APPLICATIONS OF THE PRINCIPLE COMPONENT ANALYSIS (PCA) AT GRAPE VARIETIES FROM THE SORTOGROUP COARNĂ NEAGRĂ FOR ESTABLISHING PHENOTYPICAL VARIABILITY

APLICAREA ANALIZEI ÎN COMPONENȚI PRINCIPALI (ACP) LA SOIURILE DIN SORTOGRUPUL COARNĂ NEAGRĂ, ÎN VEDEREA STABILIRII VARIABILITĂȚII FENOTIPICE A ACESTORA

ISTRATE A.[1], ROTARU Liliana[1], COLIBABA Lucia Cintia[1]
e-mail: aistrate@uaiasi.ro

***Abstract**: Multivariate analysis allows the analysis of variables of different individuals, data measured together. Therefore, o complete description of the interdependence relations that exist between groups of individuals and the measured variable is obtained. The multivariate analysis method deals with reducing the number of innitial variables by substitution with others resulting from their combination. This can be represented in a graph by points in a two dimensional or three dimensional space, while not loosing an excessive amount of information. The PCA technique can be seen from more points of view. For classical statistics, PCA is the determination of main axes of an elipsoid, indicator of a normal multivariate distribution, these axes being estimated as random samples. PCA is a graphic representation of these, having as optimal character according to some algebraic and geometrical criteria that does not presume the emission of a initial hypothesis of statistical nature on the data that is to be analyzed. PCA allows the extraction of the maximum information, in a simple and coherent form, as a data ensemble, by underlining the interrelations between variables and individuals, either by similarity or opposition.*
***Key words**: multivariate analysis, PCA technique, phenotipical variability, Coarnă neagră sortogroup*

***Rezumat**: Analiza multivariatională permite analizarea variabilelor măsurate la indivizi diferiți, dar analizate împreună. Astfel se obține o descriere completă a relațiilor de interdependență care există între grupurile de indivizi şi variabilele măsurate. Metoda de analiză multivariațională constă în reducerea numărului de variabile inițiale substituindu-se cu altele rezultate din combinația lor, astfel încât acestea pot fi reprezentate ca "roiuri" de puncte într-un spațiu cu două sau trei dimensiuni, fără ca acest lucru să ducă la pierderea excesivă a informației. Tehnica ACP poate fi privită din mai multe puncte de vedere. Pentru statistica clasica ACP constituie determinarea axelor principale ale unui elipsoid, indicator al unei distributii normale multidimensionale, aceste axe fiind estimate plecând de la un eşantion de probe aleatoriu. ACP constituie o reprezentare a acestora, având un caracter optimal conform anumitor criterii algebrice şi geometrice, şi care nu presupun emiterea*

[1] University of Agricultural Sciences and Veterinary Medicine of Iaşi, Romania

nici unei ipoteze inițiale de natura statistică asupra datelor de prelucrat. ACP permite extragerea maximei informații, sub formă simplă şi coerentă dintr-un ansamblu de date, prin reliefarea interrelațiilor dintre variabile şi indivizi, prin asemănarea sau contrarietatea lor.

Cuvinte cheie: *analiză multivariațională, tehnica ACP, variabilitate fenotipică, sortogrupul Coarnă neagră*

INTRODUCTION

Grapevine genetic variability that is characteristic of plants occurs in response to the amendment of internal and external factors and is manifested both phenotypic and genotypic (Boursiquot and This, 1997).

Variability may be appropriate to the conditions that caused it, adequacy representing response to the action of the moment a factor, or adaptive, adaptation, written in the history of the species that is reflected in the ontology individuals, giving an orientation of individual variability, orientation which is the source of intraspecific diversification (Indreas *et al.*, 2003).

The purpose of the analysis in principal components is to present the information contained in ampelometrics matrices graphical form (circle of correlations, the plane defined by the main components) to be able indicate that two individuals or two variables are more like the more close each other on the charts. To do this, proceed by first calculating the correlation matrix of variables, based on the Pearson correlation coefficient (Rotaru and Țârdea, 1999).

Pearson correlation coefficient coefficient corresponds classic linear correlation. Its value ranges from -1 to 1, and it expresses the degree of linear correlation between two variables. The square of the correlation coefficient Pearson, gives an idea of how much variability character is influenced by another character. P values are calculated for each coefficient, permit null hypothesis testing - coefficients are not significantly different from 0 (Rotaru and Țârdea, 2002).

MATERIAL ŞI METHOD

The principle component analysis was performed on 7 native vine varieties: Coarnă neagră, Coarnă neagră selecționată, Azur, Mara, Ozana, Milcov and Gelu.

For these varieties were performed ampelometric measurements for a total of 30 variables analyzed in adult leaves: the length of the main ribs (N1, N2, N3, N4); the distance between the base of the lateral sinus and the petiole point (U,O); the opening of the lateral sinuses (SS,SI) and the petiole sinus (SP); the length (ALT) and the width (AN) of the limbus; the outer contour of the leaf (ENS, ENM, ENI, NL); the inner contour of the leaf (DS1, DS2, DS); the angle between the main ribs (A, B, C); the angles that define the shape of the median lobe (F, AP); the angle between the median rib and the extremity of the inferior lateral lobe (ABE); the ratio between the length of the ribs (21a, 31a, 41a); the ratio between the base of the lateral sinuses and the rib that support the sinuses (UN2, ON3); the ratio between the length and the width of the limbus (L-A), the ratio of the angles between the main ribs and the depth of the sinuses (AU, BO, ABUO) (Fig. 1).

For mathematical processing of the ampelometric data, it was used the XL-STAT software, with MICROSOFT license.

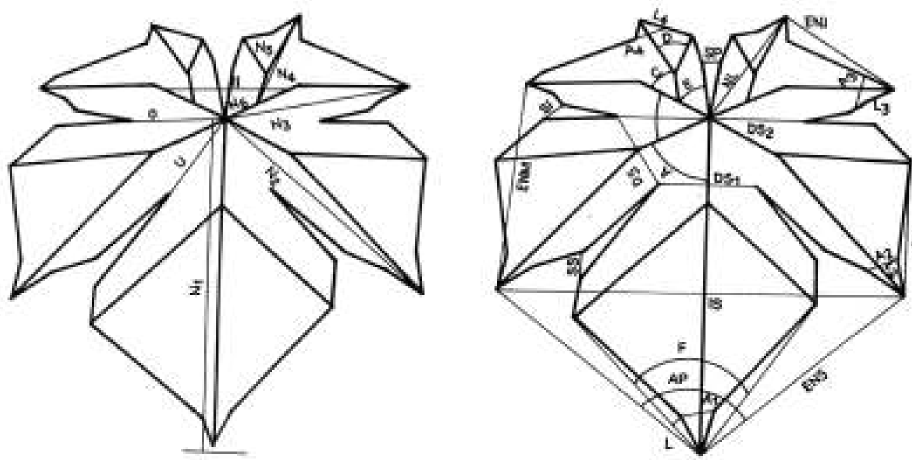

Fig. 1 - The schematic representation of characters examined on leaf vine

RESULTS AND DISCUSSIONS

A first stage in the principle component analysis is represented by the calculation of the correlation matrix between variables, based on the Pearson coefficient (simple correlation coefficient) between two variables x and y.

It is calculated as the arithmetic average of the normalized deviations product of the two variables and which reflects the intensity of the linear connections between the variables. The values can be between +1 and -1 and the closer the coefficient value is to these limits, the stronger the correlation between the two variables will be.

The next stage represents the determination of the variables and the eigenvectors of them in the created space by the first two main components. (Fig. 2).

The inertia percentage of the first two main components, in the case of the analyzed varieties, was 84,31%, of which 63.83% the first principal component (Axis 1) and 20.48% the second principal component (Axis 2).

Thus it gives away the multidimensional space of the 30 variables analyzed initially, at the bidimensional one, created by the first two principal components, while maintaining an 84,31% of the total inertia (variation) of the individuals.

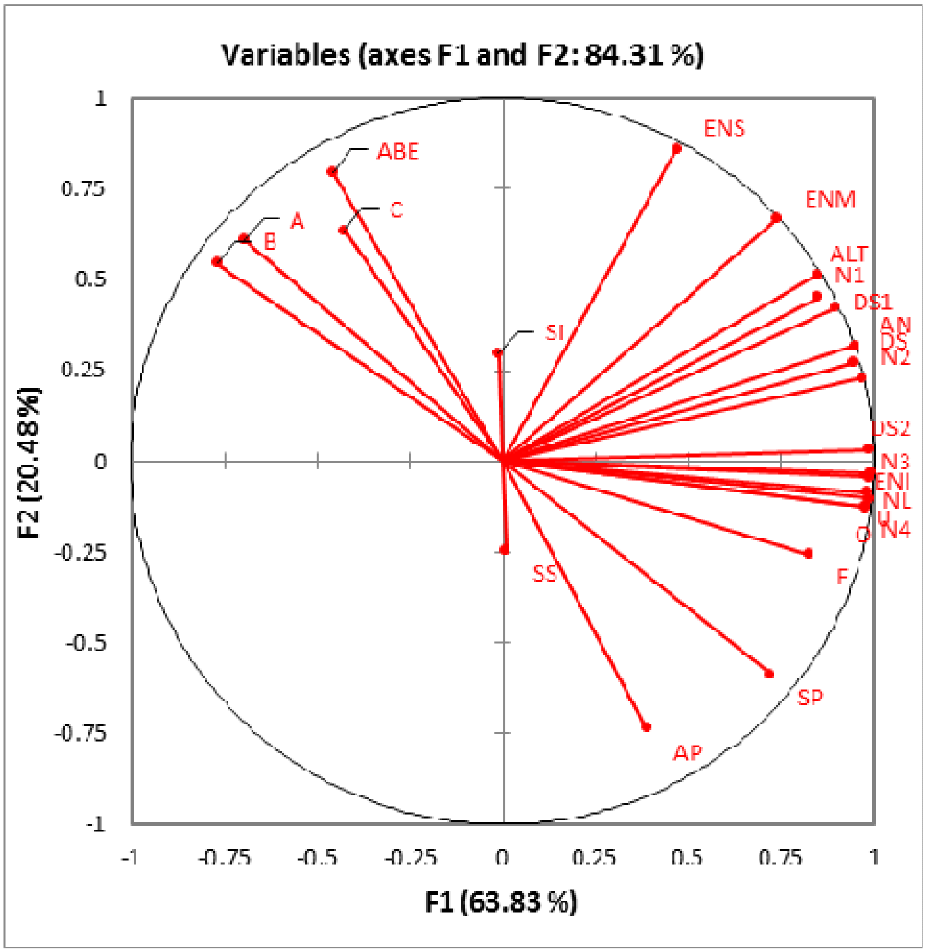

Fig. 2 –The correlation circle determined by the first two main components

The analysis of the correlation circle reveals that the factor 1 (the main component 1) is determined by the N3 (0,9850), ENI (0,9756), U (0,9743) and DS2 (0,9719) variables. At the opposite side there are the Si and SS variables with 0,0001 each, AP (0,1499) and C (0,1829). All these variables are situated to the extremity of the correlation circle on the axis of the factor 1

The second main component is determined by the ENS (0,7392), ABE (0,6364) and AP (0,5421) variables. The smallest values were found at the N3 (0,0009), DS2 (0,0011), ENI (0,0019) and NL (0,0073) variables.

The 90^0 angles between the variables shows that they are not correlated with each other, each of them having an important contribution to the architecture of the leaf, unable to be replaced with correlations of these values.

The analysis of the individuals (varieties) coordinates on the main axes are interested in those individuals who had the highest contribution in defining the main components and which, by default, are placed eccentrically on the direction of the main axes (fig. 3). It also states the densest areas that can be considered future groups of individuals, but for their determination is necessary to make a cluster analysis.

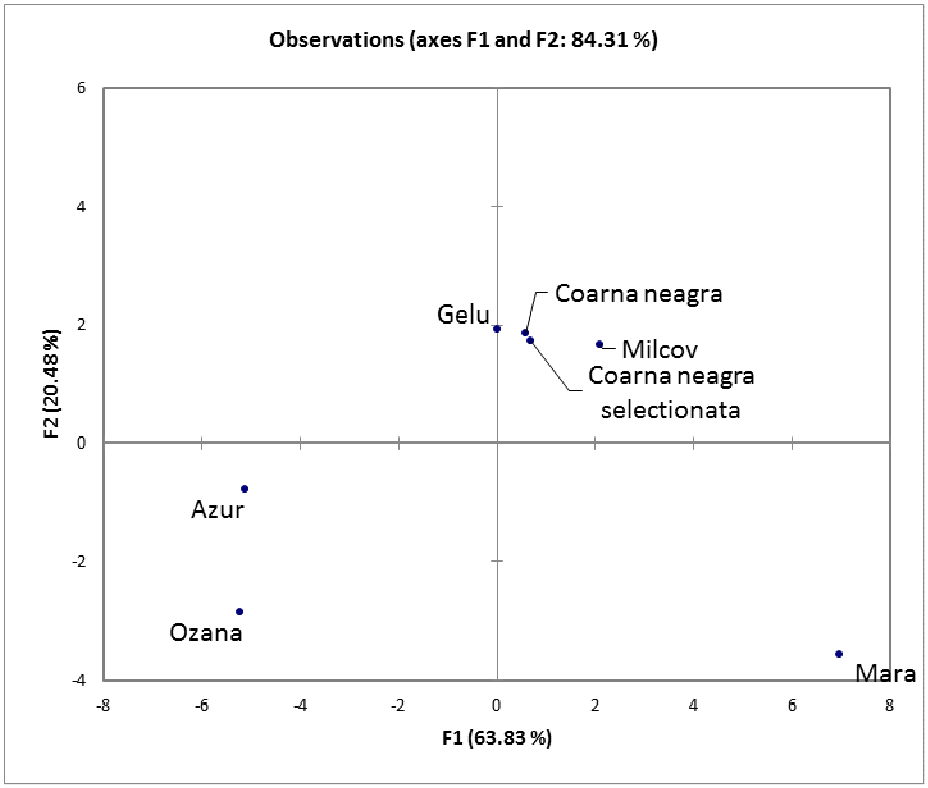

Fig. 3 - The varieties coordonates on the plan that it had been by the first two main components

In defining the main component 1, the biggest contribution had the Mara (6,9550) and Milcov (2,1038) varieties, varieties with medium leaves, trilobite, and in negative way, big contributions were from Ozana (-5,2330) and Azur (-5,1216) varieties, varieties with big leaves, pent lobate.

The main component 2 was determined mainly by the Gelu (1,9330) variety, variety with big leaf, slightly elongate, and in negative way by the Mara (-3,5619) variety.

It should be pointed the position of the Coarnă neagră and Coarnă neagră selecționată varieties which had almost the same contribution in defining the main components which indicates the high similarity of the leaf architecture of those varieties

CONCLUSIONS

1. The principle component analysis (PCA) represents a multidimensional statistical and mathematical method which can be applied in ampelography to establish in a first stage the characters that determines the differentiation of a set of varieties with similar phenotypic characteristics.

2. The inertia percentage of the first two main components was 84,31% from which 63,83% the first one and 20,48% the second one, which indicates the fact that bidimensional representation of the plan determined by them highlights the existing linear connections between the 30 analyzed variables at the 7 local vine varieties.

3. The first main component was determined by the N3, ENI, U and DS2 variables with a maximum value of 0,9850 and the main component 2 was given by the ENS, ABE and AP variables with a maximum value of 0,7392.

4. The factor 1 was characterized by Mara and Milcov varieties and the factor 2 was represented mostly by Gelu and Coarnă neagră varieties.

5. From the analysis of the factors 1 and 2 it can be observed a very big phenotypic resemblance of the Coarnă neagră and Coarnă neagră selecționată varieties fact confirmed by the cluster analysis previously established.

REFERENCES

1. **Boursiquot J.M., This P., 1997** – *Les nouvelles techniques utilisées en ampelographie: informatique et marquage*. Journal International des Sciences de la Vigne et du Vin, vol 40(1): 13-23.
2. **Indreaş Adriana, Stroe Marilena, Rotaru Liliana, Mărcuță Alina, 2003** - *The establishment of the origin for some autochtonous vine varietes by multi-variable statistical methods*. Lucrări Ştiințifice U.S.A.M.V.Bucureşti, seria B, vol. XLVI.
3. **Rotaru Liliana, Țârdea C., 1999** – *Contribuții la studiul ampelometric al soiurilor de viță de vie, prin prelucrarea datelor pe calculator-programul Microsoft EXCEL-97*. Lucr. Şt. ale UAMV Iaşi, seria Hortic. vol. 1 (42).
4. **Rotaru Liliana, Țârdea C., 2002** – *Contribuții la prelucrarea datelor ampelometrice prin analiza în componenți principali.* Analele Universității din Craiova, vol. VII (XLIII).