy, $\text{HS}_\lambda^{\omega_t}$ – $\text{HS}_\upsilon^{\gamma_t}$, led to convergence problems. Such prior shrunk all coefficients towards zero, in that sense it failed to converge in almost all cases. These findings are consistent with those reported by Kakikawa et al. (2023) and evidenced a trade-off between shrinkage priors placed on coefficients and their differences. We also observed that shrinkage properties were slightly reinforced using horseshoe prior with a fixed global shrinkage parameter ($\text{HS}_1^{\gamma_t}$) on coefficients compared to the use of a Laplace prior ($\text{NE}_\upsilon^{\gamma_t}$) (see MSE_z column in Table 3).

In the multi-group context, Figure 1 and Table B.4 in Appendix B showed that priors assuming HS distributions on both the coefficients and their differences yielded highly relevant results, regardless of the value of the residual variance and the number of groups. For priors combining a HS distribution on differences with a NE distribution on coefficients, results were number of group dependent. The use of NE distribution priors on both components led to poor results or even did not converge when number of group was high. Such priors did not sufficiently shrink parameters, leading to an overestimation of zero coefficients strongly impacting algorithm convergences. These findings confirmed the importance of using a heavy-tailed distributions on differences, as observed in the one-group context and in Kakikawa et al. (2023). Figures B.1 in the Appendix B presented the estimated coefficients profiles for all priors for a given scenario. As expected, the $\text{NE}_\lambda^{\omega_t}$ – $\text{NE}_\upsilon^{\gamma_t}$ prior resulted in noisy estimations along with wide credible intervals.

We now gain insights into the results obtained with priors using HS on differences. For a small to moderate number of groups ($G = 5$ or 10), performance in terms of variable selection was similar across all priors, with MCC values ranging from 0.89 to 0.98. False Positive Rates

σ^2	Priors	MSE_z	MSE_{nz}	MCC	TPR	Counts
1	$\text{HS}_{\lambda}^{\omega_t}$ – $\text{HS}_{1}^{\gamma_t}$	0.00004	**0.016**	**0.91**	**84**	95
1	$\text{HS}_{\lambda}^{\omega_t}$ – $\text{HS}_{\upsilon}^{\gamma_t}$	**0.00000**	2.662	0.07	1	100
1	$\text{HS}_{\lambda}^{\omega_t}$ – $\text{NE}_{\upsilon}^{\gamma_t}$	0.00068	0.050	0.90	**84**	91
1	$\text{NE}_{\lambda}^{\omega_t}$ – $\text{NE}_{\upsilon}^{\gamma_t}$	0.00273	0.046	0.21	5	100
1	$\text{NEG}_{1}^{\omega_t}$ – $\text{NE}_{\upsilon}^{\gamma_t}$	0.00326	0.047	0.04	0	100
16	$\text{HS}_{\lambda}^{\omega_t}$ – $\text{HS}_{1}^{\gamma_t}$	0.00012	**0.050**	0.82	71	89
16	$\text{HS}_{\lambda}^{\omega_t}$ – $\text{HS}_{\upsilon}^{\gamma_t}$	**0.00000**	2.562	0.07	1	94
16	$\text{HS}_{\lambda}^{\omega_t}$ – $\text{NE}_{\upsilon}^{\gamma_t}$	0.00092	0.055	**0.89**	**81**	87
16	$\text{NE}_{\lambda}^{\omega_t}$ – $\text{NE}_{\upsilon}^{\gamma_t}$	0.00516	0.054	0.15	3	96
16	$\text{NEG}_{1}^{\omega_t}$ – $\text{NE}_{\upsilon}^{\gamma_t}$	0.00612	0.056	0.03	0	100

Table 3: Mean squared errors of the true zeroes (MSE_z), the true non-zeroes (MSE_{nz}), the Matthews Correlation Coefficient (MCC), and the True Positive Rate on all non-zero coefficients (TPR) using priors defined in Table 1 with residual variance σ^2 equal to 1 or 16 in the one-group context ($G = 1$). All criteria were averaged over 100 simulated datasets. The last column, Counts, gives the number of MCMC runs that reached convergence. A value lower than 100 indicates that some runs failed to converge, mainly due to numerical instabilities.

(FPR), not reported here, were systematically equal to zero. Regarding group selection, all priors enjoyed TPR_{gr} values close to 1, highlighting their ability to accurately identify the relevant groups of variables. In terms of estimating non-zero coefficients, priors gave comparable results, with MSE_{nz} values ranging from 0.010 to 0.032. For the zero coefficients, as observed in the one-group context, the shrinkage property was slightly enhanced when using an HS prior on the coefficients (see the MSE_z column in Table B.4 in Appendix B).

When number of groups increased ($G = 30$ or 100), performances of $\text{HS}_{\lambda_g}^{\omega_{gt}}$ – $\text{HS}_{1}^{\gamma_{gt}}$ and $\text{HS}_{\lambda}^{\omega_{gt}}$ – $\text{HS}_{1}^{\gamma_{gt}}$ were not impacted and results were comparable to those obtained when the number of group was small (less or equal to 10). In contrast, as shown in Figure 1, $\text{HS}_{\lambda}^{\omega_{gt}}$ – $\text{NE}_{\upsilon_g}^{\gamma_{gt}}$, $\text{HS}_{\lambda_g}^{\omega_{gt}}$ – $\text{NE}_{\upsilon_g}^{\gamma_{gt}}$ and the fusion prior, $\text{HS}_{\lambda}^{\omega_{gt}}$, which assumed distribution only on differences (Faulkner and Minin, 2018), performances were group dependent. When the number of groups was equal to 30, over the 100 runs and whatever the value of the residual variance (1 or 16), approximately 89% of times MCMC reached convergences for $\text{HS}_{\lambda}^{\omega_{gt}}$ – $\text{NE}_{\upsilon_g}^{\gamma_{gt}}$ while $\text{HS}_{\lambda_g}^{\omega_{gt}}$ – $\text{NE}_{\upsilon_g}^{\gamma_{gt}}$ systematically failed to converge (see Table B.4 in Appendix B). For the fusion $\text{HS}_{\lambda}^{\omega_{gt}}$, all criteria were slightly deteriorated. For example MCC values were lower than 0.9 while were closed to one when number of groups was equal to 10. For $G = 100$, all priors excepted using the HS – HS type priors, failed to convergence. These results could be explained by the reduction in the number of observations within groups avoiding either an appropriate inference of global parameters, υ_g, at the group level (see Table B.4 in Appendix B) or because the distribution tail used for the coefficients was not heavy enough. These results underscored the importance of carefully defining priors on coefficients to ensure numerical regularization and reinforced trade-off shrinkage priors placed on coefficients and their differences.

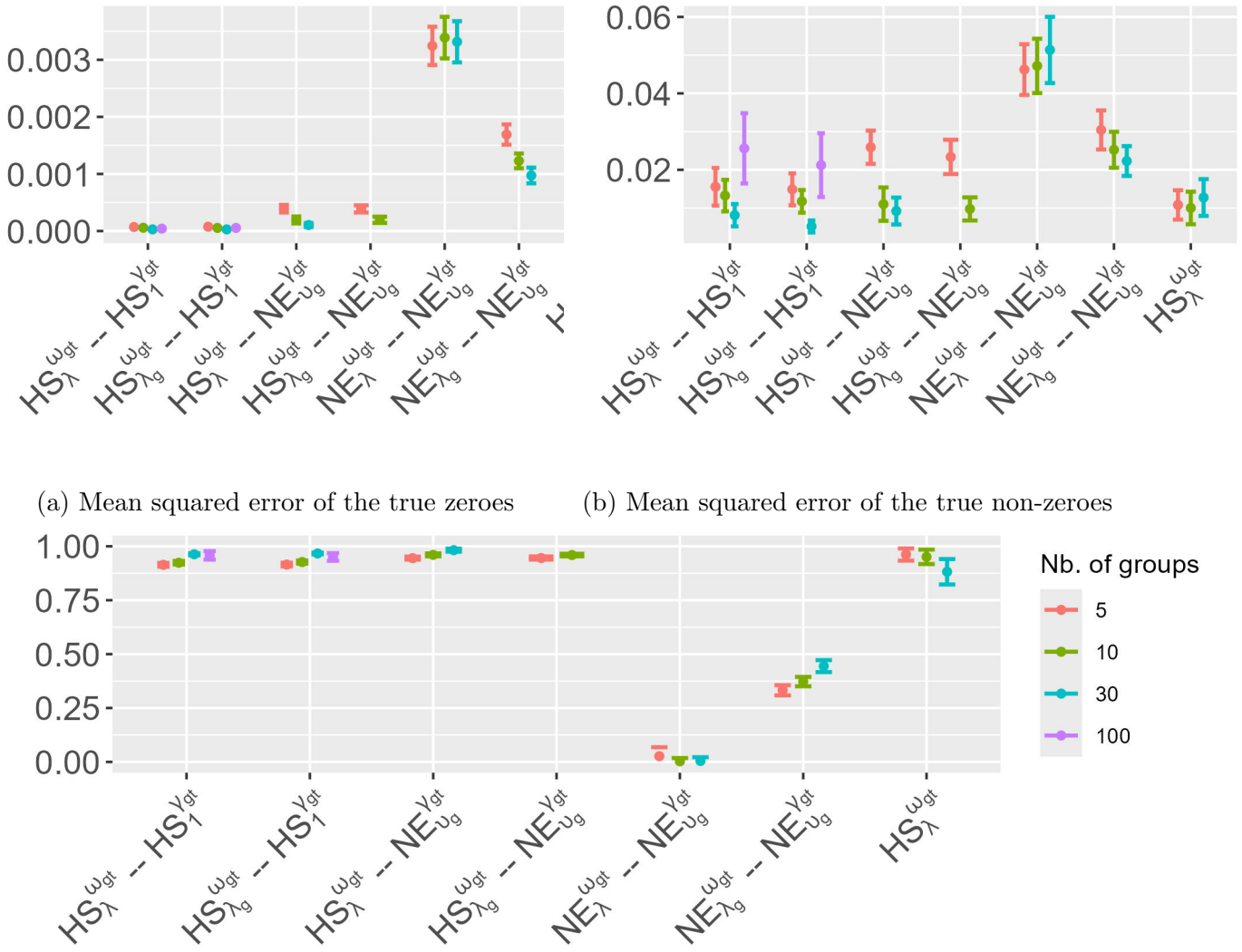
Results evidenced no differences in terms of selection and estimation performances between $\text{HS}_{\lambda}^{\omega_{gt}}$ – $\text{HS}_{1}^{\gamma_{gt}}$ and $\text{HS}_{\lambda_g}^{\omega_{gt}}$ – $\text{HS}_{1}^{\gamma_{gt}}$ (see Figure 1c) and in estimating non-zero and zero coefficients (see Figure 1b). Focusing on algorithmic performances and stability, again both priors achieved convergence across all tested scenarios. A higher residual variance had only a slight impact on the results. For instance, the *MCC* values remained close to 0.9 for $\sigma^2 = 1$ and 16, whatever the number of groups. To sum up, these findings encouraged favoring the simplified version, which assumes a single global shrinkage parameter $\lambda_g = \lambda$, resulting in a more parsimonious prior.

4 Real application: determinisms of abscission in oil palm

4.1 The abscission dataset

The objective of the application is to identify the environmental variables and time periods of the inflorescence and bunch development that affect the oil palm fruit abscission process (Tisné et al., 2020). The dataset was provided by "le Centre de Recherches Agricoles-Plantes Pérennes" (CRA-PP) of the national institute for agricultural research of Benin Republic (INRAB) which manages an oil palm seed garden involving a self-pollinated population of 140 oil palm trees planted between 2000 and 2005 in a single homogeneous field plot. Each palm tree produced between 1 and 8 bunches per year over the whole period of the experiment from 2014 to 2018 (see the examples of two individual oil palms in Figure 2c). The manual pollination date differed for each bunch. The date was recorded and the bunch was monitored up to harvest (see Figure 2c). A total of $1,173$ bunches were considered over multiple years to take advantage of the climatic seasonality and the continuous fruit production of this species. We used the number of days from pollination to fruit drop (DFD) as the response variable. DFD is the classical harvest indicator and its variation integrates different underlying abscission processes at different developmental stages.

Additionally, nine covariates were used as predictors. Five climate variables were recorded from 2014 to 2018: maximum and minimum temperature (Tmax, Tmin, in °C, see top left panel Fig. 2a), relative air humidity (RH, in %), rainfall (R, in mm) and solar radiation (SR, in $\text{cal.cm}^{-2}\text{.d}^{-1}$, see bottom left panel Fig. 2a). Note that climate variables are similar for all individuals but vary between months and years. Errorbars reflect such variations. Four ecophysiological variables were calculated using climate and individual production data: two exogenous variables, the maximum daily vapor pressure deficit (VPD), the fraction of transpirable soil water (FTSW, see right panel Fig. 2a), and two endogenous (trophic) variables: the supply–demand ratio (SD) and the daily reproductive demand (DRD, see left panel Fig. 2b). More details are available in Tisné et al. (2020). Contrary to climate variables, ecophysiological variables are fixed among years but vary between individuals. Error-bars summarize such variability. We also added the auto-correlation function (ACF) illustrating strong dependencies between successive measurements thus reinforcing regularization challenges (see right panel Fig. 2b). These variables can have punctual or cumulative effects,



(a) Mean squared error of the true zeroes

(b) Mean squared error of the true non-zeroes

(c) Matthews correlation coefficient

Figure 1: Boxplots of mean squared errors of the true zeroes 1a, the true non-zeroes 1b and of the Matthews correlation coefficient 1c over 100 replicates for priors defined in Table 2. Colors refer to number of groups varying from 5 to 100. The residual variance is set to one.

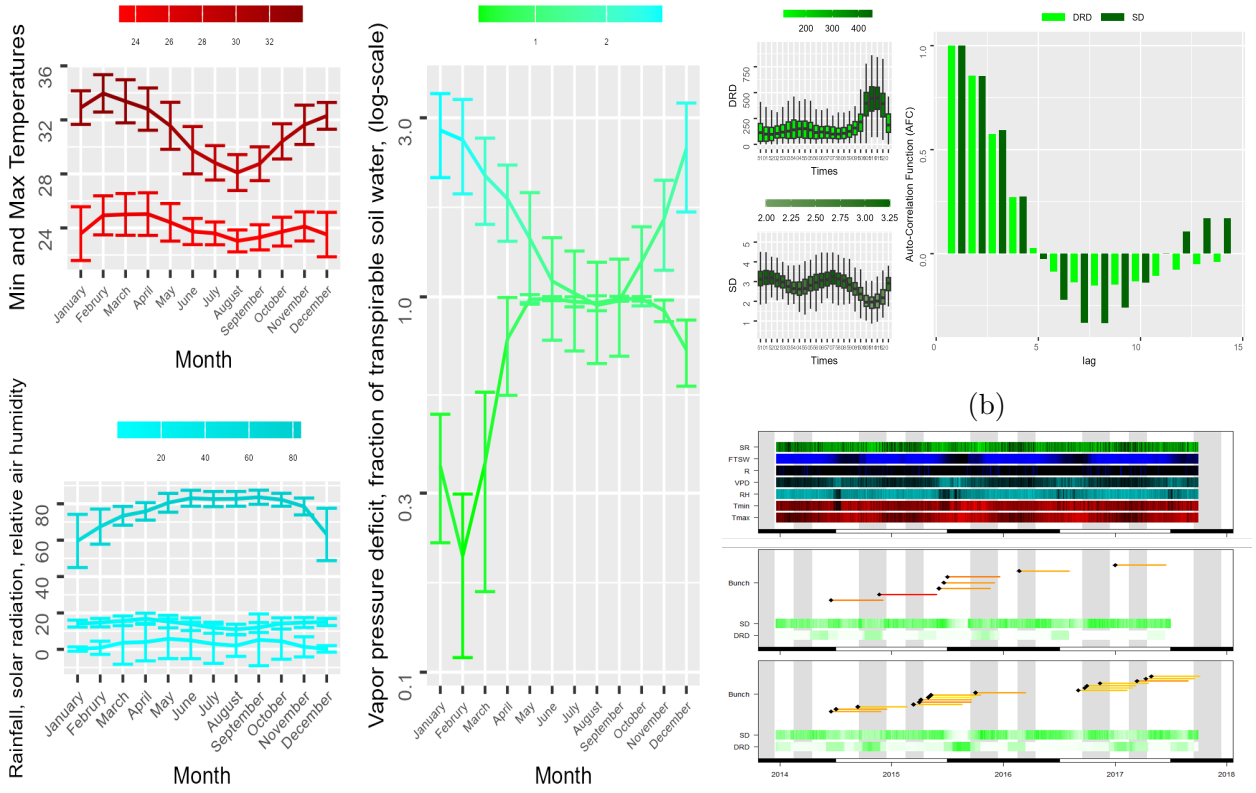


Figure 2: Panel 2a corresponds to raw (Tmin, Tmax, R, RH and SR) and calculated (RH, VPD and FTSW) climate variable over times. Error bars are the inter-annual variability. Panel 2b presents the time-varying average of the endogenous variables DRD and SD (left panels). Error bars are calculated among individuals. Auto-correlation function associated to both variables is presented as an example (right panel). Panel 2c corresponds to raw and calculated environmental variables (top panel) and example of bunch production for two oil palm individuals along with their endogenous variables (middle and bottom panels). Raw and calculated variables (heatmaps) as well as bunches (horizontal segments) are plotted over the experiment duration, from 2014 to 2018. The segments for bunches are represented from the manual pollination (black diamonds) to the harvest, the yellow to red color indicating increasing DFD. DFD: days to fruit drop; SR: Solar Radiation; FTSW: Fraction of transpirable soil water; R: rainfall; VPD: Vapor pressure deficit; RH: Relative humidity; Tmin/Tmax: minimun/maximum temperatures; SD: supply-demand ratio; DRD: Daily reproductive demand.

depending on the biological process or the developmental stage concerned. Temperature can have punctual effects such as stopping growth at low temperatures, but also cumulative effects on developmental rates that led to the thermal time development. A three-day time grid, from −180 (individualization of the floral meristem) to +180 (ripe fruit) days after pollination, was used to calculate either the average values over three days (Tmax, Tmin, RH, VPD, FTSW, DRD, and SD) or the cumulative values over 15 days (R and SR) of each variable. This experimental design led to nine groups of covariates ($G = 9$) measured $T = 121$ times. Within each matrix, the i^{th} row corresponds to the i^{th} bunch analyzed, and the t^{th} column corresponds to the value of the corresponding climatic/ecophysiological variable at time t for each bunch (see bottom panel 2c). All matrices were scaled to obtain a similar order of magnitude. All the results are based on 20 MCMC runs initialized at random starting values and 50,000 iterations with a burn-in of 20,000 and a thinning of 10. A group is considered selected if at least one regression effect within it has a credible interval that does not contain zero.

4.2 Identification of determinism of abscission and biological interpretation

The coefficient profiles estimated using $\text{HS}_{\lambda}^{\omega_{gt}}$ – $\text{HS}_{1}^{\gamma_{gt}}$ prior are very clear and allow the identification of relevant time periods of four variables (DRD, SD, SR, and Tmin, see Figure 3). Two types of patterns can be observed: smooth effects for Tmin and SR and punctual effects for DRD and SD. The Tmin variable is negatively associated with DFD during the development of the inflorescence from day −180 to −100, while the three other variables are associated with DFD at the end of the fruit bunch development. SR, the solar radiation variable, is positively associated with DFD from day 120 to 180, at the final stage before fruit drop. The DRD variable is punctually associated with DFD at days 99 and 105 after pollination, first positively before an inversion of the association direction at day 100. The SD factor is negatively associated with DFD with a peak at day 160.

The striking pattern of DRD around day 100 after pollination (see Figure 3) was also observed by Tisné et al. (2020) and corresponds to the "lag period" of the oil palm fruit bunch development between the cell division/expansion phase and the maturation phase (Tranbarger et al., 2011). The selection of the DRD variable at this key developmental stage suggests that to modulate its maturation and abscission timing, the fruit bunch concerned integrates current and future whole plant photosynthate demand due to concomitant developing bunches. Such carbohydrate-based regulation is common in fruit tree species and leads to the wave of abscission that affects fruitlets (Sawicki et al., 2015), the only difference being that the oil palm regulates the timing of ripe fruit abscission rather than dropping unripe fruits. In contrast with DRD and SD which have similar punctual patterns to those reported in the Tisné et al. (2020) study, the profile of the Tmin effect is different, with a continuous moderate effect instead of many weak effects spread out over the −180 to −100 period (see Figure 3). The SR factor was not selected by Tisné et al. (2020), but using our prior, it has a positive effect from day 120 to 180 (see Figure 3). These discrepancies may be due to

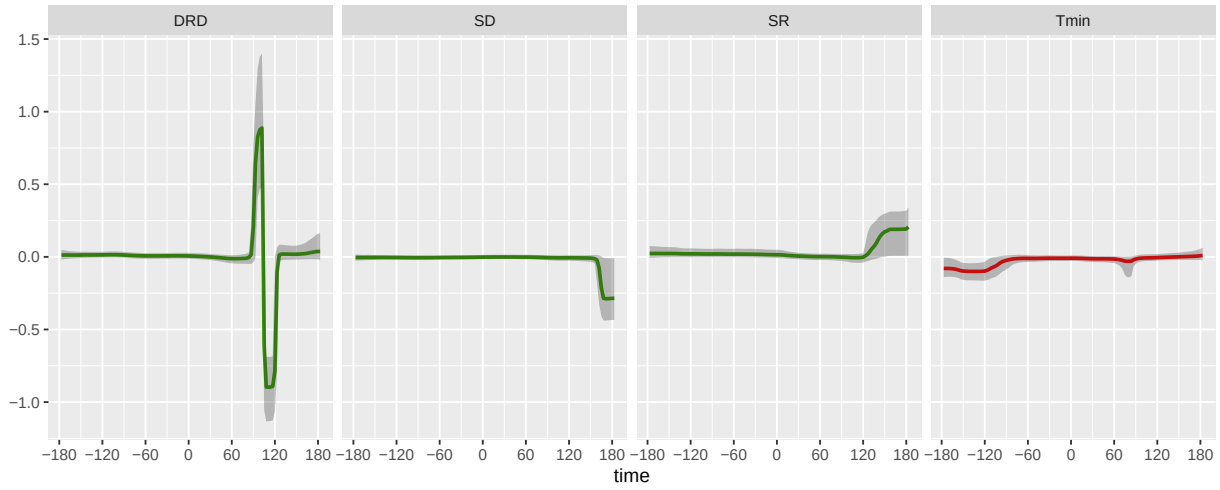


Figure 3: Estimated non-zero coefficient profile provided by the $\text{HS}_{\lambda}^{\omega_{gt}}$ – $\text{HS}_{1}^{\gamma_{gt}}$ prior using the abscission dataset. Gray shadowed areas indicate the 95% credible interval. Colors represent the different categories of environmental variables, green is for photosynthesis variables (DRD, SR, SD) and red is for temperature variables.

the cumulative nature of both Tmin and SR effects at their respective developmental stages. Hence, the Tmin effect at the early inflorescence developmental stages could be linked to thermal time, which is known to be associated with development rates. The differentiation of floral organs occurs in the period identified (Adam et al., 2011) and variations in cumulative thermal time could modulate the developmental program and ultimately the timing of fruit drop. The cumulative effect of radiation was identified throughout the final stage before fruit drop that corresponds to fruit maturation with intensive accumulation of lipids and is closely linked to photosynthate availability (Tranbarger et al., 2011). Our proposed prior, which was designed to estimate a smooth and flexible coefficient profile is thus well suited to study the effect of cumulative effect variables in addition to the punctual effect variables that were consistently identified in both approaches, in line with the group penalized approach used in Tisné et al. (2020).

5 Conclusion

In this paper, we propose a set of original fused-type priors adapted to both the one-group and the multi-group context. In particular, we call for the use of horseshoe priors on differences and on coefficients. We show that prior with heavier tail distribution compared to the usual Laplace distribution is efficient not only for selection and estimation but also for algorithmic stability mainly when the number of groups is large and the groups are small. However, we also show that there is a trade-off between prior assumptions made on differences and on coefficients. The use of a full horseshoe prior for differences and coefficients with random global shrinkage parameters, $\text{HS}_{\lambda}^{\omega_t}$ – $\text{HS}_{\upsilon}^{\gamma_{gt}}$, led to inconsistent results, shrinking

all parameters to zero. While using a NE distribution on coefficients combined with a HS distribution on differences may lead to poor and unstable results especially when the number of groups is large and their size is low. To sum up, this work promote the use of a horseshoe prior distribution with a global shrinkage parameter fixed at 1 for coefficients and a random global parameter for differences but common to all groups. Such prior appears as a appropriate trade-off between efficiency and parsimony.

From a biological point of view, the proposed prior clearly identifies four environmental variables as well as the periods during which they affect the abscission process of oil palm trees. By allowing flexibility in the estimation of regression coefficient profiles, we identified an additional environmental variable to those identified in a previous study using a group penalized approach on the same data, and improved the biological interpretation of the regression profiles. The proposed prior will help biologists identify the best time to harvest the bunches.

The proposed prior can be directly applied to a broad range of applications particularly because the considered groups may vary in size. For exemple, it can be used in the near-infrared spectroscopy context, which involves one-group of ordered variables through a spectrum, or in the genetic mapping context, where markers can be viewed as groups of ordered variables at the chromosome level. To account for multi-dimensional indexation (spatial or spatio-temporal structures) instead of only one-dimensional indexation (time structure), this prior should be extended even if it raises computational challenges.

Declarations

Conflicts of interests The authors have no conflict of interest to declare.

References

Adam, H., Collin, M., Richaud, F., Beulé, T., Cros, D., Omoré, A., Nodichao, L., Nouy, B., and Tregear, J. W. (2011). Environmental regulation of sex determination in oil palm: current knowledge and insights from other species. *Annals of botany*, 108(8):1529–1537.

Alaíz, C. M., Barbero, A., and Dorronsoro, J. R. (2013). Group fused lasso. In *International Conference on Artificial Neural Networks*, pages 66–73. Springer.

Andrews, D. F. and Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(1):99–102.

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480.

Castillo, I., Schmidt-Hieber, J., and van der Vaart, A. (2015). Bayesian linear regression with sparse priors. *The Annals of Statistics*, 43(5):1986–2018.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360.

Faulkner, J. R. and Minin, V. N. (2018). Locally adaptive smoothing with markov random fields and shrinkage priors. *Bayesian analysis*, 13(1):225–252.

George, E. I. and McCulloch, R. E. (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889.

Gill, A. L., Gallinat, A. S., Sanders-DeMott, R., Rigden, A. J., Short Gianotti, D. J., Mantooth, J. A., and Templer, P. H. (2015). Changes in autumn senescence in northern hemisphere deciduous trees: a meta-analysis of autumn phenology studies. *Annals of botany*, 116(6):875–888.

Griffin, J. E., Brown, P. J., et al. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian analysis*, 5(1):171–188.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Kakikawa, Y., Shimamura, K., and Kawano, S. (2023). Bayesian fused lasso modeling via horseshoe prior. *Japanese Journal of Statistics and Data Science*, 6(2):705–727.

Kyung, M., Gill, J., Ghosh, M., Casella, G., et al. (2010). Penalized regression, standard errors, and bayesian lassos. *Bayesian Analysis*, 5(2):369–411.

Land, S. R. and Friedman, J. H. (1997). Variable fusion: A new adaptive signal regression method. Technical report, Technical Report 656, Department of Statistics, Carnegie Mellon University.

Liquet, B., Mengersen, K., Pettitt, A., Sutton, M., et al. (2017). Bayesian variable selection regression of multivariate responses for group data. *Bayesian Analysis*, 12(4):1039–1067.

Makalic, E. and Schmidt, D. F. (2015). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters*, 23(1):179–182.

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, 405(2):442–451.

Park, T. and Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686.

Piironen, J., Vehtari, A., et al. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics*, 11(2):5018–5051.

Polson, N. G. and Scott, J. G. (2011). Shrink Globally, Act Locally: Sparse Bayesian Regularization and Prediction. In *Bayesian Statistics 9*. Oxford University Press.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reichardt, S., Piepho, H.-P., Stintzi, A., and Schaller, A. (2020). Peptide signaling for drought-induced tomato flower drop. *Science*, 367(6485):1482–1485.

Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*. Chapman and Hall/CRC press.

Sawicki, M., Aït Barka, E., Clément, C., Vaillant-Gaveau, N., and Jacquard, C. (2015). Cross-talk between environmental stresses and plant metabolism during reproductive organ abscission. *Journal of Experimental Botany*, 66(7):1707–1719.

Shimamura, K., Ueki, M., Kawano, S., and Konishi, S. (2019). Bayesian generalized fused lasso modeling via neg distribution. *Communications in Statistics-Theory and Methods*, 48(16):4132–4153.

Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2013). A sparse-group lasso. *Journal of computational and graphical statistics*, 22(2):231–245.

Song, Q. and Cheng, G. (2020). Bayesian fusion estimation via t shrinkage. *Sankhya A*, 82(2):353–385.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):91–108.

Tibshirani, R. J. et al. (2014). Adaptive piecewise polynomial estimation via trend filtering. *The Annals of Statistics*, 42(1):285–323.

Tisné, S., Denis, M., Domonhédo, H., Pallas, B., Cazemajor, M., Tranbarger, T. J., and Morcillo, F. (2020). Environmental and trophic determinism of fruit abscission and outlook with climate change in tropical regions. *Plant-Environment Interactions*, 1(1):17–28.

Tranbarger, T. J., Dussert, S., Joët, T., Argout, X., Summo, M., Champion, A., Cros, D., Omore, A., Nouy, B., and Morcillo, F. (2011). Regulatory mechanisms underlying oil palm fruit mesocarp maturation, ripening, and functional specialization in lipid and carotenoid metabolism. *Plant physiology*, 156(2):564–584.

Wilmowicz, E., Kućko, A., Pokora, W., Kapusta, M., Jasieniecka-Gazarkiewicz, K., Tranbarger, T. J., Wolska, M., and Panek, K. (2021). Epip-evoked modifications of redox, lipid, and pectin homeostasis in the abscission zone of lupine flowers. *International journal of molecular sciences*, 22(6):3001.

Xu, X., Ghosh, M., et al. (2015). Bayesian variable selection and estimation for group lasso. *Bayesian Analysis*, 10(4):909–936.

Xu, Z., Schmidt, D. F., Makalic, E., Qian, G., and Hopper, J. L. (2016). Bayesian grouped horseshoe regression with application to additive models. In *Australasian Joint Conference on Artificial Intelligence*, pages 229–240. Springer.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67.

Zhang, L., Baladandayuthapani, V., Mallick, B. K., Manyam, G. C., Thompson, P. A., Bondy, M. L., and Do, K.-A. (2014). Bayesian hierarchical structured variable selection methods with application to molecular inversion probe studies in breast cancer. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 63(4):595–620.

Zou, H. and Hastie, T. J. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B*, 67:301–320.

Appendices

A Prior details, Bayesian hierarchical models and full conditional distributions

A.1 The $\text{HS}^{\omega_t}_{\lambda}$ – $\text{HS}^{\gamma_t}_{1}$ prior for one-group

The equation (4) can be reformulated as a hierarchical model using the global-local parametrization of Gaussian scale mixture representation:

$$\begin{aligned}
\boldsymbol{\beta}|\sigma^2, \lambda, \boldsymbol{\gamma}, \boldsymbol{\omega} &\sim \mathcal{N}_T(0, \sigma^2 \boldsymbol{Q}^{-1}) \\
\lambda &\sim \mathcal{C}^+(0,1) \\
\omega_t &\sim \mathcal{C}^+(0,1) \;\; t = 1, \ldots, T \\
\gamma_t &\sim \mathcal{C}^+(0,1) \;\; t = 1, \ldots, T \\
\sigma^2 &\sim \mathcal{C}^+(0,1)
\end{aligned}$$

where $\boldsymbol{Q}$ is the matrix equal to

$$\boldsymbol{Q} = \begin{pmatrix}
\frac{1}{\gamma_1^2} + \frac{1}{\lambda^2\omega_2^2} & -\frac{1}{\lambda^2\omega_2^2} & 0 & \ldots & 0 & 0 \\
\frac{1}{\lambda^2\omega_2^2} & \frac{1}{\gamma_2^2} + \frac{1}{\lambda^2\omega_2^2} + \frac{1}{\lambda^2\omega_3^2} & -\frac{1}{\lambda^2\omega_3^2} & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \ldots & \frac{1}{\gamma_{T-1}^2} + \frac{1}{\lambda^2\omega_{T-1}^2} + \frac{1}{\lambda^2\omega_T^2} & -\frac{1}{\lambda^2\omega_T^2} \\
0 & 0 & 0 & \ldots & -\frac{1}{\lambda^2\omega_T^2} & \frac{1}{\gamma_T^2} + \frac{1}{\lambda^2\omega_T^2}
\end{pmatrix}$$

which is equivalent to

$$\boldsymbol{Q} = \left(\boldsymbol{\Upsilon}^{-1} + \boldsymbol{D}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{D} / \lambda^2 \right).$$

The first matrix, $\boldsymbol{\Upsilon}^{-1}$, refers to regression parameters, and the second, $\boldsymbol{D}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{D}/\lambda^2$ to the differences. $\boldsymbol{D}$ is the known $T \times (T-1)$-matrix associated to the finite differences operator of order 1, and $\boldsymbol{\Omega} = diag(\omega_1^2, \ldots, \omega_{T-1}^2)$ and $\boldsymbol{\Upsilon}_g = diag(\gamma_1^2, \ldots, \gamma_T^2)$ the $(T-1)\times(T-1)$-diagonal matrices of local parameters.

In order to compute the full conditional distributions, we use the scale mixture of inverse-Gamma distribution representation of the half-Cauchy distribution (Makalic and Schmidt,

2015). The resulting hierarchical model is the following:

$$
\begin{aligned}
\mathbf{y}|\mu,\boldsymbol{\beta},\alpha,\sigma^2 &\sim \mathcal{N}_n(\mu+\boldsymbol{X}\boldsymbol{\beta}+\mathbf{Z}\alpha,\ \sigma^2 I_n)\\
\mu &\sim \mathcal{U}_{(-\infty,\infty)}\\
\boldsymbol{\beta}|\sigma^2,\lambda,\boldsymbol{\gamma},\boldsymbol{\omega} &\sim \mathcal{N}_T(0,\sigma^2\boldsymbol{Q}^{-1})\\
\lambda^2|\xi &\sim \mathcal{IG}(\frac{1}{2},\frac{1}{\eta}),\ \xi\sim\mathcal{IG}(\frac{1}{2},1)\\
\omega_t^2|\phi_t &\sim \mathcal{IG}(\frac{1}{2},\frac{1}{\phi_t}),\ \phi_t\sim\mathcal{IG}(\frac{1}{2},1),\ t=1,\ldots,T\\
\gamma_t^2|\eta_t &\sim \mathcal{IG}(\frac{1}{2},\frac{1}{\eta_t}),\ \eta_t\sim\mathcal{IG}(\frac{1}{2},1),\ t=1,\ldots,T\\
\alpha|\sigma_u^2 &\sim \mathcal{N}_P(0,\sigma_\alpha^2 A),\quad \sigma_\alpha^2\sim IG\left(\frac{1}{2},\frac{1}{2}\right)\\
\sigma^2|a &\sim \mathcal{IG}(\frac{1}{2},\frac{1}{a}),\ a\sim\mathcal{IG}(\frac{1}{2},1)
\end{aligned}
$$

The corresponding full conditional distributions for the model parameters are given by:

$$
\begin{aligned}
&\mu|.\sim\mathcal{N}\left(\frac{1}{n}\mathbb{1}^\top(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{\alpha}),\ \frac{\sigma^2}{n}\right)\\
&\boldsymbol{\beta}|.\sim\mathcal{N}_T\left(\boldsymbol{\Sigma}_b\frac{\boldsymbol{X}^\top}{\sigma^2}(\boldsymbol{y}-\mu\mathbb{1}-\mathbf{Z}\alpha),\ \boldsymbol{\Sigma}_b=\sigma^2\left(\boldsymbol{X}^\top\boldsymbol{X}+\boldsymbol{\Upsilon}^{-1}+\frac{\boldsymbol{D}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{D}}{\lambda^2}\right)^{-1}\right)\\
&\lambda^2|.\sim IG\left(\frac{1+T}{2},\ \frac{1}{\xi}+\sum_{g=1}^{G}\frac{\boldsymbol{\beta}^\top\boldsymbol{D}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{D}\boldsymbol{\beta}}{2\sigma^2}\right),\quad \xi|.\sim IG(1,1+1/\lambda^2)\\
&\omega_j^2|.\sim IG\left(1,\ \frac{1}{\phi_j}+\frac{((\boldsymbol{D}\boldsymbol{\beta})_{[j]})^2}{2\sigma^2\lambda^2}\right),\quad \phi_j|.\sim IG(1,\ 1+1/\omega_j^2),\quad j=1,\ldots,T-1\\
&\upsilon_t^2|.\sim IG\left(1,\ \frac{1}{\eta_t}+\frac{\beta_t^\top\beta_t}{2\sigma^2}\right),\quad \eta_t|.\sim IG(1,\ 1+1/\upsilon_t^2),\quad t=1,\ldots,T\\
&\alpha|.\sim\mathcal{N}_P\left(\Sigma_\alpha\frac{Z^\top}{\sigma^2}(\boldsymbol{y}-\mu\mathbb{1}-\boldsymbol{X}\boldsymbol{\beta}),\ \Sigma_\alpha=\left(\frac{A^{-1}}{\sigma_\alpha^2}+\frac{\mathbf{Z}^\top\mathbf{Z}}{\sigma^2}\right)^{-1}\right)\\
&\sigma_\alpha^2|.\sim IG\left(\frac{1+P}{2},\ \frac{1}{2}+\alpha^\top A^{-1}\alpha\right)\\
&\sigma^2|.\sim IG\left(\frac{1+T+n}{2},\ \frac{1}{a}+\frac{1}{2}\boldsymbol{\beta}^\top\left(\boldsymbol{\Upsilon}^{-1}+\frac{\boldsymbol{D}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{D}}{\lambda^2}\right)\boldsymbol{\beta}+\frac{1}{2}||\boldsymbol{y}-\ \mu\mathbb{1}-\boldsymbol{X}\boldsymbol{\beta}-\mathbf{Z}\boldsymbol{\alpha}||_2^2\right)\\
&a|.\sim IG(1,1+1/\sigma^2)
\end{aligned}
$$

A.2 The $\mathrm{HS}_{\lambda_g}^{\omega_{gt}}$ – $\mathrm{HS}_1^{\gamma gt}$ prior for multi-group

As previously detailed for the $\mathrm{HS}_{\lambda}^{\omega_t}$ – $\mathrm{HS}_1^{\gamma_t}$ prior for one-group (see Appendix A.1), the $\mathrm{HS}_{\lambda_g}^{\omega_{gt}}$ – $\mathrm{HS}_1^{\gamma gt}$ prior given in equation 7 can be reformulated as a hierarchical model using

global-local parametrization of Gaussian scale mixture representation. Also, using the scale mixture of inverse-Gamma distribution representation of the half-Cauchy distribution, the resulting Bayesian hierarchical model for the MCMC implementation of the $\text{HS}^{\omega_{gt}}_{\lambda_g}$ – $\text{HS}^{\gamma gt}_{1}$ prior is given by:

$$
\begin{aligned}
&\mathbf{y}|\mu, \boldsymbol{\beta}, \alpha, \sigma^2 \sim \mathcal{N}_n(\mu + \sum_{g=1}^{G} \boldsymbol{X}_g \boldsymbol{\beta}_g + \mathbf{Z}\alpha,\ \sigma^2 I_n) \\
&\mu \sim \mathcal{U}_{(-\infty,\infty)} \\
&\boldsymbol{\beta}_g|\boldsymbol{\Upsilon}_g, \lambda_g^2, \boldsymbol{\Omega}_g, \sigma^2 \sim \mathcal{N}_T\left(0, \sigma^2\left(\boldsymbol{\Upsilon}_g^{-1} + \frac{1}{\lambda_g^2}\boldsymbol{D}_g^\top \boldsymbol{\Omega}_g^{-1}\boldsymbol{D}_g\right)^{-1}\right) \\
&\Upsilon^2_{g_t}|\eta_{g_t} \sim IG\left(\frac{1}{2}, \frac{1}{\eta_{g_t}}\right), \quad \eta_{g_t} \sim IG\left(\frac{1}{2}, 1\right), \quad g = 1, \ldots, G, \quad t = 1, \ldots, T \\
&\lambda_g^2|\psi_g \sim IG\left(\frac{1}{2}, \frac{1}{\psi_g}\right), \quad \psi_g \sim IG\left(\frac{1}{2}, 1\right), \quad g = 1, \ldots, G \\
&\omega^2_{g_j}|\phi_{g_j} \sim IG\left(\frac{1}{2}, \frac{1}{\phi_{g_j}}\right), \quad \phi_{g_j} \sim IG\left(\frac{1}{2}, 1\right), \quad g = 1, \ldots, G, \quad j = 1, \ldots, T-1 \\
&\alpha|\sigma_u^2 \sim \mathcal{N}_P(0, \sigma_\alpha^2 A), \quad \sigma_\alpha^2 \sim IG\left(\frac{1}{2}, \frac{1}{2}\right) \\
&\sigma^2|a \sim IG\left(\frac{1}{2}, \frac{1}{a}\right), \quad a \sim IG\left(\frac{1}{2}, 1\right)
\end{aligned}
$$

where $\boldsymbol{\Upsilon}_g = \text{diag}(\upsilon^2_{g_1}, \ldots, \upsilon^2_{g_T})$, $\boldsymbol{\Omega}_g = \text{diag}(\omega^2_{g_1}, \ldots, \omega^2_{g_{T-1}})$ and $\boldsymbol{D}_g$ is the known $T \times (T-1)$-matrix associated with the finite difference operator of order 1.

The corresponding full conditional distributions for the model parameters are given by:

$$\mu|. \sim \mathcal{N}\left(\frac{1}{n}\mathbb{1}^\top(\boldsymbol{y} - \sum_{g=1}^{G}\boldsymbol{X}_g\boldsymbol{\beta}_g), \frac{\sigma^2}{n}\right)$$

$$\boldsymbol{\beta}_g|. \sim \mathcal{N}_T\left(\boldsymbol{\Sigma}_{b_g}\frac{\boldsymbol{X}_g^\top}{\sigma^2}(\boldsymbol{y} - \mu\mathbb{1} - \sum_{\tilde{g}\neq g}\boldsymbol{X}_{\tilde{g}}\boldsymbol{\beta}_{\tilde{g}} - Z\alpha), \boldsymbol{\Sigma}_{b_g} = \sigma^2\left(\boldsymbol{X}_g^\top\boldsymbol{X}_g + \boldsymbol{\Upsilon}_g^{-1} + \frac{\boldsymbol{D}_g^\top\boldsymbol{\Omega}_g^{-1}\boldsymbol{D}_g}{\lambda_g^2}\right)^{-1}\right)$$

$$\lambda_g^2|. \sim IG\left(\frac{1}{2} + \frac{T}{2}, \frac{1}{\psi_g} + \frac{\boldsymbol{\beta}_g^\top\boldsymbol{D}_g^\top\boldsymbol{\Omega}_g^{-1}\boldsymbol{D}_g\boldsymbol{\beta}_g}{2\sigma^2}\right), \quad \psi_g|. \sim IG(1, 1 + 1/\lambda_g^2), \quad g = 1, \ldots, G$$

$$\omega_{g_j}^2|. \sim IG\left(1, \frac{1}{\phi_{g_j}} + \frac{((\boldsymbol{D}_g\boldsymbol{\beta}_g)_{[j]})^2}{2\sigma^2\lambda_g^2}\right), \quad \phi_{g_j}|. \sim IG(1, 1 + 1/\omega_{g_j}^2), \quad g = 1, \ldots, G, \quad j = 1, \ldots, T-1$$

$$\upsilon_{g_t}^2|. \sim IG\left(1, \frac{1}{\eta_{g_t}} + \frac{\beta_{g_t}^\top\beta_{g_t}}{2\sigma^2}\right), \quad \eta_{g_t}|. \sim IG(1, 1 + 1/\upsilon_{g_t}^2), \quad g = 1, \ldots, G, \quad t = 1, \ldots, T$$

$$\alpha|. \sim \mathcal{N}_P\left(\Sigma_\alpha\frac{Z^\top}{\sigma^2}(\boldsymbol{y} - \mu\mathbb{1} - \sum_{g=1}^{G}\boldsymbol{X}_g\boldsymbol{\beta}_g), \Sigma_\alpha = \left(\frac{A^{-1}}{\sigma_\alpha^2} + \frac{Z^\top Z}{\sigma^2}\right)^{-1}\right)$$

$$\sigma_\alpha^2|. \sim IG\left(\frac{1+P}{2}, \frac{1}{2} + \alpha^\top A^{-1}\alpha\right)$$

$$\sigma^2|. \sim IG\left(\frac{1+T+n}{2}, \frac{1}{a} + \frac{1}{2}\sum_{g=1}^{G}\beta_g^\top\left(\boldsymbol{\Upsilon}_g^{-1} + \frac{\boldsymbol{D}_g^\top\boldsymbol{\Omega}_g^{-1}\boldsymbol{D}_g}{\lambda_g^2}\right)\boldsymbol{\beta}_g + \frac{1}{2}||\boldsymbol{y} - \mu\mathbb{1} - \sum_{g=1}^{G}\boldsymbol{X}_g\boldsymbol{\beta}_g - \mathbf{Z}\alpha||_2^2\right)$$

$$a|. \sim IG(1, 1 + 1/\sigma^2)$$

A.3 The $\text{HS}_\lambda^{\omega_{gt}}$ – $\text{HS}_1^{\gamma gt}$ prior for multi-groups

As previously detailed for the $\text{HS}_\lambda^{\omega_t}$ – $\text{HS}_1^{\gamma_t}$ prior for one-group (see Appendix A.1), the $\text{HS}_\lambda^{\omega_{gt}}$ – $\text{HS}_1^{\gamma gt}$ prior giving in equation 8 can be reformulated as a hierarchical model using the global-local parametrization of Gaussian scale mixture representation. Also, using the scale mixture of inverse-Gamma distribution representation of the half-Cauchy distribution, the resulting Bayesian hierarchical model for the MCMC implementation of the Global $\text{HS}_\lambda^{\omega_{gt}}$ – $\text{HS}_1^{\gamma gt}$ prior is given by:

$$
\begin{aligned}
&\mathbf{y}|\mu, \boldsymbol{\beta}, \alpha, \sigma^2 \sim \mathcal{N}_n(\mu + \sum_{g=1}^{G} \boldsymbol{X}_g \boldsymbol{\beta}_g + \mathbf{Z}\alpha,\ \sigma^2 I_n) \\
&\mu \sim \mathcal{U}_{(-\infty,\infty)} \\
&\boldsymbol{\beta}_g|\boldsymbol{\Upsilon}_g, \lambda^2, \boldsymbol{\Omega}_g, \sigma^2 \sim \mathcal{N}_T\left(0, \sigma^2\left(\boldsymbol{\Upsilon}_g^{-1} + \frac{1}{\lambda^2}\boldsymbol{D}_g^\top \boldsymbol{\Omega}_g^{-1}\boldsymbol{D}_g\right)^{-1}\right) \\
&\Upsilon_{g_t}^2|\eta_{g_t} \sim IG\left(\frac{1}{2}, \frac{1}{\eta_{g_t}}\right), \quad \eta_{g_t} \sim IG\left(\frac{1}{2}, 1\right), \quad g = 1, \ldots, G, \quad t = 1, \ldots, T \\
&\lambda^2|\xi \sim IG\left(\frac{1}{2}, \frac{1}{\xi}\right), \quad \xi \sim IG\left(\frac{1}{2}, 1\right) \\
&\omega_{g_j}^2|\phi_{g_j} \sim IG\left(\frac{1}{2}, \frac{1}{\phi_{g_j}}\right), \quad \phi_{g_j} \sim IG\left(\frac{1}{2}, 1\right), \quad g = 1, \ldots, G, \quad j = 1, \ldots, T-1 \\
&\alpha|\sigma_u^2 \sim \mathcal{N}_P(0, \sigma_\alpha^2 A), \quad \sigma_\alpha^2 \sim IG\left(\frac{1}{2}, \frac{1}{2}\right) \\
&\sigma^2|a \sim IG\left(\frac{1}{2}, \frac{1}{a}\right), \quad a \sim IG\left(\frac{1}{2}, 1\right)
\end{aligned}
$$

where $\boldsymbol{\Upsilon}_g = \text{diag}(\gamma_{g_1}^2, \ldots, \gamma_{g_T}^2)$, $\boldsymbol{\Omega}_g = \text{diag}(\omega_{g_1}^2, \ldots, \omega_{g_{T-1}}^2)$ and $\boldsymbol{D}_g$ is the known $T \times (T-1)$-matrix associated with the finite difference operator of order 1.

The corresponding full conditional distributions for the model parameters are given by:

$$\mu|. \sim \mathcal{N}\left(\frac{1}{n}\mathbb{1}^\top(\boldsymbol{y} - \sum_{g=1}^{G}\boldsymbol{X}_g\boldsymbol{\beta}_g), \frac{\sigma^2}{n}\right)$$

$$\boldsymbol{\beta}_g|. \sim \mathcal{N}_T\left(\boldsymbol{\Sigma}_{b_g}\frac{\boldsymbol{X}_g^\top}{\sigma^2}(\boldsymbol{y} - \mu\mathbb{1} - \sum_{\tilde{g}\neq g}\boldsymbol{X}_{\tilde{g}}\boldsymbol{\beta}_{\tilde{g}} - \mathbf{Z}\alpha), \boldsymbol{\Sigma}_{b_g} = \sigma^2\left(\boldsymbol{X}_g^\top\boldsymbol{X}_g + \boldsymbol{\Upsilon}_g^{-1} + \frac{\boldsymbol{D}_g^\top\boldsymbol{\Omega}_g^{-1}\boldsymbol{D}_g}{\lambda^2}\right)^{-1}\right)$$

$$\lambda^2|. \sim IG\left(\frac{1+T}{2}, \frac{1}{\xi} + \sum_{g=1}^{G}\frac{\boldsymbol{\beta}_g^\top\boldsymbol{D}_g^\top\boldsymbol{\Omega}_g^{-1}\boldsymbol{D}_g\boldsymbol{\beta}_g}{2\sigma^2}\right), \quad \xi|. \sim IG(1, 1 + 1/\lambda^2)$$

$$\omega_{g_j}^2|. \sim IG\left(1, \frac{1}{\phi_{g_j}} + \frac{((\boldsymbol{D}_g\boldsymbol{\beta}_g)_{[j]})^2}{2\sigma^2\lambda^2}\right), \quad \phi_{g_j}|. \sim IG(1, 1 + 1/\omega_{g_j}^2), \quad g = 1, \ldots, G, \quad j = 1, \ldots, T-1$$

$$\upsilon_{g_t}^2|. \sim IG\left(1, \frac{1}{\eta_{g_t}} + \frac{\beta_{g_t}^\top\beta_{g_t}}{2\sigma^2}\right), \quad \eta_{g_t}|. \sim IG(1, 1 + 1/\upsilon_{g_t}^2), \quad g = 1, \ldots, G, \quad t = 1, \ldots, T$$

$$\alpha|. \sim \mathcal{N}_P\left(\Sigma_\alpha\frac{Z^\top}{\sigma^2}(\boldsymbol{y} - \mu\mathbb{1} - \sum_{g=1}^{G}\boldsymbol{X}_g\boldsymbol{\beta}_g), \Sigma_\alpha = \left(\frac{A^{-1}}{\sigma_\alpha^2} + \frac{Z^\top Z}{\sigma^2}\right)^{-1}\right)$$

$$\sigma_\alpha^2|. \sim IG\left(\frac{1+P}{2}, \frac{1}{2} + \alpha^\top A^{-1}\alpha\right)$$

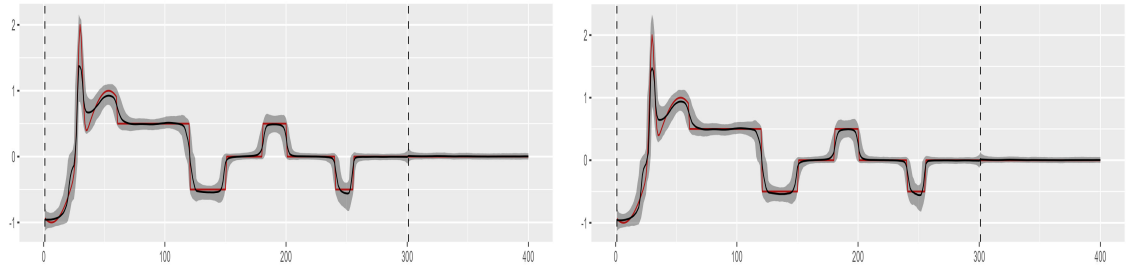
$$\sigma^2|. \sim IG\left(\frac{1+T+n}{2}, \frac{1}{a} + \frac{1}{2}\sum_{g=1}^{G}\beta_g^\top\left(\boldsymbol{\Upsilon}_g^{-1} + \frac{\boldsymbol{D}_g^\top\boldsymbol{\Omega}_g^{-1}\boldsymbol{D}_g}{\lambda^2}\right)\boldsymbol{\beta}_g + \frac{1}{2}||\boldsymbol{y} - \mu\mathbb{1} - \sum_{g=1}^{G}\boldsymbol{X}_g\boldsymbol{\beta}_g - \mathbf{Z}\alpha||_2^2\right)$$

$$a|. \sim IG(1, 1 + 1/\sigma^2)$$

B Simulation results

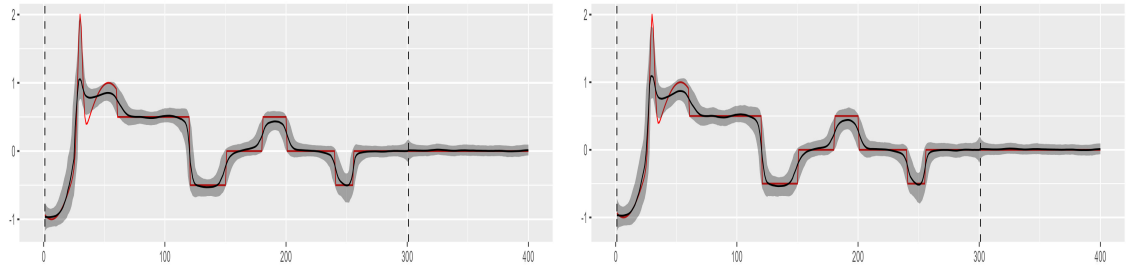
σ^2	G	Priors	MSE_z	MSE_{nz}	MCC	TPR	TPR_{gp}	Counts
1	5	$HS^{\omega_{gt}}_{\lambda} - HS^{\gamma_{gt}}_{1}$	**0.00007**	0.016	0.91	85	100	100
1	5	$HS^{\omega_{gt}}_{\lambda_g} - HS^{\gamma_{gt}}_{1}$	0.00008	**0.015**	0.92	85	100	100
1	5	$HS^{\omega_{gt}}_{\lambda} - NE^{\gamma_{gt}}_{\upsilon_g}$	0.00039	0.026	0.95	91	100	100
1	5	$HS^{\omega_{gt}}_{\lambda_g} - NE^{\gamma_{gt}}_{\upsilon_g}$	0.00039	0.023	0.95	91	100	100
1	5	$NE^{\omega_{gt}}_{\lambda} - NE^{\gamma_{gt}}_{\upsilon_g}$	0.00324	0.046	0.03	0	31	100
1	5	$NE^{\omega_{gt}}_{\lambda_g} - NE^{\gamma_{gt}}_{\upsilon_g}$	0.00169	0.030	0.33	12	100	100
1	5	$HS^{\omega_{gt}}_{\lambda}$	0.00020	0.011	**0.96**	**97**	100	99
1	10	$HS^{\omega_{gt}}_{\lambda} - HS^{\gamma_{gt}}_{1}$	0.00006	0.013	0.92	87	100	100
1	10	$HS^{\omega_{gt}}_{\lambda_g} - HS^{\gamma_{gt}}_{1}$	**0.00005**	0.012	0.93	87	100	100
1	10	$HS^{\omega_{gt}}_{\lambda} - NE^{\gamma_{gt}}_{\upsilon_g}$	0.00019	0.011	**0.96**	**93**	100	100
1	10	$HS^{\omega_{gt}}_{\lambda_g} - NE^{\gamma_{gt}}_{\upsilon_g}$	0.00020	**0.010**	**0.96**	**93**	100	99
1	10	$NE^{\omega_{gt}}_{\lambda} - NE^{\gamma_{gt}}_{\upsilon_g}$	0.00339	0.047	0.00	0	2	100
1	10	$NE^{\omega_{gt}}_{\lambda_g} - NE^{\gamma_{gt}}_{\upsilon_g}$	0.00123	0.025	0.37	16	53	100
1	10	$HS^{\omega_{gt}}_{\lambda}$	0.00023	0.010	0.95	98	100	100
1	30	$HS^{\omega_{gt}}_{\lambda} - HS^{\gamma_{gt}}_{1}$	**0.00003**	0.008	0.96	93	100	100
1	30	$HS^{\omega_{gt}}_{\lambda_g} - HS^{\gamma_{gt}}_{1}$	0.00003	**0.005**	0.97	94	100	100
1	30	$HS^{\omega_{gt}}_{\lambda} - NE^{\gamma_{gt}}_{\upsilon_g}$	0.00011	0.009	**0.98**	97	100	89
1	30	$NE^{\omega_{gt}}_{\lambda} - NE^{\gamma_{gt}}_{\upsilon_g}$	0.00332	0.051	0.00	0	1	73
1	30	$NE^{\omega_{gt}}_{\lambda_g} - NE^{\gamma_{gt}}_{\upsilon_g}$	0.00097	0.022	0.44	22	47	14
1	30	$HS^{\omega_{gt}}_{\lambda}$	0.00035	0.013	0.88	**98**	100	100
1	100	$HS^{\omega_{gt}}_{\lambda} - HS^{\gamma_{gt}}_{1}$	**0.00004**	0.026	0.96	**92**	100	100
1	100	$HS^{\omega_{gt}}_{\lambda_g} - HS^{\gamma_{gt}}_{1}$	0.00005	**0.021**	**0.95**	91	100	100
16	5	$HS^{\omega_{gt}}_{\lambda} - HS^{\gamma_{gt}}_{1}$	**0.00014**	0.029	0.89	82	100	100
16	5	$HS^{\omega_{gt}}_{\lambda_g} - HS^{\gamma_{gt}}_{1}$	**0.00014**	0.028	0.90	82	100	100
16	5	$HS^{\omega_{gt}}_{\lambda} - NE^{\gamma_{gt}}_{\upsilon_g}$	0.00055	0.032	**0.94**	90	100	100
16	5	$HS^{\omega_{gt}}_{\lambda_g} - NE^{\gamma_{gt}}_{\upsilon_g}$	0.00054	0.030	**0.94**	90	100	100
16	5	$NE^{\omega_{gt}}_{\lambda} - NE^{\gamma_{gt}}_{\upsilon_g}$	0.00384	0.049	0.03	0	37	100
16	5	$NE^{\omega_{gt}}_{\lambda_g} - NE^{\gamma_{gt}}_{\upsilon_g}$	0.00242	0.035	0.30	10	100	100
16	5	$HS^{\omega_{gt}}_{\lambda}$	0.00058	**0.017**	**0.94**	**96**	100	101
16	10	$HS^{\omega_{gt}}_{\lambda} - HS^{\gamma_{gt}}_{1}$	**0.00012**	0.025	0.91	84	100	100
16	10	$HS^{\omega_{gt}}_{\lambda_g} - HS^{\gamma_{gt}}_{1}$	**0.00012**	0.021	0.91	85	100	100
16	10	$HS^{\omega_{gt}}_{\lambda} - NE^{\gamma_{gt}}_{\upsilon_g}$	0.00040	0.022	**0.95**	92	100	100
16	10	$HS^{\omega_{gt}}_{\lambda_g} - NE^{\gamma_{gt}}_{\upsilon_g}$	0.00039	0.018	**0.95**	92	100	94
16	10	$NE^{\omega_{gt}}_{\lambda} - NE^{\gamma_{gt}}_{\upsilon_g}$	0.00405	0.049	0.01	0	5	100
16	10	$NE^{\omega_{gt}}_{\lambda_g} - NE^{\gamma_{gt}}_{\upsilon_g}$	0.00196	0.031	0.34	13	52	100
16	10	$HS^{\omega_{gt}}_{\lambda}$	0.00059	**0.017**	0.94	**96**	100	100
16	30	$HS^{\omega_{gt}}_{\lambda} - HS^{\gamma_{gt}}_{1}$	**0.00008**	0.022	0.95	91	100	100
16	30	$HS^{\omega_{gt}}_{\lambda_g} - HS^{\gamma_{gt}}_{1}$	**0.00008**	**0.012**	0.95	91	100	100
16	30	$HS^{\omega_{gt}}_{\lambda} - NE^{\gamma_{gt}}_{\upsilon_g}$	0.00036	0.021	**0.97**	96	100	89
16	30	$NE^{\omega_{gt}}_{\lambda} - NE^{\gamma_{gt}}_{\upsilon_g}$	0.00398	0.052	0.00	0	1	75
16	30	$NE^{\omega_{gt}}_{\lambda_g} - NE^{\gamma_{gt}}_{\upsilon_g}$	0.00182	0.031	0.38	16	40	15
16	30	$HS^{\omega_{gt}}_{\lambda}$	0.00115	0.022	0.87	**96**	100	99
16	100	$HS^{\omega_{gt}}_{\lambda} - HS^{\gamma_{gt}}_{1}$	**0.00011**	**0.035**	**0.92**	**85**	100	100
16	100	$HS^{\omega_{gt}}_{\lambda_g} - HS^{\gamma_{gt}}_{1}$	0.00015	0.042	0.90	81	100	100

Table B.4: Mean squared errors of the true zeroes (MSE_z), the true non-zeroes (MSE_{nz}), the Matthews Correlation Coefficient (MCC), the True Positive Rate on all non-zero coefficients (TPR) and on selected groups (TPR_{gp}) using the different priors with residual variance σ^2 equal to 1 or 16, and a number of groups equal to 5, 10, 30 or 100. All the criteria were calculated over the 100 replications and averaged. The last column gives the number of simulations that were reach convergence (Counts).



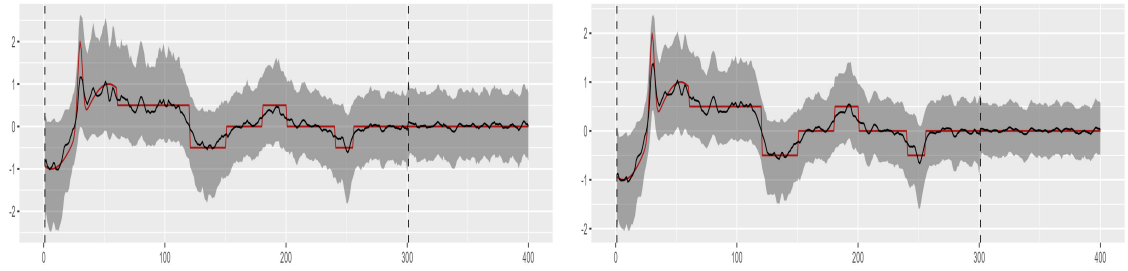
(a) $\text{HS}_{\lambda}^{\omega_{gt}}$ – $\text{HS}_{1}^{\gamma_{gt}}$

(b) $\text{HS}_{\lambda_g}^{\omega_{gt}}$ – $\text{HS}_{1}^{\gamma_{gt}}$



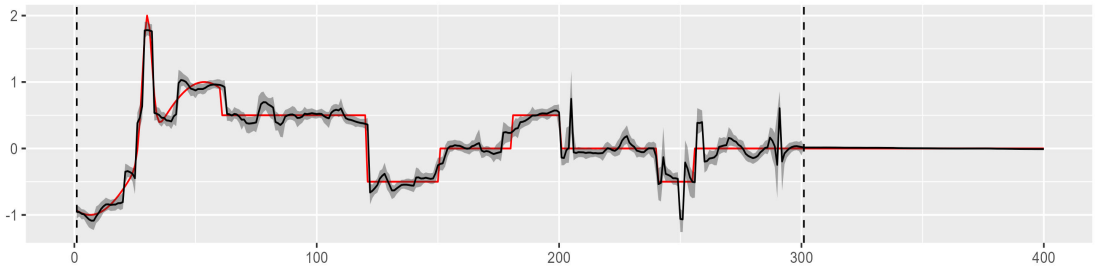
(c) $\text{HS}_{\lambda}^{\omega_{gt}}$ – $\text{NE}_{\upsilon_g}^{\gamma_{gt}}$

(d) $\text{HS}_{\lambda_g}^{\omega_{gt}}$ – $\text{NE}_{\upsilon_g}^{\gamma_{gt}}$



(e) $\text{NE}_{\lambda}^{\omega_{gt}}$ – $\text{NE}_{\upsilon_g}^{\gamma_{gt}}$

(f) $\text{NE}_{\lambda_g}^{\omega_{gt}}$ – $\text{NE}_{\upsilon_g}^{\gamma_{gt}}$



(g) $\text{HS}_{\lambda}^{\omega_{gt}}$

Figure B.1: MCMC chain for the fused priors (Fig. B.1a, B.1b, B.1c, B.1d, B.1e, B.1f) and fusion prior (Fig. B.1g) for one simulation with setting equal to: residual variance fix to one and number of groups set to five. The red line is the true profile of coefficients. The black line is the estimated profile and gray shadow is the credible interval at 95%. The vertical dashed line delimits the groups. Only the first 400 coefficients are represented for more visibility.