Vitis 38 (2), 73–77 (1999)

Elliptic Fourier Analysis (EFA) and Artificial Neural Networks (ANNs) for the identification of grapevine (*Vitis vinifera* L.) genotypes

by

S. Mancuso

Dipartimento di Ortoflorofrutticoltura, Università di Firenze, Italia

Summary: The potential application of the Elliptic Fourier Analysis (EFA) for the objective quantitative description of leaf morphology, combined with the use of a Back-propagation Neural Network (BPNN) for data modelling, was evaluated to characterize and identifiy 12 Sangiovese-related accessions (*Vitis vinifera* L.). The results enable us to distinguish, with considerable certainty, between 10 accessions. Cluster analysis revealed the existence of a uniform group for the Prugnolo (acerbo, medio and dolce) ecotypes showing a high degree of relatedness. Among all accessions only the so-called Casentino ecotype significantly diverged from all the others, indicating probably a different origin. The application of EFA coupled with the use of artificial neural networks opens interesting prospects for the characterization of varieties, allowing to study differences and/or relationships which can not be detected by standard ampelographic systems.

Key words: ampelography, backpropagation, cultivar identification, EFA, Vitis vinifera.

Introduction

In the last years ampelographic data have been used to resolve the complex problem of the definition and identification of grapevine varieties. Quantitative determination of morphological elements of the leaf (*i.e.* angles, area, teeth number, petiole length) have been extensively utilised in ampelographic research (OIV-IBPGR-UPOV charts 1983; Galet 1985, Costacurta *et al.* 1996, Silvestroni *et al.* 1996). However the origin of the varieties, their heterogeneity and the frequently observed homonymy and synonymy, often resulted in doubtful classification.

With the advent of inexpensive personal computers, alternative methods for the objective quantitative description of morphological characters have become readily available. In particular, outline-shape analysis seems well suited to study grapevine leaves.

The methods of shape analysis fall into three categories: classical shape analysis, the fitting of polynomial curves, and a group of methods based on Fourier decomposition (see literature in Rohlf and Bookstein 1990). An extensive discussion of the advantages has been published elsewhere (Foster and Kaesler 1988, and references therein). Elliptic Fourier analysis (EFA), however, has been favored in several recent studies in different fields of science from botany to palaeontology (*e.g.*, Rohlf and Archie 1984, Ferson *et al.* 1985, White *et al.* 1988, Temple 1992) and also in grapevine variety classification (Diaz *et al.* 1991). Ferson *et al.* (1985) provide a brief introduction to the theory of EFA, and a detailed account of the method and associated problems. Unlike many other methods, EFA can describe complex shapes, does not require mathematically determined centroids, does not require points on the outline to be equally spaced, and can include simple normalizations for size, position, orientation, and starting position of the trace. A further property of Fourier methods is its ability to invert the transformation and to reconstruct an outline from a set of Fourier coefficients. Hence, for example, an "average" shape can be reconstructed from the mean coefficients of a largc number of outlines (Ferson *et al.* 1985).

Further interesting approaches in the field of varietal identification have recently been suggested, *e.g.* the application of Artificial Neural Networks (ANN), to clarify grapevine (Mancuso *et al.* 1998) and olive (Mancuso and Nicese 1999) genotypes, on the basis of phyllometric parameters. The structure of ANNs makes them particularly useful to recognize different shapes (Hertz *et al.* 1991) or pattern in complex, nonlinear data, *e.g.* those derived from experimental areas of horticulture. Therefore, it seemed to be interesting to verify whether EFA can be used to describe grapevine leaves and Elliptic Fourier coefficients can be used as input in a back-propagation neural network for the identification of grapevine genotypes.

Material and Methods

Plant material and image acquisition: The study was carried out with 11 putative Sangiovese-related ecotypes and the registered clone Sangiovese R 10 as a reference (Tab. 1); the 12 ecotypes had recently been characterized by DNA marker technology (Sensi *et al.* 1996). Samples were collected from the grapevine germplasm collection of the Department of Horticulture of the University

Correspondence to: Dr. S. Mancuso, Dipartimento di Ortoflorofrutticoltura, Università di Firenze, Via Donizetti 6, I-50144 Firenze, Italy. Fax: +39-55-331497. E-mail: electro@cesit1.unifi.it

Table 1

Grapevine genotypes included in this study

#	Genotype	#	Genotype
1	Prugnolo gentile	7	Casentino
2	Brunellone	8	Chiantino
3	Brunelletto	9	Morellino
4	Prugnolo acerbo	10	Morellino di Scansano
5	Prugnolo dolce	11	Piccolo precoce
6	Prugnolo medio	12	Sangiovese R 10

of Florence. At the time of veraison, from 15 plants per accession, 65 fully expanded, healthy leaves, positioned between the 7th and the 11th node (ALLEWELDT and DETTWEILER 1986) were selected according to uniformity of appearance, growth habit and exposure.

Leaf images were acquired at 360 x 360 d.p.i., 256 gray scale, by using an optical scanner. The contour for each leaf (xy-coordinates of 1500 points equally spaced) was then obtained by image analysis.

Elliptic Fourier Analysis: The software to perform EFA is available in ROHLF and BOOKSTEIN (1990) and is written in Fortran for IBM-compatible personal computers. Required input for each outline is a string of xy-coordinates preceded by a sample number and the number of outline coordinates. EFA describes outlines in terms of harmonically related ellipses (Fig. 1), and each ellipse is, in turn, described by 4 coefficients. Because of the basically non-elliptical shape of grapevine leaves, a relatively large amount of harmonics is required to describe their outlines. The number of harmonics required to accurately describe an outline can be estimated in two ways. One can calculate the average discrepancy between the original outline and the inverse Fourier reconstruction based on n harmonics. The Fourier series is truncated at the value of n corresponding to a negligible discrepancy (*e.g.*, smaller than the resolution of digitization). Alternatively, one can sum the variance for successive harmonics and compare this sum to the total variance of the Fourier series based upon the maximum possible number of harmonics (equal to half the number of points on the digitized outline). The variance, or power, of each harmonic is equal to half the sum of the squares of the Fourier coefficients. The Fourier series is truncated at the value of n at which, say 99 % of the variance is retained. In shape analysis, the effects of specimen size (which profoundly influences harmonic amplitudes) can be removed during computation of EF coefficients. This normalization utilizes parameters of the first harmonic (*i.e.*, best fitting) ellipse and is probably appropriate in most studies. Information about relative size, however, can be reincorporated into a study during statistical analysis and is essential for an understanding of shape changes through ontogeny.

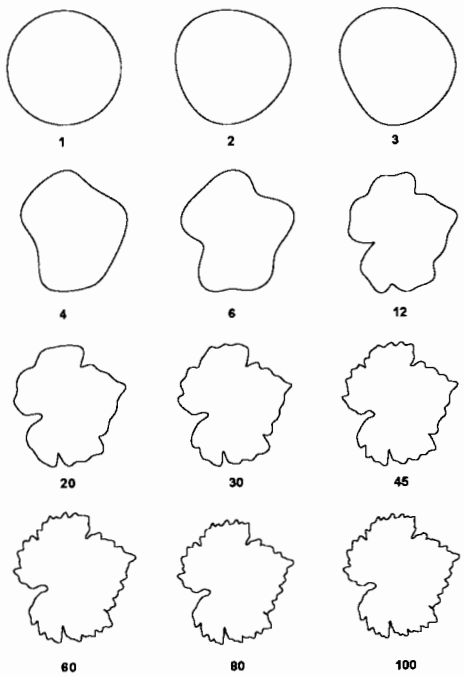

Fig. 1: Shape synthesis of a grapevine leaf by 100 harmonics obtained with Elliptic Fourier Analysis.

In the present study EFA was performed to calculate the first 100 harmonics and a total of 400 coefficients (4 per harmonic) for each leaf. By considering that this number of variables is hardly to manage, the contribution of the 400 EF coefficients was redistributed in 13 logarithmically spaced intervals (DIAZ *et al.* 1991) including the following harmonics: 1, 2, 3, 4, 5-6, 7-8, 9-12, 13-17, 18-24, 25-34, 35-49, 50-69, 70-100. The 52 resulting (4 coefficients x 13 intervals) elliptic Fourier coefficients for each outline was then treated as inputs in a back-propagation neural network.

Neural Network: A back-propagation neural network program was written and implemented in a P200 computer, following the methods previously described in MANCUSO *et al.* (1998). In brief, the network was designed using a total of 52 inputs represented by the Elliptic Fourier coefficients and 12 outputs represented by the accessions in Tab. 1. Output values were 1 or 0 (true or false). In order to optimise the neural network activity, the number of "hidden neurons" was modified. Minimum error was reached with 60 hidden neurons positioned on two levels (30 x 30). The activation function of the neurons was a sigmoidal function, $1/(1+e^{-x})$. Back-propagation of error was performed using formulas previously described by MANCUSO and NICESE (1999). Details in back-propagating errors can be found in RUMELHART *et al.* (1986).

In total, data from 720 leaves (50 per grapevine accession) were used. The learning phase in all the BPNNs tested was protracted until the RMS (root mean square) error was <0.04 and the difference between the RMS in two consecutive epochs was <0.0001. The ANNs were tested with sets of EF coefficients in inputs for which the output was known, so that the predicted and actual outputs could be compared. These data had not been used previously to train the network.

Data analysis: Neural network outputs were used to measure the dissimilarities or distances between ecotypes when forming the clusters. Euclidean distance was calculated and a dendrogram was constructed based on the distance matrix data by applying the unweighted pair group method with arithmetic averages (UPGMA) cluster analysis using the computer program Statistica version 4.0 (Statsoft Inc.).

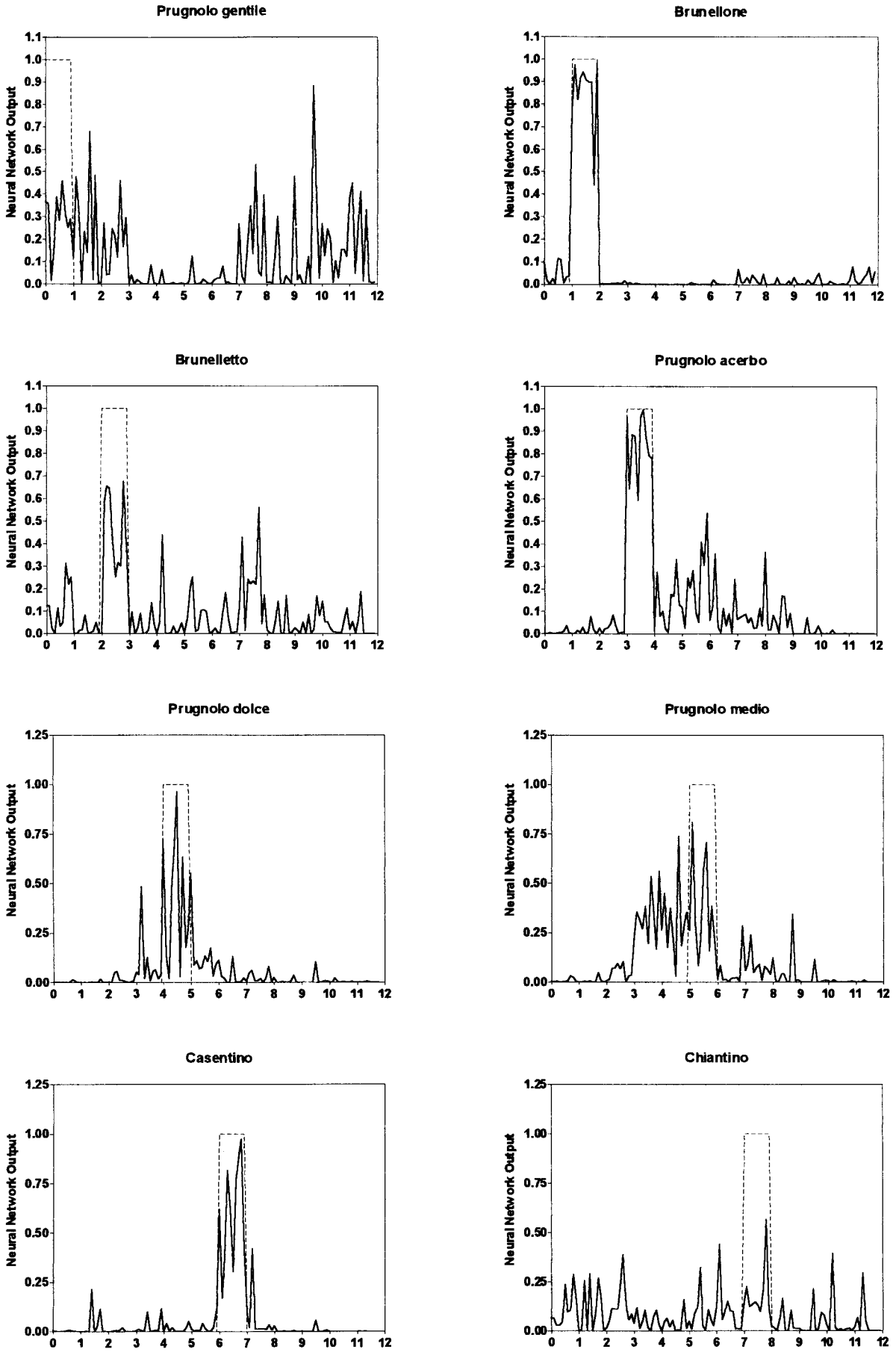

Fig. 2: Outputs of the neural network recognition phase. Each frame shows the BPNN output for the input represented by the phyllometric parameters of 15 leaves. The name of the unknown accession is given by the code number (see Tab. 1) at the abscissa, which presents the highest (closest to 1) output value.

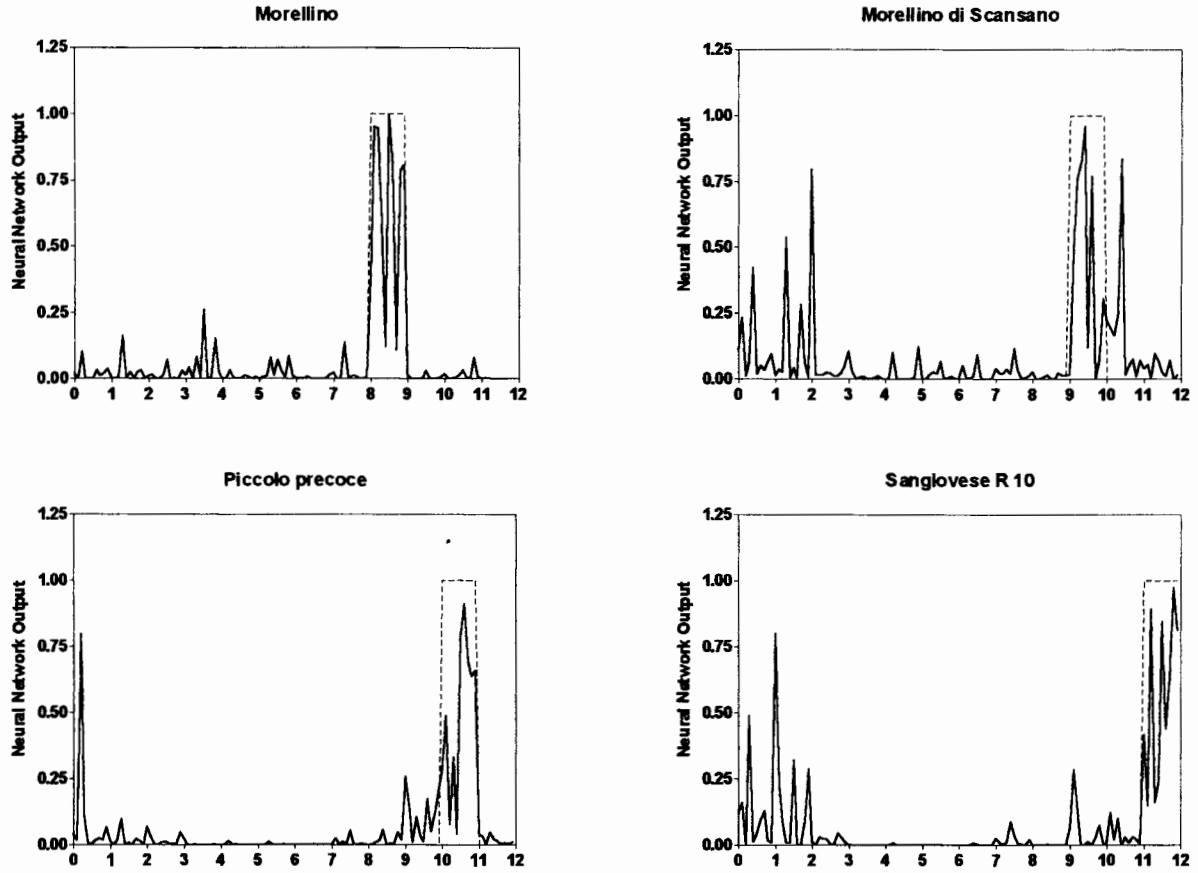

Fig. 2: continued

Results and Discussion

Neural network analysis of the 12 *Vitis vinifera* accessions revealed a successful identification of the different genotypes, except in the case of Prugnolo gentile (identified by #1 in Tab. 1) and Chiantino (#8) that showed confused output diagrams (Fig. 2).

Outputs obtained through the back-propagation neural network were analysed using the distance matrix in Tab. 2. The euclidean distances ranged from 1.74 for Chiantino (#8) and Brunelletto (#3) to 10.09 for Casentino (#7) and Prugnolo acerbo (#4). UPMGA cluster analysis of the distance matrix (Fig. 3) separated two different groups among the grapevine genotypes. The first group comprises three of the 4 Prugnolo accessions: Prugnolo dolce (#5), Prugnolo medio (#6) and, more distanced, Prugnolo acerbo (#4). The second group consists of Brunelletto (#3), Chiantino (#8) and Prugnolo gentile (#1).

The results showed a high degree of relatedness for Prugnolo acerbo, Prugnolo medio and Prugnolo dolce, in agreement with the results of studies made with molecular marker methods (SENSI *et al.* 1996). The data support the hypothesis that Prugnolo acerbo, Prugnolo medio and Prugnolo dolce could be mutations originated from the same seedling, whereas the origin of Prugnolo gentile is more distant. The same discussion is valid for Brunelletto, Chiantino and Prugnolo gentile that showed a similar pattern of relationships.

Among all the genotypes tested, Casentino appears to be most distant. As a consequence of mutations, a certain degree of variation could be expected within ancient varieties such as Sangiovese, but the high degree of divergence strongly suggests that Casentino does not share the same origin as the other Sangiovese-related ecotypes. Thus, the exclusion of this accession from the Sangiovese group can be assumed.

On the whole, the results are interesting because the use of EFA and ANNs allowed distinction of the vine accessions in agreement with the results of SENSI *et al.* (1996) obtained with the same genetic materials by PCR-based marker technologies (amplified fragment length polymorphism, AFLP; inverse sequence-tagged repeat analysis, ISTR). Moreover, it seems that the coupling of EFA and ANN demonstrates a greater possibility of distinction compared to the AFLP method, in fact, the accessions Morellino di Scansano, Prugnolo gentile and the registred clone Sangiovese R 10 were not distinct by the AFLP methods (SENSI *et al.* 1996), but clearly differentiated in the present study.

In summary, biometric methods are important tools in the study of grapevine identification and ampelography. In particular, the outline shape is a fundamental aspect of morphology which is suitable to biometric description, using either linear and angular measurements, or Fourier shape analysis. By using such methods, it is possible both to quantify and to objectively compare intra- and interpopulation variations in morphology. This information is crucial to the interpretation of apparently complex patterns of morphological changes. Besides, the results obtained in the present study show first that the Neural Network can be utilized to

Table 2

Euclidean distance matrix. Cultivar numbers correspond to those in Tab. 1

	1	2	3	4	5	6	7	8	9	10	11	12
1	—											
2	2.92	—										
3	2.29	3.30	—									
4	3.61	4.00	3.26	—								
5	2.94	3.31	2.53	3.04	—							
6	3.12	3.51	2.58	2.34	1.80	—						
7	10.02	10.08	9.95	10.09	9.92	9.97	—					
8	2.11	2.77	1.74	2.95	2.15	2.29	9.82	—				
9	3.21	3.54	2.95	3.43	2.90	3.07	10.05	2.65	—			
10	2.99	3.32	2.83	3.67	2.90	3.12	10.01	2.54	3.22	—		
11	2.74	3.36	2.66	3.55	2.72	2.97	9.97	2.40	2.99	2.72	—	
12	2.84	3.07	2.92	3.73	2.95	3.19	8.72	2.64	3.27	3.07	2.94	—

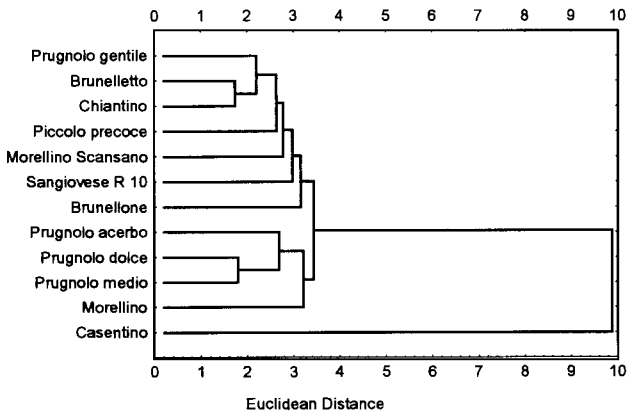

Fig. 3: Dendrogram of 12 grapevine accessions generated by UPMGA cluster analysis of the distance value shown in Tab. 2.

process EF coefficients dectecting enough differences to differentiate among grapevine genotypes and second, that EFA and ANN techniques are relatively simple methods which can be useful in current grapevine breeding programs allowing to study varietal differences and/or relationships which can not be detected using standard ampelographic systems.

References

ALLEWELDT, G.; DETTWEILLER, E.; 1986: Ampelographic studies to characterize grapevine varietes. 4° Simp. Intern. Genetica della Vite, Verona. Vignevini **13**, 56-59

COSTACURTA, A.; CALÒ, A.; CARRARO, R.; GIUST, M.; ANTONIAZZI, M.; LAZZARO, G.; 1996: Metodologie computerizzate per la caratterizzazione di vitigni (I° contributo). Riv. Viticolt. Enol. **49** (1), 27-34.

DIAZ, G.; SETZU, M.; DIANA, A.; LOI, C.; DE MARTIS, B.; PALA, M.; BOSELLI, M.; 1991: Analyse de Fourier de la forme de la feuille de vigne. Première application ampelométrique sur un échatillon de 34 cépages implantés en Sardaigne. J. Intern. Sci. Vigne Vin **25**, 37-49.

FERSON, S.; ROHLF, F. J.; KOEHN, R. K.; 1985: Measuring shape variation of two-dimensional outlines. System. Zool. **34**, 59-68.

FOSTER, D. W.; KAESLER, R. L.; 1988: Shape analysis; ideas from the Ostracoda: In: M. L. MCKINNEY (Ed.): Heterochrony in Evolution: A Multidisciplinary Approach, 53-69. Plenum Press, New York.

GALET, P.; 1985: Précis d'Ampelographie Pratique. Imprimerie Charles Déhan, Montpellier.

HERTZ, J.; KROGH, A.; PALMER, R.; 1991: Introduction to the Theory of Neural Computation. Addison-Wesley, Redwood City, California.

MANCUSO, S.; NICESE, F. P.; 1999: Identification of olive (*Olea europaea* L.) varieties using artificial neural networks. J. Amer. Soc. Hort. Sci. (in press).

- -; PISANI, P. L.; BANDINELLI, R.; RINALDELLI, E.; 1998: Application of an Artificial Neural Network (ANN) for the identification of *Vitis vinifera* L. genotypes. Vitis **37**, 27-32.

OIV-IBPGR-UPOV; 1983: Code de caractères descriptifs des variétés et espèces de *Vitis*. Office International de la Vigne e du Vin, Paris.

ROHLF, F. J.; ARCHIE, J. W.; 1984: A comparison of Fourier methods for the description of wing shape in mosquitoes (*Diptera: Culicidae*). System. Zool. **3**, 302-317.

- -; BOOKSTEIN, F. L. (Eds.); 1990: Proceedings of the Michigan Morphometrics Workshop. Ann Arbor, Michigan, University of Michigan, Museum of Zoology, Special Publication 2.

RUMELHART, D. E; HINTON, G. E.; WILLIAMS, R. J; 1986: Learning representations by back-propagating errors. Nature **323**, 533-536.

SENSI, E.; VIGNANI, R.; ROHDE, W.; BIRICOLTI, S.; 1996: Characterization of genetic biodiversity with *Vitis vinifera* L. Sangiovese and Colorino genotypes by AFLP and ISTR DNA marker technology. Vitis 35, 183-188.

SILVESTRONI, O.; INTRIERI, C.; DI DOMIZIO, N.; 1996: Impiego di metodi fillometrici per la caratterizzazione di alcuni vitigni dell'Emilia Romagna. Riv. Viticolt. Enol. **49** (1), 17-26.

TEMPLE, J. T.; 1992: The progress of quantitative methods in palaeontology. Palaeontology **35**, 475-484.

WHITE, R. J.; PRENTICE, H. C.; VERWIJST, J.; 1988: Automated image acquisition and morphometric description. Canad. J. Bot. **66**, 450-459.

Received December 9, 1998