MULTIVARIATE ANALYSIS OF PEANUT MECHANIZED HARVESTING

Rafael H. F. Noronha¹, Cristiano Zerbato², Rouverson P. da Silva², Antonio T. S. Ormond², Mailson F. de Oliveira²

¹Corresponding author. São Paulo State University - UNESP/ Jaboticabal - SP, Brasil. E-mail: rafael.noronha.agro@gmail.com

KEYWORDS
soil textural classes, 
Arachis hypogaea L., 
principal components analysis.

ABSTRACT
The peanuts harvesting mechanization is affected by the soil physical characteristics and it may increase the losses due to the production of pods in subsurface. The objective of the experiment was to identify the clusters through multivariate exploratory approaches from similarity in six soil textures (very clayey, clayey, silty clayey loam, clayey loam, sandy loam and sandy) in the state of São Paulo, Brazil, determining the main agronomic variables that most influenced the clustering division to assist the decision-making process in peanuts mechanized harvesting. The data were analyzed by the multivariate exploratory that is performed to simplify the description of a set of interrelated variables, using: yield, maturity, soil and pod moisture content, windrow width and height, visible and invisible digging losses, and gathering losses, as agronomic indicators of quality. The low and high clay content were grouped into clusters I and III, respectively, according to the agronomic traits of the peanut crop. The principal components analysis (PC) allowed a single distribution of accesses since only two eigenvalues were higher than “one”; the highest eigenvalues of 4.51 and 1.79, resulted in a Biplot that explained 70% of the original variability, 50.11% and 19.89% of which in the PC1 and PC2, respectively. The multivariate analysis indicated that high peanut yields in soils with low clay are correlated with the losses during the peanut mechanized harvesting operation.

INTRODUCTION
Peanuts (Arachis hypogaea L.) are an important oilseed in the Brazilian market, but even more in the state of São Paulo. The state produced 98% of the 418,300 tons of peanuts produced in the 2016/2017 harvest in Brazil, and peanuts are currently used as a rotational crop in sugarcane area and even pastures (CONAB, 2018).

The peanut production process is a critical time which it seeks mechanized operational excellence via new technologies to overcome the challenges and seize the opportunities of foreign markets in a sustainable way by increasing production volume, productivity and reducing production costs (Grotta et al., 2008) because in this operation losses are inevitable (Barbosa et al., 2014).

Most of the losses in the peanut harvest occurs in the digging operation and can reach high levels when the operation is not carefully managed; the values range from 3.1 to 47.1% of pod losses in relation to yield (Santos et al., 2013; Zerbato et al., 2014).

Monitoring the losses allows detecting and correcting the errors that might occur during the process so they can be minimized thus avoiding yield dropping (Bertonha et al., 2014; Câmara et al., 2007).

Multivariate analysis can be defined as exploratory statistical method that analyzes simultaneously multiple measurements in the experimental unit. The random variables must be interrelated so their effects cannot be meaningfully interpreted separately. The Clustering Analysis (CA) uses the Ward method (hierarchical) for extract statistical properties of a dataset, clustering similar vectors into classes (Hair, 2005).

The Principal Component Analysis (PCA) is performed to simplify the description of a set of interrelated variables. The created orthogonal axes are linear combinations of the original variables, starting with the eigenvalues of the covariance and/or correlation matrix of the studied variables in which the two largest eigenvalues generate the first two principal components, explain more variability than any other components (Hair et al., 2005).

The use of Multivariate exploratory techniques to control mechanized farming operations is a reality, as an innovative tool in the estimation of operating costs of
agricultural and forestry machineries (Guerrieri et al., 2017). The studies about the impact of mechanization through soil tillage on the behavior of weeds were efficient based on the similarity (CA) of the weeds (Boscardin et al., 2016; Nagahama et al., 2014).

Azevedo et al. (2015) and Silva & Lima (2012) investigated, respectively, the selection of lettuce cultivars and the nutritional status and productivity of coffee plants, and showed the efficiency of multivariate analysis, because the effects of multiple variables (random and interrelated) could not be interpreted separately.

Multivariate exploratory analysis is used in sectors of the agricultural production due to the high complexity of the produced information. Lamkowskii & Biagioni (2013) and Paredes Junior et al. (2015) reported that is used to better interpret, understand, manage and assist the decision-making process in the sugar-energy sector.

It is assumed that the peanut mechanized harvesting needs tools to assist in the effective control of loss variability and to understand the behavior of the agronomic traits in relation to soil textual classes for multiple traits.

This study aimed at identifying the clusters through multivariate exploratory approaches from similar soil textures and determining the variables that most influenced the clustering division to assist the decision-making process in mechanized agricultural operations.

MATERIAL AND METHODS

The experiment was conducted in six farms in Ribeirão Preto region, SP (Table 1). The 120-ha assessed area was planted with peanut of the Runner IAC 886 variety, sown in October 2015 with 0.90-m spaced rows. The crop was harvested in all evaluated areas in February 2016, 130 days after sowing. The farms were located between 20°58' and 21°10'S and 47°51' and 48°13'W, at 593 m average altitude. The soil in the areas presented between 7.0 and 66.8% clay (Table 1).

TABLE 1. Particle size analysis and soil textural classes.

<table>
<thead>
<tr>
<th>TREATMENTS</th>
<th>Clay</th>
<th>Silt</th>
<th>Sand</th>
<th>Textural Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>VCL</td>
<td>668</td>
<td>152</td>
<td>180</td>
<td>Very Clayey</td>
</tr>
<tr>
<td>CLA</td>
<td>496</td>
<td>143</td>
<td>361</td>
<td>Clayey</td>
</tr>
<tr>
<td>CLL</td>
<td>387</td>
<td>248</td>
<td>365</td>
<td>Clayey Loam</td>
</tr>
<tr>
<td>SCL</td>
<td>383</td>
<td>515</td>
<td>102</td>
<td>Silty Clayey Loam</td>
</tr>
<tr>
<td>SAL</td>
<td>186</td>
<td>258</td>
<td>556</td>
<td>Sandy Loam</td>
</tr>
<tr>
<td>SAN</td>
<td>70</td>
<td>40</td>
<td>890</td>
<td>Sandy</td>
</tr>
</tbody>
</table>

The digging was performed by a 680 HD Massey Ferguson tractor with maximum engine power of 127 kW (173 hp) at 2000 rpm coupled with an EA-2 Santal digger-shaker-inverter, 2 x 1 t (two harvested/digging rows, formed a windrow). However, the tractor worked at 1,500 rpm to meet the digger-shaker-inverter manufacturer’s recommendation of 340 rpm in the PTO. Although the conditions were unsuitable from the mechanical viewpoint, this rotation was used because they represented the real conditions of the equipment in the field, since it has no reduction mechanisms able to provide the indicated rotation.

The mechanized harvesting of peanuts was evaluated for six textural soil classes as follows very clayey (VCL), clayey (CLA), silty clayey loam (SCL), clayey loam (CLL), sandy loam (SAL), and sandy (SAN) with 10 repetitions per soil textural classes, totaling 60 plots in a randomized design, each plot formed a regular grid of 25 x 50 m.

The evaluated variables were yield, maturity, soil and pods moisture content, height and width of the windrow after digging, digging crop losses (visible and invisible losses), and gathering operation. The windrow formed after the passage of the digger was carefully removed to determine the visible and invisible digging losses. For this purpose, a metal frame of approximately 2 m² (1.11 x 1.80m) was placed across the windrow and the material was manually collected up to 0.25 m deep.

The frame width corresponded to the working width of the digger-shaker-inverter. After collecting, the pods were placed in paper bags, tagged, and sent to the laboratory where the samples were washed to remove the dirt from the exocarp. After that, the pods were weighed on a digital scale with 0.01 g precision and oven dried at 105 ± 3°C for 24 hours. After drying, they were weighed again to determine the losses, which were extrapolated to kg ha⁻¹ with further adjustment to 8% moisture. The losses were calculated in kg ha⁻¹ and expressed as percentage.

The previously described frame of approximately 2 m² was placed on the windrows at all sampling points to determine yield. All pods within the frame area were collected, and yield was calculated based on 8% moisture, the standard value for peanut storage in the processing companies. Subsequently, gross crop yield was calculated by adding total digging losses (sum of invisible and visible digging losses).

Pod moisture content (calculated on wet basis) was obtained following the oven method for samples collected after the digger-shaker-inverter passage, with subsequent correction for 8% water content value used for peanut storage in hulling (Martins & Lago, 2008). Soil moisture was determined is samples collected by a Dutch auger in the 0.0 to 0.2 m layer. The soil samples were placed in aluminum containers, sent to the laboratory, and oven dried at 105°C for 24 hours. Soil moisture was obtained on a dry basis, according to the methodology recommended by EMBRAPA (2006). The 0.0 to 0.2 m layer was defined as the soil-sampling layer to determine the moisture content because this region concentrates most of the peanut pods. One soil sample was collected per plot, totaling 60 samples.

Maturity (Hull scrape method) was determined by scraping the exocarp of 100 random pods for each sampling point, exposing the color of the mesocarp. The pods were classified by color, according to the Peanut Maturity Table, developed by the University of Georgia in the United States (Williams & Drexler, 1981).

The windrow width and height were measured using a graded ruler to indicate the quality of windrow reversal since they can affect the drying and gathering mechanized process.

The data were standardized (null mean and unit variance) prior to the conduction of the multivariate analysis and the variables did not present collinearity. Exploratory statistics of the data was performed using the Statistica software to analyze the hierarchical clusters, calculating the Euclidean distance between accesses by the Ward algorithm to obtain similar accesses, which was then graphically represented by a dendrogram (clustering the
The Discriminant Analysis (DA) is the oldest of the three classification methods. It was originally developed for multivariate normal distributed data. The data as a whole should not be normally distributed but within each class the data should be normally distributed. This means that if you could plot the data, each class would form an ellipsoid, but the means would differ. The Mahalanobis distance between x and the center \( c_i \) of class i is the \( S \)-weighted distance where \( S \) is the estimated variance-covariance matrix of the class.

After forming the clusters using the clustering method (Ward), the dendrogram branches were coded for technical application of principal components (Hair et al., 2005), using the same traits. The objective was to visualize the soil texture class in the two-dimensional plane formed by the principal components and interpret the discriminatory power of the variables in each major component, as:

\[
r'_{xj}(PC_h) = \frac{\alpha_{jh} \sqrt{\lambda_h}}{S_j}
\]

where,

- \( S_j \) : Standard deviation of variable \( j \);
- \( \alpha_{jh} \) : Coefficient of variable \( j \) in the \( h \)-th main principal component;
- \( \lambda_h \) : Eigenvalue \( h \);
- \( r'_{xj}(PC_h) \) : Correlation of the variable \( x_j \) to the \( h \)-th principal component.

The eigenvectors (PC1, PC2, ..., PC\( h \)) were determined from the eigenvalues of the covariance and/or correlation matrix of branch traits in descending order. Thus, PC1 is the component that explains more variability in the original dataset, while the last component explains less.

The variance in each principal component can be calculated as follows:

\[
PC_h = \frac{\lambda_h}{\text{trace}(C)} \times 100
\]

where,

- \( PC_h \) : Principal component \( h \);
- \( \lambda_h \) : Eigenvalue \( h \);
- \( C \) : covariance and/or correlation matrix;
- \( \text{trace}(C) = \lambda_1 + \lambda_2 + ... + \lambda_h \);

The principal component analysis was performed based on the diagonalization of a symmetric correlation matrix after analyzing the population variance to identify new numerical variables that explained most of the variability (Hair, 2005) by the Kaiser method with eigenvalues higher than “1”.

RESULTS AND DISCUSSION

The dendrogram that resulted from the clustering analysis is presented in Figure 1. The results corroborated with Lacerda et al. (2016) which in the evaluation on the discrimination of soil texture, the authors considered the similar soil textural groups to differentiation of the soil managements and production potential.

FIGURE 1. Dendrogram of hierarchical clustering analysis of peanut production parameters showing the clustering according to soil texture classes, very clayey (VCL), clayey (CLA), clayey loam (CLL), silty clayey loam (SCL), sandy loam (SAL), and sandy (SAN).
The processes were divided into 3 groups: I, II, and III and the clustering analysis by the discriminants analysis method (Figure 2) and the non-significant variables were excluded (IDL and VDL). The Classification and Discriminant Function Analysis Summary are presented for seven variables for three groups (Table 2 and 3). The Classification Matrix of groups (G1: 33.8%, G2: 16.7% and G3:50.0%) are represented for Observed classifications versus Predicted classifications (Table 4). Both characterized by the I and III opposing soil classes with low and high clay content, respectively, where higher yield was achieved but more advanced maturity stage caused major losses in peanut mechanized digging operation due to increased peduncle fragility. In addition, the mechanized gathering (GAL) became more difficult due to lower windrow height, and lower soil and pod moisture content, corroborating Santos et al. (2013) and Zerbato et al. (2014) that reported pod moisture contents of 35 to 45% at digging time. The GAL is a variable that most influenced the discriminant division to assist the decision-making process in mechanized agricultural operations.

TABLE 2. Classification and Discriminant Function Analysis Summary for seven variables for three groups

<table>
<thead>
<tr>
<th>Variables</th>
<th>G3</th>
<th>G1</th>
<th>G2</th>
<th>Wilks Lambda</th>
<th>Partial</th>
<th>F-remove</th>
<th>Toler.</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAL</td>
<td>-21.14</td>
<td>43.27</td>
<td>-23.11</td>
<td>0.009002</td>
<td>0.31</td>
<td>55.68</td>
<td>0.88</td>
</tr>
<tr>
<td>MAT</td>
<td>3.59</td>
<td>-1.19</td>
<td>-8.40</td>
<td>0.006780</td>
<td>0.42</td>
<td>35.64</td>
<td>0.73</td>
</tr>
<tr>
<td>SMC</td>
<td>7.02</td>
<td>-11.83</td>
<td>2.58</td>
<td>0.004489</td>
<td>0.63</td>
<td>14.98</td>
<td>0.32</td>
</tr>
<tr>
<td>YLD</td>
<td>-3.02</td>
<td>7.92</td>
<td>-6.78</td>
<td>0.003444</td>
<td>0.82</td>
<td>5.56</td>
<td>0.82</td>
</tr>
<tr>
<td>HEW</td>
<td>4.98</td>
<td>-6.69</td>
<td>-1.56</td>
<td>0.003956</td>
<td>0.71</td>
<td>10.17</td>
<td>0.41</td>
</tr>
<tr>
<td>WCP</td>
<td>0.94</td>
<td>-5.27</td>
<td>7.72</td>
<td>0.003439</td>
<td>0.82</td>
<td>5.51</td>
<td>0.85</td>
</tr>
<tr>
<td>WIV</td>
<td>-1.47</td>
<td>0.74</td>
<td>2.92</td>
<td>0.003408</td>
<td>0.83</td>
<td>5.23</td>
<td>0.68</td>
</tr>
<tr>
<td>Constant</td>
<td>-12.50</td>
<td>-46.70</td>
<td>-30.00</td>
<td>Wilks’ Lambda: 0.00283 approx. F (14,102) =129.73 p&lt;0.0000. Yield (YLD), maturity (MAT), windrow width (WIV) and height (HEW), soil (SMC) and pod (PMC) moisture content, visible and (VDL) and invisible (IDL) digging losses, and gathering losses (GAL).</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 3. Square Mahalanobis distances (upper half table) between the centroids of the group distributions (G1, G2 and G3) of the soil textural classes and respective F-values (lower half table).

<table>
<thead>
<tr>
<th>Variables</th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>251.98</td>
<td>201.62</td>
<td></td>
</tr>
<tr>
<td>G2</td>
<td>214.72</td>
<td>48.08</td>
<td></td>
</tr>
<tr>
<td>G3</td>
<td>309.26</td>
<td>46.09</td>
<td></td>
</tr>
</tbody>
</table>

Tables 4 shows the results for the three groups used in the model with agronomic variables to observed and predicted classification. The consideration concerns the percentage of well classified with the DA. In fact, all the soils of the estimation groups are well-classified (100%). Thus, there is not a situation of over-fitting: the analysis works well for the base model (Table 2 and 3), considering it is appropriate for predictions (Lucadamo & Leone, 2015) which using all the explicative and response variables, the DA works perfectly for the basic soil textural classes as the percentage of well classified, denoting better classification rates for the new model.

TABLE 4. Classification Matrix of groups (G1, G2 and G3) with Observed classifications (Rows) and Predicted classifications (Columns)

<table>
<thead>
<tr>
<th>Observed Classifications</th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th>Predicted classifications</th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th>Percent Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>33.8</td>
</tr>
<tr>
<td>G2</td>
<td>10</td>
<td>30</td>
<td>0</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>50.00</td>
</tr>
<tr>
<td>G3</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>16.7</td>
</tr>
</tbody>
</table>

FIGURE 2. Standardized coefficients for agronomic traits of peanut crop and soil texture for each group, according to the discriminants analysis. Yield (YLD), maturity (MAT), windrow width (WIV) and height (HEW), soil (SMC) and pod (PMC) moisture content, visible and (VDL) and invisible (IDL) digging losses, and gathering losses (GAL).

The principal component analysis allowed a single distribution of accesses (PC1 and PC2), since only two eigenvalues were greater than “one”, 4.51 and 1.79, respectively. The two largest principal components together enabled a two-dimensional ordering of accesses and variables, producing a Biplot graph (Figure 3). The distribution of soil textures and agronomic traits for the peanut crop showed that these components explained 70% of the variability, 50.11% and 19.89% of which in PC1 and PC2, respectively.

Engenharia Agrícola, Jaboticabal, v.38, n.2, p.244-250, mar./apr. 2018
The biplot graph (Figure 3) shows the distribution of soil classes in the peanut areas on the plan formed by the first two principal components (PC1 and PC2) and coded according to the clusters determined in the dendrogram. The x-axis (PC1) shows the contrasts of the six soil classes, three textural classes (VCL, CLA, and SCL) to the right and two (SAN and SAL) to the left, and the centralized intermediate behavior (SCL). The y-axis (PC2) shows a high direct correlation between the visible losses and windrow height, explained by the adjustment of alignment rolls and inverters of branches together with the plant mass being processed directly interfering with the windrow dimensions in the digging process (Zerbato et al., 2017).

Table 5 shows the variables with higher discriminatory power in the first principal component that had direct correlations between YLD, MAT, VDL, IDL, and GAL, enabling an efficient multivariate approach to mechanized harvesting of peanuts. Likewise, Silva et al. (2010) and Santos et al. (2016) successfully used the agronomic traits of great economic importance in the agricultural production system.

**TABLE 5. Correlation between peanut production variables and the principal components 1 and 2 for the different soil classes.**

<table>
<thead>
<tr>
<th>Variable</th>
<th>PC1</th>
<th>PC2</th>
</tr>
</thead>
<tbody>
<tr>
<td>YLD</td>
<td>-0.88</td>
<td>0.14</td>
</tr>
<tr>
<td>MAT</td>
<td>-0.64</td>
<td>-0.33</td>
</tr>
<tr>
<td>WIW</td>
<td>0.04</td>
<td>-0.33</td>
</tr>
<tr>
<td>HEW</td>
<td>0.47</td>
<td>-0.79</td>
</tr>
<tr>
<td>SMC</td>
<td>0.66</td>
<td>0.27</td>
</tr>
<tr>
<td>PMC</td>
<td>0.86</td>
<td>-0.36</td>
</tr>
<tr>
<td>VDL</td>
<td>-0.69</td>
<td>-0.61</td>
</tr>
<tr>
<td>IDL</td>
<td>-0.86</td>
<td>-0.42</td>
</tr>
<tr>
<td>GAL</td>
<td>-0.85</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Principal component 1 (PC1); principal component 2 (PC2); Yield (YLD); maturity (MAT); windrow width (WIW) and height (HEW); soil (SMC) and pod (PMC) moisture content; visible (VDL) and invisible (IDL) digging losses; and gathering losses (GAL).

The values of the correlations (Table 5) between the peanut harvest variables and the first two principal components according to soil textural class, PC1 has a high discriminatory power for the following peanut harvest variables: YLD (-0.88), MAT (-0.64), VDL (-0.69), IDL (-0.86), and GAL (-0.85). The negative associations of the variables presented their respective projections in the agglomeration directions of SAN and SAL soil textural classes (Figure 3). Then, the positive associations of SMC (0.66) and WPC (0.86) correlated with VCL, CLA and SCL soil textural classes. Thus, the peanuts digging operation could represent the PC1. It is noteworthy that the high impact of losses on the gathering of pods directly affected the final yield of the peanut crop (Ferezin et al., 2015).

Yield is the variable correlated to maturity that can directly influence peanut mechanized harvesting, in addition to the moisture content of the pods. The dry peduncles, with lower moisture content, breaking easily, cause higher losses in the gathering operation (Zerbato et al., 2014).

The PC2 showed a high discriminatory power for the HEW and VDL variables (Figure 3) with respective negative associations of -0.79 and -0.61 (Table 5). The windrow height influenced the gathering process since losses were higher for lower windrow heights because the gathering operation (PC2 representation) becomes more difficult closer to the ground.

The three multivariate statistical statistics showed, together, an efficient method of discriminating soil textural classes.
CONCLUSIONS

The hierarchical and non-hierarchical clustering analysis allowed ordering the accesses into three clusters, discriminant analysis and the principal component analysis indicating that high peanut yields in soils with low clay are correlated with losses in the mechanized harvesting operation.

Multivariate analysis procedures effectively summarized the losses and yield information and made easier to determine the GAL that most influenced the discriminant division to assist the decision-making process in mechanized agricultural operations.

ACKNOWLEDGMENTS

The authors are thankful to CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for financing this research.

REFERENCES


Engenharia Agrícola, Jaboticabal, v.38, n.2, p.244-250, mar./apr. 2018


