Soil survey and continuous classification for terroir delineation in the "Colli Orientali del Friuli" wine production area

Investigation et classification continues du sol pour une délimitation du terroir dans la zone viticole "Colli Orientali del Friuli"

Gilberto BRAGATO* and Davide MOSETTI

CRA – Centro per lo studio delle relazioni tra pianta e suolo
Via Trieste, 23 – 34170 Gorizia, Italy.

***Corresponding author:** gilberto.bragato@entecra.it

Summary

The combination of a non-parametric dissimilarity index with auger boring recordings was tested in a project of soil suitability evaluation for quality wine production in a 2000-ha hill slope portion of the "Colli Orientali del Friuli" AOC district (Italy). The morphological characteristics – horizon sequence and the characteristics of each horizon – of 236 auger borings were recorded in 2006 according to the conventional practice for detailed soil surveys. The combination of "soft" data recorded in the auger boring campaign and the unsupervised clustering procedure consistently reduced the costs of survey. In particular, it helped us to delineate three different soil-landscape units being candidate for *terroir* delineation. Viticulture trials now in progress will give a final answer at the end of 2008.

Keywords: continuous classification, dissimilarity, soil suitability, soil survey

Introduction

The delineation of soil-landscape units suitable for quality wine production requires a detailed approach to soil characterization and mapping. Nowadays, detailed predictions from field data are the aim of precision farming, which manage geographical information at the farm scale. This methodology is valuable when quantitative soil data are considered (McBratney *et al.*, 2000). However, its predictive potential decreases in the *terroir* approach to quality wines, where the nature of soil as a complex system arranged in increasing levels of organization must be considered. From this point of view, the higher levels of organization, type and sequence of soil horizons in particular, affect rizosphere and grapevine roots more than the lower ones. Unfortunately, they can be characterized with morphological attributes – colour, structure, etc. – that the conventional approach to soil investigation deals with more easiness than precision farming techniques. Conventional soil survey is however more expensive than precision farming and rest on the subjective experience of the surveyor. On the occasion of a project of soil suitability evaluation for quality wine production we tested the non-parametric dissimilarity measure developed by Goodall (1966) for the classification of plant communities to make detailed soil surveys competitive with precision farming. Goodall's method is based on the assumption that a pair of observations sharing an infrequent value are more similar than two which share a more frequent one: pairwise dissimilarities are independently determined from sample frequencies of each attribute - whether discrete or continuous - and then combined to yield an overall dissimilarity measure. We were specifically aimed at finding out an unsupervised analytical procedure capable to combine Goodall's dissimilarity with hierarchical clustering and geostatistical analysis with the purpose to: i) find out homogeneous groups of soil observations for *terroir* delineation, ii) map their probability of occurrence for soil suitability evaluation; and iii) limit survey expenses using the data from auger boring observations, usually undervalued in the soil mapping phase.

Theory

In accordance with taxonomists, who often give much weight to attribute values with a small frequency of occurrence, Goodall (1966) has proposed a probabilistic approach in which uncommon

values make greater contributions to the dissimilarity D_{ij} that a pair of observations (k,l) randomly extracted from a population will have a similarity equal or greater than the pair (i,j).
The calculation of D_{ij} is performed in two steps. In the first one, pairwise dissimilarities are separately calculated for each attribute. When equal values V occur between pairs, D_{ij} only depends on the probability of occurrence p_i within the set of attribute values

$$(V_i = V_j)\Lambda(V_k = V_l)\Lambda(p_i \leq p_k) \Rightarrow D_{ij} \leq D_{kl} \qquad (1)$$

where p_i are estimated from sample frequencies f_i.
The calculation changes with the type of attribute when pairs of observations with differing values are considered. In nominal and binary attributes, they are all regarded as equally dissimilar

$$V_i \neq V_j \Rightarrow D_{ij} = 1 \qquad (2)$$

If attributes are multistate ordered, dissimilarities are calculated according to the number of values that lie between each pair of observations: the fewer they are, the lower is D_{ij}

$$(V_i \neq V_j)\Lambda(V_k \neq V_l)\Lambda\left(\sum_{u=i}^{j} p_u \leq \sum_{u=k}^{l} p_u\right) \Rightarrow D_{ij} \leq D_{kl} \qquad (3)$$

In case of quantitative attributes, dissimilarities are ordered by the magnitude of the difference between values

$$\left|y_i - y_j\right| < \left|y_k - y_l\right| \Rightarrow D_{ij} < D_{kl} \qquad (4)$$

and

$$\left(\left|y_i - y_j\right| = \left|y_k - y_l\right|\right) \wedge \left(\sum_{u=i}^{j} p_u \leq \sum_{u=k}^{l} p_u\right) \Rightarrow D_{ij} \leq D_{kl} \qquad (5)$$

In the second step, the pairwise dissimilarities calculated for the a different attributes are combined by Fisher's transformation for continuous probabilities (1948)

$$x^2 = -2\sum_{i=1}^{a}\left(\ln D_{ij}\right)_a \qquad (6)$$

and Lancaster's transformation for discrete probabilities (1949)

$$x^2 = 2\sum_{i=1}^{a}\left(1 - \frac{D_{ij}\ln D_{ij} - D'_{ij}\ln D'_{ij}}{D_{ij} - D'_{ij}}\right)_a \qquad (7)$$

where D'_{ij} is the first smaller dissimilarity value next to D_{ij}. Both continuous and discrete x^2 are distributed as χ^2 with 2a degrees of freedom, and the probability of the χ^2 value resulting from their sum yields the overall pairwise probabilistic dissimilarity.

Material and methods

We tested the combination of auger boring recordings and Goodall's dissimilarity in a project of soil suitability evaluation planned to investigate the relationship between wine quality and soil types delineated at the 1:10.000 scale. We specifically investigated a 2000-ha hill slope portion of the "Colli Orientali del Friuli" AOC district located just to the north of Manzano (Italy) (46°00'35"N,

13°25’20”E). The surveyed area has been originated by the outcrop of Eocene turbidites ranging from 60 to 220 m a.s.l. that display alternated layers of marls and sandstones.
From March to July 2006, we made a detailed soil survey based on 236 auger borings which locations were purposively selected to homogeneously cover the investigated area, irrespective of the land use. Each soil observation was done to a depth of 100 cm, recording the vertical sequence of horizons and, for each horizon, thickness, matrix colour, percentage of redoximorphic features (*RMFs*), texture class. A data set of 14 attributes was processed with the program *Simil* (Goodall et al., 1987) and the resulting matrix of pairwise dissimilarities processed with the agglomerative hierarchical clustering tool of the *Cluster* library of *R* statistical package (The R Foundation for Statistical Computing, 2005) to find out homogeneous clusters of soil observations. The nearest observation to each centroid was chosen as the reference augering of the cluster, and the corresponding vector of dissimilarities was analysed with the geostatistical software ISATIS (Geovariances, 2000).

Results and discussion

Agglomerative hierarchical clustering was carried out with the Ward’s method and its results are summarized in the dendrogram of Figure 1.

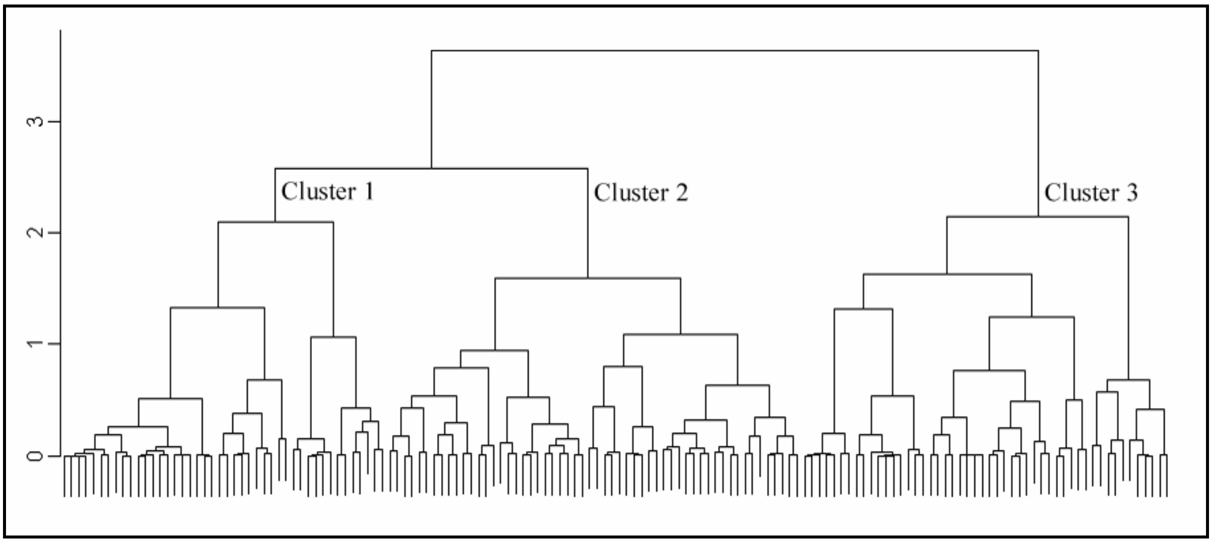

Figure 1 Dendrogram of pairwise dissimilarities.

According to a parsimony criterion, the optimal number of clusters was determined comparing cluster centroids in accordance with soil classification criteria, stopping the selection procedure when no meaningful differences between adjacent clusters in dendrogram were detected. Selection resulted in 3 clusters, which centroidal values (mode, median and mean for qualitative, multistate ordinal and quantitative attributes, respectively) are reported in Table 1.

Attributes	**Cluster**		
	1	**2**	**3**
Horizon sequence	A|AC|C	A|AC|C	A|B|C
Moist colour: - A horizon	10YR 4/3	7.5YR 4/4	7.5YR 4/4
- AC or B horizon	10YR 4/4	7.5YR 4/6	7.5YR 4/6
- C horizon	10YR 4/4	7.5YR 4/6	7.5YR 5/6
Soil depth	90	85	> 100
Depth of *RMFs*, cm	-	-	65
Effervescence to HCl	present	present	absent
Fe-Mn nodules	absent	absent	present
A horizon: - pH	7.8	7.6	6.5
- sand content, %	36	26	31

- clay content, %	23	27	22

Table 1 Centroidal values of the selected clusters.

Cluster centroids showed different horizon sequences additionally characterized by different colour, effervescence, pH and texture. They can be explained in terms of soil classification when considering the presence of a B horizon: *Cluster 1* and *Cluster 2* are formed by young soils provisionally classified as Regosols (FAO, 1998), whereas *Cluster 3* can be probably classified as a Cambisol or a more aged soil type.

The dissimilarity vectors of the nearest observation to each centroid were processed with geostatistical techniques. Since dissimilarities were far from showing a gaussian distribution, they were transformed by gaussian anamorphosis modelling (Chilés and Delfiner, 1999) before geostatistical analysis. Figure 2 summarizes the spatial variability of transformed dissimilarities reporting their variograms and the parameters of the fitted models.

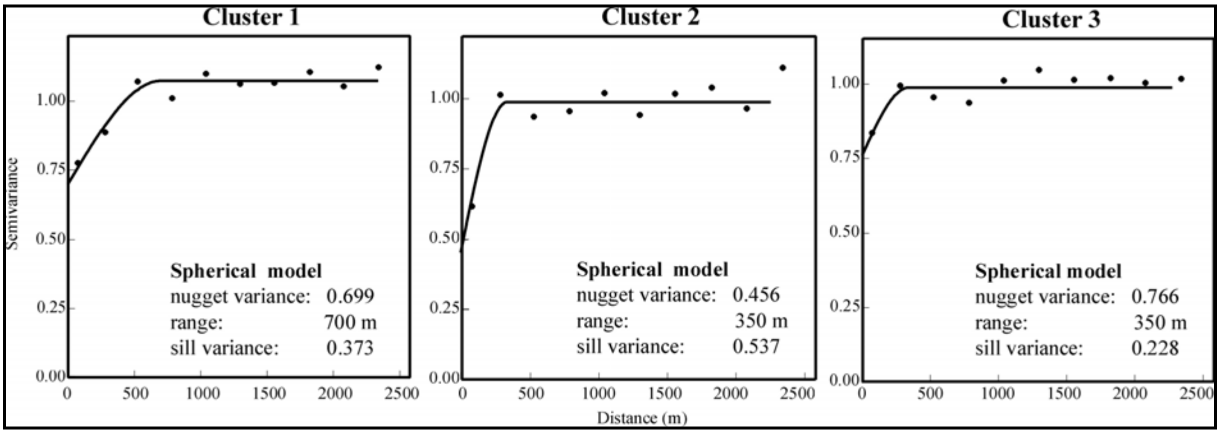

Figure 2 Variograms of dissimilarities transformed by gaussian anamorphosis.

Variogram parameters are expression of local environmental factors affecting the attribute. *Nugget* and *sill* variance are the random and spatially-related components of variance, whereas the *range* is the radius of influence of the factor of environmental variability. In our investigation, *Cluster 1* originated a variogram markedly different in *range* from the other two variograms that, on the contrary, shared the same *range* of 350 m.

The centroidal values of Tab. 1, colour and type of subsurface horizon in particular, suggest lithology as the main factor affecting soil spatial variability in the area under investigation. Grayish brown to brownish grey sandstone strata are more frequent than brown marly strata in the parent material of *Cluster 1*, whereas the situation reverses in *Cluster 2* and *Cluster 3*. According to variograms and Tab. 1, the latter clusters could represent an evolutionary sequence from young soils of *Cluster 2* to more aged soils of *Cluster 3*. Their different nugget variance also suggests that *Cluster 3* is more randomly distributed than *Cluster 2* and its presence limited to small portions of the investigated area.

Nugget variance is quite high in all variograms, increasing from 46% of relative nugget in *Cluster 2* to 77% in *Cluster 3*, and making kriging predictions less precise than desired. This random component of variability is partly related to the chaotic - hence unpredictable - sequence of strata typical of turbiditic formations. However, we think that the new terraces built with heavy machinery, which consistently altered soil organization in vineyards in the last 20 years, have exerted the most consistent influence disturbing the gradual variation usually occurring in soilscape.

After interpolation, kriging-predicted dissimilarities were back-transformed to the original [0,1] scale and the three dissimilarity vectors combined to produce the discrete mapping units reported in Figure 3. Discretization was carried out assigning each node of the interpolation grid to the cluster that displayed the lowest dissimilarity value. A further mapping unit in Fig. 3 takes into account land use, separating woods from cultivated areas.

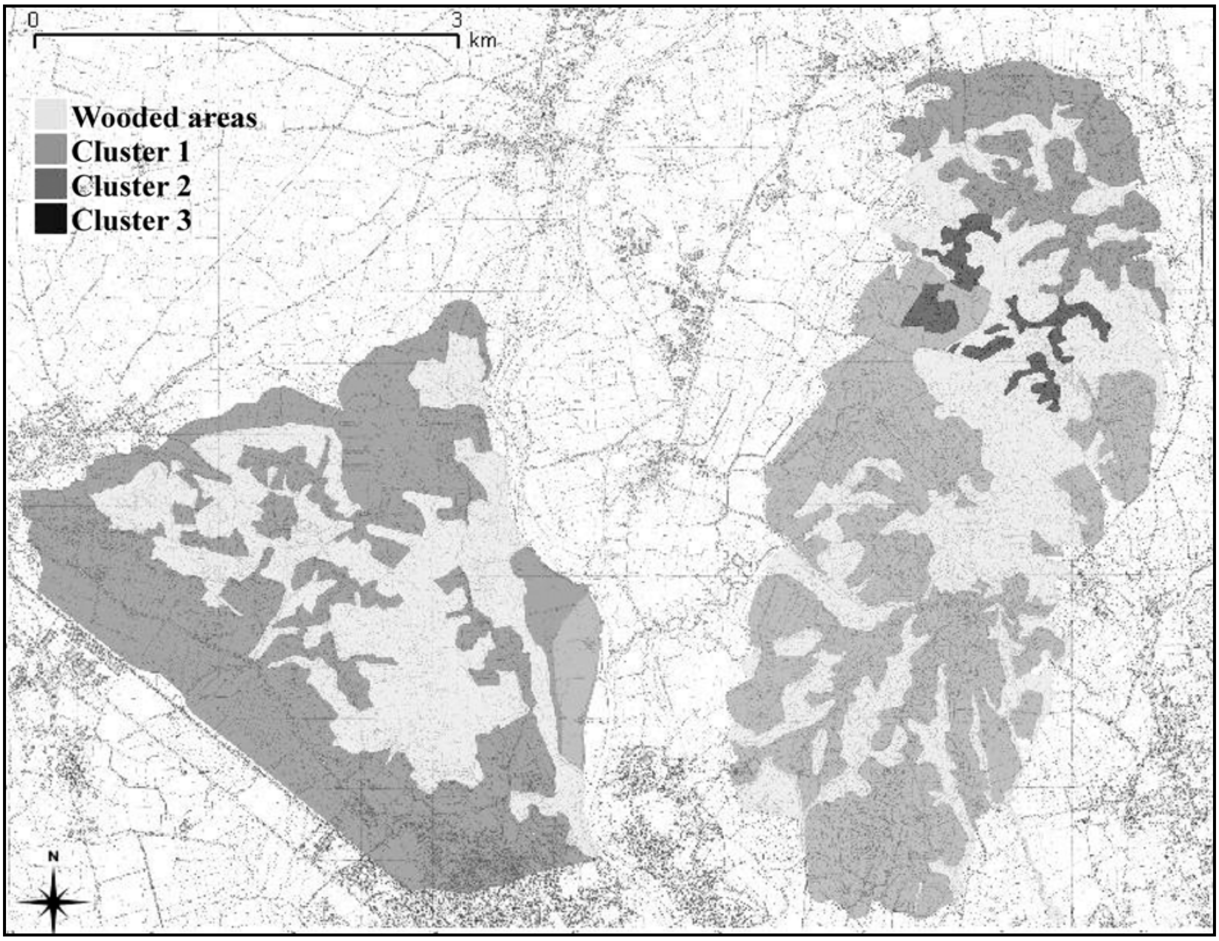

Figure 3 Spatial distribution of the three clusters in the investigated area. Wooded areas are kept apart in the map.

The map of Fig. 3 summarizes the information contained in variograms. The peripheral location of *Cluster 3* accounted for the large nugget effect observed in its variogram. Its location looks related to a sandier lithology that involves also nearby wooded areas. *Cluster 1* is homogeneously distributed in the eastern part of the investigated area, whereas *Cluster 2* largely prevails in the western hill slope area, the two clusters being separated by a fault located in between and covered by post-glacial gravely sediments. The two clusters are furthermore characterized by a slightly different landscape, with *Cluster 2* displaying more gentle slopes then *Cluster 1*.

Conclusions

The combined use of Goodall's dissimilarity, hierarchical clustering and geostatistical analysis in the analysis of auger boring recordings met three compulsory requirements to move from field observations to soil suitability evaluation: i) pedological meaning of clustered observations; ii) structured spatial variability of dissimilarity values; and, iii) maps consistent with environmental factors acting on the soilscape. The soil units we delineated with this procedure are currently under test with viticulture trials carried out in collaboration with the CRA Center for Viticulture of Conegliano Veneto to be proposed as separate *terroir* units. Since some microclimatic differences were observed between the southern and the northern mapping units of *Cluster 2*, we chose to split and test them separately.

References

CHILÉS J.P., and DELFINER, P. 1999. *Geostatistics: modelling spatial uncertainty*. Wiley, New York.

FISCHER H.R. 1948. Statistical methods for research workers. 10th Ed., Oliver and Boyd, Edinburgh.

GEOVARIANCES 2000. *ISATIS Software Manual*. Eds. Geovariances, Avon.

GOODALL D.W. 1966. A new similarity index based on probability, Biometrics, **22**, 882-907.

GOODALL D.W., GANIS P., and FEOLI E. 1987. Probabilistic methods in classification: a manual for seven computer programs. *GEAD-EQ Report no. 7*. Università degli Studi di Trieste, Trieste.

LANCASTER H.O. 1949. The derivation and partition of χ^2 in certain discrete distributions, *Biometrika*, **36**, 117-129.

FAO 1998. *World reference base for soil resources*. Roma

MCBRATNEY A.B., ODEH I.O.A., BISHOP T.F.A., DUNBAR M.S., and SHATAR T.M. 2000. An overview of pedometric techniques for use in soil survey. *Geoderma*, **97**, 293-327.

THE R FOUNDATION FOR STATISTICAL COMPUTING 2005. *R. Version 2.2.1.* ISBN 3-900051-07-0.