1   **Multi-trait and multi-environment Bayesian analysis to predict the G x E**

2   **interaction in flood-irrigated rice**

3   Antônio Carlos da Silva Júnior [(1)], Isabela de Castro Sant'Anna [(2)], Michele Jorge Silva[(1)]

4   Cosme Damião Cruz [(1)], Camila Ferreira Azevedo [(3)], Moyses Nascimento [(3)] and Plínio

5   César Soares [(4)]

6

7   [(1)]Departmento de Biologia Geral, Universidade Federal de Viçosa, CEP: 36.571-000,

8   Viçosa, Minas Gerais, Brasil

9   [(2)] Centro de Seringueira e Sistemas Agroflorestais, Instituto Agronômico (IAC), Rodovia

10   Péricles Beline, 1.481, Botafogo, Campinas, CEP: 13020-902, São Paulo, Brasil

11   [(3)]Departmento de Estatística, Universidade Federal de Viçosa, CEP: 36.571-000, Viçosa,

12   Minas Gerais, Brasil

13   [(4)] Empresa de Pesquisa Agropecuária de Minas Gerais – EPAMIG, Viçosa, Minas Gerais,
14   Brazil.

15

16   *Corresponding author: antonio.silva.c.junior@gmail.com

17

18   Tel.: +55 0313885-3243

19

20   **Abstract**

21   The biggest challenge for the reproduction of flood-irrigated rice is to identify superior

22   genotypes that present development of high-yielding varieties with specific grain

23   qualities, resistance to abiotic and biotic stresses in addition to superior adaptation to the

24   target environment.Thus, the objectives of this study were to propose a multi-trait and

25   multi-environment Bayesian model to estimate genetic parameters for the flood-irrigated

26   rice crop. To this end, twenty-five rice genotypes belonging to the flood-irrigated rice

27   breeding program were evaluated. Grain yield and flowering were evaluated in the

28   agricultural year 2017/2018. The experimental design used in all experiments was a

29   randomized block design with three replications. The Markov Chain Monte Carlo

30   algorithm was used to estimate genetic parameters and genetic values. The flowering is

31   highly heritable by the Bayesian credibility interval: $h^2 = 0.039$- 0.80, and 0.02- 0.91,

32   environment 1 and 2, respectively. The genetic correlation between traits was

33 significantly different from zero in the two environments (environment 1: -0.80 to 0.74;

34 environment 2: -0.82 to 0.86. The relationship of CV_e and CV_g higher for flowering in the

35 reduced model (CV_g/CV_e= 5.83 and 13.98, environments 1 and 2, respectively). For the

36 complete model, this trait presented an estimate of the relative variation index of: CV_e=

37 4.28 and 4.21, environments 1 and 2, respectively. In summary, the multi-trait and multi-

38 environment Bayesian model allowed a reliable estimate of the genetic parameter of

39 flood-irrigated rice. Bayesian analyzes provide robust inference of genetic parameters.

40 Therefore, we recommend this model for genetic evaluation of flood-irrigated rice

41 genotypes, and their generalization, in other crops. Precise estimates of genetic

42 parameters bring new perspectives on the application of Bayesian methods to solve

43 modeling problems in the genetic improvement of flood-irrigated rice.

44 **Keywords:** *Oryza sativa* L.; genetic parameters; multi-trait and multi-environment;

45 Bayesian model; biometric methods

46

47 **Introduction**

48 Rice (*Oryza sativa* L.) is one of the most important crops in the world and is

49 considered one of the main annual crops in Brazil [1]. Rice breeding is primarily aimed

50 at the development of high-yielding varieties with specific grain qualities, resistance to

51 abiotic and biotic stresses in addition to superior adaptation to the target environment [2,

52 3]. In this case, the breeder needs to realize mutual relationships, and knowledge of the

53 interdependence between agronomically important traits that can improve the accuracy

54 of selection [3].

55 Specifically, in the rice crop, the evaluation of multiple traits rather than a single

56 trait aims to maximize grain yield and quality. This is possible through the exploration of

57 genetic correlations between traits. In multi-trait analysis, the prediction of secondary

58 traits can be used to improve the prediction of primary traits, especially when they have

59 low heritability. Although consideration of the genetic correlation between traits is

60 essential, modeling interactions between phenotypes provides important information for

61 the development of breeding strategies that cannot be carried out with conventional

62 multivariate approaches alone [2, 4, 5].

63 Biometric methods available are useful for analyzing a single trait measured in a

64 single environment or across multiple environments with the genotype x environment (G

65 x E) interaction [6, 7, 8, 9]. This interaction can be defined as the differential response of

66   genotypes to environmental variation [10], and offers an even greater challenge for the

67   breeder [11]. The information from a network of experiments obtains a multi-trait and

68   multi-environment (MTME) structure, but present limited statistical methodology that no

69   correctly represents genetic and phenotypic variation in the data [12]. Therefore, genetic

70   correlations and G x E interaction, it requires more complex models that are difficult to

71   converge in the context of mixed linear models [11]. Therefore, Bayesian inference has

72   become a useful statistical tool for dealing with complex models [13].

73   Bayesian inference has been used successfully in studies with complex models.

74   [11] evaluated the multi-trait and multi-environment Bayesian model considering the G

75   x E interaction for nitrogen use efficiency components in tropical corn. [13], applied such

76   models through Bayesian inference applied to the breeding of *jatropha curcas* for

77   bioenergy. [14], used these models in the genetic selection of soybean progenies. [15],

78   demonstrate such models in phenotypic and genotypic data in corn and wheat. However,

79   few studies combine models of multiple traits in a multi-environment under a Bayesian

80   point of view, mainly for rice cultivation.

81   Therefore, the objectives of this study were to propose a multi-trait and multi-

82   environment Bayesian model to estimate genetic parameters for the flood-irrigated rice

83   crop. In addition to comparing: (i) the complete model (considering the interaction

84   between genotypes and environment) with the restricted model (not considering the

85   interaction); (ii) estimates of genetic parameters of models with single and multiple traits,

86   for grain yield and flowering.

87

88   **Material and Methods**

89   **Description of the experiment**

90   The experiments were carried out in the State of Minas Gerais - Brazil, in the

91   experimental fields of Agricultural Research Corporation of Minas Gerais State

92   (EPAMIG), in the cities of Lambari (21° 58' 11.24" S, 45° 20' 59.6" W) and Janaúba (15°

93   48' 77" S, 43° 17' 59.09" W). Twenty-five genotypes belonging to the flood-irrigated rice

94   breeding program of the Southeast region of the state of Minas Gerais were evaluated,

95   and five of these genotypes were used as experimental controls (Rubelita, Seleta,

96   Ourominas, Predileta, and Rio Grande). These genotypes were evaluated in comparative

97   trials after multiple generations of selection, and in addition, they are known for their high

98   yield, uniform growth rate and plant growth, resistance to major diseases, and for their

99   excellent grain quality. Grain yield (GY, Kg ha^{-1}) and flowering period in days (FL) were

100   evaluated in the agricultural year 2017/2018. The experimental design was a randomized

101   complete block design with three replications. The experiments were conducted in

102   floodplain soils with continuous flood irrigation. Grain production data in grams per

103   useful plot were used, later converted into kilograms per hectare. Management practices

104   were carried out according to recommendations for flood-irrigated rice in the region.

105   The useful area consisted of 4 m central of three internal rows (4 m x 0.9 m,

106   totaling 3.60 m^2). The soil preparation was carried out by plowing and harrowing around

107   30 days before sowing. For planting fertilization, a mixture of 100 kg ha^{-1} of ammonium

108   sulfate, 300 kg ha^{-1} of simple superphosphate, and 100 kg ha^{-1} of potassium chloride was

109   used, applied in the plot, and incorporated into the soil before planting. The fertilization

110   in the top dressing was carried out approximately 60 days after the installation of the

111   experiments, with 200 kg ha^{-1} of ammonium sulfate. The weeds were controlled with the

112   use of herbicides and manual weeding. Sowing was carried out in the planting line with

113   a density of 300 seeds m^{-2}. The irrigation started around 10-15 days after the implantation

114   of the experiments, and the water was only removed close to the maturation of the

115   materials. The harvest was carried out when the grains reached a humidity of 20-22%.

116   Grain production data were obtained by weighing all grains harvested in the useful plot,

117   after cleaning and uniform drying in the sun, until they reached a humidity of 13%. Days

118   for flowering, correspond the number of days from sowing to flowering, when the plot

119   presented approximately 50% of plants with panicles.

120

121   **Biometric analysis**

122   Grain yield (GY) and flowering period in days (FL) were analyzed using single-

123   and multi-trait models using the Bayesian Markov Chain Monte Carlo (MCMC)

124   approach. The objective was to compare: (i) the complete model (considering the

125   interaction between genotypes and environments) with the restricted model (not

126   considering the interaction); (ii) estimates of genetic parameters of models with single

127   and multiple traits, for grain yield and flowering period.

128   The multi-trait and multi-environment (MTME) model was given by:

129    $$y = X\beta + W_1 r + W_2 u + \varepsilon$$

130   which can be rewritten as:

$$
\begin{pmatrix} Y_1 \\ \cdots \\ Y_2 \end{pmatrix} = X \begin{pmatrix} \beta(E1,1) \\ \beta(E2,1) \\ \cdots \\ \beta(E1,2) \\ \beta(E2,2) \end{pmatrix} + W_1 \begin{pmatrix} r_1(E1,1) \\ r_2(E2,1) \\ \cdots \\ r_1(E1,2) \\ r_2(E2,2) \end{pmatrix} + W_2 \begin{pmatrix} u(E1,1) \\ u(E2,1) \\ \cdots \\ u(E1,2) \\ u(E2,2) \end{pmatrix} + \begin{pmatrix} \varepsilon(E1,1) \\ \varepsilon(E2,1) \\ \cdots \\ \varepsilon(E1,2) \\ \varepsilon(E2,2) \end{pmatrix},
$$

where: y is the vector of the phenotypic values of the two evaluated traits (grain yield, Y_1; and flowering period, Y_2); X is the incidence matrix of the systematic effects represented by β, assuming $\beta \sim N(\mu_\beta, I_\otimes \Sigma_\beta)$ so that $E1$ and $E2$ represent the two environments studied; W_1 it is the incidence matrix of the random effect of the environment; r is the ambient effect vector, $r \sim N(0, I_\otimes \Sigma_r)$; W_2 is the incidence matrix of the effects of the genotype x environment interaction; u is the random effects vector of the genotype x environment interaction, $u \sim N(0, I_\otimes \Sigma u)$; and ε is the residual effects vector, $\varepsilon \sim N(0, I_\otimes \Sigma_\varepsilon)$.

The (co)variance matrices are given by:

$$
\sum\nolimits_r = \begin{pmatrix} \sigma_{ry(1)}^2 & \sigma_{ry(1,2)} & \sigma_{ry,E(1)} & \sigma_{ry,E(1,2)} \\ \sigma_{ry(1,2)} & \sigma_{ry(2)}^2 & \sigma_{ry,E(2,1)} & \sigma_{ry,E(2)} \\ \sigma_{ry,E(1)} & \vdots & \sigma_{rE(1)}^2 & \cdots \\ \sigma_{ry,E(1,2)} & \cdots & \cdots & \sigma_{rE(2)}^2 \end{pmatrix}
$$

$$
\sum\nolimits_u = \begin{pmatrix} \sigma_{uy(1)}^2 & \sigma_{uy(1,2)} & \sigma_{uy,E(1)} & \sigma_{uy,E(1,2)} \\ \sigma_{uy(1,2)} & \sigma_{uy(2)}^2 & \sigma_{uy,E(1)} & \sigma_{uy,E(2)} \\ \sigma_{uy,E(1)} & \vdots & \sigma_{uE(1)}^2 & \cdots \\ \sigma_{uy,E(1,2)} & \cdots & \cdots & \sigma_{uE(2)}^2 \end{pmatrix}
$$

$$
\Sigma_\varepsilon = \begin{pmatrix} \sigma_{\varepsilon y(1)}^2 & \sigma_{\varepsilon y(1,2)} & \sigma_{\varepsilon y,E(1)} & \sigma_{\varepsilon y,E(1,2)} \\ \sigma_{\varepsilon y(1,2)} & \sigma_{\varepsilon y(2)}^2 & \sigma_{\varepsilon y,E(2,1)} & \sigma_{\varepsilon y,E(2)} \\ \sigma_{\varepsilon y,E(1)} & \vdots & \sigma_{\varepsilon E(1)}^2 & \cdots \\ \sigma_{\varepsilon y,E(1,2)} & \cdots & \cdots & \sigma_{\varepsilon E(2)}^2 \end{pmatrix},
$$

where: y represents grain yield and h represents flowering period in days; 1 and 2 represent the two environments studied. The variance-covariance matrices follow an inverted Wishart distribution, which was used as a priori to model the variance-covariance matrix [16].

The package "MCMCglmm" [17] was used. A total of 10,000,000 samples were generated and assuming a flare period and sampling interval of 500,000 and 10 iterations, respectively, this resulted in a final total of 50,000 samples. The convergence of the

151 MCMC was verified by the criterion of [18], performed in two packages R "boa" [19]

152 and "CODA" (convergence diagnosis) [20]. Even though Bayesian and frequentist

153 structures were not directly compared, especially in the field of genetics [21], the same

154 models were also adjusted based on the REML estimation method (Restricted Maximum

155 Likelihood).

156   The complete models (considering the interaction between genotypes and

157 environments) were compared with the null models (not considering the interaction) by

158 the deviation information criterion (DIC) proposed by [22]:

159 $$DIC = D(\bar{\theta}) + 2p_D$$

160 where is a point estimate of the deviation obtained by replacing the parameters with their

161 later means estimates in the probability function and p_D is the effective number of model

162 parameters. Models with a lower DIC should be preferred over models with a higher DIC.

163

164   The components of variance, broad-sense heritability, coefficient of variation

165 residual and genetic, variation index, and genotypic correlation coefficients between

166 genetic traits and values were calculated from the posterior distribution. The package

167 "boa" [19] was used to calculate the intervals of greater posterior density (HPD) for all

168 parameters. The later estimates for the broad-sense heritability of grain yield and

169 flowering period in days for each interaction were calculated from the later samples of

170 the variance components obtained by the model, using the expression:

171 $$h^{2(i)} = \frac{\sigma_g^{2(i)}}{\left(\sigma_g^{2(i)} + \sigma_r^{2(i)} + \sigma_\varepsilon^{2(i)}\right)}$$

172 where: $\sigma_g^{2(i)}$, $\sigma_r^{2(i)}$, and $\sigma_\varepsilon^{2(i)}$ are the genetic, replication, and residual variations for each

173 iteration, respectively.

174   The genetic correlation coefficients between the pairs of traits in each

175 environment were obtained, as suggested by [23], using the expression below for all

176 models:

177 $\rho_{l(1,2)} = \frac{\sigma_{gl(1,2)}}{\sqrt{\sigma_{gl(1)}^2 \sigma_{gl(2)}^2}}$: genetic correlation between environment and grain yield and $\rho_{h(1,2)}$

178 $= \frac{\sigma_{gh(1,2)}}{\sqrt{\sigma_{gh(1)}^2 \sigma_{gh(2)}^2}}$: genetic correlation between environment and flowering period in days.

179 All data analysis was conducted using the statistical software package R version 4.1.0.

180

Results

182   Geweke's convergence criterion indicated convergence for all dispersion parameters by generating 10,000,000 MCMC strings, 500,000 samples for burn-in, and a sampling interval of 10, totaling 50,000 effective samples used for estimating variance components. Similar posterior mean, median and modal estimates were obtained for variance components, suggesting normal-appearing density. However, all chains reached convergence by this criterion. According to the deviation information criteria (DIC), there was positive evidence of interactions between genotypes and environments for all analyzed models (Table 1). However, the DIC values were lower when using the complete model (considering the effects of genotype x environment interaction), in which the difference in relation to the complete model was greater than 2 (Table 1), which according to [22] the use of full model can lead to greater precision in estimating parameters (Table 1).Therefore, since the obtained DIC values were greater than two, it is possible to indicate the superiority of the complete model over the restricted models. On the other hand, as this component of the model is important, the "best" genotypes measured in diferente environments cannot be the same. However, convergence was not achieved by the AI (Average Information) and EM (Expectation-Maximization) algorithms.

198

199 **Table 1.** Deviation information criteria for the full (considering the G x E interaction)

200 and null (not considering the interaction) models.

201

202

203

204

		Deviance information criteria (DIC)	
Model	Trait	Full Model	Null Model
Mult-trait	GY, FL	-308.83	1967.79
Single-trait	GY	1867.49	1868.28
Single-trait	FL	162.19	697.67

205 GY: Grain Yield; FL: Flowering Period.

206

207 The posterior inferences for mean and higher posterior density range (HPD)

208 considering the multi-trait and multi-environment (MTME) model are described in Table

209 2. The average values for the grain yield tr ait varied from 4210.91 – 3901.56 kg ha^{-1} and

210 flowering in days of 99.40 – 76.43, in environments 1 and 2, respectively (Table 2). The

211 Bayesian credibility interval (95% probability) for average grain yield corresponds to

212 4191.70 – 4227.86 kg ha^{-1}, and 3852.60 – 3946.07 kg ha^{-1} (p<1e-05), in environments 1

213 and 2, respectively. In relation to the flowering period in days, this interval corresponds

214 to 98.09 – 100.69 e 74.56 – 78.22 (p<1e-05), environments 1 and 2, respectively.

215

216 **Table 2.** Posterior inferences for mean and highest posterior density range (HPD)

217 considering the proposed complete multi-trait multi-environment model.

			HPD 95%	
Trait	EN	post.mean	LOWER	UPPER
GY	1	4210.91***	4191.7	4227.86
FL	1	99.4***	98.09	100.69
GY	2	3901.56***	3852.6	3946.07
FL	2	76.43***	74.56	78.22

218 *** Significância estatística: p ≤ 0.001. GY: Grain Yield; FL: Flowering Period; EN:

219 environment.

220 Table 3 presents the subsequent inferences for mean and genetic variance; mode,

221 mean, median, and highest posterior density range (HPD) of heritability in the broad

222 sense; and the mode, mean, median, and greater posterior density interval (HPD) of the

223 genetic correlation, considering MTME. The grain yield trait in environment 2 was

224 considered weakly heritable with Bayesian credibility interval (95% probability): $h^2 =$

225 4.36E-07 - 9.21E-06, and in environment 1, it was moderately heritable $h^2 = 0.39 - 0.757$.

226 The low estimate of heritability observed does not depend on the number of samples

227 evaluated, since the Bayesian structure used is essentially recommended for situations

228 involving small sample sizes. In addition, the grain yield trait is largely influenced by the

229 environment as it is a quantitative character [24], which reflects this low estimate of

230 heritability. In relation to flowering period, it is highly heritable by the Bayesian

231 credibility interval (95% probability): h^2 = 0.039 - 0.80, and 0.02 - 0.91, environment 1

232 and 2, respectively. And the posterior mean of the genetic correlation between traits was

233 significantly different from zero (95% credibility intervals) in the two environments

234 (environment 1: -0.80 to 0.74; environment 2: -0.82 to 0.86) (Table 3).

235

236 **Table 3.** Resume of inferences for mean and genetic variance; mode, mean, median, and

237 highest posterior density range (HPD) of heritability in the broad sense; and the mode,

238 mean, median, and highest posterior density range (HPD) of the genetic correlation,

239 considering the complete model multi-trait and multi-environment.

Trait	EN	h^2			HPD 95%	
		Mode	Mean	Median	LOWER	UPPER
GY	1	0.11	0.28	0.18	0.39	0.757
GY	2	1.30E-06	3.32E-06	2.24E-06	4.36E-07	9.21E-06
FL	1	0.12	0.31	0.24	0.039	0.80
FL	2	0.06	0.27	0.13	0.02	0.91

		Genotypic Correlation			HPD 95%	
		Mode	Mean	Median	LOWER	UPPER
GY,FL	1	-0.0076	-0.027	-0.028	-0.80	0.74
GY,FL	2	0.037	0.018	0.019	-0.82	0.86

240 EN: environment; GY: Grain Yield; FL: Flowering Period; h^2: heritability.

241

242   Bayesian methods access the posterior density ranges of genetic parameters

243 (Figure 1). The genetic parameters of the flood-irrigated rice genotypes for each trait and

244 their larger posterior density ranges (HPD) were obtained to assist in the selection of

245 genotypes.

246
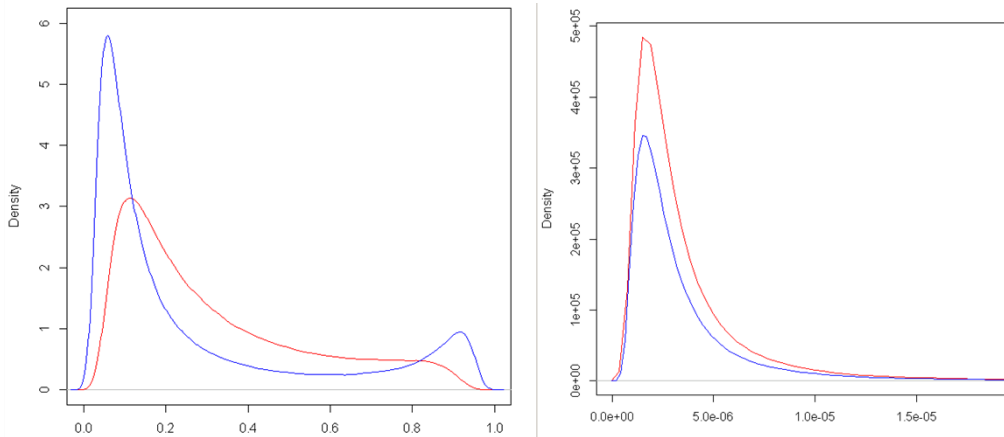
Figure 1. Posterior density for the complete model proposed by multi-traits multi-environment (left: flowering period and right: grain yield). The red line represents the posterior density for environment 1, while the blue line represents the posterior density for environment 2.

In Figure 2, there is a unimodal distribution in which there is a mixture of two distinct populations for posterior density in relation to the genotypic correlation between traits for the complete MTME model. The red line represents the posterior density for environment 1, while the blue line represents the posterior density for environment 2.

255
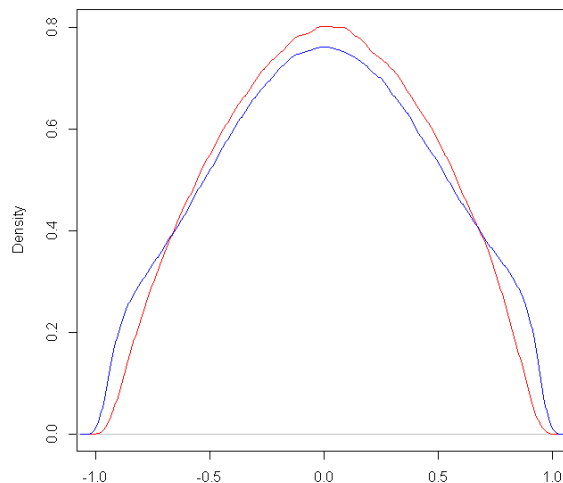
Figure 2. Posterior density for the genotypic correlation between the grain yield trait and flowering period in days for the model proposed by multi-traits and multi-environment. The red line represents the posterior density for environment 1, while the blue line represents the posterior density for environment 2.

260

Variance Estimate

262   The a posteriori estimates of the genotypic and residual variances for the reduced

263   model (MTME) were very discrepant among the environments (Table 4). The grain yield

264   trait in the reduced model obtained a greater estimate of genotypic variance in

265   environment 1, corresponding to a difference of approximately 60% greater, in relation

266   to environment 2. This indicates a greater influence of genetic components on

267   environmental components in the expression of traits. On the other hand, this estimate for

268   the complete model showed less variation between environments, especially for the

269   flowering period in days trait. This result showed the best consistency for the complete

270   model (MTME). Estimates of the variance of the greatest interaction were observed for

271   the flowering period trait.

272   **Table 4.** Genetic parameters for traits grain yield and flowering period in days, in two

273   environments, using multi-trait multi-environment (MTME) models.

Model		Trait	EM	Component		
				σ_g^2	σ_r^2	σ_{int}^2
Multi trait	Null	GY	1	38.93	8.86	-
		GY	2	23.19	6.13	-
		FL	1	14.95	4.45E-5	-
		FL	2	32.63	0.19	-
	Full	GY	1	2.63	7.24	4.64
		GY	2	2.70	7.39	4.77
		FL	1	3.35	0.18	7.57
		FL	2	5.69	0.34	15.91

274   EN: environment; GY: Grain Yield; FL: Flowering Period; σ_g^2, σ_r^2, σ_{int}^2: are the genetic,

275   replication, and interaction variations, respectively.

276

Relative variation index

278   The ratio of the coefficient of variation genotypic and the coefficient of variation

279   residual (CV_g/CV_e) corresponded to the relative variation index. When this index is greater

280   than one unit, it suggests that genetic variation is more influential than residual variation.

281   This was observed in this study for both traits, in the reduced model proposed by MTME

282   (Table 5). For the complete model in the grain yield trait, this relationship was less than

283   one unit. The relationship of CV_e and CV_g higher for flowering period trait in the reduced

284   model (CV_g/CV_e= 5.83 and 13.98, environments 1 and 2, respectively). This trait has

285   greater variability and is highly promising for selection. This is due to its complex genetic

286   inheritance resulting from the involvement of several genes with little effect on the

287 phenotype [24, 25]. For the complete model, this trace presented an estimate of the

288 relative variation index of: CV_g/CV_e= 4.28 and 4.21, environments 1 and 2, respectively

289 (Table 5). Then, this trait presented a higher posterior density interval (HPD) for 80 and

290 91% heritability, environment 1 and 2, respectively, reinforcing the possibility of genetic

291 gain (Table 3). This implies that the variability observed in these traits has genetic

292 predominance, which is interesting in the process of genetic gain in a flood-irrigated rice

293 breeding program. The coefficient of variation genotypic for the grain yield trait was low

294 in both environments. This is justified by the fact that these genotypes belong to advanced

295 comparative trials in which they have gone through several generations of selection.

296

297 **Table 5.** Coefficient of variation residual (CV_e, %), coefficient of variation genotypic (

298 CV_g, %) and relative variation index (CV_g/CV_e) for the multi-trait and multi-environment

299 model.

Model	Trait	EN	CV_g(%)	CV_e(%)	CV_g/CV_e
Null	GY	1	0.149	0.071	2.11
	GY	2	0.122	0.063	1.92
	FL	1	3.89	0.67	5.83
	FL	2	7.47	0.57	13.98
Full	GY	1	0.039	0.064	0.61
	GY	2	0.042	0.070	0.60
	FL	1	1.84	0.43	4.28
	FL	2	3.20	0.76	4.21

300 EN: environment; GY: Grain Yield; FL: Flowering Period; EN: environment.

301

302 **Discussion**

303 One of the ways to succeed in breeding programs is related to the accurate

304 prediction of genotypic values, which is closely related to the adoption of adequate

305 models. Thus, in this study, we apply a new statistical approach for estimating variance

306 components in floodplain rice breeding schemes. The implementation of multi-trait multi-

307 environment models Bayesian is straight forward and currently has been widely used due

308 to the possibility of considering a priori knowledge in modeling. In addition to its

309 application wide application in animal breeding [26, 27], Bayesian multi-trait analysis

310 has been reported in plant breeding [11, 13, 14, 15, 28].

311   Bayesian inference has been used since 1986 [29] and has been further explored

312   in recent years [2, 11, 13, 14, 15, 30, 31] due to major computational advances and new

313   applications and methodologies [32]. However, Bayesian analysis is based on knowledge

314   of the posterior distribution of the parameters to be estimated. This allows the

315   construction of exact credibility intervals for estimates of random variables and variance

316   components [33]. Values for the 95% distribution credibility interval for the broad-sense

317   heritability parameter found in this study (Table 3) were also presented in the study by

318   [11] to estimate genetic parameters for efficiency of uptake and efficiency of use of N

319   under contrasting soil N levels via MTME models. Another study, based on the estimation

320   of genetic parameters for genetic selection of segregating soybean progenies using the

321   MTME model [14]. The difference between mean, mode, and median of the broad-sense

322   heritability estimates (Table 3) reflects some lack of symmetry in the posterior

323   distribution estimates. The lack of symmetry between mean, mode, and median

324   heritability estimates in posterior distribution estimates was reported by [11] and [30].

325   The low broad-sense heritability observed in the traits does not depend on the

326   number of samples evaluated, since the Bayesian structure used is essentially

327   recommended for situations involving small samples. On the other hand, quantitative

328   traits of agronomic interest, determined by several genes, demosntre low expression and

329   significantly influenced by the environment [24], reflected in the traits grain yield and

330   flowering period in days.

331   Based on the results of heritability estimation in the broad sense for the GY and

332   FL traits varied in: h^2 = 0.39 - 0.757 and h^2 = 4.36E-07 - 9.21E-06; h^2 = 0.039 - 0.80,

333   and 0.02 - 0.91, environment 1 and environment 2, respectively (Table 3). [34] found

334   heritability estimates of 0.48 and 0.94 for GY and FL, respectively, using 198 rice

335   progenies by the ANOVA technique. [35] evaluated upland rice genotypes, by this same

336   technique, and obtained an estimate of h^2 for GY and FL traits 0.35 and 0.77, respectively.

337   [36] obtained results of h^2 ranging from 0.44 - 0.87 for GY, 0.46 - 0.94 for FL. [37]

338   applied a mixed model in studies using FL and GY traits, found an estimate of h^2 de

339   0,88, and 0,71, respectively. [38], using the ANOVA method, found h^2 of 0.76 for FL and

340   [39] estimate of 0.30 for GY. Regarding the estimate of CVs using the ANOVA method

341   for the FL and GY traits, representing 2.98% and 15.28%, respectively [34].

342   The amounts of data that breeding programs around the world are generating

343   continue to increase; consequently, there is a growing need to extract more knowledge

344   from the data being produced. For this, MTME models are commonly used to take

345 advantage of correlated traces to improve parameter estimation and prediction accuracy.
346 However, when there are a large number of features, implementing these types of models
347 is a challenge. Therefore, it is necessary to develop efficient models of multiple trait and
348 multiple environments for selection, in order to take advantage of multiple correlated
349 features. In this work, was proposed an alternative method to analyze MTME model data
350 that could be useful for genotype selection, and estimation of genetic parameters in
351 flooded rice in the context of an abundance of traits and environments.

352 Full-model multi-trait analysis tends to be powerful and provide more accurate
353 estimates than single-trait analysis because the previous method can take into account the
354 underlying correlation structure found in a multi-trait dataset. However, Bayesian and
355 non-Bayesian inferences from the MTME model analysis are complex and
356 computationally demanding [40]. [28] argue that Bayesian multi-trait analysis is more
357 appropriate than ANOVA to perform analyzes and select superior genotypes for genetic
358 improvement since the Bayesian model can capture small genetic differences between
359 families, while ANOVA cannot. In this study, was explained how to make Bayesian
360 inference using a multi-trait multi-environment Bayesian model in plant breeding. These
361 results showed that the approach was efficient in estimating genetic parameters in flooded
362 rice.

363 The correlation study revealed favorable associations for the traits in studies in
364 two settings. This result indicates that the selection of genotypes characterized by longer
365 flowering period favors grain yield in flood-irrigated rice, which is desirable for rice
366 cultivation since later plants tend to be more productive. However, late cycle cultivars
367 tend to be more productive in relation to the early cycle, since they obtain an increase in
368 the amount of photoassimilates that are translocated to grains [3]. This result justifies the
369 significant correlation between the traits. According to [24], correlations between traits
370 may be the result of pleiotropy or genetic linkage. Thus, if the undesirable correlations
371 are caused by genetic linkage, these associations can be broken by recombination caused
372 by crossing or self-fertilization; consequently, these factors do not necessarily become
373 major impediments to breeding programs [](Liersch et al., 2016; Radkowski et al., 2017;
374 Bocianowski et al, 2019).

375 Another point that we would like to highlight is that our proposed model is of
376 MTME, but with the restriction that an identity matrix is assumed for the variance-
377 covariance matrix of the environments. However, even with this restrictive assumption in
378 the variance-covariance matrix of the environments, the model has the advantage of

379 taking into account the terms of interaction x trait, genotype x trait, and the environment

380 x genotype x trait. Furthermore, it takes into account the correlated traits. The program of

381 irrigated rice program aims to obtain the desired results in a short period and with

382 precision, therefore, the choice of the model will be use as breeding strategies.

383

384 **Conclusion**

385   The multi-trait and multi-environment Bayesian model was efficient to estimate

386 genetic parameters for the flood-irrigated rice crop.

387   The estimates of genetic parameters bring new perspectives on the application of

388 Bayesian methods to solve modeling problems in the genetic improvement of flood-

389 irrigated rice.

390

391 **Compliance with ethical standards**

392   This manuscript is not under consideration for publication elsewhere and has

393 not been made publicly available online. All authors have approved the submission, and

394 all the copyrighted individuals are listed. The authors declare that they have no conflicts

395 of interest.

396

397 **References**

398 1. Silva Júnior, AC, Silva, MJ, Cruz, CD, Nascimento, M, Azevedo, CF & Soares,

399   PC. (2020). Patterns recognition methods to study genotypic similarity in flood-

400   irrigated rice. *Bragantia*, *79*(3), 356-363.https://doi.org/10.1590/1678-

401   4499.20200232.

402 2. Yu, H, Campbell, T, Zhang, Qi, Walia, H, Morota, G. Genomic Bayesian

403   Confirmatory Factor Analysis and Bayesian Network To Characterize a Wide

404   Spectrum of Rice Phenotypes. (2019). G3: GENES, GENOMES, GENETICS.

405   vol. 9 n. 6. 1975-1986; https://doi.org/10.1534/g3.119.400154.

406 3. Silva Júnior AC, Santanna IC, Silva GN, Cruz CD, Nascimento M, Bhereng LL,

407   Soares PC. (2022). Computational intelligence to study the importance of

408   characteristics in flood-irrigated rice. Acta Scientiarum-Agronomy.

409 4. Valente BD, Rosa GJ, de Los Campos G, Gianola D, Silva MA. (2010). Searching

410   for recursive causal structures in multivariate quantitative genetics mixed models.

411   Genetics;185(2):633–644. https://doi.org/10.1534/genetics.109.112979.

5. Momen M, Campbell MT, Walia H, Morota G. Predicting Longitudinal Traits Derived from High-Throughput Phenomics in Contrasting Environments Using Genomic Legendre Polynomials and B-Splines. G3 (Bethesda, Md.). (2019) Oct;9(10):3369-3380. https://doi.org/10.1534/g3.119.400346.

6. Burgueño, J., G. de los Campos, K. Weigel, and J. Crossa. (2012). Genomic prediction of breeding values when modeling genotype·environment interaction using pedigree and dense molecular markers. Crop Sci. 52(2): 707–719.

7. Heslot, N., D. Akdemir, M. E. Sorrells, and J. L. Jannink. (2014). Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. Theor. Appl. Genet. 127(2): 463–480.

8. Jarquín, D., J. Crossa, X. Lacaze, P. D. Cheyron, J. Daucourt et al. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor. Appl. Genet. 127(3): 595–607.

9. Montesinos-López, O. A., A. Montesinos-López, P. Pérez-Rodríguez, G. de los Campos, K. M. Eskridge et al. (2015). Threshold models for genome-enabled prediction of ordinal categorical traits in plant breeding. G3 (Bethesda) 5(1): 291–300.

10. Silva Júnior AC, Silva MJ, Sousa I, Costa WG, Cruz CD, Nascimento M, Soares PC. (2021). Fuzzy logic for adaptability and stability studies in irrigated rice (*Oryza Sativa* L.) genotypes. Plant Breeding. https://doi.org/10.1002/PBR.12973.

11. Torres LG, Rodrigues MC, Lima NL, Trindade TFH, Silva FFe, Azevedo CF, et al. (2018). Multi-trait multi-environment Bayesian model reveals G x E interaction for nitrogen use efficiency components in tropical maize. PLoS ONE 13(6): e0199492. https://doi.org/10.1371/journal.pone.0199492.

12. Malosetti, M., Ribaut, J.M., Vargas, M. et al. (2008). A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). Euphytica. 161: 241. https://doi.org/10.1007/s10681-007-9594-0.

13. Peixoto MA, Evangelista JSPC, Coelho IF, Alves RS, Laviola BG, Fonseca e Silva F, et al. (2021). Multiple-trait model through Bayesian inference applied to *Jatropha curcas* breeding for bioenergy. PLoS ONE 16(3): e0247775. https://doi.org/10.1371/journal.pone.0247775.

445   14. Volpato L, Alves RS, Teodoro PE, Vilela de Resende MD, Nascimento M,
446   Nascimento ACC, et al. (2019). Multi-trait multi-environment models in the
447   genetic selection of segregating soybean progeny. PLoS ONE 14(4): e0215315.
448   https://doi. org/10.1371/journal.pone.0215315.

449   15. Montesinos-López, OA, Montesinos-López A, Montesinos-López JC, Crossa J,
450   Luna-Vázquez FJ, Salinas-Ruiz JA. (2018). Bayesian Multiple-Trait and
451   Multiple-Environment Model Using the Matrix Normal Distribution. Phys.
452   Methods Stimul. Plant Mushroom Dev. 19.

453   16. Van Tassel CP, Van Vleck LD. (1996). Multiple-trait Gibbs sampler for animal
454   models: flexible programs for Bayesian and likelihood-based (co)variance
455   component inference. J Anim Sci. 74:2586–2597. pmid:8923173.
456   https://doi.org/10.2527/1996.74112586x.

457   17. Hadfield J. MCMC Methods for Multi-Response Generalized Linear Mixed
458   Models: The MCMCglmm R Package. J Stat Softw. 2010; 33:1–22.
459   https://doi.org/10.18637/jss.v033.i02.

460   18. Geweke J. Evaluating the accuracy of sampling-based approaches to the
461   calculation of posterior moments. In: Bernardo JM, Berger JO, Dawid AP, Smith
462   AFM, editors. Bayesian Statistics 4. Oxford University Press; 1992. pp 625–631.

463   19. Smith BJ. boa: an R package for MCMC output convergence assessment and
464   posterior inference. J Stat Softw. 2007; 21:1–37.
465   https://doi.org/10.18637/jss.v021.i11.

466   20. Plummer M, Best N, Cowles K, Vines K. CODA: Convergence diagnosis and
467   output analysis for MCMC. R news. 2006; 6:7–11.

468   21. Sorensen DA, Gianola D. Likelihood, Bayesian and MCMC methods in
469   quantitative genetics: statistics for biology and health. Springer-Verlag; 2002.

470   22. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian Measures of
471   Model Complexity anf Fit. J R Stat Soc Ser B. Statistical Methodol. 2002;64: 583–
472   639.

473   23. Piepho HP. Allowing for the structure of a designed experiment when estimating
474   and testing trait correlations. J Agric Sci. 2018;156: 59–70.

475   24. Falconer DS, Mackay TFC. 1996. Introduction to quantitative genetics.
476   Edinburgh, SC: Addison Wesley Longman.

477   25. Cruz, CD; Regazzi, AJ; Carneiro, PCS. (2012). Modelos biométricos aplicados ao
478   melhoramento genético. UFV. Viçosa.

479   26. Ventura HT, Silva FF, Varona L, de Figueiredo EAP, Costa EV, da Silva LP, et
480   al. Comparing multi-trait Poisson and Gaussian Bayesian models for genetic
481   evaluation of litter traits in pigs. Livestock Science. 2015;176:47–53.

482   27. Rolf MM, Garrick DJ, Fountain T, Ramey HR, Weaber RL, Decker JE, et al.
483   Comparison of Bayesian models to estimate direct genomic values in multi-breed
484   commercial beef cattle. Genetics Selection Evolution. 2015;47(1):23.

485   28. Junqueira VS, Azevedo Peixoto Ld, Galvêas Laviola B, Lopes Bhering L,
486   Mendonça S, et al. (2016) Correction: Bayesian Multi-Trait Analysis Reveals a
487   Useful Tool to Increase Oil Concentration and to Decrease Toxicity in *Jatropha*
488   *curcas*   L..   PLOS   ONE   11(8):   e0161046.
489   https://doi.org/10.1371/journal.pone.0161046.

490   29. Gianola D, Fernando RL. Bayesian Methods in Animal Breeding Theory. J Anim
491   Sci. 1986;63: 217–244.

492   30. Mora, F. & Serra, N. (2014). Bayesian estimation of genetic parameters for
493   growth, stem straightness, and survival in Eucalyptus globulus on an Andean
494   Foothill site. Tree Genetics & Genomes.10: 711. https://doi.org/10.1007/s11295-
495   014-0716-2.

496   31. Duhnen A, Gras A, Teyssèdre S, Romestant M, Claustres B, Daydé J, et al.
497   Genomic selection for yield and seed protein content in Soybean: A study of
498   breeding program data and assessment of prediction accuracy. Crop Sci. 2017;57:
499   1325–1337.

500   32. Van de Schoot, R., Depaoli, S., King, R. *et al.* Bayesian statistics and
501   modelling. *Nat   Rev   Methods   Primers* 1, 1   (2021).
502   https://doi.org/10.1038/s43586-020-00001-2.

503   33. Resende MDV, Duda LL, Guimarães PRB, Fernandes JSC. Análise de modelos
504   lineares mistos via inferência Bayesiana. Rev Mat Estat. 2001;19: 41–70.
505   Available: http://jaguar.fcav.unesp.br/RME/fasciculos/v19/A3_Artigo.pdf.

506   34. Guimarães, P. H. R., Melo, P. G. S., Cordeiro, A. C. C., Torga, P. P., Rangel, P.
507   H. N., Castro, A. P. Index selection can improve the selection efficiency in a rice
508   recurrent   selection   population.   Euphytica   (2021)   217:95.
509   https://doi.org/10.1007/s10681-021-02819-7.

510   35. Sari, W. K., Nualsri, C., Junsawang, N., Soonsuwon, W. Combining ability and
511   heritability for yield and its related traits in Thai upland rice (*Oryza sativa* L.).
512   Agr. Nat. Resour. Vol. 54 No. 3 (2020)

36. Catolos Margaret, Sandhu Nitika, Dixit Shalabh, Shamsudin Noraziya A. A., Naredo Ma E. B., McNally Kenneth L., Henry Amelia, Diaz Ma G., Kumar Arvind. Genetic Loci Governing Grain Yield and Root Development under Variable Rice Cultivation Conditions. Front. Plant Sci., 16 October 2017 | https://doi.org/10.3389/fpls.2017.01763.

37. Bhandari, A., Bartholomé, J., Cao, T.-V., Kumari, N., Frouin, J., Kumar, A., Ahmadi, N. (2019). Selection of trait-specific markers and multi-environment models improve genomic predictive ability in rice. PLoS ONE, 14(5). https://doi.org/10.1371/journal.pone.0208871.

38. Sangaré, J. R., Konaté, A. K., Cissé, F., Sanni, A. Assessment of genetic parameters for yield and yield related-traits in an intraspecific rice (Oryza sativa L.) population. J. Plant Breed. Genet. 05 (02) 2017. 45-56.

39. Xu, Y., Wang, X., Ding, X. et al. Genomic selection of agronomic traits in hybrid rice using an NCII population. Rice 11, 32 (2018). https://doi.org/10.1186/s12284-018-0223-4.

40. Mathew, B., Holand, A.M., Koistinen, P. et al. Reparametrization-based estimation of genetic parameters in multi-trait animal model using Integrated Nested Laplace Approximation. Theor Appl Genet 129, 215–225 (2016). https://doi.org/10.1007/s00122-015-2622-x.

41. Liersch A, Bocianowski J, Woś H, Szała L, Sosnowska K, Cegielska-Taras T, Nowosad K, Bartkowiak-Broda I. (2016). Assessment of genetic relationships in breeding lines and cultivars of Brassica napus and their implications for breeding winter oilseed rape. Crop Sci. 56(4):1540– 1549.

42. Radkowski A, Sosin-Bzducha E, Radkowska I. (2017). Effects of silicon foliar fertilization of meadow plants on the nutritional value of silage fed to dairy cows. J Elem. 22(4):1311–1322.

43. Bocianowski, J; Nowosad, K; Szulc, P; Tratwal, A; Bakinowska, E & Piesik, D. (2019). Genetic parameters and selection of maize cultivars using Bayesian inference in a multi-trait linear model, Acta Agriculturae Scandinavica, Section B — Soil & Plant Science, 69:6, 465-478. https://doi.org/10.1080/09064710.2019.1601764.