Row-col method associated with frequentist and Bayesian statistics in a passion fruit population

André Oliveira Souza[1], Alexandre Pio Viana[2], Fabyano Fonseca e Silva[3], Camila Ferreira Azevedo[4], Natan Ramos Cavalcante[2] and Flavia Alves da Silva[2*]

Abstract: *This study was conducted to test the significance of adding row and column factors in the frequentist and Bayesian models used in the evaluation of a population of Passiflora edulis, as well as selecting promising genotypes to form the next generation. The following parameters were evaluated: number of fruits, yield, fruit weight, transverse fruit diameter, longitudinal fruit diameter, pulp percentage, skin thickness and total soluble solids. For the Bayesian model, two priors were considered, namely, inverse gamma and a priori distribution with extended parameters. The model with a priori distribution with extended parameters showed lower root mean square error and higher correlation coefficient between observed and predicted values than the inverse gamma model. Furthermore, for a selection intensity of 37%, the mixed and Bayesian models selected practically the same progenies in both experiments. The use of the 5-fold cross-validation technique indicated that both tested models were efficient.*

Keywords: *REML, prior, post-hoc blocking Row-Col*

INTRODUCTION

Brazil is considered the world's largest passion fruit producer, generating 602 thousand tons of the fruit from a cultivated area of approximately 43,000 ha, which amounts to an average yield of 14 t ha^{-1}. These estimates indicate the production potential and the economic importance of this crop for the country. About 62% of national production comes from the northeast, followed by the southeast, south, north, and center-west regions. However, when it comes to yield, the south region is the leader, accounting for about 25% of domestic production (IBGE 2018).

The passion fruit yield in the state of Rio de Janeiro was 6,000 t ha^{-1}, which can be considered low compared with the 16,000 t ha^{-1} of Bahia state, the largest producer. This low yield is due, among other factors, to the lack of high-yielding genotypes, large variability in commercial orchards, lack of improved cultivars adapted to the soil, and climatic and environmental conditions of the growing region (Gonçalves et al. 2007). In the current Brazilian scenario, yield and fruit quality, as well as disease control in passion fruit cultivation, have become extremely important issues to ensure the sustainability of this crop. Unfortunately, these factors have contributed to a reduction in average yield in recent years in Brazil. Given the complexity of this scenario, breeding programs have employed a variety of methods to address these challenges.
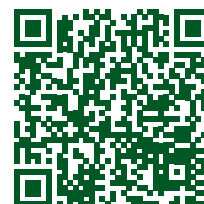
*Corresponding author:
E-mail: flavia_uems@hotmail.com
ORCID: 0000-0001-7818-707X

[1] Instituto Federal do Espírito Santo, Rodovia ES-482, km 72, Rive, 29500-000, Alegre, ES, Brazil
[2] Universidade Estadual do Norte Fluminense Darcy Ribeiro, Avenida Alberto Lamego, 2000, Parque California, 28013-602, Campos dos Goytacazes, RJ, Brazil
[3] Universidade Federal do Espírito Santo, Avenida Fernando Ferrari, 514, Goiabeiras, 29060-900, Vitória, ES, Brazil
[4] Universidade Federal de Viçosa, Avenida P H Rolfs, s/n, Campus Universitário, 36570-900, Viçosa, MG, Brazil

In the field of passion fruit breeding, several methods are used to deal with spatial variation problems, especially when it comes to experiments. An important example is the mixed model methodology (REML/BLUP), which plays a crucial role in this context (Viana and Resende 2014). This approach has become standard for estimating genetic parameters and predicting genotypic values, being widely adopted in several breeding programs. The work carried out by Santos et al. (2015) with individuals from a segregating population from an interspecific cross of passion fruit showed that the REML/BLUP method was effective in identifying superior genotypes.

In turn, the Bayesian approach has advantages such as allowing the incorporation of previous distributions into the model, especially when the availability of information is limited (Silva et al. 2018). In addition to these techniques, post-hoc blocking Row-Col emerges as another promising strategy to control the local effect. This approach involves superimposing a structure of rows and columns over the original layout, generated in estimates of model parameters with greater precision (Machado et al. 2020).

However, using effective methods goes beyond that. It is necessary to consider the combination of appropriate approaches and techniques to face the specific challenges related to yield, fruit quality and disease control in passion fruit cultivation. This integration of methods, combined with a holistic approach, is key to driving progress in improving this crop and ensuring its long-term sustainability. In this context, the present study was conducted with the objective of estimating the components of variance, heritability and selection gain through the mixed model and Bayesian methods. In addition, we sought to evaluate the application of the post-hoc blocking Row-Col technique, using chi-square statistics, for real data variables from passion fruit breeding experiments.

MATERIAL AND METHODS

A total of 97 full-sib progenies from the fourth cycle of recurrent selection of the UENF breeding program were evaluated. These progenies were obtained by crossing 23 half-sib progenies selected by Cavalcante et al. (2019). The experiment was conducted on Santo Antão farm, belonging to Instituto Federal Fluminense, in Cambuci - RJ, Brazil (lat 21° 34' S, long 41° 54' W, alt 35 m asl). The experiment was laid out in a randomized block design with two replicates, where each plot consisted of three plants (582 plants). To work with the post-hoc blocking Row-Col technique, a grid with 19 rows and 15 columns was implemented over the experimental layout.

The following variables were evaluated: number of fruits (NF), by counting the total number of fruits in each plot; total yield (YLD, kg plot^{-1}), determined by measuring the amount harvested during the experiment; fruit weight (FW, g), calculated as the arithmetic mean of the weight of 15 fruits sampled per plot; fruit longitudinal diameter (LGT, mm); fruit transverse diameter (TVS, mm); fruit pulp percentage (PP, %), obtained as the ratio between pulp weight and total fruit weight; peel thickness (PT, mm), determined in the median portion of fruits cut transversally, along the larger diameter; total soluble solids content (TSS, °Brix), determined with an ATAGO N1 digital handheld refractometer, with reading performed in the range from 0 to 95 °Brix in aliquots of pulp juice.

Models

To model the variables, two models were proposed. The first model used was a mixed linear model: $y = 1\mu + Z_1u_1 + Z_2u_2 + \varepsilon$, where Y_{nx1} is the vector of phenotypic observations; 1_{nx1} is a column matrix of ones; u_1 is the vector of random effects for blocks associated with incidence matrix Z_1, with probability distribution $u_1 \sim N(0, I\sigma^2_{block})$; u_2 is the vector of random effects for progeny associated with incidence matrix Z_2, with probability distribution $u_2 \sim N(0, I\sigma^2_{progeny})$; ε is the vector of random errors, with probability distribution $\varepsilon \sim N(0, I\sigma^2_{error})$, in which $\varepsilon | \sigma^2_{error} \sim N(0, I\sigma^2_{error})$; and the joint distribution of the observed data is given by $E(y) \sim N(1\mu + Z_1u_1 + Z_2u_2, I\sigma^2_e)$, here referred to as "reduced model".

Subsequently, the post-hoc blocking Row-Col method was applied on the reduced model by tracing a grid with 13 rows and 18 columns over the experimental layout, thus modifying the model, which was then termed "full model": $y = 1\mu + Z_1u_1 + Z_2u_2 + Z_3u_3 + Z_4u_4 + \varepsilon$, where Y_{nx1} is the vector of phenotypic observations; $1_{n \times 1}$ is a column matrix of value 1; u_1 is the vector of random effects for blocks associated with incidence matrix Z_1, with probability distribution $u_1 \sim N(0, I\sigma^2_{block})$; u_2 is the vector of random effects for progeny associated with incidence matrix Z_2, with probability $u_2 \sim N(0, I\sigma^2_{progeny})$; u_3 is the vector of random effects for row associated with incidence matrix Z_3, with probability distribution $u_3 \sim N(0, I\sigma^2_{row})$; u_4 is the vector of random effects for column associated with incidence matrix Z_4, with probability distribution $u_4 \sim N(0,$

$I\sigma^2_{column}$) ; ε is the vector of random errors, with probability distribution $\varepsilon \sim N(0, I\sigma^2_{error})$; and the joint distribution of the observed data is given by $E(y) \sim N(1\mu + Z_1u_1 + Z_2u_2 + Z_3u_3 + Z_4u_4, \sigma^2_2)$. The models were compared by the likelihood ratio test (LRT) using chi-square statistics, adopting two degrees of freedom and a 5% probability level. After both models were tested under the frequentist approach, they were also tested under the Bayesian approach. The joint distribution of the data observed under this approach was obtained from the following distributions:

$$y\,|\,u_1, u_2, u_3, u_4, \sigma^2_{progeny}, \sigma^2_{block}, \sigma^2_{row}, \sigma^2_{column}, \sigma^2_{error} \sim N(1\mu + Z_1u_1 + Z_2u_2 + Z_3u_3 + Z_4u_4, I\sigma^2_e)$$

where $\mu \sim N(\mu_0, \sigma^2_0)$; $u_1\,|\,\sigma^2_{progeny} \sim N(0, I\sigma^2_{progeny})$; $\sigma^2_{progeny} \sim IG(v = v_1, V = V_1)$; $u_2\,|\,\sigma^2_{block} \sim N(0, I\sigma^2_{block})$; $\sigma^2_{block} \sim IG(v = v_2, V = V_2)$; $u_3\,|\,\sigma^2_{column} \sim N(0, I\sigma^2_{column})$; $\sigma^2_{column} \sim IG(v = v_2, V = V_3)$; $u_4\,|\,\sigma^2_{row} \sim N(0, I\sigma^2_{row})$; $\sigma^2_{row} \sim IG(v = v_2, V = V_4)$; and $\sigma^2_{error} \sim IG(v = v_0, V = V_0)$. This was termed "full Bayesian model". For the reduced model, as in the frequentist approach, only the effects of block, progeny, and error were maintained, with the same distributions of the variances described above. To improve the estimates in the Bayesian approach, two priors were defined for the parameters, here called Bayes I (v=1; V=0.002; alpha.mu=0; alpha.V=0; Classification μ=Inverse Gamma (IG) N[0.10^8]) and Bayes II (v=1; V=1; alpha.mu=0; alpha.V=252; Classification μ= Half-Cauchy (HC) N[0.10^8]).

The choice of prior was based on the smallest square root of the mean squared error and on the correlation between what was observed and predicted by the model, so prior II was chosen for presenting such more accurate metrics. It is noteworthy that the non-informative a priori models did not converge, as some variance components are low. The work included 5-fold cross-validation. The metrics root mean square error of the cross validation $\left(predictive\ efficiency\ PE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N}}\right)$ and predictive ability coefficient $PC = 100.\left(\frac{cov(y_{predicted}; y_{observed})}{S_{ypredicted} \cdot S_{yobserved}}\right)$ were considered as described by Silva al. (2018). Bayes prior II was chosen because it showed more accurate metrics than prior I. The choice of prior distribution was based on the suggestion of Hadfield (2021), adding the extension parameters alpha.mu and alphaV. The choice of the model (prior) under the Bayesian approach was made considering the smallest root mean square error (RMSE), based on the smallest value of Deviance Information Criterion (DIC) between the models. The models were considered equal when $|\Delta| < 2$ between their DIC (Spiegelhalter et al. 2002). The predictive ability of the model was also considered in a 5-fold cross-validation. In each fold, randomly, 80% of the data was used for modeling, and 20% of the data was used to observe the discrepancy with the model prediction when choosing the prior. Thus, the average correlation between the predicted dependent variables and those observed for the cross-validation folds was considered. This same model choice approach was repeated when the frequentist and Bayesian models were compared, by performing a new cross-validation, but comparing the models by predictive ability and mean square error.

To verify the effects of row and column factors, each model was fitted with and without the technique (post-hoc blocking row-column). The comparison between the models was performed using the likelihood ratio test (LRT) using the statistic $LRT = 2[ln(Model\ of\ interest) - ln(Model\ under\ Ho)]$, where LRT under Ho has an approximate Chi-Square distribution with 2 (two) degrees of freedom. As for Bayesian models, the choice and comparison of models was made by the Deviance Information Criterion (DIC) with several iterations equal to 7000000, thinning = 20, burn-in = 200000. To evaluate the percentage (index) of coincidences of the selected progenies, the common percentage of progenies was selected by the two estimation methodologies for each of the three levels of selection intensity: 37%, 25% and 15%.

All statistical analyses were performed using the R software (R Core Team 2019) using the Sommer (Covarrubias 2016, 2018), MCMCglmm (Hadfield 2010, Hadfield and Nakagawa 2010) and Imer4 (Bates et al. 2014) libraries. The convergence of the MCMC chains was checked by the criterion of Geweke (1992) using the Coda package (convergence diagnosis and output analysis for MCMC, Plummer et al. 2006) and the DCCPs were obtained according to Hadfield (2010).

Genetic parameters

Some genetic parameters were estimated by both approaches, using the full or reduced model, according to their fit for each variable. If the full model was significantly different from the reduced model for a given variable, this full model was used to estimate the genetic parameters. Broad-sense heritability ($\hat{h}^2$) was estimated as shown below: $\hat{h}^2 = \frac{\hat{\sigma}^2_{progeny}}{\hat{\sigma}^2_{progeny} + \hat{\sigma}^2_{block} + \hat{\sigma}^2_{column} + \hat{\sigma}^2_{row} + \hat{\sigma}^2_{error}}$, where ($\hat{\sigma}^2_{error}$) is the estimate of residual variance; ($\hat{\sigma}^2_{block}$) is the estimate of variance due to the block factor; ($\hat{\sigma}^2_{column}$) is the estimate of variance due to the column factor; ($\hat{\sigma}^2_{row}$) is the estimate of

the variance due to the row factor; and $\hat{\sigma}^2_{progeny}$ is the estimate of genotypic variance.

Confidence intervals for heritability were obtained for the frequentist approach (CI-$\hat{h}^2$) and the credible interval for heritability was obtained for the Bayesian approach. At the end, a ranking was created by selecting 30 individuals for which selection gain was obtained by the delta method, based on: $G_S = 100 \cdot \left(\dfrac{\overline{X}_{30} - \overline{X}_G}{\overline{X}_G} \right)$, where selection gain is relative to the overall mean ($\overline{X}_G$) and the mean of the first 30 (thirty) progenies selected ($\overline{X}_{30}$).

RESULTS AND DISCUSSION

The significance of the factors (RCB + Row-Col) was tested by the likelihood ratio test (LRT) using chi-square statistics (Table 1). The full model was significant only for the production variables (NF and YLD), indicating that it had a better fit for these two variables (p<0.05). On the other hand, the Deviance Information Criterion (DIC) (Table 2) shows that the full model had a better fit for all variables, except PT and TSS.

Table 1. Likelihood ratio test (LRT) for comparison between the two proposed models: randomized complete blocks (RCB) and post-hoc blocking Row-Col

Variable	Model	Deviance	df	Statistics (D)	*p-value*
YLD	RCB + Row-Col	1289.00	2	33.50*	<0.001
	RCB	1322.50			
NF	RCB + Row-Col	2051.00	2	29.00*	<0.001
	RCB	2022.00			
FW	RCB + Row-Col	1789.10	2	2.20	0.3343
	RCB	1791.30			
TVS	RCB + Row-Col	1212.10	2	5.10	0.0772
	RCB	1217.20			
LGT	RCB + Row-Col	1075.80	2	1.79	0.6353
	RCB	1074.01			
PT	RCB + Row-Col	618.07	2	0.00	1.000
	RCB	618.07			
PP	RCB + Row-Col	1156.60	2	2.30	0.3257
	RCB	1158.90			
TSS	RCB + Row-Col	622.16	2	0.06	0.9722
	RCB	622.22			

Source: developed by the authors. (*) significant at the 5% probability leavel by the LRT; Yield (**YLD**), Number of Fruits (**NF**), Fruit Weight (**FW**), Fruit Transverse Diameter (**TVS**), Fruit Longitudinal Diameter (**LGT**), Peel Thickness (**PT**), Pulp Percentage (**PP**), and Total Soluble Solids (**TSS**).

Table 2. Comparison of the Bayesian models in randomized complete blocks (RCB) and Row-Col by Deviance Information Criterion (DIC) – Bayes

Variable	Model	DIC	W	Δ	Convergence	
					Geweke (1992)	Heidenberg & Welch (1983)
NF	RCB + Row-Col	2005.57	1.00	-48.30[d]	Converged	Converged
	RCB	2053.87	0.00		Converged	Converged
YLD	RCB + Row-Col	1271.72	1.00	-54.99[d]	Converged	Converged
	RCB	1326.71	0.00		Converged	Converged
FW	RCB + Row-Col	1789.65	0.85	-3.43[d]	Converged	Converged
	RCB	1793.08	0.15		Converged	Converged
TVS	RCB + Row-Col	1206.08	0.98	-7.59[d]	Converged	Converged
	RCB	1213.67	0.02		Converged	Converged
LGT	RCB + Row-Col	1073.92	0.60	0.76[i]	Converged	Converged
	RCB	1074.68	0.40		Converged	Converged
PP	RCB + Row-Col	1158.35	0.84	-3.37[d]	Converged	Converged
	RCB	1161.72	0.16		Converged	Converged
PT	RCB + Row-Col	609.87	0.17	3.22[d]	Converged	Converged
	RCB	606.65	0.83		Converged	Converged
TSS	RCB + Row-Col	671.18	0.25	2.15[d]	Converged	Converged
	RCB	669.03	0.75		Converged	Converged

Source: developed by the authors. Number of Fruits (**NF**), Yield (**YLD**), Fruit Weight (**FW**), Fruit Transverse Diameter (**TVS**), Fruit Longitudinal Diameter (**LGT**), Pulp Percentage (**PP**), Peel Thickness (**PT**), and Total Soluble Solids (**TSS**). Δ = DIC$_r$ – DIC$_{R'}$ where DIC$_c$ is the Deviance Information Criterion of the full model, DIC$_R$ Deviance Information Criterion of the reduced model. Prior as suggested by Hadfield, Jarrod (2021) with addition of the extension parameters alpha.mu and alpha.V: ([i]) Models are equal when |Δ| < 2 (Spiegelhalter et al. 2002) ([d]) Models are different when |Δ| ≥ 2 (Spiegelhalter et al. 2002) The choice between the two models was based on the lower value of DIC. W is the posterior probability of the model (Wilberg and Bence 2008).

With respect to the mixed-model approach, Silva et al. (2016) found the same results when the post-hoc blocking Row-Col technique was added to the model, that is, only the YLD and NF variables showed significance (p<0.05). Because both variables result from the total sum of the plot and are not sampled in fruits with a replicate, a variance was captured by Row-Col, which was not necessary when there were more observations in the other variables. Gezan et al. (2006) and Machado et al. (2020), compared the efficiency of Row-Col with the randomized block design and obtained more accurate estimates of heritability using the former. However, in general, the randomized block design is efficient when the variability within the replicates is relatively small. This is a rare occurrence when a high number of genotypes is evaluated, especially in large-scale experiments, as is the present case. In this study, the post-hoc blocking Row-Col technique was more efficient to estimate genetic parameters, since when the observations were modeled via Bayesian models, only the variables TVS, PT, and TSS showed no significance for the RCB + Row-Col factor.

In this case, as it is the fourth cycle of recurrent selection, the estimates of genetic parameters (Tables 3 and 4) reveal that the greatest selection gains were obtained for YLD and NF. By the frequentist method, they also reveal the non-existence of genetic variability (or not captured by the model) and little selective potential among the passion fruit progenies under study, which is confirmed by the frequentist confidence intervals for heritability containing zero. This result could also be due to high environmental variation. This makes it difficult to fit a model properly, especially when this environmental variation results in data with a large standard deviation. Thus, the passion fruit breeder who sporadically may face high environmental variability, such as long periods of drought, high temperatures, and presence of diseases like CABMV, can opt for a model that, although more complex and computationally costly, will provide more accurate estimates.

Table 3. Estimation of genetic parameters of 97 passion fruit progenies from the fourth cycle of recurrent selection via mixed models (REML/BLUP)

Variable	$\hat{\sigma}^2_{progeny}$	$\hat{\sigma}^2_{block}$	$\hat{\sigma}^2_{row}$	$\hat{\sigma}^2_{column}$	$\hat{\sigma}^2_{error}$	$\hat{h}^2$	Gs (%)	CI-$\hat{h}^2$(95%)
YLD	4.50	1.26	24.01	0.00	34.21	0.07	5.89	-0.06; 0.20
NF	197.31	89.17	947.14	16.58	1486.45	0.07	6.06	-0.06; 0.20
FW	30.98	31.89	-	-	561.98	0.05	1.30	-0.14; 0.24
LGT	6.28	1.65	-	-	25.07	0.19	2.22	0.01; 0.38
TVS	1.26	1.73	-	-	13.45	0.08	0.73	-0.10; 0.26
PP	3.50	0.00	-	-	19.87	0.15	2.24	-0.04; 0.35
PT	0.41	0.13	-	-	1.04	0.26	-6.54	0.08; 0.44
TSS	0.28	0.00	-	-	1.57	0.15	2.41	-0.04; 0.35

Yield (**YLD**), Number of Fruits (**NF**), Fruit Weight (**FW**), Fruit Longitudinal Diameter (**LGT**), Fruit Transverse Diameter (**TVS**), Pulp Percentage (**PP**), Peel Thickness (**PT**), and Total Soluble Solids (**TSS**). ($\hat{h}^2$) is the point estimate of broad-sense heritability $\left(\hat{h}^2 = \dfrac{\hat{\sigma}^2_{progeny}}{\hat{\sigma}^2_{progeny}+\hat{\sigma}^2_{block}+\hat{\sigma}^2_{column}+\hat{\sigma}^2_{row}+\hat{\sigma}^2_{error}}\right)$; ($\hat{\sigma}^2_{error}$) is the estimate of residual variance; ($\hat{\sigma}^2_{block}$) is the estimate of variance due to the block factor; ($\hat{\sigma}^2_{column}$) is the estimate of variance due to the column factor; ($\hat{\sigma}^2_{row}$) is the estimate of variance due to the row factor; $\hat{\sigma}^2_{progeny}$ is the estimate of genetic variance; CI-$\hat{h}^2$ is the approximation of the confidence interval for heritability by the delta method; and $G_s = 100. \left(\dfrac{\overline{X}_{30}-\overline{X}_G}{\overline{X}_G}\right)$ is the selection gain relative to the overall mean ($\overline{X}_G$) and the mean of the 30 (thirty) first selected progenies ($\overline{X}_{30}$).

Table 4. Estimation of genetic parameters of 97 passion fruit progenies from the fourth cycle of recurrent selection- Bayes

Variable	$\hat{\sigma}^2_{progeny}$	$\hat{\sigma}^2_{block}$	$\hat{\sigma}^2_{row}$	$\hat{\sigma}^2_{column}$	$\hat{\sigma}^2_{error}$	$\hat{h}^2$	Gs (%)	HPD-95%
NF	170.90	2593.00	1098.00	55.46	1533.00	0.05	6.06	0.00; 0.15
YLD	4.75	1332.00	32.68	0.93	34.51	0.04	5.67	0.00; 0.13
FW	51.89	2847.00	44.91	26.10	515.50	0.06	1.99	0.00; 0.18
TVS	6.39	986.30	1.14	3.08	23.37	0.11	2.22	0.00; 0.28
LGT	1.35	13.42	-	-	13.61	0.04	0.72	0.00; 0.15
PP	3.26	11.50	1.82	0.72	19.27	0.08	1.96	0.00; 0.24
PT	0.41	682.00	-	-	1.08	0.09	-6.27	0.00; 0.30
TSS	0.26	1034.00	-	-	1.63	0.07	2.04	0.00; 0.24

Number of Fruits (**NF**), Yield (**YLD**), Fruit Weight (**FW**), Fruit Transverse Diameter (**TVS**), Fruit Longitudinal Diameter (**LGT**), Pulp Percentage (**PP**), Peel Thickness (**PT**), and Total Soluble Solids (**TSS**). ($\hat{h}^2$) is the estimate of the heritability obtained by the mean of the posterior distribution; ($\hat{\sigma}^2_{error}$) is the estimate of residual variance; ($\hat{\sigma}^2_{block}$) is the estimate of variance due to the block factor; ($\hat{\sigma}^2_{column}$) is the estimate of variance due to the column factor; and ($\hat{\sigma}^2_{row}$) is the estimate of variance due to the row factor.

For example, by comparing the estimates between the frequentist model (which is relatively simpler and less costly) against the Bayesian model, one can see that the highest posterior density (HPD) values had smaller amplitudes when compared to the confidence interval (CI) (Tables 3 and 4). This occurs due to the parametric space assumed by prior II. However, it is important to note that the selection gain estimates were very similar to those of the frequentist method, so it is the responsibility of the breeder to choose which parameters are of interest to him in order to choose which model will provide an approximate estimate, and a more accurate one.

Under the mixed-model approach, heritability estimates ranged from 0.05 to 0.26, with PT showing the highest value (0.26). Using the same experimental design, Silva et al. (2016) obtained divergent estimates in full-sib families of the same crop, describing values that ranged from 0.14 to 0.53. Selection gain ranged from -6.54% to 6.06%, with the highest results achieved by NF and YLD (6.06% and 5.89%, respectively). Under the Bayesian approach, these heritability estimates ranged from 0.04 to 0.11 and selection gain from -6.27% to 6.06%. Overall, the magnitude of the heritability estimates did not coincide between the two estimation methods, but selection gain estimates were

Table 5. Ranking of the first 30 passion fruit progenies from full-sibs of the fourth cycle of recurrent selection via mixed models (REML/BLUP, the first trait) and Bayesian models (the second trait)

N	NF	NF	YLD	YLD	FW	FW	LGT	LGT	TVS	TVS	PP	PP	PT	PT	TSS	TSS
1	23	23	67	67	89	89	32	32	32	32	49	49	45	45	90	90
2	9	9	85	85	40	99	40	59	14	14	41	41	49	49	45	45
3	35	35	23	23	99	60	90	90	99	99	44	45	31	31	72	72
4	71	71	90	90	60	6	59	40	89	89	72	72	41	41	3	3
5	90	90	9	9	6	40	99	89	92	92	23	23	100	100	11	11
6	67	67	71	71	32	32	72	6	98	98	45	44	92	92	34	34
7	70	70	40	4	72	72	6	72	72	72	31	89	9	9	21	21
8	88	88	4	40	92	14	89	38	40	40	51	51	29	29	55	55
9	49	49	35	35	59	92	76	95	76	76	38	30	47	47	19	19
10	48	48	48	46	58	59	95	99	95	95	43	31	73	73	41	41
11	85	85	46	48	14	4	33	21	3	3	30	38	40	40	31	31
12	12	46	88	19	4	58	21	14	6	6	59	27	3	3	100	100
13	46	19	58	58	16	39	31	76	51	51	92	92	23	23	98	98
14	19	63	19	88	85	3	38	39	81	81	61	91	91	91	92	92
15	63	12	62	62	41	41	58	4	85	85	9	61	69	69	27	27
16	65	65	25	25	3	85	39	98	41	41	89	84	1	1	37	37
17	40	91	49	63	39	16	61	8	59	59	91	59	59	59	52	52
18	91	40	65	65	35	48	4	58	96	96	27	43	80	80	68	68
19	55	55	63	49	90	21	3	9	25	25	98	54	30	30	64	64
20	28	28	12	28	48	50	14	3	69	16	100	52	12	12	18	18
21	4	4	89	89	98	100	98	65	16	69	54	9	68	68	43	43
22	20	52	28	27	50	98	8	31	60	60	4	98	87	87	53	53
23	25	25	27	52	100	76	9	33	4	4	84	11	22	22	89	89
24	52	58	22	12	21	90	88	61	58	58	47	4	84	84	96	96
25	58	20	3	22	76	35	65	66	83	83	11	100	24	24	33	33
26	62	62	52	70	52	52	34	23	66	22	55	93	98	79	78	78
27	15	15	39	3	64	64	92	88	22	66	93	24	79	98	87	87
28	8	8	70	39	83	83	50	85	33	33	24	57	61	61	35	35
29	72	93	6	6	96	56	23	50	100	100	57	20	89	89	47	47
30	22	68	55	55	56	96	75	92	39	39	58	47	70	70	51	51
$\bar{X}_G$	116.4	116.4	18.4	18.4	158.6	159.2	81.0	81.0	72.2	72.3	46.8	46.9	7.3	7.3	13.4	13.4
$\bar{X}_{30}$	123.5	123.5	19.4	19.5	160.6	162.3	82.8	82.8	72.8	72.8	47.9	47.8	7.8	6.9	13.7	13.7

Number of Fruits (**NF**), Yield (**YLD**), Fruit Weight (**FW**), Fruit Longitudinal Diameter (**LGT**), Fruit Transverse Diameter (**TVS**), Pulp Percentage (**PP**), Peel Thickness (**PT**), and Total Soluble Solids (**TSS**). Progeny ranking formed from the mean value added to the predicted genotypic value (u + g); $\bar{X}_G$ is the overall mean and $\bar{X}_{30}$ is the mean of the first 30 (thirty) selected progenies.

Table 6. Estimates of metrics obtained by the models in 5-fold cross-validation

Variable	Model	$PA_{training}$	$PA_{validation}$	$RMSE_{training}$	$RMSE_{validation}$
Number of fruits	Mixed	0.74	0.39	33.44	**44.43**
	Bayesian	0.75	0.39	33.74	44.23
Yield	**Mixed**	0.74	0.40	5.15	6.69
	Bayesian	0.75	0.40	5.07	6.68
Fruit weight	**Mixed**	0.70	0.07	20.43	**24.85**
	Bayesian	0.71	0.13	20.21	24.52
FLD/Length	**Mixed**	0.83	0.26	4.05	**5.46**
	Bayesian	0.82	0.26	4.10	5.46
FTD/Diameter	**Mixed**	0.51	0.20	3.50	**3.82**
	Bayesian	0.57	0.22	3.44	3.81
Peel thickness	Mixed	0.81	0.31	0.89	**1.18**
	Bayesian	0.78	0.31	0.92	1.18
Pulp percentage	**Mixed**	0.78	0.10	3.75	4.99
	Bayesian	0.79	0.06	3.72	5.01
Total soluble solids	**Mixed**	0.79	0.14	1.19	1.35
	Bayesian	0.78	0.12	1.19	1.34

Predictive ability of training ($PA_{training}$); Predictive ability of validation ($PA_{validation}$); Root means square error of training ($RMSE_{training}$); Root mean square error of validation ($RMSE_{validation}$).

close. The heritability estimates found for all traits are much lower than those described in the literature, which may be due to high environmental/experimental management influence or the polygenic nature of these variables. Heritability estimates were also lower than those reported by Viana et al. (2004), which is explained by the fact that this is the fourth cycle of recurrent selection, when the variables are homogeneous for the desirable traits.

The analysis of the ranking of the first 30 selected progenies, according to Table 5, reveals consistency between the two winning methods. Differences in the selection of these progenies were not compatible. The coincidence rate between the progenies selected by the two methods ranged from 88.33% to 100%. When analyzing the 30 best progenies selected for each characteristic, progenies 89, 98, 92, 72, 58, 40, 4, 3, 100, 90, 59, 41, 23 and 9 form a set that shows a high genetic probability of being the best progenies in all evaluated traits. It is noteworthy that these results differ from those found by Silva et al. (2020) in guava crop, for which distinct sets of individuals were selected through the methodologies employed.

To evaluate the accuracy, reliability, and efficiency of the models, the study included the investigation of some statistics (metrics) that provide information on the selective and predictive accuracies of the models by the 5-fold method. Overall, the predictive ability (PA) and predictive efficiency (RMSE) metrics for each variable (Table 6) have practically the same magnitude by the two models, i.e., both have the same generalization ability. In the validation and training stages of cross-validation (5-fold), the models exhibited similar predictive ability for all variables, except FW PP, and TSS. The predictive ability of the models for the variables ranged from 6% to 40%.

CONCLUSION

From the standpoint of the frequentist and Bayesian models, the post-hoc blocking Row-Col technique captured the existing variability for the variables yield and number of fruits, for which it had a direct impact on experimental precision.

The tested models select practically the same individuals, which shows that the two estimation methods are equivalent for this structure of observations.

Overall, the parameterization of priors reduced the root mean square error, increased the correlation between observed and predicted values, and decreased the deviance information criterion for the models, compared with prior distribution I (Inverse Gamma).

FUNDING

REFERENCES

Bates D, Mächler M, Bolker B and Walker S (2014) Fitting linear mixed-effects models using lme4. **Journal of Statistical Software 67**: 1-48.

Cavalcante NR, Viana AP, Filho JEA, Pereira MG, Ambrósio M, Santos EA, Ribeiro RM, Rodrigues DL and Sousa CMB (2019) Novel selection strategy for half-sib families of sour passion fruit *Passiflora edulis* (Passifloraceae) under recurrent selection. **Genetics and Molecular Research 18**: 1-12.

Covarrubias PG (2016) Genome-assisted prediction of quantitative traits using the R package sommer. **PLOS ONE 11**: e0156744.

Covarrubias PG (2018) Software update: moving the R package sommer to multivariate mixed models for genome-assisted prediction. BioRxiv bioRxiv [Preprint] 10.1101/354639.

Geweke J (1992) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In Bernardo JM, Berger JO, Dawid AP and Smith AFM (eds) **Bayesian statistics.** Clarendon Press, Oxford, p. 169-193.

Gezan SA, Huber DA and White TL (2006) Post hoc blocking to improve heritability and precision of best linear unbiased genetic predictions. **Canadian Journal of Forest Research 36**: 2141-2147.

Gonçalves GM, Viana AP, Bezerra Neto FV, Pereira MG and Pereira TNS (2007) Seleção e herdabilidade na predição de ganhos genéticos em maracujá-amarelo. **Pesquisa Agropecuária Brasileira 42**: 193-198.

Hadfield JD (2010) "MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package." **Journal of Statistical Software 33**: 1-22.

Hadfield JD (2021) MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R Package. **Journal of Statistical Software 33**: 1-22.

Hadfield JD and Nakagawa PD (2010) General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. **Journal of Evolutionary Biology 23**: 494-508.

IBGE - Instituto Brasileiro de Geografia e Estatística (2018) Produção agrícola municipal. Available at <https://sidra.ibge.gov.br/tabela/5457#resultado>. Accessed on January 29, 2021.

Machado RFC, Krause W, Silva CA, Fachi LR, Silva FHL and Viana AP (2020) Implications of the post hoc blocking row–col technique on the intrapopulational improvement of the passion fruit. **Euphytica 216**: 62.

Plummer M, Best N, Cowles K and Vines K (2006) CODA: convergence diagnosis and output analysis for MCMC. **R News 6**: 7-11.

R Core Team (2019) R: A language and environment for statistical computing. Viena: R Foundation for Statistical Computing. Available at <https://www.r-project.org/>. Accessed on January 29, 2021.

Santos EA, Viana AP, Freitas JCO, Rodrigues DL, Tavares RF, Paiva CL and Souza MM (2015) Genotype selection by REML/BLUP methodology in a segregating population from an interspecific Passiflora spp. crossing. **Euphytica 204**: 1-11.

Silva FA, Viana AP, Corrêa CCG, Carvalho BM, de Sousa CMB, Amaral BD, Ambrósio M and Glória LS (2020) Impact of Bayesian inference on the selection of *Psidium guajava*. **Scientific Reports 10**: 1-9.

Silva FHL, Muñoz PR, Vincent CI and Viana AP (2016) Generating relevant information for breeding *Passiflora edulis*: genetic parameters and population structure. **Euphytica 208**: 609-619.

Silva JL, Sanches A, Andrade ACB, Oliveira GHF and Mauro AOD (2018) Bayesian approach, traditional method, and mixed models for multienvironment trials of soybean. **Pesquisa Agropecuária Brasileira 53**: 1093-1100.

Spiegelhalter DJ, Best NG, Carlin BP and Van der Linde A (2002) Bayesian measures of model complexity and fit (with discussion). **Journal of the Royal Statistical Society: Series B 64**: 583-639.

Viana AP and Resende MDV (2014) **Genética quantitativa no melhoramento de fruteiras**. Interciência, Rio de Janeiro, 296p.

Viana AP, Pereira TNS, Pereira MG, Amaral Júnior AT, Souza MM and Maldonado JFM (2004) Parâmetros genéticos em populações de maracujazeiro-amarelo. **Revista Ceres 51**: 545-555.

Wilberg MJ and Bence JR (2008) Performance of deviance information criterion model selection in statistical catch-at-age analysis. **Fisheries Research 93**: 212-221.