

# MATHEMATICAL MODEL TO ESTIMATE THE GENETIC VALUE OF PROGENITORS AND CROSSINGS IN SUGAR CANE

## Modelo matemático para estimar el valor genético de progenitores y cruces en caña de azúcar

Reynaldo Rodríguez Gross<sup>1✉</sup>, Yaquelin Puchades Isaguirre<sup>1</sup>, Wilfre Aiche Maceo<sup>1</sup> and María T. Cornide Hernández<sup>2</sup>

**ABSTRACT.** Parental selection is one of the most crucial steps to improve genetic gain in any breeding program. The objective of the present work was to propose a mathematical model to estimate, from the selection information, the genetic value of parents and crosses with the processing of various variables in their progenies in sugarcane in Cuba. For this, information was used from the first three stages of selection (postures, clonal propagation 1 and 2) of the improvement program developed in the south-eastern region of Cuba, from 2000 to 2007. Likewise, the computer system was used SASEL, used in the genetic improvement of sugarcane, to program the proposed mathematical model. The model allowed estimating, in a multivariate way, the genetic value of the parents and crosses used in the genetic improvement program of sugarcane in Cuba. It was possible to establish, from the prediction of this value, the most outstanding parents and crosses in the analyzed period, which constitutes a useful tool to contribute to the improvement of the improvement program.

*Key words:* genetic, software, selection

### INTRODUCTION

The selection of parents and the prediction of their value is one of the most important actions in any selection program (1). Therefore, increased selection and genetic gain is an important measure of the success of breeding programs (2).

**RESUMEN.** La selección de progenitores es una de las etapas más importantes en cualquier programa de mejora para obtener ganancia genética. El objetivo del presente trabajo fue proponer un modelo matemático para estimar, a partir de la información de selección, el valor genético de progenitores y cruces con el procesamiento de diversas variables en sus progenies en caña de azúcar en Cuba. Para esto se utilizó la información de las primeras tres etapas de selección (posturas, propagación clonal 1 y 2) del programa de mejora desarrollado en la región sur-oriental de Cuba, desde el año 2000 al 2007. Asimismo, se recurrió al sistema informático SASEL, utilizado en el mejoramiento genético de la caña de azúcar, para programarle el modelo matemático propuesto. El modelo permitió estimar, de forma multivariada, el valor genético de los progenitores y cruces utilizados en el programa de mejoramiento genético de la caña de azúcar en Cuba. Se pudo establecer, a partir de la predicción de este valor, los progenitores y cruces más destacados en el período analizado, lo que constituye una herramienta útil para contribuir al perfeccionamiento del programa de mejora.

*Palabras clave:* genética, informática, selección

In the sugar cane selection process, particularly in the early stages, various traits that make up agricultural and industrial performance are evaluated, as well as the incidence of diseases. A simultaneous selection methodology is necessary for these early stages, which is the basis for selecting an index based on multiple characters rather than intensive selection on a particular character (3).

Currently, it is widely accepted that mathematical models can provide valuable tools for studying the behavior of plants, while also saving field experiments, resources and time (4-6).

<sup>1</sup> Grupo de Genética y Protección de Plantas. Estación Territorial de Investigaciones de la Caña de Azúcar Oriente Sur, Cuba. (22) 502254

<sup>2</sup> Grupo de Genética y Protección de Plantas. Instituto de Investigaciones de la Caña de Azúcar, Cuba

✉ rodriguez@inicasc.azcuba.cu

Statistical analysis of correlation and linear regression are frequently used for the evaluation of models (7).

The Research Institute of the Sugar Cane of Cuba develops a program of genetic improvement to respond to the obtaining of new cultivars. This means that annually it manages a large population of clones in different stages and years of selection. For this purpose, a computer program (SASEL) is available to capture, store and process the information obtained in the selection process (8). However, this program does not have a model that allows estimating the genetic value of parents and crosses from the selection information.

The objective of the present work is to propose a mathematical model to estimate, from the selection information, the genetic value of parents and crosses with the processing of various variables in their progenies in sugarcane.

## MATERIALS AND METHODS

### INFORMATION USED IN THE MODEL

Selection data from the first three stages of the sugarcane breeding program developed in the

south-eastern region of Cuba were used. The period of genetic selection included the years from 2000 to 2007 (Table 1).

The trials were established in the experimental area of América Libre town of Contramaestre municipality in the Santiago de Cuba province (-76.2° longitude and 20.3° latitude) on a Brown sialitic soil. The conduction of the studies was carried out according to the methodological norms of the genetic improvement program of sugarcane in Cuba (9).

For the construction of the model, a total of 18 variables were taken into account, which in turn constitutes the selection criteria in each of the stages considered (Table 2). This information obtained in the field was captured and validated by the SASEL software. The three stages of genetic selection and the years of study comprised a database composed of the information of the progeny of 242 and 110 female and male parents respectively, as well as 640 double-parent crosses, which allowed establishing the mathematical model.

**Table 1. Stages of the genetic selection scheme in sugar cane, series and cycles used for the preparation of the mathematical model**

Study stages	Series (years)	Month of planting	Cycle evaluated	Age of assessment (months)
Seedlings	2000 al 2007	September	First sprout	7
Clonal propagation 1	2000 al 2007	September	First sprout	12
Clonal propagation 2	2000 al 2007	April	First sprout	12

**Table 2. Stages of the selection scheme and evaluated variables considered in the relationship functions of the mathematical model**

Stage	Evaluated variables considered (relationship function)	Stage	Evaluated variables considered (relationship function)
Seedling	- Selection percentage	LC2	- Selection percentage
	- % of individual affected by coal		- % of individual affected by coal
	- % of individual affected by rust		- % of individual affected by rust
	- Selection percentage		- % of individual affected by low Brix
LC1	- % of individual affected by coal	General (from the stage of postures to LC2)	- Number of stems
	- % of individual affected by rust		- % of individuals selected from LC2 with respect to total LC1
	- % of individual affected by low Brix		- % individuals selected from LC2 regarding the postures (final selection)
	- Refractometric Brix		
	- Stem diameter		
	- Stem length		
	- Number of stems		

LC1- Clonal propagation 1; LC2- Clonal propagation 2; % - Percentage

**Mathematical model to estimate the genetic value of parents and crosses**

A linear model was used to quantify the existing interaction between the genotype (parent) or the crossing and the environment (genotype-environment interaction) through the following function F, which depends on the parent or cross 'x' and the environment 'y'.

$$F(x,y) = \left[ \frac{\sum_{i=1}^n \alpha_i C_i(x,y)}{\sum_{i=1}^n \alpha_i} \right] * 100$$

where:

C<sub>i</sub>(x, y), 0 ≤ C<sub>i</sub>(x, y) ≤ 1 for (i = 1,2, ..., N) are the relationship functions between the parent or cross "x" with the environment "y".

α<sub>i</sub>, 0 ≤ α<sub>i</sub> ≤ 1 para (i = 1,2,...,N) are the weighting coefficients of the relationship functions.

i, is the specific relationship function that are included in the mathematical model

N, is the total number of relationship functions that are included in the mathematical model.

The number of relationship functions (variables evaluated to the progenies of the parents or crosses) to be used in the mathematical model is a function of the interests of the breeder and of the stage or stages that it is decided to use. The mathematical model is flexible and allows orientation depending on the objectives of improvement that are pursued. Likewise, the weighting coefficients can be modified depending on the selection objectives and the relative importance of the variables to be considered.

To convert the evaluated values of the variables used in the C<sub>i</sub>(x, y) function with a range between zero and one (0 ≤ C<sub>i</sub>(x, y) ≤ 1, for (i = 1,2, ..., N) ) the following procedure was used:

$$C_i(x,y) = \left[ \frac{X_i Y}{Max(X_i Y)} \right]$$

X<sub>i</sub>Y - Value of the evaluated variable of the parent or crossing i in the Y environment  
 Max X<sub>i</sub>, Y - Maximum value found of the variable used by the parent or crossing i in the Y environment

**RESULTS Y DISCUSIÓN**

The proposed mathematical model offers the possibility of using in its conformation up to 18 variables that constitute selection criteria in the three selection stages considered (Figure 1).

It also allows modifying the weighting coefficient (called Imp.) of each of these variables, based on the relative importance of the variables.

This coefficient oscillates between zero and one, and can be considered, in some variables, as an interpretation of heritability. In this case, there are variables related to Brix, components of agricultural yield and resistance to diseases.

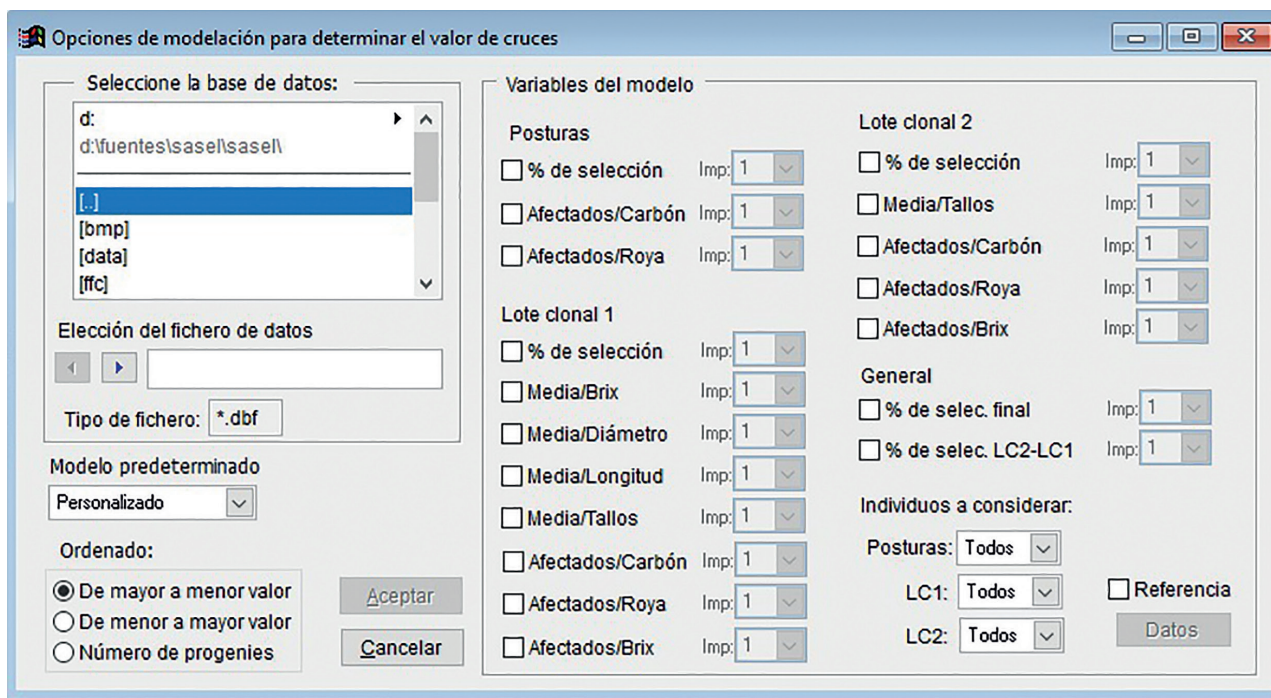


Figure 1. SASEL software options to choose variables and weighting coefficient (Imp) to estimate the genetic value of the crosses

When defined by the breeder the variables to be considered and the relative weight of the same (weighting coefficient), the mathematical model estimates the genetic value of the crosses contained in the selection database or of the parents as chosen. This value is estimated from the evaluation made to their progenies and allows the breeder to select the crosses or parents with the best response according to the objectives pursued.

Also, it can restrict the number of crosses to be considered by the model, by discriminating the crosses that have low number of progenies evaluated ("Individuals to consider" option). This possibility allows increasing the accuracy of the mathematical model because it estimates the genetic value of parents and crosses with a greater number of evaluated progenies.

Other options of the model menu allow inserting, before calculating the genetic value of the crosses, the ideal or "reference" values of the variables considered. That is to say, from these introduced values, the model determines the ideal or expected value of the crossing and allows establishing a limit to consider the crossings as outstanding or to discard.

From the objectives of selection in the cultivation of sugarcane (9) and to facilitate the work of the breeder, five predetermined models were established that estimate the genetic value of parents and crosses (Table 3). As can be seen for each model, the variables that define the improvement objectives in sugarcane were chosen.

This does not mean that the breeder can include or exclude other variables in the model according to the desired objectives.

**Table 3. Variables and weighting coefficients used in pre-established models to estimate the genetic value of parents and crosses**

Selection stage	Variables considered in each given model	Objectives of the predetermined models and definition of the weighting coefficient				
		Selection	Agricultural yield	Brix	Resistance to rust	Resistance to coal
Seedlings	- % of selection	0,5			0,7	0,7
	-% of individuals affected by coal					1
	-% of individuals affected by rust				1	
LC1	- % of selection	1	0,7	0,7	0,7	0,7
	-% of individuals affected by coal					1
	-% of individuals affected by rust				1	
	-% of individuals affected by low brix			0,8		
	- Brix refractometric			1		
	- Diameter of the stem		1			
	- Stem length		1			
- Number of stems		1				
LC2	- % of selection	0,8	0,7	0,7	0,7	0,7
	-% of individuals affected by coal					1
	-% of individuals affected by rust				1	
	-% of individuals affected by low brix			0,8		
	- Number of stems		1			
General (since seedling to LC2)	-% of individuals selected from LC2 with respect to total LC1	1			0,7	0,7
	-% of individuals selected from LC2 with respect to total positions (final selection)	0,8		0,7	0,7	0,7

LC1- Clonal propagation 1; LC2- Clonal propagation 2; % - Percentage

The predetermined models are aimed at obtaining an estimate of the value of parents and crosses for different improvement objectives such as: contribution of selection, sugar content (Brix), resistance to diseases such as rust and sugar cane, as well as agricultural yield. In all cases, the variable that was systematically considered in the five models was the percentage of selection in the stages of clonal propagation 1 and 2. These variables are important because they summarize the qualities of the individual as a whole.

When choosing a model oriented to obtain an estimate of the genetic value of crosses and parents for agricultural yield, the variables that make up the agricultural yield are enabled such as diameter, length and number of stems with a maximum weighting coefficient. At the same time, the selection variables (percentage of selection in clonal propagation stage 1 and 2) with weighting coefficients of 0.7 were included.

Thus, when the objective of the model was not to estimate the genetic value of parents and cross for their contribution to selection, the weighting coefficients of the selection percentage variables were reduced to maximize the coefficients of the other model variables. These weighting values were determined from the selection pressure established by selection stages (9) and the numerous simulations that were carried out with the series studied.

In the case of estimating the genetic value of the male and female parents, the menu options of the mathematical model are the same, with the difference that there is an option to choose whether to work with the female or male parents.

By executing the predetermined mathematical model (selection contribution), the genetic value of the crosses could be estimated (Table 4). The first five crosses with the highest value reached were: C229-84 x CP70-11333, CP70-1527 x C144-79, C86-503 x CP70-1527, C323-68 x C86-506 and C90-501 x B6368. These crosses achieved a value higher than 40 %, which could indicate a high specific combinatorial ability.

**Table 4. Estimation of the genetic value of the first 22 crosses for a model based on percentage of selection in all stages evaluated**

Nu.	Cross	Value of the model variables						Final value
		Individuals LC1	Seedling Selection	LC1 Selection	LC2 Selection	General Selection 1-2    Final selection		
1	C229-84 X CP70-1133	65	1	100	24	76	12	52,0
2	CP70-1527 X C144-79	227	46	21	25	22	77	46,6
3	C86-503 X CP70-1527	84	11	39	27	39	72	45,7
4	C323-68 X C86-503	41	2	43	32	80	20	42,9
5	C90-501 X B6368	68	1	29	80	48	11	41,6
6	CP70-1527 X C86-502	226	6	32	38	44	43	39,8
7	CP52-43 X B45181	160	4	22	51	46	28	36,9
8	C1616-75 X C86-503	91	13	25	44	45	24	36,7
9	CP52-43 X CP70-1133	405	27	55	11	10	45	35,9
10	C323-68 X CP72-2086	169	5	36	20	39	33	32,3
11	Na63-90 X C86-502	50	3	60	20	33	15	31,8
12	CP72-2086 X My56180	100	18	23	16	16	49	29,8
13	C323-68 X C84-474	191	3	33	29	34	15	27,8
14	C86-253 X B45181	167	5	33	18	29	24	26,6
15	C323-68 X Co421	85	3	50	12	29	13	26,1
16	Mex69-257 X C88-519	42	9	30	16	20	28	24,9
17	Co312 X Co6806	42	8	30	16	20	25	24,0
18	Eros X CP72-2086	58	12	30	13	14	28	23,9
19	Pomex72 X Mex60-1459	483	20	20	12	10	34	23,5
20	Co740 X CP70-1133	308	7	28	16	21	24	23,4
21	CP48-103 X C85-277	48	6	26	27	17	17	22,7
22	C86-12 X CP70-1133	595	4	35	19	21	14	22,5

LC1- Clonal propagation 1; LC2- Clonal propagation 2



In the same way, the rest of the crosses also reached a high genetic value with respect to the 640 crosses managed in the study database. In particular, we can point out crosses that reached a high value with a high number of progenies evaluated. These crosses were: C86-253 x B45181 (26.6 %), Pomex72 x Mex60-1259 (23.5 %), Co470 x CP70-1133 (23.4 %), Co740 x CP70-1133 (23.5 %) and C86-12 x CP70-1133 (22.5 %).

This last crossing is the progenitors of the cultivar C00-575 obtained and recommended for environmental conditions from Santiago de Cuba province (10). The lowest values of the variables related to the percentage of selection correspond to the stage of seedlings and final selection. This may be due to several factors such as: high population of individuals in the initial stage with respect to those who are finally selected in clonal propagation 2, competition effects in the posture phase, as well as criteria and selection pressure in the three stages of genetic improvement.

As for the crosses, the model estimates the genetic value of female and male parents (Table 5 and 6). The female parents with the highest selection contribution and value greater than 40 % were: B80-250, CP70-1133, Na63-90, C1616-75, CP45-105, C86-407 and C90-501. These progenitors stood out fundamentally for their contribution of selection in

the stages of clonal propagation 1 and 2, as well as percentage of selection in the clonal propagation 1 with respect to the 2. These results show a high general combinatorial ability of these progenitors. Within this group of female parents, the C90-501 is noteworthy for its high estimate of genetic value and at the same time high number of progenies evaluated (1045). Also, with a high number of progenies evaluated and high estimated genetic value compared to the total number of parents studied: C323-68 (36.7 %), C86-456 (35.8 %), CP52-43 (34.2 %), C229-84 (30.4 %), C1051-73 (29.6 %), CP70-1527 (29 %) and C88-553 (27.9 %).

Regarding male parents according to the estimated genetic value of the model, the first four genotypes with a percentage higher than 40%, the following highlighted are: C87-506, CP70-1527, My5724 and C86-502. Other genotypes with a high number of evaluated progenies are also highlighted, which gives significance to the results obtained by the model. In this regard, the following parents stand out: CP52-43 (38.1 %), CP70-1133 (27.7 %), CP72-2086 (26.9 %), C86-503 (25.6 %), C86-531 (22.5 %) and B45181 (21.4%). As for the female parents, the high estimates of the genetic value of these male parents show a high overall combinatorial ability.

**Table 5. Estimation of the genetic value of the first 20 female progenitors for a model based on percentage of selection in all stages evaluated**

Nu.	Femenine progenitor	Value of the model variables						Final value
		Individuals LC1	Seedlings Selection	LC1 Selection	LC2 Selection	General Selection 1-2    Final selection		
1	B80250	315	50	79	40	10	49	55,6
2	CP70-1133	255	18	54	35	42	77	55,2
3	Na63-90	50	2	100	30	61	15	50,9
4	C1616-75	134	2	37	65	80	5	46,0
5	CP45-105	96	7	74	21	48	34	45,0
6	C86-407	127	8	53	28	48	40	43,3
7	C90-501	1045	4	59	35	48	18	40,1
8	C323-68	786	1	49	34	59	7	36,7
9	C86-502	53	2	47	30	58	10	35,9
10	C86-456	948	11	62	16	27	30	35,8
11	CP52-43	1336	5	59	26	32	18	34,2
12	C86-253	241	3	52	25	44	12	33,0
13	Co740	317	5	47	23	39	21	32,9
14	Mex60-1459	94	11	80	7	16	18	32,3
15	B4327	184	4	20	80	17	7	31,2
16	C229-84	408	4	55	22	30	13	30,4
17	Co312	55	9	38	24	28	25	30,1
18	C1051-73	691	5	45	25	29	18	29,6
19	CP70-1527	838	3	38	35	35	9	29,1
20	C88-553	1587	1	45	28	35	5	27,9

**Table 6. Estimation of the genetic value of the first 20 male parents for a model based on percentage of selection in all stages evaluated**

Nu.	Masculine Progenitor	Value of the model variables						Final value
		Individuals LC1	Seedling Selection	LC1 Selection	LC2 Selection	General Selection 1-2	Final selection	
1	C87-506	69	15	41	46	80	100	68,7
2	CP70-1527	104	13	51	27	53	57	48,9
3	My5724	81	4	70	25	68	20	45,5
4	C86-502	307	3	58	35	72	18	45,2
5	CP52-43	498	27	42	15	22	50	38,1
6	CP65-315	75	1	22	80	37	4	35,1
7	My56180	138	9	41	23	40	29	34,5
8	C89-509	45	5	72	9	31	13	31,5
9	C84-474	262	3	42	29	42	11	31,1
10	C144-79	290	19	29	25	29	16	28,8
11	CP70-1133	2772	3	56	17	30	8	27,7
12	C79-54	129	4	56	11	32	10	27,6
13	C88-553	110	7	59	9	25	14	27,6
14	Ja64-20	380	10	30	28	25	20	27,5
15	B37141	216	16	52	13	13	17	27,3
16	CP72-2086	618	3	59	12	27	10	26,9
17	C85-277	199	25	24	15	14	29	26,2
18	C86-503	427	6	31	26	32	10	25,6
19	C86-531	578	4	48	15	19	6	22,5
20	B45181	1696	1	37	19	27	3	21,4

The use of the BLUP models (Better Non-Skewed Linear Predictor) is the most common procedure to estimate the components of variance and the value of crosses (11,12). However, they require, for their best estimation, information about the parents' family tree (1,13).

The model proposed in this study offers an estimate of the genetic value of a parent or crossbreed, based on the response of its progenies in the genetic selection program of sugar cane in Cuba. Thus, it allows evaluating the breeders, through the first three stages of selection, the use or not of crosses and parents according to the improvement objectives, while also making it possible to perfect the established crossing program in order to obtain a new cultivar with superior characteristics to the witnesses.

Other authors use the evaluation of plant breeding and genotype classification by means of phenotypic ranges with the use of decision tree and the logistic regression method in early stages of selection (14,15). However, the use of one or the other method or their combination is indispensable to evaluate the genetic gain and efficiency in breeding programs.

## CONCLUSIONS

The proposed mathematical model allowed estimating, in a multivariate way, the genetic value of progenitors and crosses used in the program of genetic improvement of sugarcane in Cuba. In this way, it is possible to assess, through the first three stages of selection, the use or not of crosses and parents according to the improvement objectives, while at the same time it makes it possible to perfect the established cross-breeding program in order to obtain a new cultivar with superior characteristics to the witnesses.

## BIBLIOGRAPHY

1. Atkin FC, Dieters MJ, Stringer JK. Impact of depth of pedigree and inclusion of historical data on the estimation of additive variance and breeding values in a sugarcane breeding program. *Theoretical and Applied Genetics*. 2009;119(3):555–65. doi:10.1007/s00122-009-1065-7
2. Zhou MM, Gwata ET. Location effects and their implications in breeding for sugarcane yield and quality in the midlands region in South Africa. *Crop Science*. 2015;55(6):2628–38. doi:10.2135/cropsci2015.02.0101

3. Tahir M, Hussain Khalil I, McCord PH, Glaz B, Todd J. Phenotypic Selection of Sugarcane (*Saccharum* spp.) Genotypes using best linear unbiased predictors. *Journal of American Society of Sugarcane Technologists*. 34:44–56.
4. Herndl M, Shan C-G, Pu W, Graeff S, Claupein W. A model based ideotyping approach for wheat under different environmental conditions in North China plain. *Agricultural Sciences in China*. 2007;6(12):1426–36. doi:10.1016/S1671-2927(08)60004-8
5. Letort V, Cournède P-H, Mathieu A, De Reffye P, Constant T. Parametric identification of a functional–structural tree growth model and application to beech trees (*Fagus sylvatica*). *Functional plant biology*. 2008;35(10):951–63.
6. Rodríguez Gross R, Puchades Izaguirre Y, Abiche Maceo W, Rill Martínez S, Suarez HJ, Salmón Cuspineda Y, *et al*. Estudio del rendimiento y modelación del periodo de madurez en nuevos cultivares de caña de azúcar. *Cultivos Tropicales*. 2015;36(4):134–43.
7. Gauch HG, Hwang JT, Fick GW. Model evaluation by comparison of model-based predictions and measured values. *Agronomy Journal*. 2003;95(6):1442–6.
8. Rodríguez R, Puchadez Y, Abiche Maceo W, Tamayo M, Vázquez L, García H, *et al*. Software para la gestión de la información obtenida en el programa de mejoramiento de la caña de azúcar. *Revista Cuba&Caña*. 2012;(1):35–45.
9. Jorge H, González R, Casas M, Jorge I. Normas y procedimientos del programa de mejoramiento genético de la caña de azúcar en Cuba. *PUBLINICA*. La Habana, Cuba. 2011;308.
10. Rill S, Rodríguez S, Puchades D, Abiche Maceo W. C00-575 nueva variedad de caña de azúcar con alto potencial productivo. *Revista Cuba&Caña*. 2014;(1):7–11.
11. Wei X, Jackson P, Piperidis G, Parfitt R, Atkin F, Cox M. How well are the predicted breeding values of parental clones associated with family performance in the BSES-CSIRO sugarcane breeding program? *International Sugar Journal*. 2012;115(1371):208–11.
12. Hodge G. KeyNote 1: Use of best linear unbiased prediction in forest tree improvement and potential applications in plant breeding [Internet]. CAMCORE Cooperative, North Carolina State University, Campus Box 7626, Raleigh, NC 27695-7626, USA . [cited 2018 Apr 3]. Available from: [www.sapba.co.za/uploads/files/conferences/2004/abstract\\_booklet.pdf](http://www.sapba.co.za/uploads/files/conferences/2004/abstract_booklet.pdf)
13. Piepho HP, Möhring J, Melchinger AE, Büchse A. BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica*. 2008;161(1–2):209–28. doi:10.1007/s10681-007-9449-8
14. Zhao X, Liu Z, Dan F, Wang K. Plant breeding evaluation with rank entropy-based decision tree. *IFAC-PapersOnLine*. 2016;49(16):336–40. doi:10.1016/j.ifacol.2016.10.062
15. Zhou MM, Kimbeng CA, Tew TL, Gravois KA, Pontif M, Bischoff KP. Logistic regression models to aid selection in early stages of sugarcane breeding. *Sugar Tech*. 2014;16(2):150–6. doi:10.1007/s12355-013-0266-1

Received: October 5<sup>th</sup>, 2017

Accepted: January 19<sup>th</sup>, 2018