



# NEW INSIGHTS INTO ADJUSTMENT FOR SPATIAL DEPENDENCE OF SOIL ATTRIBUTES AND USE OF AERIAL IMAGES IN THE INITIAL SELECTION STAGE OF SUGARCANE FAMILIES

Matheus de Paula Ferreira<sup>1</sup>, Marco Luís Conde Machado de Souza<sup>1</sup>, Luiz Alexandre Peternelli<sup>1\*</sup>, Marcio Henrique Pereira Barbosa<sup>2</sup>, Danilo Pereira Barbosa<sup>3</sup>, Antônio Policarpo Souza Carneiro<sup>1</sup>

Submitted: 06/05/2024

Accepted: 31/07/2024

<sup>1</sup>Department of Statistics, Federal University of Viçosa, Av. Peter Henry Rolfs, s/n – Campus Viçosa, Viçosa – MG, 36570-900, Brazil. ORCID: 0000-0002-4110-9147; 0009-0001-2716-1731; 0000-0002-7683-0583; 0000-0002-9043-3242

<sup>2</sup>Department of Agronomy, Federal University of Viçosa, Av. Peter Henry Rolfs, s/n – Campus Viçosa, Viçosa – MG, 36570-900, Brazil. ORCID: 0000-0002-5838-9579

<sup>3</sup>Mathematics and Statistics Centre, Federal Institute of the State of Goiás, Rodovia Sul Goiana, Km 01, Zona Rural, Campus Rio Verde, Rio Verde – GO, 75901-970, Brazil. ORCID: 0000-0001-5117-4009

\*Correspondence email: [peternelli@ufv.br](mailto:peternelli@ufv.br)

**Abstract:** One of the challenges of sugarcane breeding programs is the initial selection of genotypes aimed at developing varieties. Following the hybridization of genitors and family production, the process generally consists of selecting the best families and individuals within those families. Furthermore, the traditional selection model based on experimental designs assumes that the experimental field is well-prepared and homogeneous within blocks. If the allocation of blocks is incorrect, the ordering of families will be compromised, directly impacting the selection process and the program's success. Researchers seek quick, non-destructive alternatives to contribute to a less biased family selection process. These alternatives include modeling and statistical analysis or alternative data collection through images by unmanned aerial vehicles. This work proposes adjusting the ranking of sugarcane families by incorporating soil attributes in the statistical model and evaluating how some vegetation indices (VI) derived from the visible spectrum are associated with sugarcane yield (TSH). The experiment consisted of 60 families in a randomized complete block design with four replications. We also collected 36 soil samples and aerial images. The construction of the field map based on soil attributes and geostatistics indicated that the spatial position of the experimental blocks was incorrect. The correction implemented in this work allowed the ranking of families without the influence of the systematic variation in soil attributes. Additionally, the VI used showed a linear association with TSH, indicating the possibility of using aerial images to select or discard families in the initial stages of a breeding program.

**Keywords:** RGB; plant selection; vegetation indices; geostatistics; early selection.

## 1. Introduction

Brazil is the world's largest producer of sugarcane, with a planting area estimated at approximately 8.5 million hectares, and is also one of the world leaders in sugar and ethanol production, according to data from the third harvest survey 2023/24 (CONAB, 2023). Because of the search for sustainable energy sources, ethanol fuel derived from sugarcane is considered one of the primary renewable and clean energy sources (Canabarro et al., 2023), which makes Brazil a global reference in fuel production (Paiva et al., 2021). Therefore, there is great interest in developing new sugarcane cultivars with greater productivity and richness in sucrose, the raw material for the manufacture of ethanol (Yang et al., 2019).

In this context, genetic improvement programs play a fundamental role in developing new cultivars and methodologies for the more efficient selection of genotypes. The selection of the first genotypes begins by screening the families and then separating the superior genotypes within the best families since the heritability of family characteristics is greater than in individual plants (Moreira and Peternelli, 2015).

Improvements in the statistical methodology for selecting individuals are always an issue. In this sense, information from the soil can contribute to the quality of this process since soil fertility expresses the capacity to supply essential nutrients for crop development and, therefore, the families' performance knowledge of these attributes is fundamental for implementing field experiments so that soil characteristics can be used as a correction factor for experimental models (Dalchiavon et al., 2017).

The traditional selection model is based on a randomized block design (Ferreira et al., 2022), where only the effects of repetition and individual genotypes or families are considered, thus ignoring other possible systematic factors, such as soil information. However, these factors can be crucial in modeling the adjustment of main effects, such as the averages of the families under evaluation. One way to incorporate such information is to use soil characteristics as covariates in the experimental design model (Dalchiavon et al., 2017). Including these covariates in the statistical model can be understood as a combination of ANOVA and regression analysis, allowing the covariates to correct for the response variable so that families can be further classified (Rencher and Schaalje, 2008).

Another aspect of great interest is the use of aerial image data associated with individual selection characteristics. Using unmanned aerial vehicles (UAVs) with onboard sensors makes obtaining this data fast, efficient, and flexible (Yang et al., 2017). Information from aerial images in the agricultural context can help reduce labor and input use by digitizing large areas quickly and allowing for vegetation indices to be obtained that can be linked to valuable agronomic traits and serve as auxiliary variables for decision-making (Araus and Cairns, 2014). This situation may happen frequently in the case of sugarcane when, after ordering the families, the researcher needs to quickly decide whether a family or clone should be selected or discarded (Brasileiro et al., 2015).



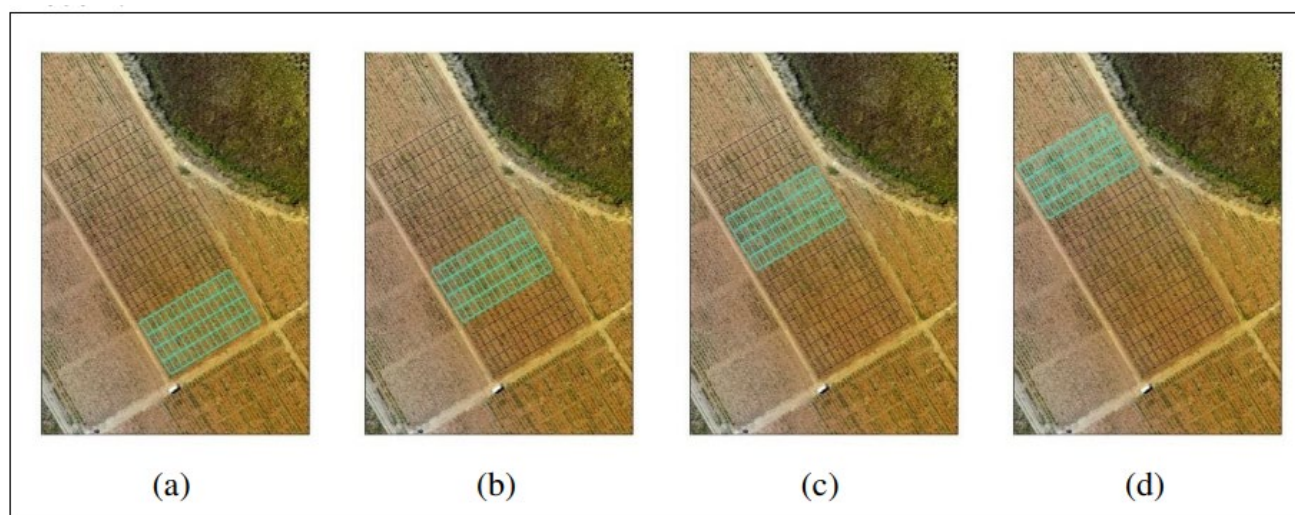
Several vegetation indices are created for specific purposes (Barbosa et al., 2019). In particular, the indices VARI (Visible Atmospherically Resistant Index), GLI (Green Leaf Index), and NGRDI (Normalized Green Red Difference Index) are often found in plant breeding work (Barbosa et al., 2021; Ranjan et al., 2019; Volpato et al., 2021).

This work aimed to compare the traditional methodological analysis for ranking sugarcane families with the methodology of incorporating soil variables into the design model, using geostatistics techniques to correct the ranking and better select families in the initial stage of genetic breeding programs. Additionally, we will evaluate how vegetation indices obtained from RGB images relate to productivity.

## 2. Materials and Methods

### 2.1. Experimental data

The experiment was conducted at the Sugarcane Research and Improvement Center (CECA) at the Federal University of Viçosa (UFV), located in the municipality of Oratórios, MG (latitude 20°25' S, longitude 42°48' W and 494 m altitude). Ferreira et al. (2022) provide a detailed description of the data's origin. Sixty full-sib families were evaluated through a randomized complete block design (RCBD) experiment with four blocks, totaling 240 plots. Each plot comprised two 5.0 meters long furrows spaced at 1,4 m. The distance between plants in each furrow was 0,5 m, totaling 20 plants per plot. For each plot, after harvest, the variable tons of stalks per hectare (TSH) was estimated (Ferreira et al., 2022). Figure 1 shows the position of the four blocks in the experiment.



**Figure 1** – Layout of the experiment, with its four blocks and their plots. (a) Block 1. (b) Block 2. (c) Block 3. (d) Block 4.

### 2.2. Soil Attributes

To obtain attributes related to soil fertility in the experimental area, we considered 36 soil samples collected in a regular 4×9 grid at 0 - 20 cm depth on 08/21/2019 after fully harvesting the experiment and collecting the agronomic traits. The soil chemical attributes analyzed from the samples are presented in Table 1.

**Table 1** – Soil chemical variables' respective acronyms and measurement units/scale.

Variable	Acronym	Unit of measurement
Hydrogen potential	pH	0-14
Phosphor	P	mg/dm <sup>3</sup>
Potassium	K	mg/dm <sup>3</sup>
Magnesium	Mg	cmolc/dm <sup>3</sup>
Calcium	Ca	cmolc/dm <sup>3</sup>
Aluminium	Al	cmolc/dm <sup>3</sup>
Potential acidity	H+Al	cmolc/dm <sup>3</sup>
Sum of bases	SB	cmolc/dm <sup>3</sup>
Cation exchange capacity	CTC(t)	cmolc/dm <sup>3</sup>
Cation exchange capacity at pH 7	CTC(T)	cmolc/dm <sup>3</sup>
Aluminium saturation index	m	%
Base saturation index	V	%



### 2.3. Imaging

The RGB images that originated the orthomosaics were obtained from a flight of an RPA (Remotely Piloted Aircraft) UAV, model Phantom 4 from the DJI brand, with four rotors and equipped with an RGB camera, whose FC330 sensor has a resolution of 12 megapixels, focal length equal to 3.61 mm,  $1.56 \times 1.56 \mu\text{m}$  pixel, and CMOS sensor. The flight occurred on 03/15/2019, after 12 months of transplanting the plants into the experimental area and when they were fully developed. The following flight parameters were adopted: height of 60 meters; 425 images collected; 2 seconds interval between photographs; Ground Sample Distance (GSD) of 1.58 cm; flight started at 11:30 am. The photographic device was fixed on a 3-axis stabilizing platform, enabling rotation, stability, and independent movement of the UAV. The camera movement angle was from  $0^\circ$  to  $-90^\circ$  for panoramic or nadir photography (Fonseca Neto et al., 2018).

### 2.4. Statistical Analysis

Statistical analyses were performed using the software R (R Development Core Team, 2023), and ArcGIS (ESRI, 2004), version 10.2.2. To propose a correction in the ranking model of sugarcane families, we compared two models, M1 and M2. Model M1 corresponds to the usual family selection method (Ferreira et al., 2022) based on a randomized block design analysis, while in model M2, we implemented the soil information in the design model.

#### 2.4.1. Geostatistical procedures

To fit theoretical semivariogram models and create kriging maps (Li et al., 2022), four methods were used: ordinary least squares (OLS), weighted least squares (WLS), maximum likelihood (ML), and restricted maximum likelihood (REML). Furthermore, three theoretical semivariogram models were analyzed: exponential, Gaussian, and spherical. The best model was chosen from the visual analysis of the adjustment of the theoretical model to the sampling points of the experimental semivariogram, followed by self-validation (Pasini et al., 2021). With the kriging maps for each soil variable that showed spatial dependence, the average value of the attributes was extracted within each of the 240 plots of the experiment, whose spatial positioning in the field was known.

#### 2.4.2. Analysis of covariance and correction of means

A two-step analysis was performed to determine which variables showing spatial dependence would be used as covariates in the explanatory model of the TSH response variable (Stich et al., 2008). The ANOVA residuals for the RCBD model were obtained in the first stage. In the second stage, the residues obtained in the previous stage were used as a dependent variable and as a function of the soil variables that showed spatial dependence. Still, in the second stage, the soil variables that significantly affected the decomposition of residues were selected using the stepwise method. Stepwise is a widely known variable selection method used in several works (Anshori et al., 2021; Liu et al., 2021; Sant'Anna et al., 2020; Yassein et al., 2020).

The modeling involved in the two-step selection mentioned above can be understood as follows. The initial model is given by:

$$y_{ij} = \mu + f_i + b_j + \varepsilon_{ij} \quad (1)$$

where  $y_{ij}$  is the value of the family  $i$  on the block  $j$ ,  $\mu$  is the model constant,  $f_i$  is the family effect  $i$  ( $i = 1, 2, \dots, 60$ ),  $b_j$  is the block effect  $j$  ( $j = 1, 2, 3, 4$ ),  $\varepsilon_{ij}$  is the residue in the family  $i$  on the block  $j$ .

To insert the effect of soil covariates, the residues  $\varepsilon_{ij}$  obtained in equation (1) were decomposed as:

$$\varepsilon_{ij} = \beta_k X_{ijk} + \gamma_{ij} \quad (2)$$

where  $\beta_k$  is the regression coefficient of the covariate  $k$  ( $k = 1, 2, \dots, p$ ),  $X_{ijk}$  is the value of the covariate  $k$  in the family  $i$  on the block  $j$ ,  $\gamma_{ij}$  is the new residues now disregarding the effect of covariates. This way, we can obtain the corrected value for each plot,  $z_{ij}$  doing:

$$z_{ij} = y_{ij} - (\beta_k X_{ijk}) = \mu + f_i + b_j + \gamma_{ij} \quad (3)$$

The term  $\hat{\beta}_k X_{ijk}$ , represents the correction values of the TSH response variable at the plot level.

#### 2.4.3. Vegetation indices and linear correlation

Vegetation indices generated from RGB images are obtained through algebraic expressions between the red (R), green (G), and blue (B) bands. For this work, we have considered the indices presented in Table 2.



**Table 2** – Vegetation indices (VI) used in this study. B: blue, R: red, and G: green.

IV	Name	Objective	Equation	Reference
VARI	Visible Atmospherically Resistant Index	Highlight vegetation, mitigate the influence of atmospheric effects	$\frac{G - R}{G + R - B}$	(GITELSON et al., 2002)
GLI	Green Leaf Index	Indicate chlorophyll, classify where there are live, dead plants, and exposed soil	$\frac{2G - R - B}{2G + R + B}$	(LOUHAICHI et al., 2013)
NGRDI	Normalized Green Red Difference Index	Identify areas of vegetation and green biomass	$\frac{G - R}{G + R}$	(TUCKER, 1979)

For each plot  $ij$  ( $i = 1, \dots, 60$  families, and  $j = 1, \dots, 4$  blocks) of the experiment, we obtained the values related to vegetation indices VARI, GLI e NGRDI. From these values, the means of each index were calculated at the family level ( $\overline{VARI}_i$ ,  $\overline{GLI}_i$ ,  $\overline{NGRDI}_i$ ). With these average values and the corresponding family-level TSH averages, a graphical analysis was performed to infer the type of association between each VI and TSH.

### 3. Results and discussion

Analysis of variance based on a randomized complete block design for the TSH response variable resulted in a significant family effect ( $p$ -value  $< 0,01$ ). This result indicates that there is a difference between family averages and, therefore, reinforces the idea that any possible adjustment must be made on the residues of this first analysis (Peternelli and de Resende, 2015), as shown in equation 2, to better compare the family averages. As the objective of this work is only to compare the ranking of families and not to carry out tests to compare means, it was decided not to evaluate the assumptions of the analysis of variance (Rencher and Schaalje, 2008). Any values considered outliers were also kept in the analysis, as the presence of outliers for this study can be an important indicator, as they may be associated with some anomaly in the experimental plots, such as the death of individuals or different environmental conditions due to variation in soil fertility.

#### 3.1. Adjustment of the semivariogram model

From the soil samples, geostatistical analysis was carried out for the twelve soil variables (Table 1), of which ten, Al, Ca, H+Al, Mg, P, pH, SB, t, and V, showed signs of spatial dependence (Table 3). For each of these attributes, we obtained the spatial dependence index – SDI (Cezana et al., 2024), the adjusted model and its parameter estimates, and the method that allowed the best adjustment of the semivariogram (Table 3). There was a predominance of moderate spatial dependence ( $25 < SDI < 75$ ), indicating that the spatial dependence was well captured and that the point sampling grid was adequate (Guimarães, 2013).

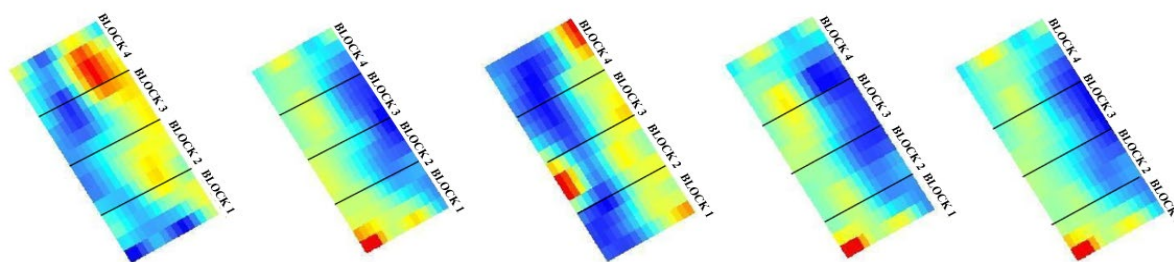
**Table 3** – Adjusted semivariogram parameters and estimation methods that describe the spatial variability of soil chemical variables that showed spatial dependence. Identification of variables: see Table 1;  $C_0$ : nugget effect;  $C_1$ : contribution; SDI: spatial dependence index; WLS: weighted least squares; ML: maximum likelihood; and REML: restricted maximum likelihood.

Variable	$C_0$	$C_1$	Range(m)	Model	SDI (%)	Class	Method
Al	0,06	0,086	23,51	Spherical	0,61	Moderate	WLS
Ca	0,35	0,180	24,76	Spherical	0,34	Moderate	WLS
H+Al	0,30	0,390	32,01	Spherical	0,57	Moderate	REML
m	65,32	247,410	21,00	Exponential	0,79	High	ML
Mg	0,006	0,018	40,36	Gaussian	0,75	High	WLS
P	2,00	9,250	24,71	Spherical	0,82	High	ML
pH	0,042	0,037	32,24	Gaussian	0,47	Moderate	WLS
SB	0,12	0,220	36,53	Gaussian	0,65	Moderate	WLS
t	0,05	0,069	13,22	Exponential	0,58	Moderate	ML
V	25,00	44,460	36,36	Gaussian	0,64	Moderate	WLS



### 3.2. Analysis of covariance and correction of means

From the selection of variables using the stepwise method, the variables m, Mg, P, pH, and SB showed greater significance in explaining the variation in residues (equation 2), with regression coefficients equal to -0.94, -350.32, 2.16, -99.97, and 96.97 respectively. This way, the variables m, Mg, and pH penalize the values observed in the plots, while P and SB bonus the response variable values. The selected penalizing variables seek to equalize the conditions for plant development. The map of average values at the plot level for each selected variable was obtained through the spatial distribution of the variables m, Mg, P, pH, and SB and interpolated by kriging (Figure 2). These maps indicate that the blocks were incorrectly defined when planning the experiment, as they do not determine homogeneous regions for these attributes. Therefore, when ranking families using the M1 model, the response variable may have been directly influenced by the source of systematic variation arising from these soil variables.



**Figure 2** – Map of the average value of the most significant soil variables and original marking of the blocks in the experiment. (a): Aluminum, (b): Magnesium, (c): Phosphorus, (d): Hydrogen potential, (e): Sum of bases.

To compare the results from model M1 (based on the original design) and model M2 (based on adjustment for soil covariates), we assume that the breeder is interested in selecting the 20% best sugarcane families that make up this study, that is, the families in positions 1 to 12 out of the 60 families in ranking. Comparing model M1 and model M2, there was an agreement of 83,33% and, therefore, a disagreement of approximately 16% in selecting the best families (Table 4), indicating the possibility of losing any good families throughout the selection process.

**Table 4** – Ranking of the 20% of the best families using the M1 and M2 models. M1: a model based on the adjustment of family means from the original design; M2: a model based on family means adjustment after considering corrections due to soil variations. Families that did not coincide in the selection from both models are in bold.

Rank	1	2	3	4	5	6	7	8	9	10	11	12
M1	289	499	285	470	497	473	219	<b>466</b>	514	424	<b>539</b>	384
M2	289	499	473	285	424	497	<b>439</b>	219	514	470	<b>348</b>	384

The divergences regarding the selection percentage of the 20% best families occurred to the sugarcane families identified by codes 466, 539, 439, and 348, appearing in bold in Table 4. In model M1, families 466 and 539 were ranked eighth and 11<sup>th</sup>, respectively, within the selection percentage of 20%, while families 348 and 439 occupied the 25<sup>th</sup> and 13<sup>th</sup> positions, respectively. On the other hand, in the ranking carried out by the M2 model, families 348 and 439 are within the selection of 20% of the best families, being classified in 11<sup>th</sup> and 7<sup>th</sup>, respectively. In contrast, families 466 and 539 are classified in positions of 17<sup>th</sup> and 13<sup>th</sup> in that order.

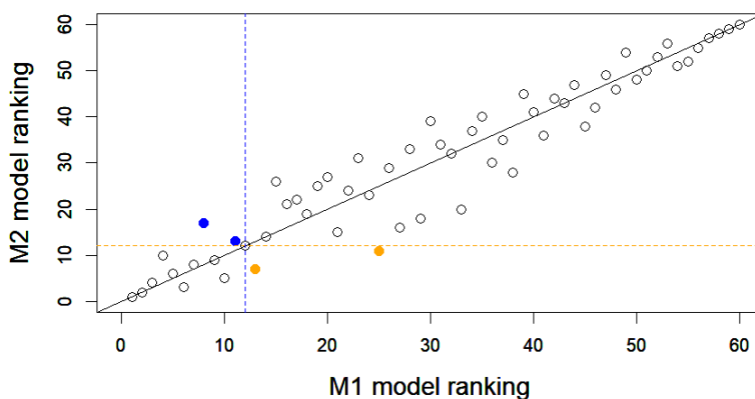
Given the incorrect implementation of the blocks in the field, one may question if the randomization of plots within each block favored some families over others, so the location of their plots was privileged concerning the combination of soil attribute levels at that plot. For example, aluminum's value is associated with soil toxicity (Shetty et al., 2021), while pH greatly influences crop development (Msimbira and Smith, 2020). Phosphorus, in turn, is an essential macronutrient in plant metabolism (Lambers, 2022) and is of great importance in crop development.

Therefore, by classification via model M1, families 466 and 539 may have been favored due to the positioning of their plots in the field. In contrast, families 348 and 439 were disadvantaged since this model did not consider soil information. In short, as information on soil attributes was not considered and when choosing a blocking design did not guarantee homogeneity within the blocks, unequal conditions may occur for the evaluation of families and, consequently, a comparison with a greater possibility of error in the selection process. Incorporating soil variables into the model allows a correction in the response variable, making it possible to reduce discrepant conditions caused by the influence of soil variables, thus allowing a fairer comparison between families.

It is important to comment that, in general, the installation of field experiments is carried out based on ease of execution and the expectation that the eventual soil fertility correction process has been previously carried out correctly and homogeneously (De

Campos et al., 2022) which does not always happen. Due to the costs of sampling and analysis, soil fertility information is not routinely obtained in the field before the experiments' installation, making the block allocation process in the field risky and always challenging.

Considering all 60 families evaluated, we observed that 49 families were classified in different positions (Figure 3) and that the point cloud moves away from direct identity (a 45-degree straight line), indicating dissimilarity between the classifications. This inconsistency in ranking the families following the analysis reinforces that the inclusion of the effects of soil variables in the modeling can generate changes in family ranking, probably enabling better family selection if we assume the blocks are not homogeneous. In Figure 3, families 466 and 539, possibly favored by the positions of their plots in the field, are identified in blue, within the selection percentage of the 20% best families according to the M1 model ranking. Families 348 and 439, possibly harmed by the M1 model, are identified by orange and are within the selection percentage based on the M2 model ranking.

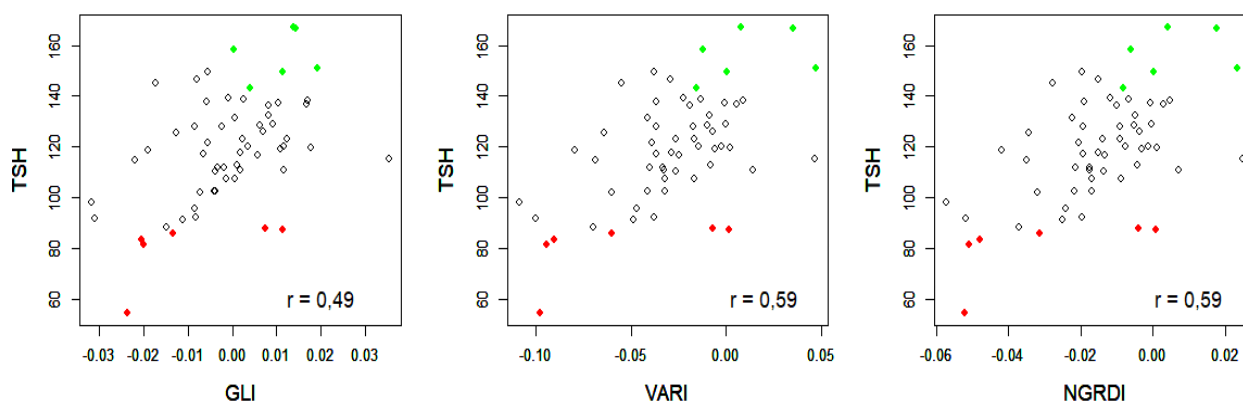


**Figure 3** – Scatter diagram of family classification according to models M1 and M2. The points to the left of the blue vertical line correspond to the 20% of families selected by the M1 model. The points below the orange horizontal line correspond to the 20% of families selected by the M2 model. The highlighted points show the four families that did not coincide in the selection by both models, as shown in Table 4.

In genetic breeding programs, besides choosing the best crossings among genitors, selection at T1 is crucial for the continuity of the program (Moreira et al., 2021; Peternelli et al., 2018; Verissimo et al., 2020). Therefore, when homogeneous conditions within experimental blocks are not guaranteed, a good amount of genetic material can be wasted or wrongly selected, resulting in valuable losses in the following breeding process.

### 3.3. Vegetation indices

Using the average productivity ( $TSH_i$ ) and the average vegetation indices ( $\overline{VARI}_i$ ,  $\overline{GLI}_i$ ,  $\overline{NGRDI}_i$ ) at the family level ( $i = 1, \dots, 60$ ), it is possible to graphically observe a positive linear association trend between the indices and the TSH (Figure 4). Pearson correlation coefficient values (Steel and Torrie, 1980) between the variables  $GLI \times TSH$  ( $r = 0,49$ ),  $VARI \times TSH$  ( $r = 0,59$ ) e  $NGRDI \times TSH$  ( $r = 0,59$ ) indicate moderate correlations (Dancey and Reidy, 2018), or can be classified as medium magnitude for  $GLI \times TSH$ , and strong for  $VARI \times TSH$  e  $NGRDI \times TSH$ , accordingly to (Cohen, 2013).



**Figure 4** – Scatter diagrams between vegetation indices and the TSH variable. Each point corresponds to the average taken at a family level.



It is also observed that in all graphs (Figure 4), the 10% best families selected by the M2 model are located at the upper end of the point cloud, identified by the green color, while the 10% worse families selected by the M2 model are at the lower end of the point cloud, identified by the color red. These results reinforce a linear association between vegetation indices (VARI, GLI, and NGRDI) and the TSH. Khuimphukhieo et al. (2023), found significant correlations between the percentage of sucrose and the NGRDI and VARI indices in a study using different vegetation indices to evaluate sugarcane productivity. The author concluded that using vegetation indices obtained by UAV can be helpful in sugarcane breeding programs. Todd et al. (2022), highlight the potential of using remote sensing methods using RGB images to evaluate sugarcane seedlings. The author carried out selection and prediction work on characteristics related to the productivity of sugarcane seedlings using various vegetation indices, including the NGRDI.

These results highlight the potential of vegetation indices as an alternative for monitoring the production cycle and as a characteristic for the prior selection of families (or even individuals represented by experimental plots) in cases where many families are to be selected (initial screening) in the field. It is worth mentioning that using sensors that include the RE (red-edge) and NIR (Near InfraRed) bands instead of exclusively RGB sensors allows the application of other vegetation indices (Araus and Cairns, 2014) and the possibility of better association results. Vasconcelos et al. (2023) obtained a predictive model for sugarcane production using the VARI and also the NDVI vegetation indices (which include the NIR band) as explanatory variables, concluding that vegetation indices can be reliably used to predict sugarcane productivity.

#### 4. Conclusions

From the geostatistical analysis of soil attributes, it was possible to verify non-homogeneity within the experimental blocks, indicating that they were incorrectly defined. In this way, the proposed model, M2, with the incorporation of soil chemical attributes, was able to promote adjustments in the response variable, enabling the ranking of sugarcane families without the influence of the source of systematic variation caused by the spatial dependence of the soil attributes, thus producing corrections in the ranking of families when compared to the ranking made by the traditional M1 model. Additionally, the use of the VARI, GLI, and NGRDI vegetation indices obtained through images in the RGB standard showed the existence of a positive linear association between each of the indices and sugarcane productivity at the family level, demonstrating the potential for using these indices for future research related to the prior selection of sugarcane families in the initial phases of genetic improvement programs.

**Acknowledgments:** The authors are thankful to the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG) for the financial support, and the Inter-University Network for the Development of Sugarcane Industry (RIDESA) for providing the dataset.

#### 5. References

- Anshori, M.F., Purwoko, B.S., Dewi, I.S., Ardie, S.W., Suwarno, W.B. A New Approach to Select Doubled Haploid Rice Lines under Salinity Stress Using Indirect Selection Index. *Rice Sci* 28, 368–378, 2021. <https://doi.org/10.1016/j.rsci.2021.05.007>.
- Araus, J.L., Cairns, J.E. Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci* 19, 52–61, 2014. <https://doi.org/10.1016/j.tplants.2013.09.008>.
- Barbosa, B.D.S., Ferraz, G.A.E.S., Dos Santos, L.M., Santana, L.S., Marin, D.B., Rossi, G., Conti, L. Application of RGB images obtained by UAV in coffee farming. *Remote Sens (Basel)* 13, 2397, 2021. <https://doi.org/10.3390/rs13122397>.
- Barbosa, B.D.S., Ferraz, G.A.S., Gonçalves, L.M., Marin, D.B., Maciel, D.T., Ferraz, P.F.P., Rossi, G. RGB vegetation indices applied to grass monitoring: A qualitative analysis. *Agronomy Research* 17, 349–357, 2019. <https://doi.org/10.15159/AR.19.119>.
- Brasileiro, B.P., Marinho, C.D., Costa, P.M. de A., Cruz, C.D., Peternelli, L.A., Barbosa, M.H.P. Selection in sugarcane families with artificial neural networks. *Crop Breeding and Applied Biotechnology* 15, 72–78, 2015. <https://doi.org/10.1590/1984-70332015v15n2a14>.
- Canabarro, N.I., Silva-Ortiz, P., Nogueira, L.A.H., Cantarella, H., Maciel-Filho, R., Souza, G.M. Sustainability assessment of ethanol and biodiesel production in Argentina, Brazil, Colombia, and Guatemala. *Renewable and Sustainable Energy Reviews* 171, 113019, 2023. <https://doi.org/10.1016/j.rser.2022.113019>.
- Cezana, D.C., Gontijo, I., Cavalcanti, A.C., Silva, M.B. da, Santos, E.O. de J., Partelli, F.L. Spatio-temporal variability of leaf macronutrients in a conilon coffee crop. *Brazilian Journal of Production Engineering* 10, 178–187, 2024. <https://doi.org/10.47456/bjpe.v10i1.42755>.
- Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2<sup>ed</sup>. Routledge, New York, 2013. 567 p. <https://doi.org/10.4324/9780203771587>.
- CONAB - COMPANHIA NACIONAL DE ABASTECIMENTO. Acompanhamento da safra brasileira de cana-de-açúcar, Brasília, DF, 12(1), 2024. ISSN: 2318-7921. Available at: <http://www.conab.gov.br>.
- Dalchiavon, F.C., Rodrigues, A.R., Lima, E.S. de, Lovera, L.H., Montanari, R. Variabilidade espacial de atributos químicos do solo cultivado com soja sob plantio direto. *Revista de Ciências Agroveterinárias* 16, 144–154, 2017. <https://doi.org/10.5965/223811711622017144>.
- Dancey, C., Reidy, J. *Estatística Sem Matemática para Psicologia*, 7<sup>ed</sup>. Penso Editora, Porto Alegre, 2018. 624 p. ISBN: 9788584291434.



- De Campos, M., Rossato, O.B., Marasca, I., Martello, J.M., de Siqueira, G.F., Garcia, C.P., Rossetto, R., Calonego, J.C., Cantarella, H., Crusciol, C.A.C. Deep tilling and localized liming improve soil chemical fertility and sugarcane yield in clayey soils. *Soil Tillage Res* 222, 105425, 2022. <https://doi.org/10.1016/j.still.2022.105425>.
- ESRI. ArcGIS Desktop: Release 10. Environmental Systems Research Institute, Redlands, 2004.
- Ferreira, P.H.S., Gonçalves, M.T.V., Teixeira, G., Ferreira, M. de P., de Oliveira, R.L., Barbosa, M.H.P., Peternelli, L.A. Comparison of family selection methodologies used in the initial phase of sugarcane breeding. *Crop Sci* 62, 679–689, 2022. <https://doi.org/10.1002/csc2.20685>.
- Fonseca Neto, F. de D., De Oliveira, J.C., Botelho, M.F., Dos Santos, A. de P., Nascimento, L.A., Fonseca, A.L.B. Avaliação da qualidade posicional de ortofotos geradas por SISVANT com diferentes alturas de voo para aplicações cadastrais. *Revista Brasileira de Cartografia* 69, 1247–1261, 2018. <https://doi.org/10.14393/rbcv69n7-43986>.
- Guimarães, W.D. Geoestatística para o mapeamento da variabilidade espacial de atributos físicos do solo. Doctor Scientiae thesis in Civil Engineering. Federal University of Viçosa, Viçosa, 60p. 2013.
- Khuimphukhieo, I., Marconi, T., Enciso, J., da Silva, J.A. The use of UAS-based high throughput phenotyping (HTP) to assess sugarcane yield. *J Agric Food Res* 11, 100501, 2023. <https://doi.org/10.1016/j.jafr.2023.100501>.
- Lambers, H. Phosphorus Acquisition and Utilization in Plants. *Annu Rev Plant Biol* 73, 17–42, 2022. <https://doi.org/10.1146/annurev-arplant-102720-125738>.
- Li, Y., Baorong, Z., Xiaohong, X., Zijun, L. Application of a semivariogram based on a deep neural network to Ordinary Kriging interpolation of elevation data. *PLoS One* 17, e0266942, 2022. <https://doi.org/10.1371/journal.pone.0266942>.
- Liu, Y., Heuvelink, G.B.M., Bai, Z., He, P., Xu, X., Ding, W., Huang, S. Analysis of spatio-temporal variation of crop yield in China using stepwise multiple linear regression. *Field Crops Res* 264, 108098, 2021. <https://doi.org/10.1016/j.fcr.2021.108098>.
- Moreira, É.F.A., Barbosa, M.H.P., Peternelli, L.A. Can statistical learning models make early selection among sugarcane families easier and still efficient? *Crop Sci* 61, 456–465, 2021. <https://doi.org/10.1002/csc2.20334>
- Moreira, É.F.A., Peternelli, L.A. Sugarcane families selection in early stages based on classification by discriminant linear analysis. *Brazilian Journal of Biometrics* 33, 484–493, 2015. Available at: <http://www.biometria.ufpr.br/index.php/BBJ/article/view/28>.
- Msimbira, L.A., Smith, D.L. The Roles of Plant Growth Promoting Microbes in Enhancing Plant Tolerance to Acidity and Alkalinity Stresses. *Front Sustain Food Syst* 4(106), 2020. <https://doi.org/10.3389/fsufs.2020.00106>.
- Paiva, I.C., Carvalho, L.M., Sanches, S.L., Moreira Silva, G. Estudo da cana de açúcar na matriz energética brasileira. *Casos de Marketing Público y No Lucrativo* 8, 3–10, 2021. <https://doi.org/https://doi.org/10.5281/zenodo.6640412>.
- Pasini, M.P.B., Engel, E., Lúcio, A.D., Nora, S.L.D. Selection of Interpolators to Predict Populations of *Tibraca limbativentris* in Irrigated Rice. *Brazilian Archives of Biology and Technology* 64, 1–12, 2021. <https://doi.org/10.1590/1678-4324-2021180601>.
- Peternelli, L.A., Bernardes, D.P., Brasileiro, B.P., Barbosa, M.H.P., Silva, R.H.T. Decision Trees as a Tool to Select Sugarcane Families. *Am J Plant Sci* 09, 216–230, 2018. <https://doi.org/10.4236/ajps.2018.92018>.
- Peternelli, L.A., de Resende, M.D.V. Experimental Designs for Next Generation Phenotyping, in: *Phenomics*. Springer International Publishing, Cham, pp. 15–32, 2015. [https://doi.org/10.1007/978-3-319-13677-6\\_2](https://doi.org/10.1007/978-3-319-13677-6_2)
- Ranjan, R., Chandel, A.K., Khot, L.R., Bahlol, H.Y., Zhou, J., Boydston, R.A., Miklas, P.N., 2019. Irrigated pinto bean crop stress and yield assessment using ground based low altitude remote sensing technology. *Information Processing in Agriculture* 6, 502–514, 2019. <https://doi.org/10.1016/j.inpa.2019.01.005>.
- R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Version 4.1.0. Vienna, 2023. Available at: <https://www.R-project.org/>.
- Rencher, A.C., Schaalje, G.B., 2008. *Linear models in statistics*, 2<sup>o</sup>ed. John Wiley & Sons, New Jersey, 2008. 672 p. ISBN: 978-0-471-75498-5.
- Sant'Anna, I. de C., Silva, G.N., Nascimento, M., Cruz, C.D. Subset selection of markers for the genome-enabled prediction of genetic values using radial basis function neural networks. *Acta Sci Agron* 43, e46307, 2020. <https://doi.org/10.4025/actasciagron.v43i1.46307>.
- Shetty, R., Vidya, C.S.-N., Prakash, N.B., Lux, A., Vaculík, M. Aluminum toxicity in plants and its possible mitigation in acid soils by biochar: A review. *Science of The Total Environment* 765, 142744, 2021. <https://doi.org/10.1016/j.scitotenv.2020.142744>.
- Steel, R.G.D., Torrie, J.H. *Principles and procedures of statistics, a biometrical approach*, 2<sup>o</sup>ed. McGraw-Hill Book Company, New York, 1980. 512 p. ISSN: 0070609268.
- Stich, B., Möhring, J., Piepho, H.-P., Heckenberger, M., Buckler, E.S., Melchinger, A.E. Comparison of Mixed-Model Approaches for Association Mapping. *Genetics* 178, 1745–1754, 2008. <https://doi.org/10.1534/genetics.107.079707>.
- Todd, J., Johnson, R., Verdun, D., Richard, K. Identification of Selection Preferences and Predicting Yield Related Traits in Sugarcane Seedling Families Using RGB Spectral Indices. *Agriculture* 12, 1313, 2022. <https://doi.org/10.3390/agriculture12091313>.
- Vasconcelos, J.C.S., Speranza, E.A., Antunes, J.F.G., Barbosa, L.A.F., Christofoletti, D., Severino, F.J., de





- Almeida Cançado, G.M. Development and Validation of a Model Based on Vegetation Indices for the Prediction of Sugarcane Yield. *AgriEngineering* 5, 698–719, 2023. <https://doi.org/10.3390/agriengineering5020044>.
- Verissimo, M.A.A., Azeredo, A.A.C. de, Silva, S.D. dos A. e, Brasileiro, B.P., Besspalhok Filho, J.C., Daros, E., Oliveira, R.A. de. Selection of sugarcane families and clones under cold stress. *Ciência Rural* 50(1), 2020. <https://doi.org/10.1590/0103-8478cr20181043>.
- Volpato, L., Dobbels, A., Borem, A., Lorenz, A.J. Optimization of temporal UAS-based imagery analysis to estimate plant maturity date for soybean breeding. *The Plant Phenome Journal* 4(1), e20018, 2021. <https://doi.org/10.1002/ppj2.20018>.
- Yang, G., Liu, J., Zhao, C., Li, Zhenhong, Huang, Y., Yu, H., Xu, B., Yang, X., Zhu, D., Zhang, X., Zhang, R., Feng, H., Zhao, X., Li, Zhenhai, Li, H., Yang, H. Unmanned Aerial Vehicle Remote Sensing for Field-Based Crop Phenotyping: Current Status and Perspectives. *Front Plant Sci* 8(1111), 2017. <https://doi.org/10.3389/fpls.2017.01111>.
- Yang, X., Song, J., Todd, J., Peng, Z., Paudel, D., Luo, Z., Ma, X., You, Q., Hanson, E., Zhao, Z., Zhao, Y., Zhang, J., Ming, R., Wang, J. Target enrichment sequencing of 307 germplasm accessions identified ancestry of ancient and modern hybrids and signatures of adaptation and selection in sugarcane (*Saccharum* spp.), a 'sweet' crop with 'bitter' genomes. *Plant Biotechnology Journal* 17, 488–498, 2019. <https://doi.org/10.1111/pbi.12992>.
- Yassein, A.A.M., Khalaf, A.E.A., Mohdaly, A.A.A., Roby, M.H.H. Selections of donors depending on agronomic traits, seed yield components, and fatty acid profile for genetic improvement of *Carthamus* using stepwise multiple regression. *OCL* 27(66), 2020. <https://doi.org/10.1051/ocl/2020063>.