Bayesian group fused priors

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

Bayesian group fused priors

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 abscission process is strongly involved in a series of physiological events; an optimal execution of this process is of major importance for species survival. Environmental variation impacts species development and abscission process with varying effects over developmental stages. The identification of the environmental factors as well as the time periods at which they modulate the abscission process is crucial to deal with climate changes. Considering environmental variables as time series, i.e. groups of correlated variables, poses a statistical challenge in selecting relevant groups and temporally correlated variables within them. In this study, we address these objectives by introducing and discussing four Bayesian group

fused priors through a general parametrization. In particular, we highlighted that the horseshoe prior on differences with a unique global parameter over groups combined with a normal half-Cauchy distribution on coefficients outperformed the extension of the usual fused priors, which consists of assuming Laplace distributions on coefficients and their differences. The fruit abscission of oil palm trees motivated this development. This application, based on an impressive experimental design in Benin Republic, illustrated performances of the proposed prior to select environmental variables as well as successive past environmental variations involved in the timing of bunch harvesting.

Keywords: Bayesian variable selection, Fusion and Fused priors, Horseshoe prior, Structured variables.

1 Introduction

Understanding the impact of environmental variables on development and adaptation processes is crucial in facing to climate changes. Abscission consists of the shedding of various parts of organisms, such as leaves during autumn or flowers after fertilization. It is one of the most important adaptation process. This biological mechanism is highly sensitive to climate conditions and to their variations over the growing seasons and years. The abscission process could be illustrated by the well-known leaf senescence and fall of deciduous trees, which was delayed in response to an increase in temperature between 1931 and 2010 in the northern hemisphere (Gill et al., 2015). Environmental stress may severely impact abscission processes due to complex regulations involving exogenous and endogenous signals (Sawicki et al., 2015). For instance, drought stress can induce activation and premature flower abscission in lupine (Wilmowicz et al., 2021) or tomato plants (Reichardt et al., 2020) and so

negatively impact crop productivity.

In many contexts, while it is clear that environmental variables have consequences on organ losses, it is not yet clear which one, either exogenous (climate, soil,...) or endogenous (development, carbon status,...), is responsible for the responses observed and at which stages of the organ development or the abscission process the regulation occurs. For the oil palm trees, the abscission time is critical because fruit bunches are harvested when the first fruits detach and fall to the ground. A premature abscission of fruits can lower the oil yield if the optimal maturity is not reached, while too much abscission leads to extra work in collecting detached fruits on the ground. A recent study has shown that environmental variables, such as temperature or solar radiation, alter the oil palm tree reproductive development by modulating the timing of fruit drop (Tisné et al., 2020). In this paper, we aim to identify over the environment experienced by the fruit bunch the relevant environmental variables and the periods at which they have an effect in the phenotypic variations of fruit abscission.

Considering environmental variables as time series, i.e groups of temporally correlated variables, raises at least two challenges both related to model regularization and variable selection. The first one is the selection of groups (environmental variables). The second one is the selection of correlated variables within groups (time periods). Natural ordering of variables within groups can lead to potentially high correlation between consecutive variables. 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.

Considerable attention has been paid in the last decades to variable and group selection. Developed methods 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 yet the Elastic-Net (Zou and Hastie, 2005) are classically used. Note that Elastic-Net is well adapted when 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. We may cite among others, the spike-and-slab prior (Mitchell and Beauchamp, 1988; George and McCulloch, 1993, 1997), 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). Nevertheless, these methods do not take into account a potential group structure within covariates. Lasso extensions to group selection have been 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 such a prior 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.

The above methods do not allow serial correlations between successive variables within groups to be taken into account. These dependencies may lead to identifiability problems impacting the estimation task which aims to assign similar effects for 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) introduce the fusion and fused Lasso. The fusion Lasso penalizes the ℓ_1-norm of successive differences of 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, various studies pointed out that the Bayesian Lasso prior, which uses Laplace distribution on coefficients, does not shrink enough each coefficient or differences towards zero, leading to biased (Carvalho et al., 2010; Polson and Scott, 2011) 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 Faulkner and Minin (2018) and Kakikawa et al. (2023) placed a HS prior on differences. Note that, all of these approaches place a Laplace distribution on regression coefficients. These methodologies show good properties in terms of prediction accuracy but also to estimate smooth functions with potentially abrupt changes. However, they have been designed for only one group. A direct extension of Bayesian fused Lasso to multi-group context has been proposed by Alaíz et al. (2013). Zhang et al. (2014) used this multi-group version for the slab part of a group spike-and-slab prior. Theses methods can suffer from low shrinkage properties of the Bayesian Lasso, leading to poor estimations when the number of covariates within groups is large.

In this paper, we propose via a thorough simulation study, to investigate the trade-off between strong shrinkage prior on coefficients and their differences. Results evidence the interest of considering distributions on coefficients with heavier tails than the usual Laplace distribution resulting to propose the horsehoe normal half-Cauchy (HS-NhC) fused prior. Two extensions of this prior to the multi-group context assuming either global shrinkage parameters at the group level or one global shrinkage parameter common to all groups have been developed. These priors are compared to multi-group extensions of priors proposed by

Faulkner and Minin (2018) and Kakikawa et al. (2023). This paper is organized as follows. Section 2 is dedicated to the construction of Bayesian group fused priors in linear regression context. Using simulated data, section 3 is devoted to compare and evaluate the efficiency of the proposed priors according to the number of groups, their size, and the signal-to-noise ratio. Section 4 aims at identifying environmental variables and time periods affecting the oil palm fruit abscission process.

2 Model

2.1 Notation and model

Let $\mathbf{y} = (y_1, \ldots, y_n)'$ be a n-continuous response vector and $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_G]$ a $(n \times GT)$-matrix concatenating G known groups of covariates measured at T regular spaced times. For all $g = 1, \ldots, G$, $\mathbf{X}_g = \left[\mathbf{x}'_{g1}, \ldots, \mathbf{x}'_{gT}\right]$ denotes a $(n \times T)$-matrix concatenating vectors $\mathbf{x}_{gt} = (x_{1gt}, \ldots, x_{ngt})$ for $t = 1, \ldots, T$. In this paper, $\mathbf{y}$ corresponds to the abscission time measured on $n = 1,173$ bunches from $l = 140$ oil palm trees. Each $\mathbf{X}_g$ describes environmental variable. As many bunches are from the same palm tree, we use a linear mixed model such that:

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

where μ is an intercept, $\mathbb{1}$ a n-vector of 1, $\boldsymbol{\beta}_g$ a T-vector of regression coefficients associated to group g, and $\alpha = (\alpha_1, \ldots, \alpha_l)'$ a l-vector of random effects assumed Gaussian distributed with zero expectation and variance equal to σ^2_α. This random effect allows to take into account the dependence between observations done on a same oil palm tree. Finally, $\boldsymbol{\varepsilon}$ is a n-vector of independent Gaussian residuals with zero mean and variance equal to $\sigma^2 Id_n$ independent from α.

2.2 Prior construction

In Bayesian framework and when $G = 1$, a usual approach to take into account time structure within covariate matrix and to impose sparsity on coefficients relies on the use of the Bayesian fused prior. This prior consists in placing independent shrinkage priors on both regression coefficients and their successive differences. A general formulation is given by:

$$\prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\upsilon^2\gamma_t^2\sigma^2}} \exp\left(-\frac{\beta_t^2}{2\upsilon^2\gamma_t^2\sigma^2}\right) \quad \text{and} \quad \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\lambda^2\omega_t^2\sigma^2}} \exp\left(-\frac{(\beta_t-\beta_{t-1})^2}{2\lambda^2\omega_t^2\sigma^2}\right). \tag{2}$$

This formulation considers a global-local parametrization with local shrinkage parameters γ_t and ω_t as well as global shrinkage parameters, υ and λ, respectively specific to coefficients and their differences. Global parameters perform shrinkage on all coefficients and their differences, whereas local parameters allow true large effects to escape from overall shrinkage. Fused-type priors are originally based on the use of the Laplace distribution, also called normal-Exponential (NE) distribution, for regression parameters and their differences. However, Laplace prior faces posterior inconsistency notably due to its exponentially light tail (Castillo et al., 2015). To overcome such drawbacks, alternative priors with heavier tails, such as t-Student, NEG distributions or yet HS priors on differences combined with a Laplace distribution on coefficients have been proposed (see Table 1). The use of Laplace distribution for coefficients has not yet been questioned. In this paper, according to investigations using simulations, we propose a normal half-Cauchy (NhC) distribution on regression coefficients with a horseshoe distribution on their differences as an alternative steady prior face to dimension complexity and signal-to-noise ratio. This prior, denoted HS-NhC prior, is defined as follows:

$$\prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\gamma_t^2\sigma^2}} \exp\left(-\frac{\beta_t^2}{2\gamma_t^2\sigma^2}\right) \quad \text{and} \quad \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\lambda^2\omega_t^2\sigma^2}} \exp\left(-\frac{(\beta_t-\beta_{t-1})^2}{2\lambda^2\omega_t^2\sigma^2}\right), \tag{3}$$

where γ_t $(t = 1, \ldots, T)$, λ, and ω_t $(t = 2, \ldots, T)$ follow a half-Cauchy distribution.

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

Table 1: Fused priors in the one group context ($G = 1$).

A natural extension of the proposed prior to the multi-group context consists in assuming that parameters controlling sparsity on coefficients and differences are group specific, such that:

$$\prod_{t=1}^{T_g} \frac{1}{\sqrt{2\pi\gamma_{gt}^2\sigma^2}} \exp\left(-\frac{\beta_{gt}^2}{2\gamma_{gt}^2\sigma^2}\right) \quad \text{and} \quad \prod_{t=2}^{T_g} \frac{1}{\sqrt{2\pi\lambda_g^2\omega_{gt}^2\sigma^2}} \exp\left(-\frac{(\beta_{gt}-\beta_{gt-1})^2}{2\lambda_g^2\omega_{gt}^2\sigma^2}\right). \tag{4}$$

This prior (see Eq. 4) relies on a large set of parameters and it is well-know that the inference of the global parameters ($\lambda_g,\ g = 1, \ldots, G$) is complex and can lead to poor results in terms of selection (Piironen et al., 2017). In the multi-group context, the number of groups as well as their size may reinforce such difficulties. We therefore suggest an alternative parametrization assuming one global parameter ($\lambda_g = \lambda$) to control shrinkage over all groups

while keeping local parameters ω_{gt}. This prior is defined by:

$$\prod_{t=1}^{T_g} \frac{1}{\sqrt{2\pi\gamma_{gt}^2\sigma^2}} \exp\left(-\frac{\beta_{gt}^2}{2\gamma_{gt}^2\sigma^2}\right) \quad \text{and} \quad \prod_{t=2}^{T_g} \frac{1}{\sqrt{2\pi\lambda^2\omega_{gt}^2\sigma^2}} \exp\left(-\frac{(\beta_{gt}-\beta_{g(t-1)})^2}{2\lambda^2\omega_{gt}^2\sigma^2}\right). \tag{5}$$

In the following, priors assuming group specific shrinkage parameters λ_g will be referred to specific priors, while priors assuming $\lambda_g = \lambda$ will be denoted global priors (see Table 2).

2.3 MCMC implementation

Bayesian inference of the proposed models is achieved using Markov chain Monte Carlo (MCMC) algorithm sampling. As the full conditional distributions of each parameter have a closed form, an efficient Gibbs sampler algorithm (Gilks et al., 1995) is used.

Following Kyung et al. (2010), combining independant priors (Eq. 5) on regression coefficients and their differences leads to a synthetic multivariate expression:

$$\boldsymbol{\beta}_g|\boldsymbol{\gamma}_g, \lambda^2, \boldsymbol{\omega}_g, \sigma^2 \sim \mathcal{N}_T(0, \sigma^2\boldsymbol{Q}_g^{-1}), \quad g = 1, \ldots, G \tag{6}$$

where Q_g is equal to

$$\boldsymbol{Q}_g = \left(\boldsymbol{\Upsilon}_g^{-1} + \boldsymbol{D}_g^\top \boldsymbol{\Omega}_g^{-1} \boldsymbol{D}_g/\lambda^2\right). \tag{7}$$

The first matrix, $\boldsymbol{\Upsilon}_g^{-1}$, refers to regression parameters, and the second, $\boldsymbol{D}_g^\top \boldsymbol{\Omega}_g^{-1} \boldsymbol{D}_g/\lambda^2$ to the differences. $\boldsymbol{D}_g$ is the known $T \times (T-1)$-matrix associated to the finite differences operator of order 1, and $\boldsymbol{\Omega}_g = diag(\omega_{g1}^2, \ldots, \omega_{gT-1}^2)$ and $\boldsymbol{\Upsilon}_g = diag(\gamma_{g1}, \ldots, \gamma_{gT_g})$ the $(T-1) \times (T-1)$-diagonal matrices of local parameters. According to this multivariate formulation and conjugacy properties, the posterior distribution of regression parameters $\boldsymbol{\beta}$, is a multivariate Gaussian distribution (see Appendix A).

Shrinkage parameters follow half-Cauchy distribution, which may be rewritten as a scale

mixture of inverse-Gamma distributions as introduced by Makalic and Schmidt (2015):

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

This simplifies the computation of full conditional distributions. Thus, for instance for the global shrinkage parameter λ^2 defined in Eq. (5), its full conditional distribution is given by :

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

The remaining full conditional distributions along with the hierarchical representation of the proposed models are provided in Appendix. Code to implement the proposed models is available in the R language (R Core Team, 2023) on GitHub: `https://github.com/Heuclin/GroupFusedHorseshoe`.

Prior names	**Difference prior**	**Coefficient prior**
Specific HS-NE	$\lambda_g \sim \mathcal{C}^+(0,1)$ $\omega_{gt} \sim \mathcal{C}^+(0,1)$	$\upsilon_g^2 \sim \mathcal{IG}(s,r)$ $\gamma_{gt}^2 \sim \mathcal{E}xp(1/2)$
Global HS-NE	$\lambda_g = \lambda$ $\lambda \sim \mathcal{C}^+(0,1)$ $\omega_{gt} \sim \mathcal{C}^+(0,1)$	 $\upsilon_g^2 \sim \mathcal{IG}(s,r)$ $\gamma_{gt}^2 \sim \mathcal{E}xp(1/2)$
Specific HS-NhC	$\lambda_g \sim \mathcal{C}^+(0,1)$ $\omega_{gt} \sim \mathcal{C}^+(0,1)$	 $\gamma_{gt} \sim \mathcal{C}^+(0,1)$
Global HS-NhC	$\lambda_g = \lambda$ $\lambda \sim \mathcal{C}^+(0,1)$ $\omega_{gt} \sim \mathcal{C}^+(0,1)$	 $\gamma_{gt} \sim \mathcal{C}^+(0,1)$

Table 2: Fused priors in the multi-group context.

3 Simulation study

This section aims to demonstrate performances of the HS-NhC group fused priors, and in particular its global version ($\lambda_g = \lambda$), in terms of shrinkage properties, parameter estimation and algorithmic stability according to the number of groups, their size, and the signal-to-

noise ratio. We compare results obtained using HS-NhC group fused priors with the three alternative priors: the global and specific HS-NE and the global HS-HS. In this simulation study, $p = 1500$ covariates were generated. We assumed that p was divided into $G = 1, 10, 30$ or 100 groups. Covariates within each group were generated from a $\frac{p}{G}$-multivariate Gaussian distribution with zero mean and a covariance matrix defined by a first-order autoregressive (AR1) structure with a parameter fixed to 0.95. Functional effects were defined as the combination of different smooth functions: a continuous smooth function as proposed by Faulkner and Minin (2018) and piece-wise functions (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 $n = 150$ observations were sampled with residual standard deviations equal to 1 or 4. We assessed the relative performances of *priors* using mean squared errors either of the non-zero coefficients or only of the true zeroes, and the Matthews Correlation Coefficient (MCC) to summarize the selection (shrinkage) property of the different priors (Matthews, 1975). By denoting $\mathcal{C}_{nz} = \{t : \beta_t \neq 0\}$ and $\mathcal{C}_z = \{t : \beta_t = 0\}$ the sets of indices of non-zero and true zero coefficients respectively, the performances are

calculated through:

$$MSE_{nz} = \frac{1}{|\mathcal{C}_{nz}|} \sum_{t \in \mathcal{C}_{nz}} (\beta_t - \hat{\beta}_t)^2; \; MSE_z = \frac{1}{|\mathcal{C}_z|} \sum_{t \in \mathcal{C}_z} (\beta_t - \hat{\beta}_t)^2; \text{ and}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

where TP, TN, FP, FN correspond to True (T) and False (F) negatives (N) and positives (P), respectively. Each criterion was averaged over 30 repetitions.

In the one group context ($G = 1$), results showed that HS-NhC and HS-NE had very close performances whatever the signal-to-noise ratio and clearly outperformed the usual NE-NE prior (see Table 3). We also observed that shrinkage properties have been reinforced using NhC on coefficients (see MSE_z column in Table 3). These results tended to confirm that using priors with heavier tails compared to Laplace distribution on coefficients could improve results. However assuming a horseshoe prior on coefficients leads to erroneous results by shrinking towards zero all the coefficients.

Priors	σ^2	MCC	MSE_z	MSE_{nz}
HS_NhC	1	0.90688	0.00003	0.01485
HS_NE	1	0.90018	0.00067	0.05000
HS_HS	1	0.06456	0.00000	2.33525
NE_NE	1	0.21203	0.00273	0.04602
HS_NhC	16	0.81803	0.00012	0.05151
HS_NE	16	0.88696	0.00091	0.05639
HS_HS	16	0.06620	0.00000	2.45363
NE_NE	16	0.13722	0.00517	0.05561

Table 3: Matthews Correlation Coefficient (MCC), mean squared errors of the true zeroes (MSE_z), and mean squared errors of the non-zero coefficients MSE_{nz} using the different priors with residual variance σ^2 equal to 1 or 16, and $G = 1$.

In the multi-group context, the HS-NhC priors outperformed the HS-NE priors whatsoever the global or specific versions (see Figure 1). Moreover, we observed that global and specific HS-NhC versions performed similarly well in terms of selection (see Figure 1-a) and

estimation of non-zero coefficients (see Figure 1-b). We also noted that results obtained using global and specific HS-NE versions presented close patterns. These results highlighted that the simplified versions assuming a global shrinkage parameter $\lambda_g = \lambda$ is efficient while being more parsimonious. However, when looking at the convergence evaluation, we observed that the specific HS-NE prior could fail when the number of groups increased, which in part, can be explained by the reduction of the number of observations within groups (see Table B1 in Appendix B). Those convergence failures had never been observed for the specific HS-NhC prior, however we noted that a large number of groups may slightly impacted HS-NhC priors. For instance, when the number of groups was set to 100 leading to a group size equal to 15 and a residual variance set to 1, MCC values were equal to 0.959 and 0.951 for global and specific HS-NhC priors, respectively. This can be explained by the difficulties in estimating global shrinkage hyperparameters at the group level with few measurements as already been noted in the literature (Piironen et al., 2017). A higher signal-to-noise ratio slightly impacted the results, for example the MCC values were equal to 0.92 for $\sigma^2 = 1$ and to 0.89 for $\sigma^2 = 16$. Finally, to evaluate the use of prior on coefficients, we compared the global HS-NhC fused prior with a global fusion prior. While results were close for moderate number of groups ($G = 5$ or 10), all criteria were highly impacted when the number of groups was greater than 30. Even for $G = 100$ the fusion did not converge, demonstrating therefore the importance of priors on coefficients for numerical regularization.

To sum up, the global HS-NhC prior appeared highly stable and insensitive to the number of groups, the magnitude of signal-to-noise ratio, and is clearly more parsimonious than its specific version. Moreover, the use of the NhC distribution on coefficients compared to the Laplace distribution do not require the specification of hyperparameters, as it is the case in Inverse Gamma distribution. In the following, all results obtained on the real dataset are thus based on the global HS-NhC prior.

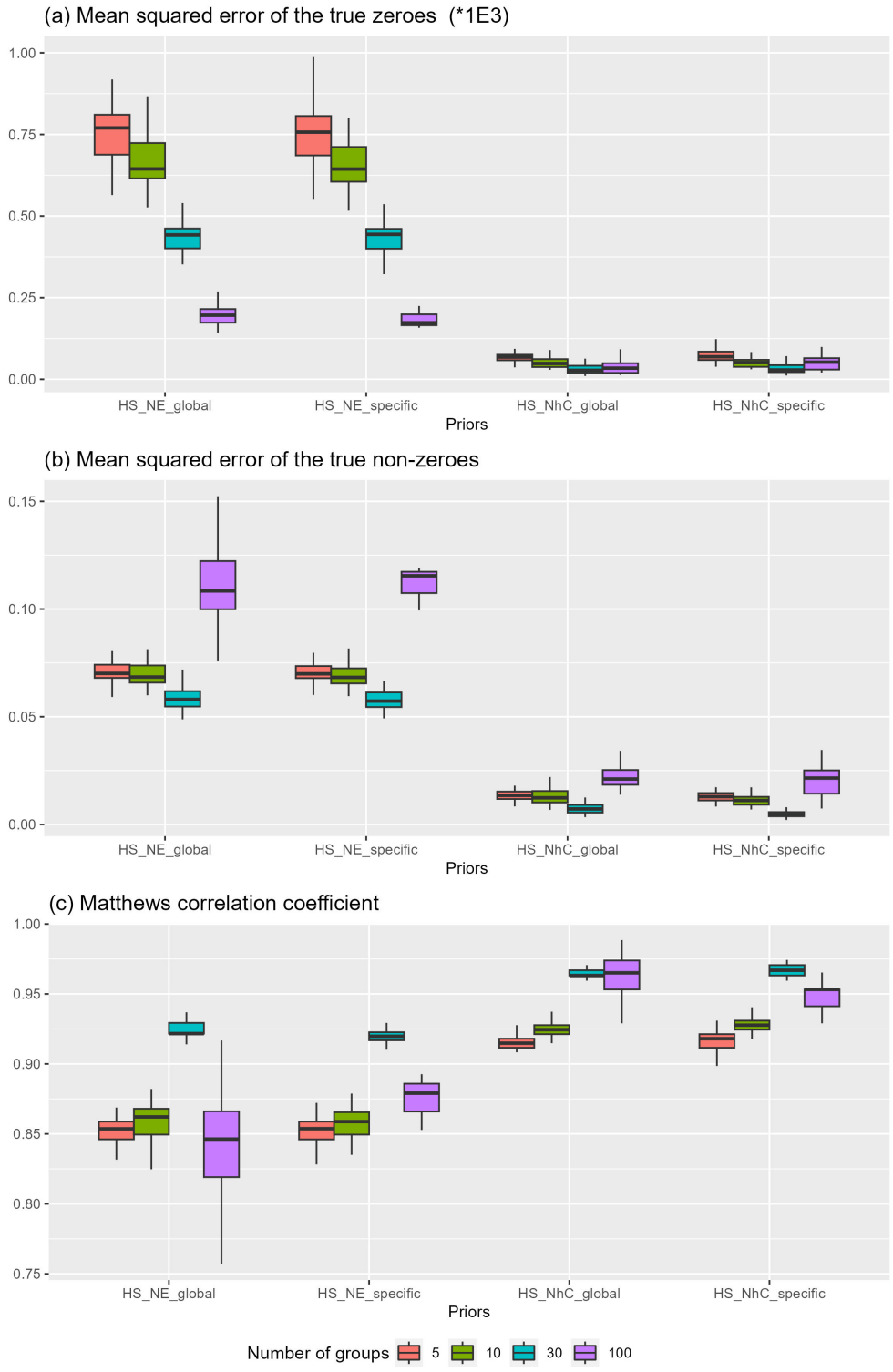

Figure 1: (a) Mean squared errors of the true zero coefficients MSE_z, (b) Mean squared errors of the non-zeroes (MSE_{nz}), and (c) Matthews Correlation Coefficient (MCC) using the different priors with residual variance σ^2 equal to 1, and $G = 5, 10, 30, 100$.

4 The abscission dataset

This application aims at identifying environmental variables and time periods affecting the oil palm fruit abscission process (Tisné et al., 2020). The dataset is 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 all the experiment time from 2014 until 2018. The manual pollination date, different for each bunch, was recorded and the bunch was monitored up to its harvest. A total of $1,173$ bunches were considered over multiple years, taking 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 time indicator and its variation integrates different underlying abscission processes at different developmental stages.

Additionally, nine environmental variables were used. Five climatic variables were recorded from 2014 until 2018: the maximum and minimum temperature (Tmax, Tmin, in °C), the relative air humidity (RH, in %), the rainfall (R, in mm) and the solar radiation (SR, in $\text{cal.cm}^{-2}\text{.d}^{-1}$). Four ecophysiological variables were calculated using climate and individual production data: two exogenous variables including the maximum daily vapor pressure deficit (VPD), the fraction of transpirable soil water (FTSW), and two endogenous (trophic) variables: the supply–demand ratio (SD) and the daily reproductive demand (DRD) (see Tisné et al. (2020) for further details of the calculations). These variables can have ponctual or cumulative effects, depending on the biological process or the developmental stage. Temperature can have ponctual effects as the arrest of growth at low temperatures, but also cumulative effects on developmental rates that led to the thermal time development.

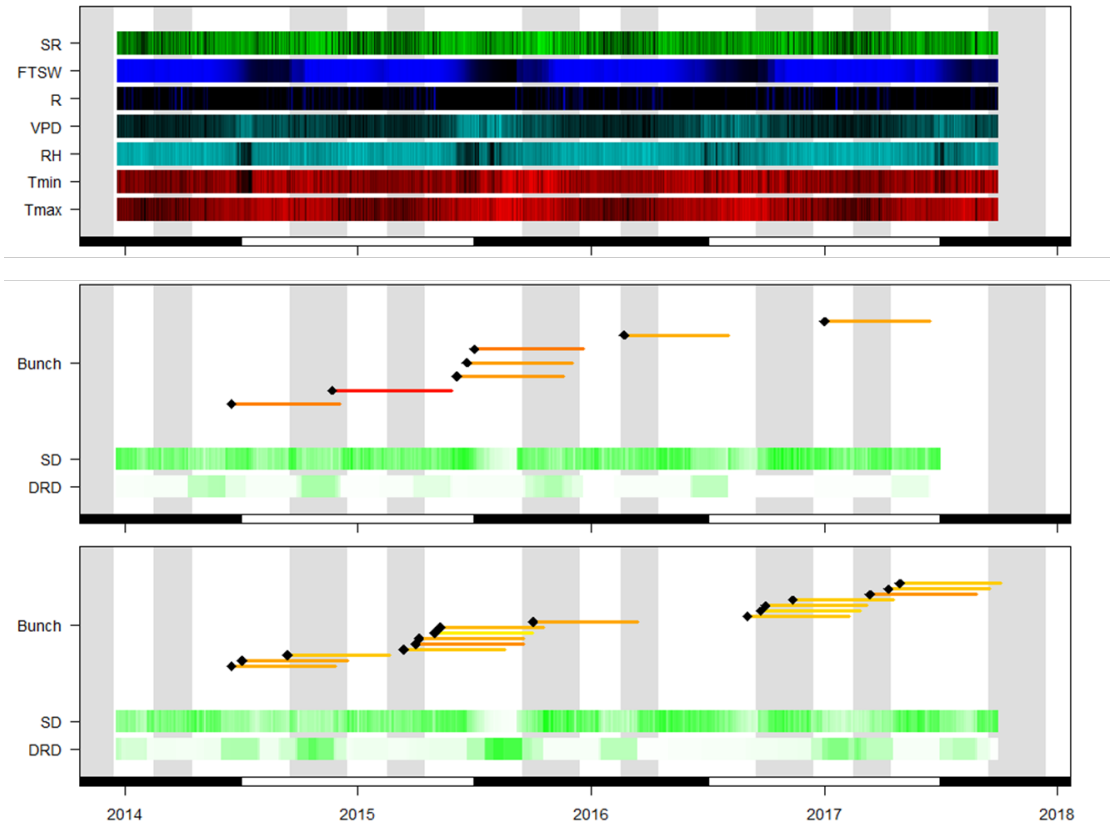

Figure 2: Raw and calculated environmental variables (top panel) and example of bunch production for two oil palm individuals along with their endogenous variables (middle and bottom panels). Raw and calculated variables (heatmaps) as well as bunches (horizontal segments) are plotted over the experiment duration, from 2014 to 2018. The segments for bunches are represented from the manual pollination (black diamong) to the harvest, the yellow to red color indicating increasing DFD. 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.

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 thus leads to nine groups of covariates ($G = 9$) measured at $T = 121$ 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. All matrices have been scaled to obtain a similar order of magnitude. All 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 thining of 10. A group is considered selected if at least one regression effect within it has a credible interval that does not contain zero.

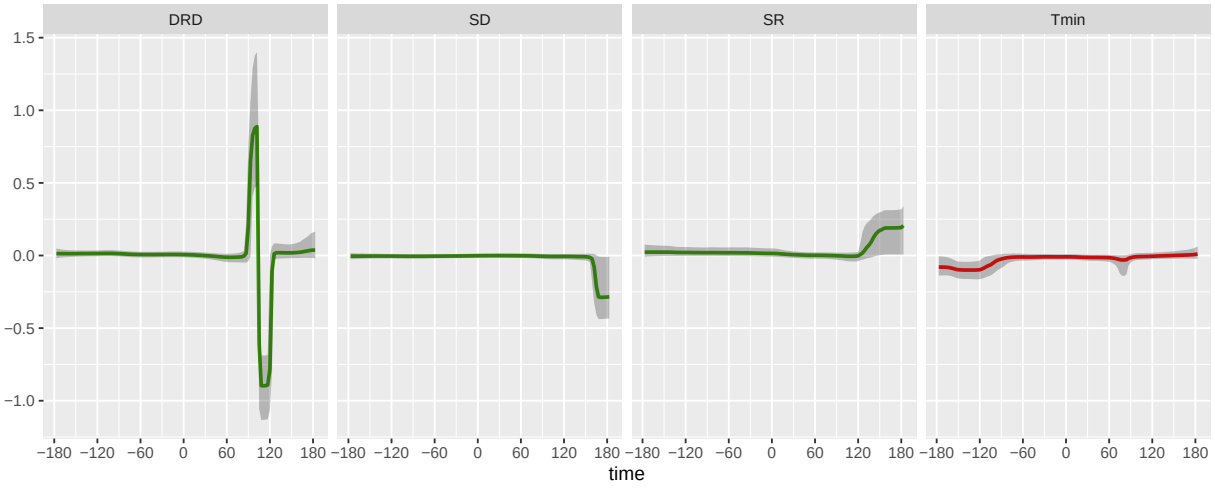

Figure 3: Non-zero coefficient profile estimation provided by the global HS-NhC prior on the abscission dataset. Gray shadows represent the 95% credible interval. Colors represent the different categories of environmental variables, green is for photosynthesis variables (DRD, SR, SD) and red is for temperature variable.

Comparison with previous studies and biological interpretation

Estimated coefficient profiles provided by the the global HS-NhC prior are very clear

and allow identification of relevant time periods of four variables (Tmin, SR, DRD, and SD). Two types of patterns are observed: with smooth effects for Tmin and SR and with punctual effects for DRD and SD. The Tmin variable is negatively associated with DFD during the inflorescence development 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 the 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 observed in Tisné et al. (2020), is thus confirmed 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 the considered fruit bunch integrates current and future whole plant photosynthate demand due to concomitant developing bunches, to modulate its maturation and abscission timing. Such carbohydrate-based regulation is commonly found in fruit tree species and leads to the wave of abscission concerning fruitlets (Sawicki et al., 2015), the only difference being that the oil palm regulates ripe fruit abscission timing instead of dropping unripe fruits. In contrast with DRD and SD that have similar punctual patterns with those of Tisné et al. (2020) study, Tmin effect profile is different, showing a continuous moderate effect instead of many weak effects spread over the −180 to −100 period. The SR factor was not selected by Tisné et al. (2020) but it has a positive effect from day 120 to 180 using our prior. These discrepancies may be due to 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 related to thermal time which is known to be associated with developmental rates. In the period identified, the differentiation of floral

organs occurs (Adam et al., 2011) and variation in cumulative thermal time could modulate the developmental program and ultimately the fruit drop timing. Concerning the cumulative effect of the radiation, it was identified all over the final stage before fruit drop that corresponds to the fruit maturation with intensive lipid accumulation, which is highly related to photosynthate availability (Tranbarger et al., 2011). Our proposed prior, which has been designed to estimate smooth and flexible coefficient profile is then well suited to study the effect of cumulative effect variables in addition to the punctual effect variables that were identified consistently between both approaches.

5 Conclusion

We proposed four Bayesian fused priors in the multi-group context (see Table 2). We showed that the combination of a normal half-Cauchy on coefficients and a horseshoe on their differences is more efficient and stable than the natural extension of the HS-NE fused prior whatever the number of groups ($G \geq 1$), the size of groups, and the signal-to-noise ratio. The proposed general formulation (see Eq. 2) encompasses most priors already developed in the literature as shown in Table 1. Through simulations, we evidenced the importance of placing heavy-tailed distributions with a spike at zero on the differences (Kakikawa et al., 2023) and considering a distribution with a heavier tail property than the usual Laplace prior on coefficients. However, we showed that using a horseshoe prior on both coefficients and their differences resulted in poor performances, as it tended to shrink all parameters towards zero. We noted the advantage of fused-type priors on fusion-type priors mainly when the number of groups was large with a moderate to small size.

From a biological point of view, the proposed prior clearly identifies four environmental variables as well as periods at which they affect the oil palm abscission process. By providing

flexibility in the estimation of regression coefficient profiles, we identify one supplementary environmental variable than the previous study, and improve the interpretability of the regression profiles. Moreover, by giving high predictive performances, the proposed prior may be a useful tool to assist biologists in identifying the best time to harvest the bunches.

The global HS-NhC prior may be directly applied to a broad type of applications such as in the near infrared spectroscopy context, which involves one group of ordered variables through a spectrum, or in the genetic mapping context, where markers may be viewed as groups of ordered variables at the chromosome level. To take into account 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 conflicts 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.

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.

George, E. I. and McCulloch, R. E. (1997). Approaches for bayesian variable selection. *Statistica Sinica*, 7(2):339–373.

Gilks, W., Richardson, S., and Spiegelhalter, D. (1995). *Markov Chain Monte Carlo in Practice*. CRC press.

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.

Mitchell, T. J. and Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the american statistical association*, 83(404):1023–1032.

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.

Appendix

A. Bayesian hierarchical models and full conditional distributions

A1. The group fused HS-NhC global.

The Bayesian hierarchical model used for the MCMC implementation of the group fused HS-NhC global is given by:

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

where $\boldsymbol{\Upsilon}_g = \text{diag}(\upsilon^2_{g_1},\ldots,\upsilon^2_{g_T})$ and $\boldsymbol{\Omega}_g = \text{diag}(\omega^2_{g_1},\ldots,\omega^2_{g_{T-1}})$.

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

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

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

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

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

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

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

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

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

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

A2. The group fused HS-NhC global.

The Bayesian hierarchical model for the MCMC implementation of the group fused HS-NhC specific is given by:

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

where $\boldsymbol{\Upsilon}_g = \text{diag}(\upsilon^2_{g_1}, \ldots, \upsilon^2_{g_T})$ and $\boldsymbol{\Omega}_g = \text{diag}(\omega^2_{g_1}, \ldots, \omega^2_{g_{T-1}})$. The corresponding full conditional distributions for the model parameters are given by:

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

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

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

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

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

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

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

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

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

B. Convergence evaluation.

σ^2	g	HS-NE global	HS-NE specific	HS-NhC global	HS-NhC specific
1	5	30	30	30	30
1	10	30	30	30	30
1	30	30	20	30	30
1	100	26	3	30	30
16	5	30	30	30	30
16	10	30	30	30	30
16	30	30	17	30	30
16	100	26	6	30	30

Table B1: Number of Monte Carlo Markov chains that have converged over 30 replicates in the multi-group context for each scenario.