



**HAL**  
open science

## Bayesian group fused priors

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

► **To cite this version:**

Benjamin Heuclin, Frédéric Mortier, Sébastien Tisné, Julien Gibaud, Catherine Trottier, et al..  
Bayesian group fused priors. 2024. hal-04486172

**HAL Id: hal-04486172**

**<https://hal.science/hal-04486172>**

Preprint submitted on 1 Mar 2024

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

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

# Bayesian group fused priors

B. Heuclin<sup>1,2</sup>, F. Mortier<sup>4,5</sup>, S. Tisné<sup>2,3</sup>, J. Gibaud<sup>1</sup>, C. Trottier<sup>1,6</sup>,  
and M. Denis<sup>2,3</sup>

<sup>1</sup> IMAG, Univ Montpellier, CNRS, Montpellier, France,

<sup>2</sup> CIRAD, UMR AGAP Institut, F-34398 Montpellier, France

<sup>3</sup> UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, F-34398 Montpellier, France

<sup>4</sup> Forêts et Sociétés, Cirad, F-34398 Montpellier, France,

<sup>5</sup> Forêts et Sociétés, Univ Montpellier, Cirad, Montpellier, France,

<sup>6</sup> 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

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

21

22 **Keywords:** Bayesian variable selection, Fusion and Fused priors, Horseshoe prior, Struc-  
23 tured variables.

24

## 25 **1 Introduction**

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

37 negatively impact crop productivity.

38

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

51 Considering environmental variables as time series, i.e groups of temporally correlated  
52 variables, raises at least two challenges both related to model regularization and variable  
53 selection. The first one is the selection of groups (environmental variables). The second one  
54 is the selection of correlated variables within groups (time periods). Natural ordering of vari-  
55 ables within groups can lead to potentially high correlation between consecutive variables.  
56 These dependencies have to be taken into account to avoid ill-conditioned problems and  
57 over-fitting, but also to better reflect reality and detect successive meaningful time periods.

58

59 Considerable attention has been paid in the last decades to variable and group selection.  
60 Developed methods are mainly related to penalized likelihood techniques in a frequentist  
61 context, or to the use of appropriate priors reflecting desired penalties in a Bayesian context.  
62 Among others, the Least absolute shrinkage and selection operator (Lasso, Tibshirani, 1996),

63 the Smoothly Clipped Absolute Deviation (Fan and Li, 2001) penalty or yet the Elastic-Net  
64 (Zou and Hastie, 2005) are classically used. Note that Elastic-Net is well adapted when vari-  
65 ables are correlated. This approach is based on the combination of  $\ell_1$ - and  $\ell_2$ -norms on the  
66 penalization term, combining shrinkage properties from Lasso and regularization capacities  
67 from Ridge regression (Hoerl and Kennard, 1970). In a Bayesian multiple linear regression  
68 context, the set of priors for variable selection has also been extensively developed. We may  
69 cite among others, the spike-and-slab prior (Mitchell and Beauchamp, 1988; George and  
70 McCulloch, 1993, 1997), the Bayesian Lasso prior (Park and Casella, 2008), the Elastic-Net  
71 prior (Kyung et al., 2010), the normal-Gamma prior (Griffin et al., 2010) and the horse-  
72 shoe (HS) prior (Carvalho et al., 2010; Piironen et al., 2017). Nevertheless, these methods  
73 do not take into account a potential group structure within covariates. Lasso extensions to  
74 group selection have been developed in frequentist (Yuan and Lin, 2006) or Bayesian (Kyung  
75 et al., 2010; Liquet et al., 2017) contexts. In order to select groups as well as variables within  
76 groups, Xu et al. (2015) proposed the sparse group Lasso prior. This approach mimics the  
77 frequentist sparse group Lasso penalty introduced by Simon et al. (2013). Xu et al. (2016)  
78 extended such a prior considering a horseshoe prior and a scale mixture of independent  
79 Gaussian distributions with three levels of variance parameters: one global and common to  
80 all coefficients, one specific to each group and one for each coefficient.

81

82 The above methods do not allow serial correlations between successive variables within  
83 groups to be taken into account. These dependencies may lead to identifiability problems  
84 impacting the estimation task which aims to assign similar effects for two adjacent variables.  
85 In a linear regression context, to allow the integration of this information and to constrain  
86 estimation, Land and Friedman (1997) and Tibshirani et al. (2005) introduce the fusion and  
87 fused Lasso. The fusion Lasso penalizes the  $\ell_1$ -norm of successive differences of parameters,  
88 and the fused Lasso combines the fusion Lasso with the usual Lasso penalization on each

89 coefficient. Kyung et al. (2010) proposed a Bayesian fused Lasso with a Bayesian Lasso prior  
90 on differences and also on each coefficient. However, various studies pointed out that the  
91 Bayesian Lasso prior, which uses Laplace distribution on coefficients, does not shrink enough  
92 each coefficient or differences towards zero, leading to biased (Carvalho et al., 2010; Polson  
93 and Scott, 2011) and smooth estimations without possible abrupt changes (Faulkner and  
94 Minin, 2018). To allow more flexibility and sparser estimations, other continuous shrinkage  
95 priors, with stronger mass on zero and heavier tails, have been investigated on differences.  
96 For instance, Rue and Held (2005) and Song and Cheng (2020) used a Student distribution  
97 on the differences, Shimamura et al. (2019) considered normal-Exponential-Gamma (NEG)  
98 distribution, while Faulkner and Minin (2018) and Kakikawa et al. (2023) placed a HS prior  
99 on differences. Note that, all of these approaches place a Laplace distribution on regression  
100 coefficients. These methodologies show good properties in terms of prediction accuracy but  
101 also to estimate smooth functions with potentially abrupt changes. However, they have  
102 been designed for only one group. A direct extension of Bayesian fused Lasso to multi-group  
103 context has been proposed by Alaíz et al. (2013). Zhang et al. (2014) used this multi-group  
104 version for the slab part of a group spike-and-slab prior. These methods can suffer from low  
105 shrinkage properties of the Bayesian Lasso, leading to poor estimations when the number of  
106 covariates within groups is large.

107

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

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

## 121 2 Model

### 122 2.1 Notation and model

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

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

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

135 **2.2 Prior construction**

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

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

140 This formulation considers a global-local parametrization with local shrinkage parameters  
 141  $\gamma_t$  and  $\omega_t$  as well as global shrinkage parameters,  $v$  and  $\lambda$ , respectively specific to coeffi-  
 142 cients and their differences. Global parameters perform shrinkage on all coefficients and  
 143 their differences, whereas local parameters allow true large effects to escape from overall  
 144 shrinkage. Fused-type priors are originally based on the use of the Laplace distribution, also  
 145 called normal-Exponential (NE) distribution, for regression parameters and their differences.  
 146 However, Laplace prior faces posterior inconsistency notably due to its exponentially light  
 147 tail (Castillo et al., 2015). To overcome such drawbacks, alternative priors with heavier  
 148 tails, such as t-Student, NEG distributions or yet HS priors on differences combined with a  
 149 Laplace distribution on coefficients have been proposed (see Table 1). The use of Laplace  
 150 distribution for coefficients has not yet been questioned. In this paper, according to investi-  
 151 gations using simulations, we propose a normal half-Cauchy (NhC) distribution on regression  
 152 coefficients with a horseshoe distribution on their differences as an alternative steady prior  
 153 face to dimension complexity and signal-to-noise ratio. This prior, denoted HS-NhC prior,  
 154 is defined as follows:



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

155 where  $\gamma_t$  ( $t = 1, \dots, T$ ),  $\lambda$ , and  $\omega_t$  ( $t = 2, \dots, T$ ) follow a half-Cauchy distribution.

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

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

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

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

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

164 while keeping local parameters  $\omega_{gt}$ . This prior is defined by:

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

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

## 167 2.3 MCMC implementation

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

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

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

173 where  $\mathbf{Q}_g$  is equal to

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

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

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

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

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

182 This simplifies the computation of full conditional distributions. Thus, for instance for the  
 183 global shrinkage parameter  $\lambda^2$  defined in Eq. (5), its full conditional distribution is given by  
 184 :

$$\lambda^2|\cdot \sim IG\left(\frac{1+T}{2}, \frac{1}{\xi} + \sum_{g=1}^G \frac{\beta_g^\top \mathbf{D}^\top \Omega_g^{-1} \mathbf{D} \beta_g}{2\sigma^2}\right), \text{ and } \xi|\cdot \sim IG(1, 1 + 1/\lambda^2).$$

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

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

### 189 3 Simulation study

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

193 noise ratio. We compare results obtained using HS-NhC group fused priors with the three  
194 alternative priors: the global and specific HS-NE and the global HS-HS. In this simulation  
195 study,  $p = 1500$  covariates were generated. We assumed that  $p$  was divided into  $G = 1, 10, 30$   
196 or 100 groups. Covariates within each group were generated from a  $\frac{p}{G}$ -multivariate Gaussian  
197 distribution with zero mean and a covariance matrix defined by a first-order autoregressive  
198 (AR1) structure with a parameter fixed to 0.95. Functional effects were defined as the  
199 combination of different smooth functions: a continuous smooth function as proposed by  
200 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:

$$MSE_{nz} = \frac{1}{|C_{nz}|} \sum_{t \in C_{nz}} (\beta_t - \hat{\beta}_t)^2; \quad MSE_z = \frac{1}{|C_z|} \sum_{t \in C_z} (\beta_t - \hat{\beta}_t)^2; \quad \text{and}$$

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

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

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

Priors	$\sigma^2$	$MCC$	$MSE_z$	$MSE_{nz}$
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  $\sigma^2$  equal to 1 or 16, and  $G = 1$ .

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

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

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

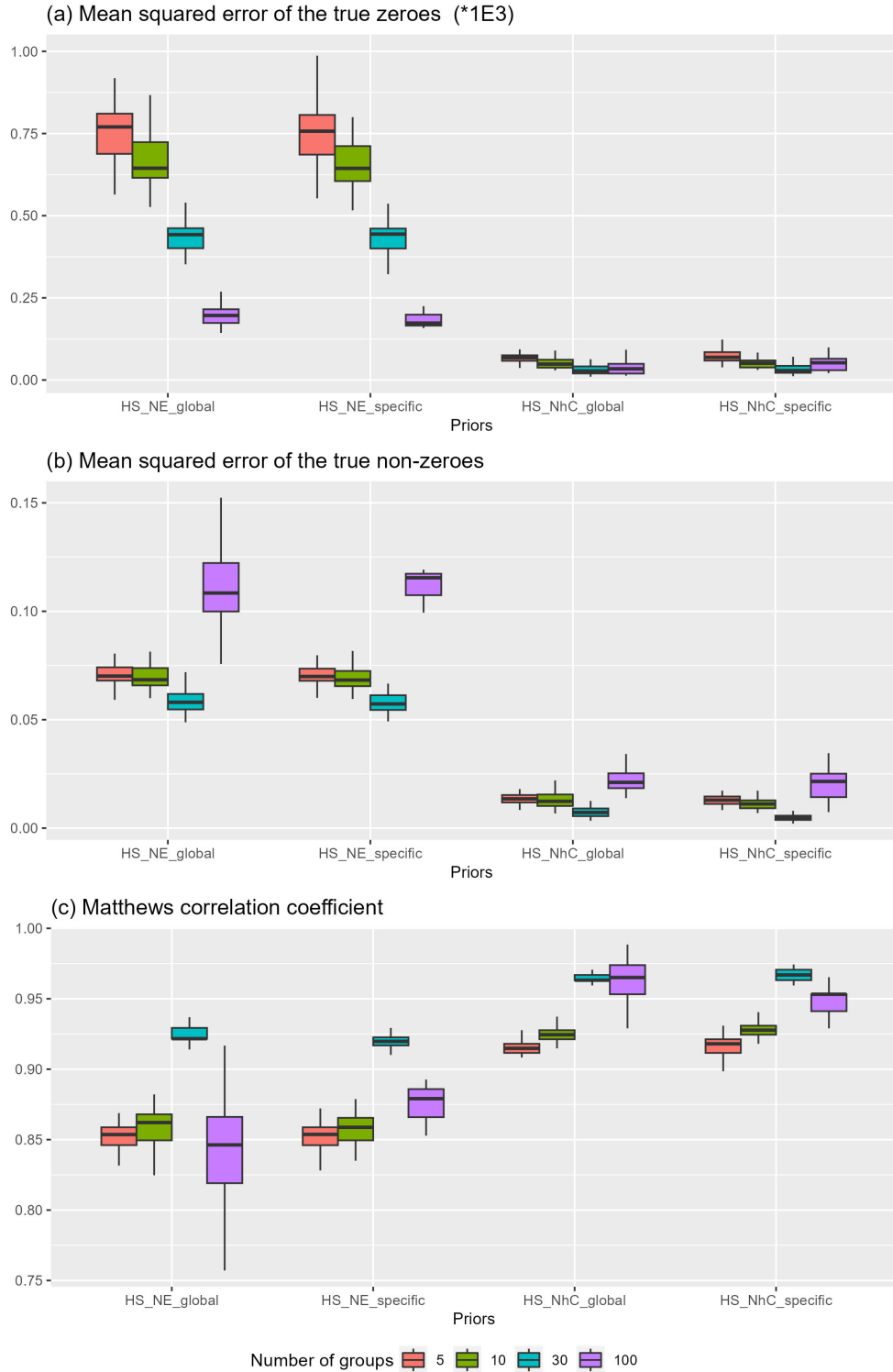


Figure 1: (a) Mean squared errors of the true zero coefficients  $MSE_z$ , (b) Mean squared errors of the non-zeroes ( $MSE_{nz}$ ), and (c) Matthews Correlation Coefficient ( $MCC$ ) using the different priors with residual variance  $\sigma^2$  equal to 1, and  $G = 5, 10, 30, 100$ .

## 238 4 The abscission dataset

239 This application aims at identifying environmental variables and time periods affecting the  
240 oil palm fruit abscission process (Tisné et al., 2020). The dataset is provided by “le Centre de  
241 Recherches Agricoles-Plantes Pérennes” (CRA-PP) of the national institute for agricultural  
242 research of Benin Republic (INRAB) which manages an oil palm seed garden involving a  
243 self-pollinated population of 140 oil palm trees planted between 2000 and 2005 in a single  
244 homogeneous field plot. Each palm tree produced between 1 and 8 bunches per year over  
245 all the experiment time from 2014 until 2018. The manual pollination date, different for  
246 each bunch, was recorded and the bunch was monitored up to its harvest. A total of 1,173  
247 bunches were considered over multiple years, taking advantage of the climatic seasonality and  
248 the continuous fruit production of this species. We used the number of days from pollination  
249 to fruit drop (DFD) as the response variable. DFD is the classical harvest time indicator and  
250 its variation integrates different underlying abscission processes at different developmental  
251 stages.

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



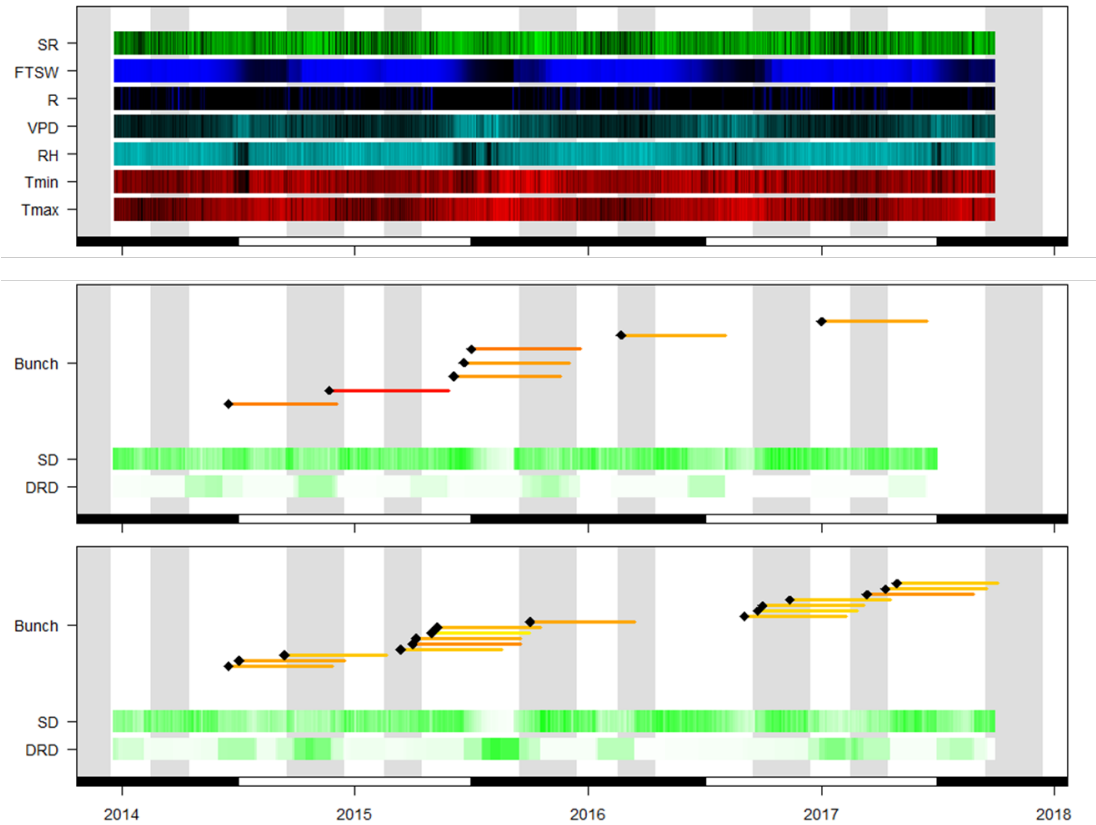


Figure 2: Raw and calculated environmental variables (top panel) and example of bunch production for two oil palm individuals along with their endogenous variables (middle and bottom panels). Raw and calculated variables (heatmaps) as well as bunches (horizontal segments) are plotted over the experiment duration, from 2014 to 2018. The segments for bunches are represented from the manual pollination (black diamond) 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: minimum/maximum temperatures; SD: supply-demand ratio; DRD: Daily reproductive demand.

263 A three-day time grid, from  $-180$  (individualization of the floral meristem) to  $+180$  (ripe  
 264 fruit) days after pollination, was used to calculate either the average values over three days  
 265 ( $T_{\max}$ ,  $T_{\min}$ , RH, VPD, FTSW, DRD, and SD) or the cumulative values over 15 days (R  
 266 and SR) of each variable. This experimental design thus leads to nine groups of covariates  
 267 ( $G = 9$ ) measured at  $T = 121$  times. Within each matrix the  $i^{\text{th}}$  row corresponds to the  
 268  $i^{\text{th}}$  bunch analyzed and the  $t^{\text{th}}$  column corresponds to the value of the corresponding cli-  
 269 matic/ecophysiological variable at time  $t$  for each bunch. All matrices have been scaled to  
 270 obtain a similar order of magnitude. All results are based on 20 MCMC runs initialized at  
 271 random starting values and 50,000 iterations with a burn-in of 20,000 and a thinning of 10. A  
 272 group is considered selected if at least one regression effect within it has a credible interval  
 273 that does not contain zero.

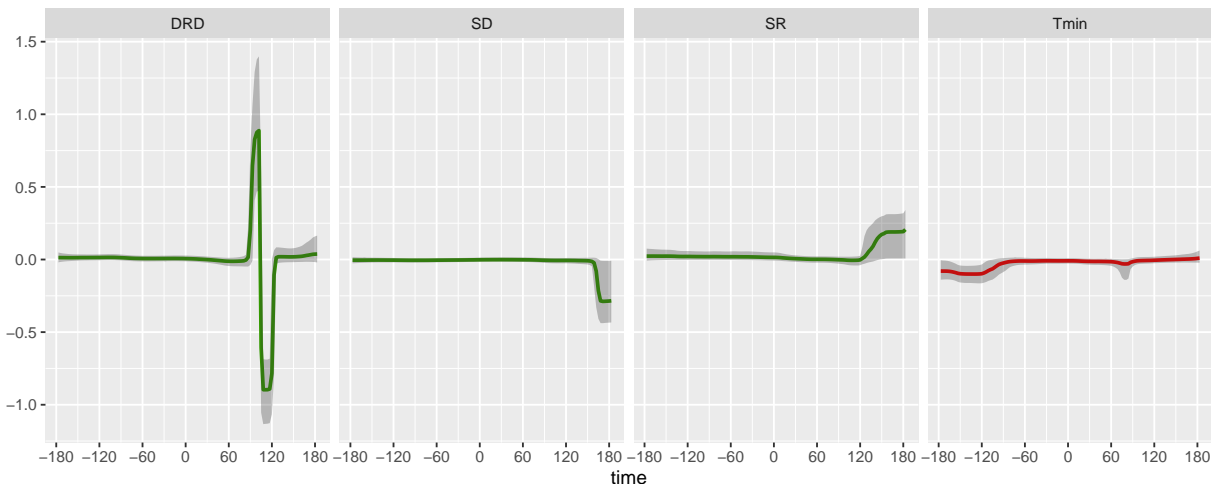


Figure 3: Non-zero coefficient profile estimation provided by the global HS-NhC prior on the abscission dataset. Gray shadows represent the 95% credible interval. Colors represent the different categories of environmental variables, green is for photosynthesis variables (DRD, SR, SD) and red is for temperature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

## 274 Comparison with previous studies and biological interpretation

275

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

277 and allow identification of relevant time periods of four variables (Tmin, SR, DRD, and  
278 SD). Two types of patterns are observed: with smooth effects for Tmin and SR and with  
279 punctual effects for DRD and SD. The Tmin variable is negatively associated with DFD  
280 during the inflorescence development from day  $-180$  to  $-100$ , while the three other variables  
281 are associated with DFD at the end of the fruit bunch development. SR, the solar radiation  
282 variable, is positively associated with DFD from day 120 to 180, at the final stage before the  
283 fruit drop. The DRD variable is punctually associated with DFD at days 99 and 105 after  
284 pollination, first positively before an inversion of the association direction at day 100. The  
285 SD factor is negatively associated with DFD with a peak at day 160.

286 The striking pattern of DRD around day 100 after pollination observed in Tisné et al.  
287 (2020), is thus confirmed and corresponds to the “lag period” of the oil palm fruit bunch de-  
288 velopment between the cell division/expansion phase and the maturation phase (Tranbarger  
289 et al., 2011). The selection of the DRD variable at this key developmental stage suggests  
290 that the considered fruit bunch integrates current and future whole plant photosynthate  
291 demand due to concomitant developing bunches, to modulate its maturation and abscission  
292 timing. Such carbohydrate-based regulation is commonly found in fruit tree species and  
293 leads to the wave of abscission concerning fruitlets (Sawicki et al., 2015), the only difference  
294 being that the oil palm regulates ripe fruit abscission timing instead of dropping unripe  
295 fruits. In contrast with DRD and SD that have similar punctual patterns with those of  
296 Tisné et al. (2020) study, Tmin effect profile is different, showing a continuous moderate  
297 effect instead of many weak effects spread over the  $-180$  to  $-100$  period. The SR factor  
298 was not selected by Tisné et al. (2020) but it has a positive effect from day 120 to 180  
299 using our prior. These discrepancies may be due to the cumulative nature of both Tmin  
300 and SR effects at their respective developmental stages. Hence, the Tmin effect at the early  
301 inflorescence developmental stages could be related to thermal time which is known to be  
302 associated with developmental rates. In the period identified, the differentiation of floral

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

## 311 5 Conclusion

312 We proposed four Bayesian fused priors in the multi-group context (see Table 2). We showed  
313 that the combination of a normal half-Cauchy on coefficients and a horseshoe on their dif-  
314 ferences is more efficient and stable than the natural extension of the HS-NE fused prior  
315 whatever the number of groups ( $G \geq 1$ ), the size of groups, and the signal-to-noise ratio.  
316 The proposed general formulation (see Eq. 2) encompasses most priors already developed  
317 in the literature as shown in Table 1. Through simulations, we evidenced the importance of  
318 placing heavy-tailed distributions with a spike at zero on the differences (Kakikawa et al.,  
319 2023) and considering a distribution with a heavier tail property than the usual Laplace prior  
320 on coefficients. However, we showed that using a horseshoe prior on both coefficients and  
321 their differences resulted in poor performances, as it tended to shrink all parameters towards  
322 zero. We noted the advantage of fused-type priors on fusion-type priors mainly when the  
323 number of groups was large with a moderate to small size.

324

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

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

331

332 The global HS-NhC prior may be directly applied to a broad type of applications such  
333 as in the near infrared spectroscopy context, which involves one group of ordered variables  
334 through a spectrum, or in the genetic mapping context, where markers may be viewed as  
335 groups of ordered variables at the chromosome level. To take into account multi-dimensional  
336 indexation (spatial or spatio-temporal structures) instead of only one dimensional indexation  
337 (time structure), this prior should be extended even if it raises computational challenges.

## 338 **Declarations**

339 **Conflicts of interests** The authors have no conflicts of interest to declare.

## 340 **References**

- 341 Adam, H., Collin, M., Richaud, F., Beulé, T., Cros, D., Omoré, A., Nodichao, L., Nouy, B.,  
342 and Tregear, J. W. (2011). Environmental regulation of sex determination in oil palm:  
343 current knowledge and insights from other species. *Annals of botany*, 108(8):1529–1537.
- 344 Alaíz, C. M., Barbero, A., and Dorronsoro, J. R. (2013). Group fused lasso. In *International*  
345 *Conference on Artificial Neural Networks*, pages 66–73. Springer.
- 346 Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse  
347 signals. *Biometrika*, 97(2):465–480.

- 348 Castillo, I., Schmidt-Hieber, J., and van der Vaart, A. (2015). Bayesian linear regression  
349 with sparse priors. *The Annals of Statistics*, 43(5):1986–2018.
- 350 Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its  
351 oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360.
- 352 Faulkner, J. R. and Minin, V. N. (2018). Locally adaptive smoothing with markov random  
353 fields and shrinkage priors. *Bayesian analysis*, 13(1):225–252.
- 354 George, E. I. and McCulloch, R. E. (1993). Variable selection via gibbs sampling. *Journal*  
355 *of the American Statistical Association*, 88(423):881–889.
- 356 George, E. I. and McCulloch, R. E. (1997). Approaches for bayesian variable selection.  
357 *Statistica Sinica*, 7(2):339–373.
- 358 Gilks, W., Richardson, S., and Spiegelhalter, D. (1995). *Markov Chain Monte Carlo in*  
359 *Practice*. CRC press.
- 360 Gill, A. L., Gallinat, A. S., Sanders-DeMott, R., Rigden, A. J., Short Gianotti, D. J.,  
361 Mantooth, J. A., and Templer, P. H. (2015). Changes in autumn senescence in northern  
362 hemisphere deciduous trees: a meta-analysis of autumn phenology studies. *Annals of*  
363 *botany*, 116(6):875–888.
- 364 Griffin, J. E., Brown, P. J., et al. (2010). Inference with normal-gamma prior distributions  
365 in regression problems. *Bayesian analysis*, 5(1):171–188.
- 366 Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthog-  
367 onal problems. *Technometrics*, 12(1):55–67.
- 368 Kakikawa, Y., Shimamura, K., and Kawano, S. (2023). Bayesian fused lasso modeling via  
369 horseshoe prior. *Japanese Journal of Statistics and Data Science*, 6(2):705–727.

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

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

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

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

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

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

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

385 Pironen, J., Vehtari, A., et al. (2017). Sparsity information and regularization in the horse-  
386 shoe and other shrinkage priors. *Electronic Journal of Statistics*, 11(2):5018–5051.

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

389 R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foun-  
390 dation for Statistical Computing, Vienna, Austria.

- 391 Reichardt, S., Piepho, H.-P., Stintzi, A., and Schaller, A. (2020). Peptide signaling for  
392 drought-induced tomato flower drop. *Science*, 367(6485):1482–1485.
- 393 Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*.  
394 Chapman and Hall/CRC press.
- 395 Sawicki, M., Aït Barka, E., Clément, C., Vaillant-Gaveau, N., and Jacquard, C. (2015).  
396 Cross-talk between environmental stresses and plant metabolism during reproductive or-  
397 gan abscission. *Journal of Experimental Botany*, 66(7):1707–1719.
- 398 Shimamura, K., Ueki, M., Kawano, S., and Konishi, S. (2019). Bayesian generalized fused  
399 lasso modeling via neg distribution. *Communications in Statistics-Theory and Methods*,  
400 48(16):4132–4153.
- 401 Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2013). A sparse-group lasso. *Journal*  
402 *of computational and graphical statistics*, 22(2):231–245.
- 403 Song, Q. and Cheng, G. (2020). Bayesian fusion estimation via t shrinkage. *Sankhya A*,  
404 82(2):353–385.
- 405 Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal*  
406 *Statistical Society: Series B (Methodological)*, 58(1):267–288.
- 407 Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005). Sparsity and  
408 smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statis-*  
409 *tical Methodology)*, 67(1):91–108.
- 410 Tibshirani, R. J. et al. (2014). Adaptive piecewise polynomial estimation via trend filtering.  
411 *The Annals of Statistics*, 42(1):285–323.



- 412 Tisné, S., Denis, M., Domonhédó, H., Pallas, B., Cazemajor, M., Tranbarger, T. J., and  
413 Morcillo, F. (2020). Environmental and trophic determinism of fruit abscission and outlook  
414 with climate change in tropical regions. *Plant-Environment Interactions*, 1(1):17–28.
- 415 Tranbarger, T. J., Dussert, S., Joët, T., Argout, X., Summo, M., Champion, A., Cros, D.,  
416 Omore, A., Nouy, B., and Morcillo, F. (2011). Regulatory mechanisms underlying oil palm  
417 fruit mesocarp maturation, ripening, and functional specialization in lipid and carotenoid  
418 metabolism. *Plant physiology*, 156(2):564–584.
- 419 Wilmowicz, E., Kućko, A., Pokora, W., Kapusta, M., Jasieniecka-Gazarkiewicz, K., Tran-  
420 barger, T. J., Wolska, M., and Panek, K. (2021). Epip-evoked modifications of redox, lipid,  
421 and pectin homeostasis in the abscission zone of lupine flowers. *International journal of*  
422 *molecular sciences*, 22(6):3001.
- 423 Xu, X., Ghosh, M., et al. (2015). Bayesian variable selection and estimation for group lasso.  
424 *Bayesian Analysis*, 10(4):909–936.
- 425 Xu, Z., Schmidt, D. F., Makalic, E., Qian, G., and Hopper, J. L. (2016). Bayesian grouped  
426 horseshoe regression with application to additive models. In *Australasian Joint Conference*  
427 *on Artificial Intelligence*, pages 229–240. Springer.
- 428 Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped  
429 variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*,  
430 68(1):49–67.
- 431 Zhang, L., Baladandayuthapani, V., Mallick, B. K., Manyam, G. C., Thompson, P. A.,  
432 Bondy, M. L., and Do, K.-A. (2014). Bayesian hierarchical structured variable selection  
433 methods with application to molecular inversion probe studies in breast cancer. *Journal*  
434 *of the Royal Statistical Society: Series C (Applied Statistics)*, 63(4):595–620.

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

## 437 Appendix

### 438 A. Bayesian hierarchical models and full conditional distributions

#### 439 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:

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

440 where  $\boldsymbol{\Upsilon}_g = \text{diag}(v_{g_1}^2, \dots, v_{g_T}^2)$  and  $\boldsymbol{\Omega}_g = \text{diag}(\omega_{g_1}^2, \dots, \omega_{g_{T-1}}^2)$ .

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

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

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

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

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

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

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

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

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

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

441 **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:

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

where  $\Upsilon_g = \text{diag}(v_{g1}^2, \dots, v_{gT}^2)$  and  $\Omega_g = \text{diag}(\omega_{g1}^2, \dots, \omega_{gT-1}^2)$ . The corresponding full conditional distributions for the model parameters are given by:

$$\begin{aligned} \mu | \cdot &\sim \mathcal{N}\left(\frac{1}{n} \mathbf{1}^\top (\mathbf{y} - \sum_{g=1}^G \mathbf{X}_g \boldsymbol{\beta}_g), \frac{\sigma^2}{n}\right) \\ \boldsymbol{\beta}_g | \cdot &\sim \mathcal{N}_T\left(\boldsymbol{\Sigma}_{b_g} \frac{\mathbf{X}_g^\top}{\sigma^2} (\mathbf{y} - \mu \mathbf{1} - \sum_{\tilde{g} \neq g} \mathbf{X}_{\tilde{g}} \boldsymbol{\beta}_{\tilde{g}} - Z\alpha), \boldsymbol{\Sigma}_{b_g} = \sigma^2 \left( \mathbf{X}_g^\top \mathbf{X}_g + \Upsilon_g^{-1} + \frac{\mathbf{D}_g^\top \Omega_g^{-1} \mathbf{D}_g}{\lambda_g^2} \right)^{-1}\right) \\ \lambda_g^2 | \cdot &\sim IG\left(\frac{1}{2} + \frac{T}{2}, \frac{1}{\psi_g} + \frac{\boldsymbol{\beta}_g^\top \mathbf{D}_g^\top \Omega_g^{-1} \mathbf{D}_g \boldsymbol{\beta}_g}{2\sigma^2}\right), \quad \psi_g | \cdot \sim IG(1, 1 + 1/\lambda_g^2), \quad g = 1, \dots, G \\ \omega_{gj}^2 | \cdot &\sim IG\left(1, \frac{1}{\phi_{gj}} + \frac{((\mathbf{D}_g \boldsymbol{\beta}_g)_{[j]})^2}{2\sigma^2 \lambda_g^2}\right), \quad \phi_{gj} | \cdot \sim IG(1, 1 + 1/\omega_{gj}^2), \quad g = 1, \dots, G, \quad j = 1, \dots, T-1 \\ v_{gt}^2 | \cdot &\sim IG\left(1, \frac{1}{\eta_{gt}} + \frac{\beta_{gt}^\top \beta_{gt}}{2\sigma^2}\right), \quad \eta_{gt} | \cdot \sim IG(1, 1 + 1/v_{gt}^2), \quad g = 1, \dots, G, \quad t = 1, \dots, T \\ \alpha | \cdot &\sim \mathcal{N}_P\left(\boldsymbol{\Sigma}_\alpha \frac{Z^\top}{\sigma^2} (\mathbf{y} - \mu \mathbf{1} - \sum_{g=1}^G \mathbf{X}_g \boldsymbol{\beta}_g), \boldsymbol{\Sigma}_\alpha = \left( \frac{A^{-1}}{\sigma_\alpha^2} + \frac{Z^\top Z}{\sigma^2} \right)^{-1}\right) \\ \sigma_\alpha^2 | \cdot &\sim IG\left(\frac{1+P}{2}, \frac{1}{2} + \alpha^\top A^{-1} \alpha\right) \\ \sigma^2 | \cdot &\sim IG\left(\frac{1+T+n}{2}, \frac{1}{a} + \frac{1}{2} \sum_{g=1}^G \boldsymbol{\beta}_g^\top \left( \Upsilon_g^{-1} + \frac{\mathbf{D}_g^\top \Omega_g^{-1} \mathbf{D}_g}{\lambda_g^2} \right) \boldsymbol{\beta}_g + \frac{1}{2} \|\mathbf{y} - \mu \mathbf{1} - \sum_{g=1}^G \mathbf{X}_g \boldsymbol{\beta}_g - Z\alpha\|_2^2\right) \\ a | \cdot &\sim IG(1, 1 + 1/\sigma^2) \end{aligned}$$

442 **B. Convergence evaluation.**

$\sigma^2$	$g$	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.