Pesquisa Agropecuária Brasileira

ISSN 1678-3921

Journal homepage: www.embrapa.br/pab

For manuscript submission and journal contents, access: www.scielo.br/pab

Genetics/ Original Article

Incorporating prior knowledge into Bayesian models for genetic evaluation in soybean breeding

Jeniffer Santana Pinto Coelho Evangelista[1],
Igor Ferreira Coelho[1],
Marco Antônio Peixoto[1],
Rodrigo Silva Alves[1],
Marcos Deon Vilela de Resende[2],
Felipe Lopes da Silva[3], and
Leonardo Lopes Bhering[1 ✉]

(1) Universidade Federal de Viçosa, Departamento de Biologia Geral, Avenida P.H. Rolfs, s/nº, Campus Universitário, CEP 36570-900 Viçosa, MG, Brazil.
E-mail: jenifferspce@gmail.com, igor.agronomia@gmail.com, marco.peixotom@gmail.com, rodrigo.alves@cenibra.com.br, leonardo.bhering@ufv.br

(2) Embrapa Café, Parque Estação Biológica, Avenida W3 Norte (Final), CEP 70770-901 Brasília, DF, Brazil.
E-mail: marcos.deon@gmail.com

(3) Universidade Federal de Viçosa, Departamento de Agronomia, Avenida P.H. Rolfs, s/nº, Campus Universitário, CEP 36570-900 Viçosa, MG, Brazil.
E-mail: felipe.silva@ufv.br

✉ Corresponding author

Received
October 23, 2023

Accepted
April 03, 2024

How to cite
EVANGELISTA, J.S.P.C.; COELHO, I.F.; PEIXOTO, M.A.; ALVES, R.S.; RESENDE, M.D.V. de; SILVA, F.L. da; BHERING, L.L. Incorporating prior knowledge into Bayesian models for genetic evaluation in soybean breeding. **Pesquisa Agropecuária Brasileira**, v.59, e03557, 2024. DOI: https://doi.org/10.1590/S1678-3921.pab2024.v59.03557.

Abstract – The objective of this work was to compare the use of noninformative and informative priors in Bayesian models, as well as to evaluate the viability of including informative priors in the estimation of variance components and genetic values in soybean breeding programs. The used phenotypic data refer to the evaluation of 80 soybean genotypes in ten environments over three years. For each evaluated crop year, informative and noninformative priors were used, and the parameters were estimated using the Gibbs sampler algorithm. Parameter estimates from the previous crop year were used as prior information for the next evaluated crop year. The goodness-of-fit was calculated using the deviance information criterion (DIC). Selective accuracy showed the highest values for the models chosen through DIC for both crop years. However, the intervals of the highest posterior density are narrower for all models that adopted informative priors. Adding information into Bayesian inference does not always result in a better model fitting.

Index terms: *Glycine max*, Bayesian inference, Gibbs sampler, HPD, MCMC.

Inclusão de conhecimento prévio em modelos bayesianos para avaliação genética no melhoramento de soja

Resumo – O objetivo deste trabalho foi comparar o uso de a priori não informativas e informativas em modelos bayesianos, bem como avaliar a viabilidade da inclusão de a priori informativas na estimativa dos componentes de variância e dos valores genotípicos em programas de melhoramento de soja. Os dados fenotípicos utilizados referem-se à avaliação de 80 genótipos de soja, em dez ambientes, ao longo de três anos. Para cada safra avaliada, foram utilizadas a priori informativas e não informativas, e os parâmetros foram estimados com uso do algoritmo de amostragem de Gibbs. As estimativas dos parâmetros da safra anterior foram utilizadas como informação prévia para a próxima safra avaliada. A qualidade do ajuste foi calculada com uso do critério de informação de desvio (DIC). A acurácia seletiva apresentou maiores valores nos modelos escolhidos por meio do DIC, para ambas as safras. No entanto, os intervalos de maior densidade a posteriori são menores para todos os modelos que adotaram a priori informativas. Adicionar informações à inferência bayesiana nem sempre resulta em melhor ajuste ao modelo.

Termos para indexação: *Glycine max*, inferência bayesiana, amostrador de Gibbs, HPD, MCMC.

Introduction

In recent decades, a vast number of statistical methods have been applied for the genetic selection of soybean [*Glycine max* (L.) Merr.] genotypes (Dalló et al., 2019; Woyann et al., 2020; Rezende et al., 2021), especially for grain yield, the main trait for this crop. However, the analyses for the selection of superior soybean genotypes are uncertain, probably due to several factors, such as unbalanced data, unsuitable blocking, limited number of replicates in early generation tests, and, particularly, the spatial variation in multienvironment trial data (Bernardo, 2020).

In the literature, the incorporation of informative priors is widely mentioned as a positive strategy for variance component and genetic parameter estimates (Silva et al., 2013), as long as it is done carefully. Regarding the possibility of incorporating informative priors, Bayesian inference is more advantageous than Fisherian inference, in addition to including flexibility in distribution choice for unknown parameters (Blasco, 2001; Sorensen & Gianola, 2002; Sorensen, 2009; Silva et al., 2013).

In the Bayesian context, the deviance information criterion (DIC) is the most adopted for model selection (Spiegelhalter et al., 2002), being widely used in Bayesian analyses of plant breeding (Torres et al., 2018; Volpato et al., 2019; Nascimento et al., 2020; Silva et al., 2020).

In genetic evaluations, the estimation of variance components and the prediction/estimation of genetic values are important steps in the selection process, mainly regarding quantitative traits, which are controlled by several genes and largely affected by environmental effects (Huang & Mackay, 2016). In soybean breeding, restricted maximum likelihood/best linear unbiased prediction (REML/BLUP) has been widely applied for such end (Rezende et al., 2021). However, other methods based on Bayesian inference have raised interest due to their statistical robustness, especially when used for annual crops (Torres et al., 2018; Montesinos-López et al., 2019; Volpato et al., 2019; Silva et al., 2020). Similarly to the REML/BLUP method, Bayesian inference allows of estimating variance components and genetic values.

For these reasons, Bayesian inference has been indicated as a suitable statistical method for the genetic evaluation of crop species (Silva et al., 2013). This inference overcomes some problems found in REML/BLUP, such as approximations in variance component estimation and assumptions of asymptotic normality (Resende, 2002). In addition, in Bayesian inference, the variance of the estimators is known, which improves the reliability of selection practices. Furthermore, under this inference, the combination of likelihood function (from the data under analysis) and prior distribution (previous information regarding the parameter) results in a posterior distribution for the parameters of interest. In this sense, the means of posterior distributions are suitable estimates for variance components and genetic values, mainly when phenotypic data are scarce (Sorensen & Gianola, 2002).

Currently, a large amount of information from previous surveys is available, and incorporating prior information into modeling is reasonable and may increase the knowledge of plant breeders (Nascimento et al., 2020), as well as ultimately improve genetic evaluation (Couto et al., 2015). However, this type of information is not always used in Bayesian inference, and its implementation is the main impediment for its exploration (Resende, 2000). Since the premise of prior distribution is that all knowledge about a given parameter is represented by prior information, the latter should be classified according to its informativeness as: vague prior, there is no knowledge about the parameter; or informative prior, there is some knowledge about the studied parameter, which can be incorporated into Bayesian information through specialist knowledge about the parameters, reference and prospective studies, and empirical Bayes methods (Wakefield, 2013). In Bayesian inference, prior information aims to reduce the uncertainty regarding the parameter under analysis in order to proceed with the estimation process.

The objective of this work was to compare the use of noninformative and informative priors in Bayesian models, as well as to evaluate the viability of including informative priors in the estimation of variance components and genetic values in soybean breeding programs.

Materials and Methods

The phenotypic data used in the present study refer to the evaluation of 80 soybean genotypes in ten environments allocated in soybean macroregion 2, covering microregions 201, 202, and 204 (Kaster & Farias, 2012). The genotypes were evaluated over three consecutive crop years (2012/2013, 2013/2014, and 2014/2015), as follows: 30 genotypes in 2012/2013 and 2013/2014, 27 genotypes in 2014/2015, and 1 genotyped in all three seasons (Table 1).

In each environment, the experiment was arranged in a completely randomized block design with three replicates. Each plot consisted of four 5.0 m lines, with 0.5 m spacing between lines and between plots. At maturity, the two central lines were harvested, totaling a usable area of 5.0 m^2. The grain yield trait was evaluated in kg ha^{-1}, and humidity was corrected to 13%. Crop management followed the technical recommendations for soybean cultivation in each site (Silva et al., 2022).

The used Bayesian statistical model considered genotypes as random effects and, therefore, does not rely on the particularities of specific genotypes. This means that the model is designed to capture the genetic variation within the germplasm rather than respond to the unique characteristics of individual genotypes, presenting a generalizability that makes its application to other similar datasets easier. The Bayesian statistical model, associated to the genotypic evaluation in the randomized complete block design in several environments, was obtained using the following equation:

$$y = X_f + Z_g + T_i + e,$$

where y is the vector of phenotypic data; g is the vector of genotype effects, assumed as $g|G \sim N(0, Z \otimes I)$; i is the vector of the genotype x environment (G x E) interaction effect, assumed as $i|E \sim N(0, T \otimes I)$; e is the residual vector; and X, Z, and T are the incidence matrices for effects f, g, and i, respectively. The conditional distribution of the phenotypic data was given by $y|\ f, g, i, G, E, R \sim N(X_f + Z_g + T_i, R \otimes I)$, where G is the genotypic variance, E is the G x E matrix of the (co)variance, R is the residual variance, and I are the replicates into trial effects, assumed as $f \sim N(f, \Sigma_f \otimes I)$. Furthermore, the posterior density for all parameters followed the joint posterior distribution, according to Bayes' theorem, as follows: $P(f, g, i, G, E, R \mid y) = P(y|\ f, g, i, G, E, R) \times P(f) \times P(g|G) \times P(i|E) \times P(G) \times P(E) \times P(R)$, where $P(f, g, i, G, E, R \mid y)$ is the joint posterior distribution provided by the multiplication of the likelihood function ($P(y|\ f, g, i, G, E, R)$) and prior distributions P(f), P(g|G), P(i|E), P(G), P(E), and P(R).

For all three crop years, the variance components and genetic values were estimated using Bayesian inference accounting for a vague prior. For such, the degree of reliability parameter was defined as 0.02 (Hadfield, 2010). This parameter was assumed to be equivalent to the precision parameter of a scaled inverse Wishart distribution, assumed as a prior distribution for G, E, and R.

The informative priors were then added to the estimation process. The posterior mean of the variance components of the 2012/2013 crop year was used as prior information to analyze the 2013/2014 crop year. Similarly, the posteriori mean of the variance components of the 2013/2014 crop year was used as prior information for the analysis of the 2014/2015 crop

Table 1. Geographic coordinates for each environment and number of soybean (*Glycine max*) genotypes evaluated in each crop year.

Environment	Altitude (m)	Latitude (S)	Longitude (W)	Number of genotypes per crop year		
				2012/2013	2013/2014	2014/2015
1	590	22°59'48"	51°11'26"	30	30	27
2	563	24°37'5"	53°19'18"	30	30	27
3	430	22°13'16"	54°48'20"	30	30	27
4	762	24°19'10"	52°31'48"	30	30	27
5	364	23°3'55"	54°11'26'	30	30	27
6	333	24°17'02"	53°50'24"	30	30	27
7	655	22°32'11"	55°43'36"	30	30	27
8	736	23°18'38"	51°22'10"	30	30	27
9	384	23°3'35"	51°1'60"	30	30	27
10	560	24°42'49"	53°44'35"	30	30	27

year. In these analyses, the reliability parameter was considered 15 (Hadfield, 2010).

For each crop year, assuming informative and noninformative priors, model fitting was tested using the DIC. As proposed by Spiegelhalter et al. (2002), DIC is described as DIC = D(θ) + 2pD, where D(θ) is a point estimate of the deviance obtained by replacing the parameters by their posterior mean estimates in the likelihood function, and pD is the effective number of parameters in the model. The lower the DIC value, the better the adjusted model.

Phenotypic variance (σ^2_{phen}), individual broad-sense heritability (h^2_g), the coefficient of determination of the G×E interaction effects (c^2_i), the coefficient of determination of the residual effects (c^2_{res}), and selective accuracy ($r_{\hat{g}g}$) were obtained using the mean of the posterior distribution, according to the following equations (Resende et al., 2014), respectively:

$$\hat{\sigma}^2_{phen} = \hat{\sigma}^2_g + \hat{\sigma}^2_i + \hat{\sigma}^2_{res},$$

$$h^2_g = \frac{\hat{\sigma}^2_g}{\hat{\sigma}^2_{phen}},\ c^2_i = \frac{\hat{\sigma}^2_i}{\hat{\sigma}^2_{phen}},\ c^2_{res} = \frac{\hat{\sigma}^2_{res}}{\hat{\sigma}^2_{phen}},\ \text{and}\ r_{\hat{g}g} = \sqrt{1 - \frac{PEV}{\hat{\sigma}^2_g}},$$

where PEV is the prediction error variance, extracted from the diagonal of the solution matrix of the mixed-model equations.

The variance components and the highest posterior density (HPD) were estimated by Gibbs sampling via the Markov chain Monte Carlo (MCMC) algorithm, using the MCMCglmm package in the R software system (Hadfield, 2010). The number of iterations was 4,000,000, and a burn-in period of 400,000 and sampling interval (thin) of 40 iterations were assumed, which provided a total of 90,000 chains. The boa package (Smith, 2007) was used to test the convergence methods of Geweke (Geweke, 1992) and Raftery & Lewis (Raftery & Lewis, 1992).

Results and Discussion

The absolute values of Z statistics for all estimated variance components and parameters fell between -1.96 and 1.96 according to the Geweke convergence criterion (Geweke, 1992), at $p<0.05$. In addition, the dependency factor using the Raftery & Lewis convergence criterion (Raftery & Lewis, 1992) was below 5.0 for all variance components (Table 2). These results are an indicative that all Gibbs sampler chains achieved the desired convergence. Based on DIC, for the analyses of the 2013/2014 crop year, the model using the noninformative prior overcomes the one using the informative prior, showing lower DIC value. The posterior values of the variance components, estimated by the model using the noninformative prior for the 2013/2014 crop year, were adopted as the informative prior for the analysis of the 2014/2015 crop year. For 2014/2015, the model using the informative prior showed goodness-of-fit to the data.

DIC indicated that the 2013/2014 crop year using the noninformative prior was the best-fitted model. This

Table 2. Convergence diagnostic using the criteria of Geweke (1992) and Raftery & Lewis (1992) for the variance components and the deviance information criterion (DIC) for model selection, using informative (Prior null) and noninformative (Prior inf) priors for the soybean (*Glycine max*) grain yield trait evaluated in the 2012/2013, 2013/2014, and 2014/2015 crop years⁽¹⁾.

Parameter		2012/2013	2013/2014		2014/2015	
		Prior null	Prior null	Prior inf	Prior null	Prior inf
Geweke	$\hat{\sigma}^2_g$	0.26 (0.80)	0.59 (0.55)	-0.23 (0.82)	-0.08 (0.94)	0.56 (0.58)
	$\hat{\sigma}^2_i$	-0.91 (0.37)	-0.27 (0.79)	-1.70 (0.09)	-0.83 (0.40)	0.30 (0.76)
	$\hat{\sigma}^2_{res}$	0.60 (0.55)	-0.61 (0.55)	-0.02 (1.00)	1.36 (0.17)	-0.27 (0.78)
Raftery & Lewis	$\hat{\sigma}^2_g$	1.003	0.999	1.001	0.995	1.001
	$\hat{\sigma}^2_i$	1.086	1.014	1.101	1.051	0.995
	$\hat{\sigma}^2_{res}$	1.004	0.998	1.017	1.00	1.002
DIC		-	13,460.57^{sm}	13,472.11	12,142.54	12,136.7^{sm}

(1)$\hat{\sigma}^2_g$, genotypic variance; $\hat{\sigma}^2_i$, genotype × environment interaction variance; $\hat{\sigma}^2_{res}$, residual variance; and ^{sm}, selected model. The values between parenthesis are the p-value.

finding shows that the information provided by the 2012/2013 crop year was inadequate for the estimates of the variance components and genetic values in the 2013/2014 crop year. Silva et al. (2020) also observed that the use of the informative prior led to worse results. These findings confirm the importance of considering DIC in model selection to allow of breeders to consider the relevance of previous information for the current analysis in the Bayesian approach.

Regarding the variance components (Table 3), the analyses using the noninformative prior in the 2013/2014 and 2014/2015 crop years overcame the model using the informative prior for the same years. The heritability estimates ranged from 0.05 to 0.19,

Table 3. Variance components obtained using the Markov chain Monte Carlo algorithm and the intervals for the highest posterior density for the estimate of the variance component, at a significance level of α = 95%, for the soybean (*Glycine max*) grain yield trait evaluated in the 2012/2013, 2013/2014 and 2014/2015 crop years, accounting for informative (Prior inf) and noninformative (Prior null) priors.

Parameter(1)		2012/2013	2013/2014		2014/2015	
		Prior null	Prior null	Prior inf	Prior null	Prior inf
$\hat{\sigma}^2_g$	Mean	8,541.55	46,953.31	28,663.26	10,199.29	27,499.78
	Lower	2,954.30	20,400.22	14,230.17	2,805.52	15,873.43
	Upper	15,254.43	79,209.96	45,234.15	19,281.60	40,754.48
	Upper-lower	12,300.13	58,809.74	31,003.98	16,476.09	24,881.05
$\hat{\sigma}^2_i$	Mean	3,414.44	53,100.32	40,899.04	42,668.73	47,996.13
	Lower	904.00	34,435.53	22,746.82	23,249.89	32,528.14
	Upper	7,272.62	71,892.32	59,788.44	62,246.00	64,026.41
	Upper-lower	6,368.63	37,456.79	37,041.62	38,996.11	31,498.27
$\hat{\sigma}^2_{res}$	Mean	189,872.28	148,232.82	154,332.32	157,295.03	153,553.00
	Lower	171,535.49	131,517.69	135,587.72	137,605.69	135,787.08
	Upper	208,627.24	165,606.04	173,284.39	177,250.30	171,786.03
	Upper-lower	37,091.74	34,088.35	37,696.67	39,644.61	35,998.95
σ^2_{phen}	Mean	201,828.27	248,286.45	223,894.62	210,163.05	229,048.90
	Lower	183,376.00	213,782.90	200,279.90	188,936.00	205,227.60
	Upper	221,275.80	285,961.80	248,290.80	232,676.80	253,465.20
	Upper-lower	37,899.80	72,178.90	48,010.90	43,740.80	48,237.60
h^2_g	Mean	0.04	0.19	0.13	0.05	0.12
	Lower	0.02	0.09	0.07	0.01	0.07
	Upper	0.07	0.29	0.19	0.09	0.17
	Upper-lower	0.05	0.20	0.12	0.08	0.10
c^2_i	Mean	0.02	0.21	0.18	0.20	0.21
	Lower	0.004	0.14	0.11	0.12	0.15
	Upper	0.04	0.29	0.26	0.29	0.27
	Upper-lower	0.04	0.15	0.15	0.17	0.12
c^2_{res}	Mean	0.94	0.60	0.69	0.75	0.67
	Lower	0.89	0.43	0.55	0.62	0.56
	Upper	0.98	0.77	0.82	0.87	0.78
	Upper-lower	0.09	0.34	0.27	0.25	0.22
$r_{\hat{g}g}$		0.74	0.89	0.85	0.70	0.84

(1)$\hat{\sigma}^2_g$, genotypic variance; $\hat{\sigma}^2_i$, genotype × environment (G x E) interaction variance; $\hat{\sigma}^2_{res}$, residual variance; h^2_g, individual broad-sense heritability; c^2_i, coefficient of determination of the G × E interaction effects; c^2_{res}, coefficient of determination of residual effects; $r_{\hat{g}g}$, selective accuracy; and upper-lower, range between the upper and lower boundaries of the interval of the highest posterior density for each variance component estimate.

showing increases when the noninformative prior was used in the 2013/2014 crop year and when the informative prior was adopted in 2014/2015 (Figure 1).

The highest accuracy estimate was obtained in the 2013/2014 crop year, using a noninformative prior of 0.89. This value decreased when information from the 2012/2013 crop year was incorporated into the model (Table 3). Conversely, the accuracy rates of the model for the 2014/2015 crop year increased from 0.70 to 0.84 when information from the previous crop was used. HPD produced significance for all variance components and showed narrower intervals when informative priors were adopted.

The h_g^2 estimates presented low ($h_g^2 \leq 0.15$) to moderate ($0.15 < h_g^2 < 0.50$) magnitudes for the evaluated crops (Resende & Alves, 2020). These values are expected for grain yield in soybean, which is controlled by several genes and highly affected by the environmental effect (Assefa et al., 2019). However, the estimated heritability of 0.04 for the 2012/2013 crop year was below the expected when an informative prior is used (Azevedo et al., 2023). Moreover, the residual coefficient of determination was 0.94, evidencing the low experimental precision of the 2012/2013 crop year, which exhibited the highest phenotypic variance due to residual variance.

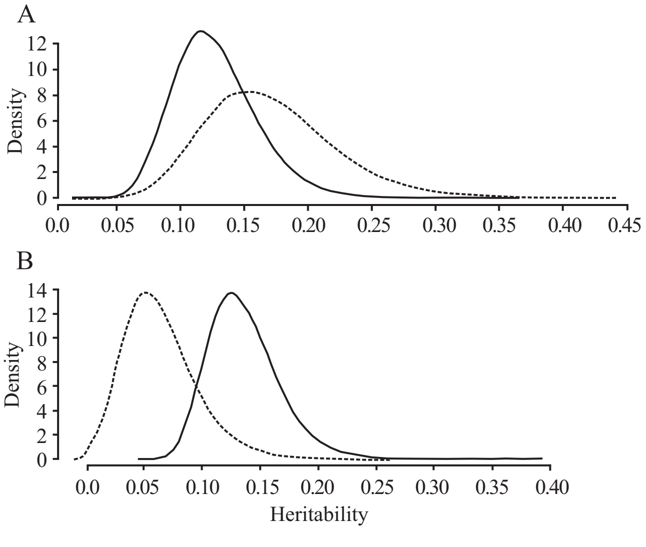

Figure 1. Posterior density of heritability for the 2013/2014 (A) and 2014/2015 (B) soybean (*Glycine max*) crop years. The solid line refers to the posterior density for the informative prior, whereas the dotted line refers to the posterior density for the noninformative prior.

As already discussed, the 2012/2013 crop year was not useful as prior information for the analysis of 2013/2014, as evidenced by DIC. Conversely, when the 2013/2014 mean posteriori variance components were used as prior information for the analysis of the 2014/2015 crop year, they increased the heritability value in comparison to the noninformative prior. In general, better results were found for variance components when an adequate prior information was adopted (Carneiro Junior et al., 2005).

Since the genetic evaluation was carried out under a genetic-statistic perspective, selective accuracy was adopted as the reliability parameter, which may inform inference reliability by measuring the correlation between estimated and real genetic values (Resende & Duarte, 2007). According to Resende & Duarte (2007), the accuracy rates of the results were classified as high ($0.70 \leq r_{\hat{g}g} < 0.90$). If decreased accuracy values are observed when an informative prior is adopted, the information added to the analysis is considered inadequate. Another scenario was observed for the 2014/2015 crop year, which showed an increased accuracy value when the informative prior using previous crop information was adopted. This result confirms that adequate informative priors may improve the reliability of genetic selection.

In Bayesian inference, the Gibbs sampling method belongs to the MCMC class and is widely used for the estimation of variance components, genetic parameters, and genotypic values. This method consists in the creation of Markov chains, in which the user defines the iteration number. In the beginning of the chain, the Gibbs sampling produces mean estimates with a considerable variation between one iteration and the iteration that follows, which will decrease as the chain extends (Hadfield, 2010). However, when an informative prior is used, the variation among iterations decreases, as well as the size of the chains necessary for convergence (Resende, 2002). Therefore, starting the process with an informative prior will reduce the number of chains in the MCMC method and result in a higher consistency (Silva et al., 2020). This may help narrow the HPD interval of the posteriori distribution when compared with that of the noninformative prior. However, in the present study, the reduced HPD interval was not evidence of the suitability of the adopted prior, since the informative prior always showed narrower intervals. Moreover, in some cases, these models were not indicated as best fit for the data. Silva et al. (2020) found that the prior adopted for yield traits was inadequate, but also observed a reduced Bayesian interval for genetic parameter estimates, as in the present study.

Conclusions

1. The inclusion of prior information into the Bayesian model does not always provide better results for the estimation of variance components and genetic values in soybean (*Glycine max*) breeding programs.
2. The addition of prior information into the Bayesian model increases its reliability.
3. The addition of prior information into the Bayesian inference framework narrows the interval of the highest posterior density.

Acknowledgments

To Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), for financing, in part, this study (Finance Code 001); to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and to Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), for support; and to the company GDM Seeds, for partnership.

References

AZEVEDO, C.F.; BARRETO, C.A.V.; SUELA, M.M.; NASCIMENTO, M.; SILVA JÚNIOR, A.C. da; NASCIMENTO, A.C.C.; CRUZ, C.D.; SORAES, P.C. Updating knowledge in estimating the genetics parameters: multi-trait and multi-environment bayesian analysis in rice. **Scientia Agricola**, v.80, e20220056, 2023. DOI: https://doi.org/10.1590/1678-992X-2022-0056.

ASSEFA, Y.; PURCELL, L.C.; SALMERON, M.; NAEVE, S.; CASTEEL, S.N.; KOVÁCS, P.; ARCHONTOULIS, S.; LICHT, M.; BELOW, F.; KANDEL, H.; LINDSEY, L.E.; GASKA, J.; CONLEY, S.; SHAPIRO, C.; ORLOWSKI, J.; GOLDEN, B.R.; KAUR, G.; SINGH, M.; THELEN, K.; LAURENZ, R.; DAVIDSON, D.; CIAMPITTI, I.A. Assessing variation in US soybean seed composition (protein and oil). **Frontiers in Plant Science**, v.10, art.298, 2019. DOI: https://doi.org/10.3389/fpls.2019.00298.

BERNARDO, R. Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE. **Heredity**, v.125, p.375-385, 2020. DOI: https://doi.org/10.1038/s41437-020-0312-1.

BLASCO, A. The Bayesian controversy in animal breeding. **Journal of Animal Science**, v.79, p.2023-2046, 2001. DOI: https://doi.org/10.2527/2001.7982023x.

CARNEIRO JÚNIOR, J.M.; ASSIS, G.M.L. de; EUCLYDES, R.F.; LOPES, P.S. Influência da informação *a priori* na avaliação genética animal utilizando dados simulados. **Revista Brasileira de Zootecnia**, v.34, p.1905-1913, 2005. DOI: https://doi.org/10.1590/S1516-35982005000600014.

COUTO, M.F.; NASCIMENTO, M.; AMARAL JR., A.T. do; SILVA, F.F. e; VIANA, A.P.; VIVAS, M. Eberhart and Russel's Bayesian method in the selection of popcorn cultivars. **Crop Science**, v.55, p.571-577, 2015. DOI: https://doi.org/10.2135/cropsci2014.07.0498.

DALLÓ, S.C.; ZDZIARSKI, A.D.; WOYANN, L.G.; MILIOLI, A.S.; ZANELLA, R.; CONTE, J.; BENIN, G. Across year and year-by-year GGE biplot analysis to evaluate soybean performance and stability in multi-environment trials. **Euphytica**, v.215, art.113, 2019. DOI: https://doi.org/10.1007/s10681-019-2438-x.

GEWEKE, J. Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments. In: BERNARDO, J.M.; BERGER, J.O.; DAWID, A.P.; SMITH, A.F.M. (Ed.). **Bayesian Statistics**: proceedings of the Fourth Valencia International Meeting, April 15-20, 1991. Oxford: Clarendon Press; New York: Oxford University Pres, 1992. v.4, p.641-649. DOI: https://doi.org/10.1093/oso/9780198522669.003.0010.

HADFIELD, J.D. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. **Journal of Statistical Software**, v.33, p.1-22, 2010. DOI: https://doi.org/10.18637/jss.v033.i02.

HUANG, W.; MACKAY, T.F.C. The genetic architecture of quantitative traits cannot be inferred from variance component analysis. **PLoS Genetics**, v.12, e1006421, 2016. DOI: https://doi.org/10.1371/journal.pgen.1006421.

KASTER, M.; FARIAS, J.R.B. **Regionalização dos testes de Valor de Cultivo e Uso e da indicação de cultivares de soja**: Terceira aproximação. Londrina: Embrapa Soja, 2012. (Embrapa Soja. Documentos, 330).

MONTESINOS-LÓPEZ, O.A.; MONTESINOS-LÓPEZ, A.; VARGAS HERNÁNDEZ, M.; ORTIZ-MONASTERIO, I.; PÉREZ-RODRÍGUEZ, P.; BURGUEÑO, J.; CROSSA, J. Multivariate Bayesian analysis of on-farm trials with multiple-trait and multiple-environment data. **Agronomy Journal**, v.111, p.2658-2669, 2019. DOI: https://doi.org/10.2134/agronj2018.06.0362.

NASCIMENTO, M.; NASCIMENTO, A.C.C.; SILVA, F.F. e; TEODORO, P.E.; AZEVEDO, C.F.; OLIVEIRA, T.R.A. de; AMARAL JUNIOR, A.T. do; CRUZ, C.D.; FARIAS, F.J.C.; CARVALHO, L.P. de. Bayesian segmented regression model for adaptability and stability evaluation of cotton genotypes. **Euphytica**, v.216, art.30, 2020. DOI: https://doi.org/10.1007/s10681-020-2564-5.

RAFTERY, A.E.; LEWIS, S. **How many iterations in the Gibbs sampler**? In: BERNARDO, J.M.; BERGER, J.O.; DAWID, A.P.; SMITH, A.F.M. (Ed.). **Bayesian Statistics**: proceedings of the Fourth Valencia International Meeting, April 15-20, 1991. Oxford: Clarendon Press; New York: Oxford University Pres, 1992. v.4, p.763-774. DOI: https://doi.org/10.1093/oso/9780198522669.003.0053.

RESENDE, M.D.V. de. **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília: Embrapa Informação Tecnológica; Colombo: Embrapa Florestas, 2002.

RESENDE, M.D.V. de. **Inferência bayesiana e simulação estocástica (amostragem de Gibbs) na estimação de componentes de variância e valores genéticos em plantas perenes**. Colombo: Embrapa Florestas, 2000. (Embrapa Florestas. Documentos, 46).

RESENDE, M.D.V. de; ALVES, R.S. Linear, generalized, hierarchical, Bayesian and random regression mixed models in genetics/genomics in plant breeding. **Functional Plant Breeding Journal**, v.3, art.11, 2020. DOI: https://doi.org/10.35418/2526-4117/v2n2a1.

RESENDE, M.D.V. de; DUARTE, J.B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical**, v.37, p.182-194, 2007.

RESENDE, M.D.V. de; SILVA, F.F. e; AZEVEDO, C.F. **Estatística matemática, biométrica e computacional**: modelos mistos, multivariados, categóricos e generalizados (REML/BLUP), inferência bayesiana, regressão aleatória, seleção genômica, QTL-QWAS, estatística espacial e temporal, competição, sobrevivência. Viçosa: UFV, 2014. 882p.

REZENDE, W.S.; CRUZ, C.D.; BORÉM, A.; ROSADO, R.D.S. Half a century of studying adaptability and stability in maize and soybean in Brazil. **Scientia Agricola**, v.78, e20190197, 2021. DOI: https://doi.org/10.1590/1678-992x-2019-0197.

SILVA, F.A. da; VIANA, A.P.; CORRÊA, C.C.G.; CARVALHO, B.M.; SOUSA, C.M.B. de; AMARAL, B.D.; AMBRÓSIO, M.; GLÓRIA, L.S. Impact of Bayesian inference on the selection of *Psidium guajava*. **Scientific Reports**, v.10, art.1999, 2020. DOI: https://doi.org/10.1038/s41598-020-58850-6.

SILVA, F.F. e; VIANA, J.M.S.; FARIA, V.R.; RESENDE, M.D.V. de. Bayesian inference of mixed models in quantitative genetics of crop species. **Theoretical and Applied Genetics**, v.126, p.1749-1761, 2013. DOI: https://doi.org/10.1007/s00122-013-2089-6.

SILVA, F.L.; BORÉM, A.; SEDIYAMA, T.; CÂMARA, G. (Org.). **Soja**: do plantio à colheita. 2.ed. São Paulo: Oficina de Textos, 2022. 312p.

SMITH, B.J. boa: an R package for MCMC output convergence assessment and posterior inference. **Journal of Statistical Software**, v.21, p.1-37, 2007. DOI: https://doi.org/10.18637/jss.v021.i11.

SORENSEN, D. Developments in statistical analysis in quantitative genetics. **Genetica**, v.136, p.319-332, 2009. DOI: https://doi.org/10.1007/s10709-008-9303-5.

SORENSEN, D.; GIANOLA, D. **Likelihood, bayesian, and MCMC methods in quantitative genetics**. New York: Springer Science and Business Media, 2002.

SPIEGELHALTER, D.J.; BEST, N.G.; CARLIN, B.P.; VAN DER LINDE, A. Bayesian measures of model complexity and fit. **Journal of the Royal Statistical Society: Series B: Statistical Methodology**, v.64, p.583-639, 2002. DOI: https://doi.org/10.1111/1467-9868.00353.

TORRES, L.G.; RODRIGUES, M.C.; LIMA, N.L.; TRINDADE, T.F.H.; SILVA, F.F. e; AZEVEDO, C.F.; DeLIMA, R.O. Multi-trait multi-environment Bayesian model reveals G x E interaction for nitrogen use efficiency components in tropical maize. **PloS One**, v.13, e0199492, 2018. DOI: https://doi.org/10.1371/journal.pone.0199492.

VOLPATO, L.; ALVES, R.S.; TEODORO, P.E.; RESENDE, M.D.V. de; NASCIMENTO, M.; NASCIMENTO, A.C.C.; LUDKE, W.H.; SILVA, F.L. da; BORÉM, A. Multi-trait multi-environment models in the genetic selection of segregating soybean progeny. **PLoS ONE**, v.14, e0215315, 2019. DOI: https://doi.org/10.1371/journal.pone.0215315.

WAKEFIELD, J. **Bayesian and frequentist regression methods**. New York: Springer, 2013. DOI: https://doi.org/10.1007/978-1-4419-0925-1.

WOYANN, L.G.; MEIRA, D.; MATEI, G.; ZDZIARSKI, A.D.; DALLACORTE, L.V.; MADELLA, L.A.; BENIN, G. Selection indexes based on linear-bilinear models applied to soybean breeding. **Agronomy Journal**, v.112, p.175-182, 2020. DOI: https://doi.org/10.1002/agj2.20044.