Industrial Crops & Products 130 (2019) 558–561

Contents lists available at ScienceDirect

Industrial Crops & Products

journal homepage: www.elsevier.com/locate/indcrop

Multiple-trait BLUP in longitudinal data analysis on *Jatropha curcas* breeding for bioenergy

Rodrigo Silva Alves[a], Paulo Eduardo Teodoro[a], Leonardo de Azevedo Peixoto[b],
João Romero do Amaral Santos de Carvalho Rocha[a], Lidiane Aparecida Silva[a],
Bruno Galveas Laviola[c], Marcos Deon Vilela de Resende[a,d], Leonardo Lopes Bhering[a,*]

[a] *Universidade Federal de Viçosa, Avenida Peter Henry Rolfs, Campus Universitário, Viçosa, MG, 36570-900, Brazil*
[b] *Monsanto, Avenida das Nações Unidas, 12901, Alto de Pinheiros, São Paulo, SP, 04578-000, Brazil*
[c] *Embrapa Agroenergia, Parque Estação Biológica, Asa Norte, Brasília, DF, 70770-901, Brazil*
[d] *Embrapa Florestas, Estrada da Ribeira, Km 111, Guaraituba, Colombo, PR, 83411-000, Brazil*

ARTICLE INFO

Keywords:
Mixed model methodology
Repeated measures
Selection index
Genetic selection
Biofuel

ABSTRACT

Despite being a species with great potential for biodiesel production, little research has been done on the breeding of *Jatropha curcas*, mainly with respect to its yield across harvests. Thus, the present study was carried out to analyze longitudinal data via multiple-trait Best Linear Unbiased Prediction (BLUP) for the genetic improvement of *Jatropha curcas*. The experiment was set up as a randomized block design with two blocks and five plants per plot. The seed yield of 730 individuals of 73 half-sib families was evaluated over six years. Variance components and genetic parameters were estimated via Restricted Maximum Likelihood (REML). The Additive Index was used for ranking and selection purposes. Genetic correlations of low to moderate magnitude were observed between pairs of harvests. The Multiple-trait BLUP / Additive Index procedure allowed for the selection of superior families based on the predicted genetic values, considering all the harvests. Therefore, it can be efficiently applied in the breeding of *Jatropha curcas*.

1. Introduction

The search for sustainable fuels has become a frequent topic in numerous world conferences. In this scenario, *Jatropha curcas* L. stands out as one of the main species from which biodiesel can be produced by virtue of its adaptability to different environments; its tolerance to drought, which enable its cultivation in low-fertile and non-irrigated soils (Laviola et al., 2017); and production life of more than 50 years (Nithiyanantham et al., 2012). Moreover, it has a higher capacity to transform oil into biodiesel compared with other species (Pu et al., 2011).

Despite being a species with great potential for biodiesel production, little research has been done on the breeding of *Jatropha curcas*. Recently, studies have been undertaken with a view to selecting genotypes using information of only one harvest (Junqueira et al., 2016); using several harvests independently (Bhering et al., 2013); or based on temporal stability and adaptability of genetic values (Alves et al., 2018a, 2018b). However, because *Jatropha curcas* is a perennial species that produces for many years, statistical methodologies that efficiently evaluate longitudinal data are necessary to allow for a more accurate genetic selection (Peixoto et al., 2016).

Longitudinal data analysis has a few peculiarities, as the many harvests are correlated with each other and because there may be heterogeneity of variances and covariances between them (Mrode, 2014). The interest in analyses of this sort of data usually lies in the prediction of genetic values across all harvests as well as in the identification of a covariance structure over time (Resende et al., 2014). In plant breeding, longitudinal data associated with harvests are usually analyzed via repeatability models (Resende et al., 2014).

However, the multiple-trait Best Linear Unbiased Prediction (BLUP) — also known as the model with an unstructured covariance matrix — can be a very efficient alternative, because it uses all pieces of information simultaneously and treats repeated measures as different and correlated traits, considering their heritabilities and genetic and residual correlations (Mrode, 2014; Resende et al., 2014). This covariance structure is applied to all random factors of the statistical model such as genotype effects, plot effects, and residual effects (Gilmour et al., 2004). In this scenario, the present study proposes to analyze longitudinal data

---

* Corresponding author.
E-mail address: leonardo.bhering@ufv.br (L. Lopes Bhering).

https://doi.org/10.1016/j.indcrop.2018.12.019
Received 1 July 2018; Received in revised form 25 November 2018; Accepted 6 December 2018

via multiple-trait Best Linear Unbiased Prediction (BLUP) for the genetic improvement of *Jatropha curcas*.

2. Materials and methods

2.1. Experimental data

The experiment was implemented in November 2008 as a randomized-block design with two blocks and five plants per plot that were arranged in rows spaced 4 m apart, with 2 m spacing between plants. The experiment was conducted in the experimental field of Embrapa Cerrados, located in Planaltina, DF - Brazil (15°35′30″ S and 47°42′30″ W; 1007 m asl). All management practices were based on Carels et al. (2012). The experiment consisted of an evaluation of the performance of 730 individuals of 73 half-sib families of *Jatropha curcas*. These individuals were evaluated for seed yield (kg plant^{-1}) in six harvests (2010–2015).

2.2. Statistical analysis

The Restricted Maximum Likelihood/Best Linear Unbiased Prediction (REML/BLUP) procedure was adopted to estimate the variance components and predict the genetic values, in accordance with Patterson and Thompson (1971) and Henderson (1975).

The multiple-trait statistical model, associated with the evaluation of genotypes in a randomized-block design with several plants per plot, is given by the following equation:

$$y = Xr + Zf + Wp + e,$$

where *y* is the phenotypes vector; *r* is the vector of fixed effects of blocks added to the overall mean; *f* is the vector of random effects of families, $f \sim N(0, \sum_a \otimes I)$; *p* is the vector of random error effects between plots, $p \sim N(0, \sum_p \otimes I)$; *e* is the vector of random residual effects within plots, $e \sim N(0, \sum_e \otimes I)$; and *X*, *Z*, and *W* represent the incidence matrices for the said effects. $\sum_a$ is a covariance matrix of random genetic effects of families; $\sum_p$ is a covariance matrix of random error effects between plots; $\sum_e$ is a covariance matrix of random residual effects within plots; *I* is an identity matrix of order appropriate to the respective random effect; and $\otimes$ denotes the Kronocker product. In addition, $\sum_a$, $\sum_p$, and $\sum_e$ are assumed unstructured covariance structures (US).

The additive genetic variance between family means (σ_a^2) was given by the following expression (Resende, 2015):

$$\sigma_a^2 = 4\sigma_g^2$$

where σ_g^2 is the estimate of the genetic variance between family means.

The heritability between family means (h_g^2) was given by the following expression (Resende, 2015):

$$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{1}{r}\sigma^2_{plot} + \frac{1}{nr}\sigma^2_{res}}$$

where σ^2_{plot} is the estimate of the variance between plots; σ^2_{res} is the estimate of the residual variance within plots; *r* is the number of blocks; and *n* is the number of plants per plot.

The standard errors of the estimates of variance components and of the genetic parameters were obtained by the mixed model output and post-processing "pin" function of ASReml software.

Genetic correlations (Pearson's correlation) between pairs of harvests ($r_{i,j}$) were obtained based on the following expression:

$$r_{i,j} = \frac{\hat{\sigma}_{i,j}}{\sqrt{\hat{\sigma}_i^2 \hat{\sigma}_j^2}}$$

where $\hat{\sigma}_{i,j}$ is the genetic covariance estimated between the family means for the pair of harvests *i* and *j*; $\hat{\sigma}_i^2$ is the genetic variance estimated between family means for harvest *i*; and $\hat{\sigma}_j^2$ is the genetic variance estimated between family means for harvest *j*. The genetic covariances were obtained directly by the mixed model output.

The Additive Index (Resende, 2007), used to identify superior families to be selected in the *Jatropha curcas* breeding program, is given by the following expression:

$$AI_k = \sum_{h=1}^{6} w_h \frac{(u_h + f_{kh})}{\sigma_h}$$

where w_h is the weight assigned for harvest *h*; $u_h + f_{kh}$ is the overall mean for harvest *h* added to the predicted genetic value of family *k* at harvest *h*; and σ_h is the standard deviation for $u_h + f_{kh}$. For the Additive Index, weights equal to u_h/u, where *u* is the overall mean, were assigned, and the selection direction was "higher". The gain predicted with the selection was obtained directly by the Additive Index output.

Statistical analyses were performed using ASReml 4.1 (Gilmour et al., 2015), Selegen REML/BLUP (Resende, 2016), and R (R Core Team, 2018) software.

3. Results

The standard errors of the estimates of the variance components revealed that all the random effects of the statistical model are statistically significant (i.e., they differ from zero) (Table 1). An increase was observed in the magnitude of genetic variance between family means and, consequently, in additive genetic variance between family means throughout the harvests, except for that of 2013. On the other hand, heritability between family means did not follow any pattern, ranging from 0.33 to 0.71 (2013 and 2014 harvests, respectively) (Table 1).

Genetic correlations between pairs of harvests ranged from 0.11 (2011 and 2013) to 0.80 (2014 and 2015) (Fig. 1), and 60% of these correlations were lower than 0.6. Overall, the highest genetic correlations were obtained for the last pairs of harvests.

Predicted genetic values in each harvest, Additive Index, and predicted selection gain across all harvests for the 73 half-sib families of *Jatropha curcas* are shown in the supplementary material - Table S1. By adopting a selection intensity of 27%, according to Additive Index, the 20 selected families were 41, 10, 15, 37, 36, 6, 54, 16, 9, 11, 72, 56, 18, 65, 34, 67, 73, 53, 70, and 39 (Fig. 2), and the predicted selection gain across all harvests was 12%.

Table 1

Estimates of variance components and genetic parameters and their respective standard error, between parenthesis, for the seed yield trait, evaluated in 730 individuals of 73 half-sib families of *Jatropha curcas* in six harvests (2010–2015).

Harvest	σ_g^2 [a]	σ_a^2 [b]	σ^2_{plot} [c]	σ^2_{res} [d]	σ^2_{phen} [e]	h_g^2 [f]
2010	0.0043	0.0171	0.0036	0.0056	0.0135	0.65
	(0.0012)	(0.0047)	(0.0008)	(0.0003)	(0.0012)	(0.08)
2011	0.0099	0.0397	0.0284	0.0314	0.0697	0.36
	(0.0055)	(0.0218)	(0.0057)	(0.0018)	(0.0056)	(0.15)
2012	0.1017	0.4069	0.1070	0.1427	0.3514	0.60
	(0.0308)	(0.1234)	(0.0228)	(0.0084)	(0.0313)	(0.10)
2013	0.0693	0.2772	0.2512	0.1372	0.4577	0.33
	(0.0428)	(0.1712)	(0.0457)	(0.0081)	(0.0429)	(0.16)
2014	0.2562	1.0249	0.1696	0.2012	0.6271	0.71
	(0.0637)	(0.2548)	(0.0356)	(0.0119)	(0.0638)	(0.07)
2015	0.2922	1.1689	0.2235	0.3099	0.8257	0.67
	(0.0772)	(0.3089)	(0.0496)	(0.0186)	(0.0785)	(0.08)

[a] genetic variance between family means.
[b] additive genetic variance between family means.
[c] variance between plots.
[d] residual variance within plots.
[e] phenotypic variance, and.
[f] heritability between family means.

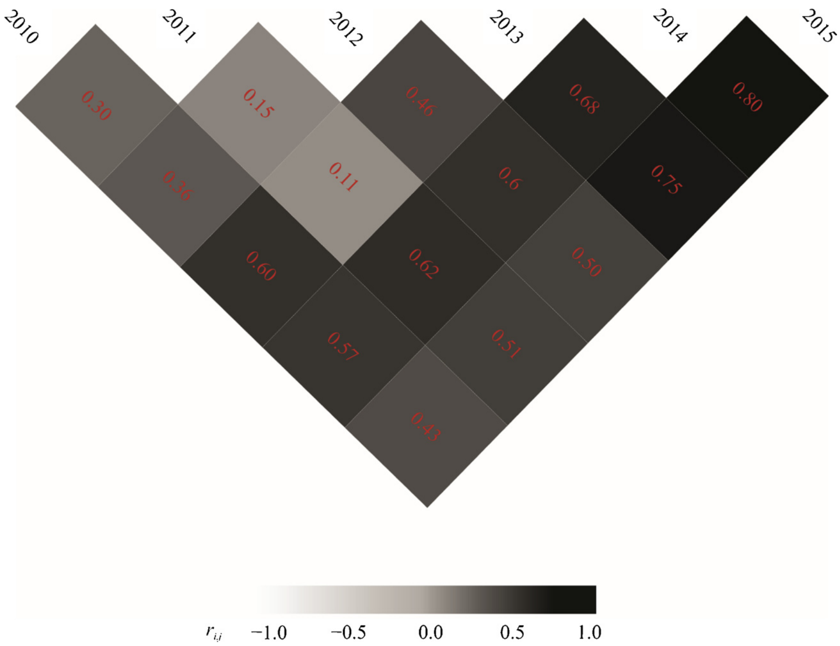

Fig. 1. Genetic correlations between pairs of harvests for the seed yield trait, evaluated in 730 individuals of 73 half-sib families of *Jatropha curcas*, in six harvests (2010–2015).

4. Discussion

BLUP assumes that variance components are known (Henderson, 1975). However, in practice, variance components are unknown and should thus be estimated in the most reliable way possible so that estimates can properly replace the parameters. In this case, the standard procedure for estimating variance components, under the approach of linear mixed models, is REML (Patterson and Thompson, 1971). Therefore, the genetic evaluation consists of using these methodologies together — the so-called REML/BLUP or mixed model methodology (Resende et al., 2014).

The main advantages of using the mixed model methodology are that it allows for the incorporation of kinship information; comparison of individuals or varieties over time and space; correction of environmental effects, simultaneous estimation of variance components and prediction of genetic values; and dealing with complex data structures. Moreover, the mixed model methodology can be applied to unbalanced data and non-orthogonal designs (Mrode, 2014; Resende, 2016).

In the identification and selection of superior families, genetic variance between families is characterized as one of the main components to quantify the breeding potential of the genotypes under study. Additionally, this parameter makes it possible to direct the selection

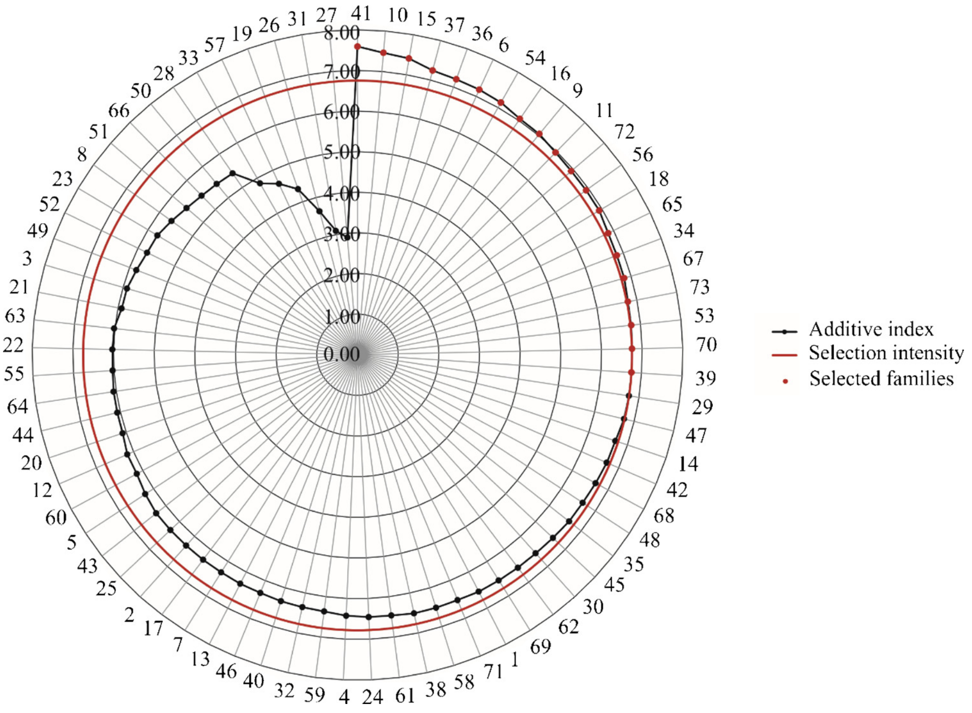

Fig. 2. Ranking of *Jatropha curcas* families and selected families by the Additive Index.

strategy to be used, increasing the frequency of favorable alleles (Kerr, 1998).

The heritable proportion of total variability is named ‘heritability’ (Falconer and Mackay, 1996). This is a measure of genetic influence that informs that part of the population variation in a phenotype may be attributed to the variation in the genotype, allowing for estimates such as the genetic gain expected from selection (Allard, 1999). According to a scale proposed by Resende (2015), the heritabilities of the 2010, 2012, 2014, and 2015 harvests were characterized as high magnitude, whereas those of the 2011 and 2013 harvests were classified as moderate magnitude. These heritability estimates indicate a moderate to high correlation between the genotype and the phenotype, constituting a favorable scenario for the selection of families.

In the breeding of perennial plants, the number of measurements taken typically varies from three to six, since a higher number of harvests would compromise the efficiency of breeding programs per time unit (Resende, 2015). Laviola et al. (2013) stated that the efficiency of early selection in *Jatropha curcas* is small and that a minimum of 4 and 7 measurements are required to achieve reliability levels of 70 and 80%, respectively, in predicting the genetic values.

Repeatability models (the most simple and parsimonious) are very efficient when the genetic correlation between successive harvests is high (above 0.80) and are greatly advantageous because few parameters are to be estimated (Resende et al., 2014). However, in the current study, most genetic correlations between harvests were lower than 0.80, indicating that more complex models must be used. Thus, the multiple-trait BLUP is a more suitable strategy for the analysis of longitudinal data in *Jatropha curcas*. The predicted selection gain, with the recombination of the selected families, is of high magnitude and, therefore, confirms the efficiency of this procedure in the genetic improvement of *Jatropha curcas*.

Imai et al. (2016) applied multiple-trait BLUP in citrus and demonstrated the usefulness of the method in predicting genetic values when the information of kinship among genotypes is known and in the presence of unbalanced data. The authors recommended multiple-trait BLUP as a tool to select genotypes. Greater accuracy and efficiency of multiple-trait BLUP were also reported by Costa et al. (2002) in rubber tree (*Hevea brasiliensis*) breeding, by Kerr (1998) in the genetic selection of open and controlled cross-pollination cultivars in forestry breeding, and by Alves et al., (2018a, 2018b) in the genetic selection of *Eucalyptus*.

Few studies exist in the literature evaluating longitudinal data in *Jatropha curcas*. Thus, the results obtained in the present study using the multiple-trait BLUP / Additive Index procedure contribute to filling a gap in the breeding of *Jatropha curcas*. In addition, this procedure can be used for genetic selection in other species. In future research, multiple-trait BLUP should be considered in association with genomic prediction analysis to improve accuracy and reduce the cycle time in the genetic improvement programs of *Jatropha curcas*.

5. Conclusion

Genetic correlations of low to moderate magnitude were observed between pairs of harvests.

The Multiple-trait BLUP / Additive Index procedure allowed for the selection of superior families based on the predicted genetic values, considering all the harvests. Therefore, it can be efficiently applied in the breeding of *Jatropha curcas*.

Acknowledgments

The authors thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for their financial support.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.indcrop.2018.12.019.

References

Allard, R.W., 1999. Principles of Plant Breeding, second edition. John Wiley & Sons, New York, USA.

Alves, R.S., Peixoto, L.A., Teodoro, P.E., Silva, L.A., Rodrigues, E.V., Resende, M.D.V., Laviola, B.G., Bhering, L.L., 2018a. Selection of *Jatropha curcas* families based on temporal stability and adaptability of genetic values. Ind. Crop Prod. 119, 290–293.

Alves, R.S., Rocha, J.R.A.S.C., Teodoro, P.E., Resende, M.D.V., Henriques, E.P., Silva, L.A., Carneiro, P.C.S., Bhering, L.L., 2018b. Multiple-trait BLUP: a suitable strategy for genetic selection of Eucalyptus. Tree Genet. Genomes 14, 77.

Bhering, L.L., Barrera, C.F., Ortega, D., Laviola, B.G., Alves, A.A., Rosado, T.B., Cruz, C.D., 2013. Differential response of *Jatropha* genotypes to different selection methods indicates that combined selection is more suited than other methods for rapid improvement of the species. Ind. Crop Prod. 41, 260–265.

Carels, N., Sujatha, M., Bahadur, B., 2012. *Jatropha*, Challenges for a New Energy Crop: Farming, Economics and Biofuel. Springer, New York, USA.

Costa, R.B., Resende, M.D.V., Gonçalves, P.D.S., Silva, M.D.A., 2002. Individual multivariate REML/BLUP in the presence of genotype x environment interaction in rubber tree (*Hevea*) breeding. Crop Breed. Appl. Biot. 131–140.

Falconer, D., Mackay, T., 1996. Introduction to Quantitative Genetics. Longman Scientific & Technical, Harlow, UK.

Gilmour, A.R., Cullis, B.R., Welham, S.J., Gogel, B.J., Thompson, R., 2004. An efficient computing strategy for prediction in mixed linear models. Comput. Stat. Data An. 44, 571–586.

Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J., Thompson, R., 2015. ASReml User Guide Release 4.1 Structural Specification. VSN International, Hemel Hempstead, UK.

Henderson, C.R., 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31 (2), 423–447.

Imai, A., Kuniga, T., Yoshioka, T., Nonaka, K., Mitani, N., Fukamachi, H., Hiehata, N., Yamamoto, M., Hayashi, T., 2016. Evaluation of the best linear unbiased prediction method for breeding values of fruit-quality traits in citrus. Tree Genet. Genomes 12, 119.

Junqueira, V.S., Peixoto, L.A., Laviola, B.G., Bhering, L.L., Mendonça, S., Costa, T.S.A., Antoniassi, R., 2016. Bayesian multiple-trait analysis reveals a useful tool to increase oil concentration and to decrease toxicity in *Jatropha curcas* l. PLoS One 11, e0157038.

Kerr, R.J., 1998. Asymptotic rates of response from forest tree breeding strategies using best linear unbiased prediction. Theor. Appl. Genet. 96, 484–493.

Laviola, B.G., Oliveira, A.M.C., Bhering, L.L., Alves, A.A., Rocha, R.B., Gomes, B.E.L., Cruz, C.D., 2013. Estimates of repeatability coefficients and selection gains in *Jatropha* indicate that higher cumulative genetic gains can be obtained by relaxing the degree of certainty in predicting the best families. Ind. Crop Prod. 51, 70–76.

Laviola, B.G., Rodrigues, E.V., Teodoro, P.E., Peixoto, L.A., Bhering, L.L., 2017. Biometric and biotechnology strategies in *Jatropha* genetic breeding for biodiesel production. Renew. Sust. Energy Rev 76, 894–904.

Mrode, R.A., 2014. Linear Models for the Prediction of Animal Breeding Values, third edition. CAB International, Wallingford, Oxon, UK.

Nithiyanantham, S., Siddhuraju, P., Francis, G., 2012. Potential of *Jatropha curcas* as a biofuel, animal feed and health products. J. Am. Oil Chem. Soc. 89, 961–972.

Patterson, H.D., Thompson, R., 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58, 545–554.

Peixoto, L.A., Laviola, B.G., Bhering, L.L., Mendonça, S., Costa, T.S.A., Antoniassi, R., 2016. Oil content increase and toxicity reduction in *Jatropha* seeds through family selection. Ind. Crop Prod. 80, 70–76.

Pu, Y., Treasure, T., Gonzalez, R.W., Venditti, R., Jameel, H., 2011. Autohydrolysis pretreatment of mixed hardwoods to extract value prior to combustion. Bioresources 6, 4856–4870.

R Core Team, 2018. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Resende, M.D.V., 2007. SELEGEN REML/BLUP: Sistema Estatístico E Seleção Genética Computadorizada Via Modelos Lineares Mistos. Embrapa Florestas, Colombo, Brasil.

Resende, M.D.V., 2015. Genética Quantitativa E De Populações. Suprema, Visconde do Rio Branco, Brasil.

Resende, M.D.V., 2016. Software Selegen-REML/BLUP: a useful tool for plant breeding. Crop Breed. Appl. Biot. 16, 330–339.

Resende, M.D.V., Silva, F.F., Azevedo, C.F., 2014. Estatística Matemática, Biométrica E Computacional. Suprema, Visconde do Rio Branco, Brasil.