Vitis **40** (2), 59–63 (2001)

Clustering of grapevine (*Vitis vinifera* L.) genotypes with Kohonen neural networks

S. MANCUSO

Dipartimento di Ortoflorofrutticoltura, Università di Firenze, Italia

Summary

Self-organizing maps generated by Kohonen neural networks provide a method to transform multidimensional problems common in ampelography into lower dimensional problems. In this study the clustering efficiency of Kohonen neural networks was evaluated to characterize and identify 10 Sangiovese-related and 10 "coloured" (fruit gives intense red colour to the wine) grapevine accessions, on the basis of the elliptic Fourier coefficients of the leaves. The non-supervised learning algorithm used allowed *a priori* classification of the accessions. The results enabled us to distinguish between 16 accessions and to denote two pairs of synonyms. To obtain quantitative information regarding relationships among these accessions, Kohonen neural networks were trained with different numbers of neurons in the Kohonen output layer permitting the graphical representation of the similarity by construction of a dendrogram. In agreement with previous studies based on molecular markers and neural network technology, a high similarity was found for the ecotypes (1) Prugnolo acerbo, Prugnolo dolce and Prugnolo medio and (2) Brunelletto and Prugnolo gentile. Among the Sangiovese-related accessions the so-called Casentino ecotype diverged from all the others, probably indicating a different origin. Producing easily comprehensible low-dimensional maps, the Kohonen neural networks approach proposed here allows to study complex ampelographic data elucidating relationships that can not be detected by traditional data analysis tools.

K e y w o r d s : ampelography, Kohonen network, cultivar identification, neural network, *Vitis vinifera.*

Introduction

The number of methods to classify and identify grapevine varieties has increased rapidly in the last two decades. Ranging from classical OIV-IBPGR-UPOV charts (*e.g.* ANONYMOUS 1983) to isoenzymatic markers (SUBDEN *et al*. 1987; BENIN *et al*. 1988) or molecular characterization by DNA analysis (BOWERS *et al*. 1993; THOMAS *et al*. 1994; XU *et al*. 1995), numerous methods to distinct between the different grapevine genotypes have been proposed.

Recently the use of Backpropagation Neural Networks (BPNN) on the basis of phyllometric parameters has been proposed for grapevine, olive and chestnut genotypes (MANCUSO *et al*. 1998, 1999; MANCUSO 1999 a; MANCUSO and NICESE 1999). The internal representation used by these networks are non-linear and are built up by a learning process based on examples. Moreover, the learning method used is a supervised learning process, and relies upon an existing structural classification. In other words, the BPNN learn to classify examples in *a priori* defined structural classes; in the case of phyllometric parameters these classes are most often represented by varieties. However, the definition of structural classes may disregard possible relationships or similarities between the accessions which could be important in ampelography.

Conventionally, at least in viticulture, reduction of multivariate data is normally carried out using principal components analysis or hierarchical clustering analysis (EVERITT 1993). Nevertheless, it is difficult to handle the high interdependence of phyllometric variables by statistical methods. Methods based on decision trees are also not suitable for ampelographic interpretation, since they lead to a sharp division of the populations analysed. Kohonen neural networks seem not to suffer from the problem encountered by the other methods and is used in many domains for data classification (KOHONEN 1984; JONGMAN *et al.* 1995).

There are many different types of Artificial Neural Networks (ANN) and a common feature is that once structured for a particular application they must be trained. There are two approaches to training, supervised and unsupervised. The most often used ANN is a fully connected supervised network with a backpropagation learning rule, which works excellently for prediction and/or classification tasks (see for example MANCUSO *et al*. 1998). Another extensively used ANN is the Kohonen network (or Self Organising Map) that uses an unsupervised learning process: it requires no *a priori* information on classes and therefore classifies examples only by intrinsic characteristics.

After a short introduction to Kohonen networks and its learning process, the present work shows results obtained by using the elliptic Fourier coefficients (MANCUSO 1999 a) of leaves as input in a Kohonen neural network for the clustering of grapevine genotypes.

Material and Methods

P l a n t m a t e r i a l a n d i m a g e a c q u i s i t i o n : The study was carried out with 9 putative Sangiovese-related ecotypes, the registered clone Sangiovese R 10 and 10 accessions of "coloured" grapevines (fruit gives intense red colour to the wine). The 20 ecotypes (Tab. 1) which were

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

Table 1

Grapevine accessions included in this study

Coloured accessions		Sangiovese-type accessions	
1	Abrostine	11	Prugnolo gentile
2	Abrusco	12	Brunelletto
3	Colorino americano	13	Prugnolo acerbo
4	Colorino di Lucca	14	Prugnolo dolce
5	Colorino di Pisa	15	Prugnolo medio
6	Grand noir	16	Casentino
7	Granoir	17	Chiantino
8	Morone	18	Morellino
9	Nereto	19	Morellino di Scansano
10	Tinturié	20	Sangiovese R10

utilised in previous studies (MANCUSO *et al*. 1998; MANCUSO 1999 a, b) and characterised by DNA marker technology (SENSI *et al*. 1996), were selected because they offered the possibility to verify the Kohonen neural network technique.

Samples were collected from the grapevine germplasm collection of the Department of Horticulture of the University of Florence. At veraison, from 15 plants per accession 65 fully expanded, healthy looking leaves, located between the 7th and 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 (EFA): 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 using the method previously described in MANCUSO (1999 a). 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 Kohonen neural network.

Kohonen neural network: This section will recall only the basic principles of Kohonen networks in order to give a short account of Kohonen's SOM (Self-Organising Maps). For a complete definition and discussion, see KOHONEN (1984) and SMITH (1994).

In the brain of mammals, *i.e.* in areas of the neurocortex, neurons are organised in a way that reflects some physical characteristics of the signals that stimulate them. Of all ANN architectures and learning schemes, the Kohonen ANN resembles the biological NN probably most closely (KOHONEN 1984). Self-organizing feature maps or Kohonen networks are designed to map or project input signal vectors of arbitrary dimension onto a structured set of processing units, "neurons". These units interact in such a way that the final trained network produces an output pattern which exhibits topological relationships of the set of input vectors. In simpler terms, input signals originating from similar cases (which can be thought of as neighbouring points in the multidimensional space spanned by the components of the input vectors) are projected onto neighbouring neurons.

Fig. 1 depicts schematically the principle of such a network. There are input neurons (organized in a linear array sometimes called "retina") which establish the interface between the network and the world outside. They serve to enter the data vectors into the model. For a specific problem one uses as many input neurons as are needed; in the present study, processing 52 elliptic Fourier coefficients on each leaf, 52 such neurons were used. The number of leaves processed in the training phase was 60 for each genotype.

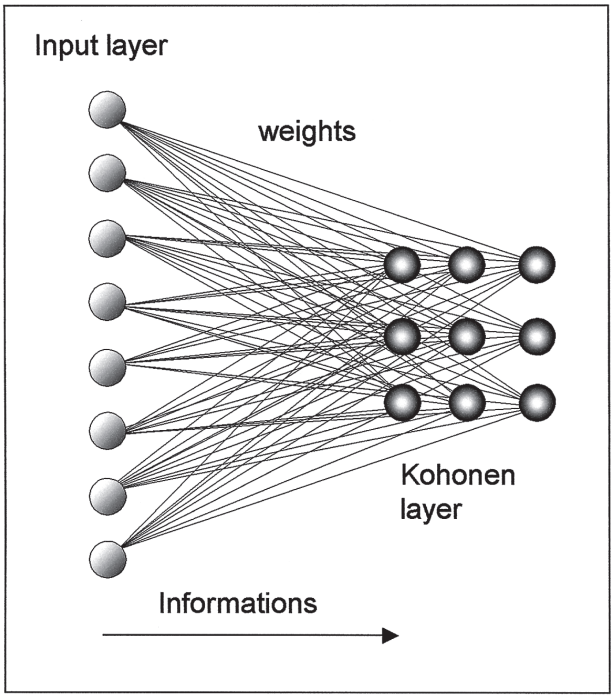

Fig. 1: An example of Kohonen-type neural network architecture. The elliptic Fourier coefficients describing leaves of different genotypes are applied to the input layer (represented here by 8 neurons; in the present work the number of neurons was actually 52) which activates a neuron or a group of neighbouring neurons in the Kohonen layer (represented here as having 3 x 3 neurons).

The input neurons send the incoming information, modified according to the specific synaptic weight factors, onto a two-dimensional array of neurons which is somehow topologically related to each other in a way that neighbourhood relationships are defined (simple rectangular scheme was used). These neurons constitute the "Kohonen layer" and they, upon receiving information from the input neurons, become activated. The Kohonen technique applies a "winner take all strategy", *i.e.* the neuron responding maximally to a given input is allowed to adjust its weight factors such that its response to a repeated stimulation with the same input will be even stronger. Few surrounding neurons are allowed to participate in this learning step; all other neurons do not adjust their synaptic weights during this step.

Thus, the network adjusts its internal connections (via the weight factors) autonomously, without reference to an external "teacher". The whole process is driven only by re-

peated representations of the input vectors and the applied learning rule. Interestingly, by doing so, an internal structure of the network emerges which allows visualisation of topological relationships hidden in the input data set.

There are many variations of basic algorithm of competitive learning. In the present study the one implemented in the "Stuttgart Neural Network Simulator, SNNS" was used. A technically oriented description of the algorithm can be found in the SNSS user manual.

Of the 65 leaves per accession studied, 60 were used in the Kohonen learning phase and 5 for the verification phase.

Results and Discussion

The result of the Kohonen network is a 2-dimensional map of neurons each of which carries a "label" that has excited it at the final verification test. It is expected that neurons excited by leaves of the same class (accession) will form clusters of small regions on such a map. If the test leaf falls into such a cluster it can be classified as belonging to the group that forms this cluster. The regions with neurons not excited by any object between two or more excited regions are called empty spaces. Empty spaces do not appear only between classes, they can be located within the region of a class as well.

After trying several Kohonen architectures: 9 x 9, 10 x 10, etc., a good separation between all 20 accessions studied was obtained with a Kohonen architecture having a layout equal to or larger than 13 x 13. After the Kohonen learning was completed, the EFA parameters of 100 leaves (5 for each accession) were sent through the 13 x 13 final Kohonen neural network and the excited neurons were marked by the label specific for each grapevine accession. In this way the map shown in Fig. 2 was obtained.

The separation and identification of all accessions was very good, except for Grand noir and Granoir (labeled by #6 and #7) which are obviously synonyms (Mancuso *et al*. 1998) and Abrostine and Abrusco (#1 and #2) which were supposed to be two clones of the same vine variety (Mancuso *et al*. 1998).

Although a single Kohonen network provides no quantitative information on the similarity of the accessions within groups, they do provide qualitative information about the groups. By using Kohonen layers of increasing sizes, finer discrimination may be sought and therefore some quantitative information can be obtained. Thus, networks with Kohonen layers of 1 x 1, 1 x 2, 2 x 2, 3 x 3, ….. 13 x 13 neurons were used to group the accessions. The details of the cluster formed at the 14 different discrimination levels are given in Tab. 2. When analysing many samples these tables are often difficult to interpret and it is therefore necessary to display the results in a more simplified graphical representation. It is evident from Tab. 2 that quantitative information on accession relationships can be elucidated, thus it should be possible to depict these details in a dendrogram format. The construction of the dendrogram (Fig. 3) begins when only a single neuron is used in the Kohonen layer and all 20 accessions necessarily group together; by increasing the number of neurons in the output layer of the network more

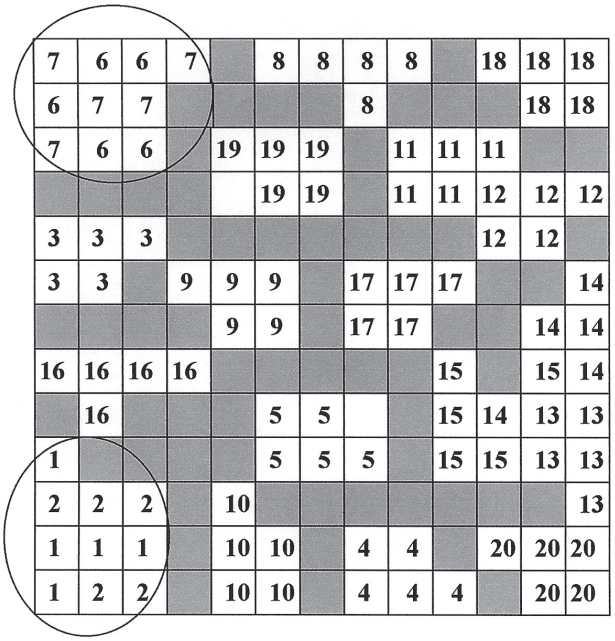

Fig. 2: The final 13 x 13 self-organizing map. There is a clear separation of each of the 20 accessions, except in the case of probable synonymy of Abrostine (#1) and Abrusco (#2) and Grand noir (#6) and Granoir (#7) (circled). For labels of the grapevine accessions see Tab. 1.

detailed discrimination is found and, finally, if the number of neurons in the output layer was 169 (13 x 13) all accessions were recovered separately except for the synonymy cases (Grand noir and Granoir; Abrusco and Abrostine).

The dendrogram separated early two different groups among the grapevine genotypes. The first group comprises the coloured accessions while the second group consists of all the Sangiovese-related ecotypes. In each of the two groups an intruder was found: Casentino in the first group and Morone in the second group. However, the affiliation of Casentino to the Sangiovese group is quite doubtful and previous works showing the high degree of divergence of Casentino from the other Sangiovese-related ecotypes, suggested a different origin for this accession (Mancuso 1999 a, b).

Regarding the presence of Morone in the Sangiovese group, it must be pointed out that the coloured accessions do not have the same origin, being linked only by the intense red colour of fruit and wine, thus the presence of the Morone in the Sangiovese group could be due either to a common origin or to a mere morphological similarity.

The results showed a high degree of relatedness both for Prugnolo acerbo, Prugnolo medio, Prugnolo dolce and for Brunelletto and Prugnolo gentile which agrees with the results of studies based on molecular markers (Sensi *et al*. 1996) and on backpropagation neural networks (Mancuso 1999 a).

The method of identification and classification of grapevine accessions using unsupervised learning with artificial neural networks has proved to be a reliable and effective analysis tool. Kohonen mapping of EFA parameters allows to distinct between vine accessions and to obtain natural grouping inside the data set. Applied to unlabeled data that contain only input data directly from EFA parameters, it pro-

Table 2

Groups produced by Kohonen neural networks having a different number of neurons in the Kohonen layer

Accession	Size of the Kohonen layer														
	#	1x1	1x2	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10	11x11	12x12	13x13
Abrostine	1	1	1	2	3	3	3	3	4	4	5	5	5	5	5
Abrusco	2	1	1	2	3	3	3	3	4	4	5	5	5	5	5
Colorino americano	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Colorino di Lucca	4	1	1	2	3	3	3	4	5	6	7	7	8	8	8
Colorino di Pisa	5	1	1	2	3	3	3	3	4	5	6	6	7	7	7
Grand noir	6	1	1	1	1	1	1	1	1	1	2	2	2	2	2
Granoir	7	1	1	1	1	1	1	1	1	1	2	2	2	2	2
Morone	8	1	2	3	4	4	5	6	7	8	9	10	11	11	12
Nereto	9	1	1	1	2	2	2	2	2	2	3	3	3	3	3
Tinturié	10	1	2	2	3	3	3	3	4	4	5	5	6	6	6
Prugnolo gentile	11	1	2	3	4	4	4	5	6	7	8	9	10	10	10
Brunelletto	12	1	2	3	4	4	4	5	6	7	8	9	10	10	11
Prugnolo acerbo	13	1	2	4	5	6	7	8	9	11	13	14	16	17	18
Prugnolo dolce	14	1	2	4	5	6	7	8	9	11	13	14	15	15	16
Prugnolo medio	15	1	2	4	5	6	7	8	9	11	13	14	15	16	17
Casentino	16	1	1	1	2	2	2	2	3	3	4	4	4	4	4
Chiantino	17	1	2	4	5	5	6	7	8	10	11	12	13	13	14
Morellino	18	1	2	3	4	4	4	5	6	7	8	8	9	9	9
Scansano	19	1	2	3	4	4	5	6	7	9	10	11	12	12	13
Sangiovese R10	20	1	2	4	5	6	7	8	9	11	12	13	14	14	15

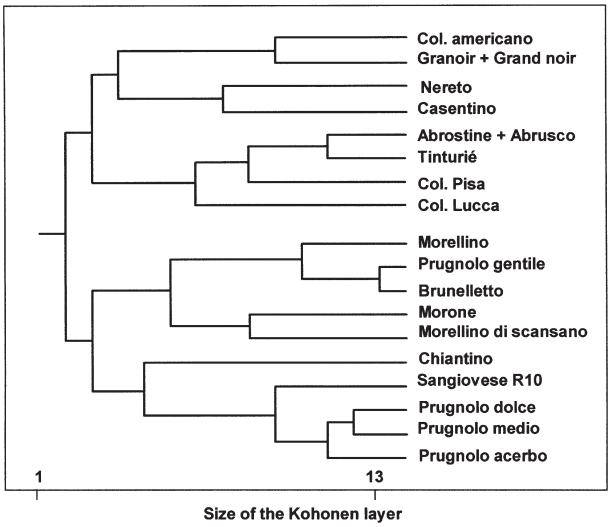

Fig. 3: Dendrogram of 20 grapevine accessions generated through the training of the Kohonen neural networks with different numbers of neurons in the Kohonen output layer.

vides unambiguous clustering on a two-dimensional plane. In term of grouping the input data, the Kohonen network could resemble to conventional multivariate statistical methods. From the aspect of reducing dimension it is similar to Principal Component Analysis (PCA). MELSENN *et al.* (1993) mentioned, however, that using a great number of data variables may result in a large number of significant principal components in PCA so that it may not retain sufficient information if only a few principal components are used for visualising the multidimensional data space. LOHNINGER and STANEL (1992) compared the Kohonen mapping and the k-nearest neighbour clustering in classification of mass spectral data in chemical compositions. They reported that the former was superior in all cases tested. This might be due to the fact that the classes did not form distinct clusters but had rather large areas of data space in common, and this overlap might be a disadvantage for k-nearest clustering.

In conclusion, the construction of self-organized maps allows both identification of "unknown" grapevine accessions and close examination of the relationships existing among them. The effectiveness of Kohonen networks in producing easily comprehensible low-dimensional maps extracting essential features out of complex data sets, may be especially helpful in comprehensive understanding of ampelographic data allowing to elucidate relationships that can not be detected by traditional data analysis tools.

Acknowledgement

This work was carried out within a "Giovani Ricercatori" research project financed by the Università di Firenze.

References

ALLEWELDT, G.; DETTWEILLER, E.; 1986: Ampelographic studies to characterize grapevine varietes. 4° Simp. Intern. di Genetica della Vite, Verona. Vignevini **13** (12, suppl.), 56-59

ANONYMOUS; 1983: Code de caractères descriptifs des variétés et espèces de *Vitis*. Office International de la Vigne et du Vin, Paris.

BENIN, M.; GASQUEZ, J.; MAHFOUDI, A.; BESSIS, R.; 1988: Caractérisation biochimique des cépages de *Vitis vinifera* L. par électrophorèse

d'isoenzymes foliaries: Essai de classification de variétés. Vitis **27**, 157-172.

Bowers, J. E.; Bandman, E. B.; Meredith, C. P.; 1993: DNA fingerprint characterization of new polymorphic cultivars. Am. J. Enol. Vitic. **44**, 266-274.

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 ampélométrique sur un échantillon de 34 cépages implantés en Sardaigne. J. Int. Sci. Vigne Vin **25**, 37-49.

Everitt, B. S.; 1993: Cluster Analysis. Edward Arnold, London, UK.

Jongman, R. H. G; Ter Braak, C. I. F.; Van Tongeren, O. F. R.; 1995: Data Analysis in Community and Landscape Ecology. Cambridge University Press, Cambridge, UK.

Kohonen, T.; 1984: Self-Organization and Associative Memory. Springer-Verlag, Berlin, Germany.

Lohninger, H.; Stanel, F.; 1992: Comparing the performance of neural networks to well-established methods of multivariate data analysis: The classification of mass spectral data. J. Anal. Chem. **344**, 186-189.

Mancuso, S.; 1999 a: Elliptic Fourier analysis and artificial neural networks for the identification of grapevine (*Vitis vinifera* L.) genotypes. Vitis **38**, 73-77.

- -; 1999 b: Fractal geometry-based image analysis of grapevine leaves using the box counting algorithm. Vitis **38**, 97-100.

- -; Ferrini F.; Nicese, F. P.; 1999: Chestnut (*Castanea sativa* Mill.) genotype identification: An artificial neural network approach. J. Hort. Sci. Biotechnol. **74**, 777-784.

- -; Nicese, F. P.; 1999: Identification of olive (*Olea europaea* L.) varieties using artificial neural networks. J. Am. Soc. Hort. Sci. **124**, 527-531.

- -; 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.

Mellssen, W. J.; Smits, J. R. M.; Rolf, G. H.; Kateman, G.; 1993: Two-dimensional mapping of IR spectra using a parallel implemented self-organising feature map. Chemom. Intell. Lab. Syst. **18**, 195-204.

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.

Smith, M.; 1994: Neural Networks for Statistical Modelling. Van Nostrand Reinhold, N.Y., USA.

Subden, R. E.; Krizus, A.; Lougheed, S. C.; Carey, K.; 1987: Isozyme characterization of *Vitis* species and some cultivars. Am. J. Enol. Vitic. **38**, 176-181.

Thomas, M. R.; Cain, P.; Scott, N. S.; 1994: DNA typing of grapevine: A universal methodology and database for describing cultivars and evaluating genetic relatedness. Plant Mol. Biol. **25**, 939-949.

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

White, R. J.; Prentice, H. C.; Verwijst, J.; 1988: Automated image acquisition and morphometric description. Can. J. Bot. **66**, 450-459.

Xu, H.; Wilson, D. J.; Arulsekar, S.; Bakalinsky, A. T.; 1995: Sequence-specific polymerase chain reaction markers derived from randomly amplified polymorphic DNA markers for fingerprinting grape (*Vitis*) rootstocks. J. Amer. Soc. Hort. Sci. **120**, 714-720.

Received January 23, 2001