Combination of heterogeneous data sets in Precision Viticulture

J-N. Paoli[1], B. Tisseyre[1], O. Strauss[2], J-M. Roger[3], S. Guillaume[3]
[1]*Agricultural Engineering university, 2 place Viala, 34070 Montpellier Cedex 1*
[2]*LIRMM, 161 Rue Ada, 34392 Montpellier Cedex 5*
[3]*Cemagref, 361 rue Breton, 34033 Montpellier Cedex 1*
paoli@ensam.inra.fr

Abstract

This paper deals with the combination of Precision Viticulture data. It proposes a method to treat qualitative and quantitative data sets according to the same formalism, and to take into account their inaccuracies. The originality of our approach lies in the use of uncertainty theories (fuzzy sets, possibility theory, Choquet integral). Its efficiency is shown by the treatment of a Precision Viticulture data set for a field diagnostic.

Keywords: data aggregation, fuzzy logic, viticulture.

Introduction

Precision agriculture and linked technologies have recently been applied on grapes to characterize the within field variability in the production system (Hall et al., 2003; Johnson et al., 2003; Bramley, 2001; Ortega et al., 2003; Tisseyre et al., 2001).

Spatial data analysis, at the within field level, leads to aggregation of spatial information were resolution and precision are not necessarily the same. In most of the cases, problems with data heterogeneity have been solved using geostatistics and kriging (Journel & Huijgbregts, 1978; Frogbrook et al., 1999). The goal of kriging is then to shift all the data onto the same grid (same resolution) which allows for classical data analysis or queries at different layers (one layer for each data set).

Nevertheless, in some particular cases, the suitability of kriging procedures and geostatistics for viticulture has to be discussed. Indeed, wine growers have a more or less rough knowledge of the within field variability with regards to some parameters like soil, climate, vigor, disease, etc. During the process of spatial data analysis, it seems unavoidable to take this expert knowledge into account. The problem is that this knowledge leads to linguistic data such as “the soil is deep”, or “the top of the field”. The combination of linguistic expert data with data provided by conventional sensing systems is not straightforward and leads to uncertainty in the reliability of the conclusions.

In this work we propose to use a method, which can be applied regardless of data type (expert or numeric) or location specifics (zone roughly drawn by an expert or DGPS). The method considers, within the same formalism:

- the inaccuracy of the data, whether they are provided by an expert (linguistic) or by a sensing system (numeric),
- the inaccuracy of the location specifics and/or the boundary of the zone where the measurements are located.

The method was applied on a heterogeneous precision viticulture data set (including grape yield and sugar content from a monitoring system as well as elevation, vigor, soil resistivity and the winegrower's soil knowledge) in order to analyze and to understand the within field variability. The goal of this project was also to test the methods as a tool, which could be used "interactively" by the grower in order to analyze his own data.

Materials and method

Experimental data sets

Our experimental Merlot vineyard is located in Navarra (Julian Chivite winery, in Spain). The field area is 1.5 ha. Grape yield and sugar content were measured and mapped in September 2003 using an on-line sensor mounted on a grape-harvester (Pellenc S.A.) located by a DGPS. The collected data had a resolution of 2400 points/ha.

Complementary measurements were done in March 2004. The recorded parameters are elevation (157 measurements) with a laser tacheometer Leica (TPS 1100), trunk diameters (45 manual measurements), and soil resistivity (45 measurements). Also, two expert zones were delimited within the field by the wine grower. The first one represents a deep soil and the second one a superficial soil zone.

Description of the information sources

Fuzzy description of the spatial data

Each data is considered as an "information source". Locations and values of the information sources are described by two fuzzy subsets:

- the first one is a fuzzy area, which describes the location of the information, and its inaccuracy,
- the second one is a fuzzy number that describes the associated value, and the sensor inaccuracy or the expert vagueness.

The shape and the size of the fuzzy subsets are different for each data set. They are chosen according to technical aspects (such as the known inaccuracy of a sensor). This paper will show how the fuzzy description is done for only two different data types: yield and soil depth. Note that sugar, elevation, trunk diameters, and soil resistivity data are described similarly to the yield data.

Yield data set

Previous works have been used to describe and model the inaccuracies of the DGPS and the yield sensor (Tisseyre et al., 2001; Ehrl, 2003). Figure 1 shows how the spatial accuracy and the measurement accuracy are measured for the yield. The spatial inaccuracy of the DGPS is modeled by a support of 1.5 m and a kernel of 0.5 m around the location given by the DGPS. A yield value is described by a support of more or less 2% around the measured value and a kernel of more or less 1% around the measured value.

a) Spatial inaccuracy of the DGPS b) Inaccuracy of the yield sensor

Figure 1. Description of a yield information source.

Soil depth data set

The boundaries of the soil zones are drawn manually by the winegrower. The boundary of each zone is used to generate a fuzzy region. The support of the region is given by a dilation of the boundary and the kernel by an erosion. The distance between the kernel and the support is 10 m. The general shape of these information sources is described in figure 2.

a) Location b) Qualitative label

Figure 2. Description of a soil depth information source.

Qualitative levels

The values are replaced with qualitative labels, such as "low, "medium" and "high". This choice seems to be relevant since our growers consider only two to five qualitative labels within field values. The choice of the threshold values between the labels is determined by technical considerations (grower knowledge) or by the data distributions. Figure 3 shows an example of yield labels based on the data distribution, high yield values are 50% of the upper distribution and low yield values are below the median.

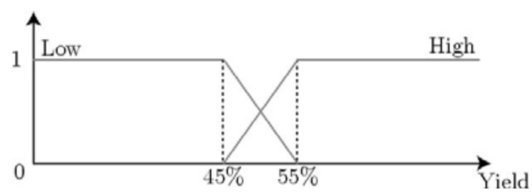


Figure 3. Partitions of the yield referential.

The association of a fuzzy measurement M and a fuzzy label A_i provides two confidence measures (Dubois & Prade, 1988) :

- The possibility degree $\Pi\left(M, A_i\right)$, indicates that the measurement M can be A_i,
- The necessity degree $\mathrm{N}\left(M, A_i\right)$, indicates with certainty that the measurement M is A_i and nothing else.

<u>Principle of the diagnostic</u>

Since the goal of the grower was to understand the yield variability, the analysis first involves yield data processing. The method required the following steps :

- computation of a yield map using our fuzzy approach,
- zones of interest manually drawn on the yield map by the expert (the grower),
- aggregation of all the complementary data on each zone delineated by the expert.

The aggregation process considers the measurement inaccuracy as well as the spatial relevancy of the information sources, or results interpretation.

Computation of the yield map

The vineyard is partitioned into a thousand fuzzy request zones R_i, called management cells. The size of a management cell is defined according to the smallest manageable area (16 m^2 in our case). Management cells have to be fuzzy in order to consider spatial continuity of the within field variations. Possibility and necessity degrees computed from the yield information sources are aggregated on the different management units. The aggregation process provides an estimate of the possibility and the necessity degrees of each label (low or high yield) on each management unit.

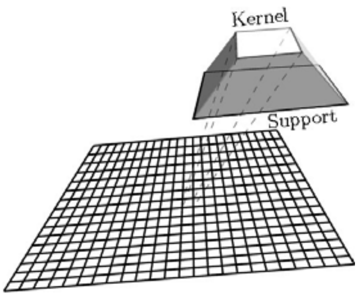


Figure 4. Partition of the vineyard.

The aggregation process has to consider several specificities:

- The spatial relevancy of the information sources on the management cells,
- Conflicts between information sources on a management cell,
- Outliers.

All of the specificities are considered with an aggregation operator based on a Choquet integral (Grabish et al., 2000; Paoli et al., 2004). The resulting aggregation gives each management cell :

- The possibility that the cell is a high yield $\Pi\left(R_i, high\right)$, and the necessity (certainty) that the cell is a high yield $\mathrm{N}\left(R_i, high\right)$
- The possibility that the cell is a low yield $\Pi\left(R_i, low\right)$, and the necessity (certainty) that the cell is a low yield $\mathrm{N}\left(R_i, low\right)$

Fuzzy partition of the yield map

The resulting yield map is divided into zones by the expert. This step corresponds to generating "request zones"' or "zones of interest", according to a high resolution data. In addition, soil map, erosion and dilation were applied on the boundaries drawn by the expert in order to consider the spatial inaccuracy of the zone delineation. A distance of 10 m between the kernel and the support was selected for this study. Zones delineated by the experts correspond to geographical fuzzy subsets of the field. In this particular study, the zoning is performed by the expert according to the yield. The delineation tends to highlight high yield zones and low yield zones.

Validation of the fuzzy partition
In order to assess the relevancy of the field delineation, an aggregation of the yield information sources is performed on each expert zone. The necessity degrees for the two labels are then computed for each zone. These degrees are a tool to assess the relevancy of the zoning. They allow the expert to refine the boundaries to improve the delineation of the field according to objective information.

Characterization of the yield zones
The zones drawn according to the yield map are considered as "zones of interest" by the expert. They are used to aggregate the information sources from the other variables. The aggregation operator used for this step is the same as the one used for the aggregation of the yield information sources.

Results and discussion

Comparison of the fuzzy yield map with a kriged map
Figure 5 shows two yield maps corresponding to the same vineyard. Figure 5)a is a kriged map, generated by the method usually applied on precision viticulture data sets (Bramley, 2001). Figure 5)b shows a fuzzy yield map resulting from our method. Both maps show the same patterns of low and high yield. In order to make the comparison possible, the same thresholds were applied on both these maps. It is important to note that these thresholds are:

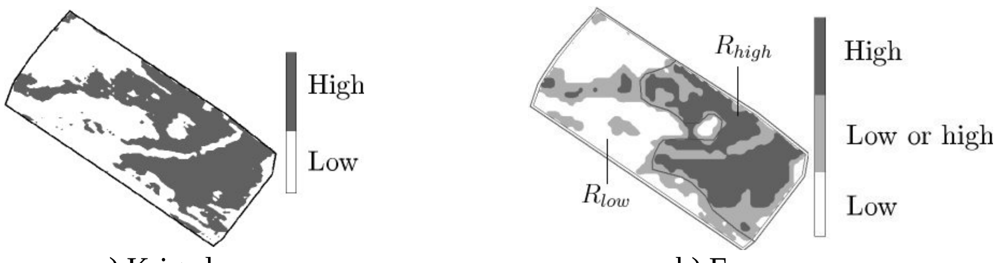
- crisp and applied on the interpolated data for the kriged map,
- fuzzy and applied on the raw data for the fuzzy map.

In spite of this difference, both maps look similar. This result confirms the relevancy of our approach. Note that our approach allows three types of the management cells to be considered:

- Dark cells, where the yield is high ($\mathrm{N}(R_i, high) = 1$).
- Gray cells, where the yield may be high or low ($\Pi(R_i, high) > 0$).
- White cells, where the yield is low ($\Pi(R_i, high) = 0$).

Our fuzzy approach presents indeterminate areas (cells where the yield may be high or low) in a comprehensive manner. This information is obvious to the expert and is helpful to delineate (and to refine) the zones, which correspond to high or low yields.

Figure 5.b shows how the expert considered two zones on this particular vineyard. A high yield zone (R_{high}) and a low yield zone (R_{low}) were drawn manually. Both these zones are considered as request zones in the next steps of the process.



a) Kriged map b) Fuzzy map
Figure 5. Yield maps.

Validation of the yield zones

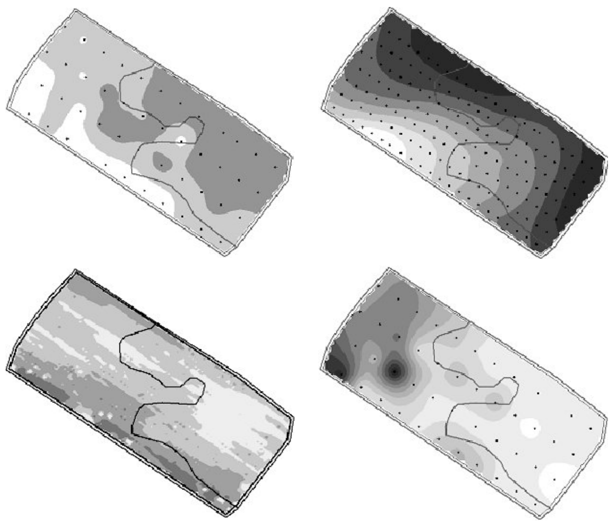
Aggregation of the information sources on the request zones gives an assessment of the zoning relevancy. In this particular case (figure 5)b), a “high” label is necessary on the high yield zone as defined by the expert ($\mathrm{N}\left(R_{high}, high\ yield\right) = 1$), and the low label is necessary on the low yield zone ($\mathrm{N}\left(R_{low}, low\ yield\right) = 1$). It shows that the expert zones are relevant. Note that this relevancy is assessed considering:

- the inaccuracy of the data provided by the sensor,
- the inaccuracy of the locations of the data and the expert boundaries,
- the labels’ inaccuracy.

During this step, our approach constitutes a powerful tool, which can be helpful for the winegrowers to delineate their field.

Diagnostic of the fuzzy yield zones

Figure 6 shows a basic representation of the spatial variations of the complementary variables. These maps are generated by an ordinary kriging approach. The sampling locations are also presented. Note that the problem with data aggregation is not obvious in this case since the resolution is different for each complementary data. For each map, the interest zones corresponding to the high yield zone and the low yield zone delineated by the expert are also shown. Figure 7 shows the location of the soil expert zones (drawn on the soil resistivity map).



a) Vigor b) Elevation

c) Sugar content d) Soil resistivity

Figure 6. Complementary measurements.

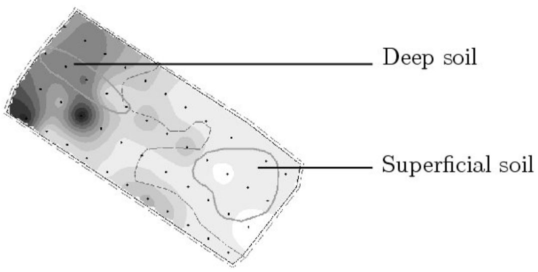


Figure 7. Qualitative information about soil.

On the low yield zone, the aggregation leads to:

- medium or high soil resistivity (but not low $\Pi\left(R_{low}, low\ resistivity\right)=0$)
- medium or high sugar content (but not low $\Pi\left(R_{low}, low\ sugar\ content\right)=0$)
- low or medium vigor (but not high $\Pi\left(R_{low}, high\ vigor\right)=0$)
- low or medium elevation (but not high $\Pi\left(R_{low}, high\ elevation\right)=0$)
- Superficial soil (certainty $\mathrm{N}\left(R_{low}, deep\ soil\right)=0.3$)

This result provides significant information, which highlights the main characteristic of the vineyard. It provides a simplification of the vineyard behavior, which makes the expert understand where the low yield results are from. In this particular case, it was used by the expert to make new assumptions on the vineyard. The low yield corresponds to particular conditions (soil resistivity, sugar, elevation) that all together confirm there are time, stable soil and water availability problems. The winegrower plans to manage this zone specifically by irrigation.

This example shows how our method allows for aggregate heterogeneous data on a zone of interest that makes sense to the expert. A similar process was performed on the high yield zone. The aggregation leads to:

- low or medium soil resistivity (but not high $\Pi\left(R_{high}, high\ resistivity\right)=0$)
- low or medium sugar content (but not high $\Pi\left(R_{high}, high\ sugar\ content\right)=0$)
- medium or high vigor (but not low $\Pi\left(R_{high}, low\ vigor\right)=0$)
- medium or high elevation (but not low $\Pi\left(R_{high}, low\ elevation\right)=0$)
- Superficial soil (certainty $\mathrm{N}\left(R_{high}, superficial\ soil\right)=0.3$)

Once again, high yield zones correspond to particular conditions that all together confirm there is time, stable soil and high water availability. Although this phenomenon was highlighted by this study, the winegrower does not plan to undertake any specific management in this zone.

Acknoledgements

This work was founded by the French ministry of research within the framework of the Eureka VITIS Project.

Conclusions

In this paper, we used a method to fuse together data between linguistic expert data and data provided by conventional sensing systems. This method is based on fuzzy sets, possibility theory and Choquet integral. It takes into account the inaccuracy of the data, (whether they are provided by an expert or by a sensing system), and the inaccuracy of their locations. It can be applied whatever the resolution of the available data sets. The efficiency of our approach was shown by the treatment of a precision viticulture data set. The example which was detailed in this paper showed the ability of this method to analyze spatial data in a straightforward way for an expert.

Our future work will focus on testing and improving this method. The comparison with usual interpolation methods has to be more detailed in order to highlight the advantages and the disadvantages of such an approach. Some theoretical aspects require further work. The spatial distribution of the data within a request zone is of particular interest. It could be taken into account to refine the estimate of possibility and the necessity degrees.

References

Bramley, R.G.V. and Lamb, D.W. 2003. Making sense of vineyard variability in Australia. Proceedings of the International Symposium on Precision Viticulture, 9th Latin American Congress on Viticulture and Oenology, pp. 35-54.

Dubois, D. and Prade, H. 1988. Possibility theory - An Approach to the Computerized Processing of Uncertainty. (Ed.) Plenum Press, New-York.

Ehrl, M., Stempfhuber, W., Auernhammer, H. and Demmel, M. 2003. Quality assessment of agricultural positioning and communication system, In: Precision Agriculture, edited by J. Stafford and A. Werner, Wageningen Academic Publishers, The Netherlands, pp. 205-210

Frogbrook, Z. L., Oliver, M.A., Salahi, M.and Ellis, R.H. 1999. Comparing the relations in the spatial variation of soil and crop attributes. In: Precision Agriculture '99, edited by J.V. Stafford, Sheffield Academic Press, Sheffield, UK, pp. 397-405

Grabisch, M., Murofushi, T.and Sugeno, M. 2000. Fuzzy measures and integrals - Theory and applications. New York, Physica-Verlag Heidelberg.

Hall, A., Louis, J. and Lamb, D. 2003. Characterizing and mapping vineyard canopy using high-spatial-resolution aerial multispectral images, Computers & Geosciences, 29, 813-822.

Johnson, L. F., Roczen, D. E., Youkhana, S. K., Nemani, R. R., and Bosch, D. F., 2003. Mapping vineyard leaf area with multispectral satellite imagery, Computers and Electronics in Agriculture, 38, 33-44.

Journel A.G., Huijbregts C.J. 1978. Mining Gesotatistics. (Ed.) Academic Press, London.

Ortega, R., Esser, A. and Santibañez O. 2003. Spatial variability of wine grape yield and quality in Chilean vineyards : economic and environmental impacts. In: Precision Agriculture, edited by J. Stafford and A. Werner, Wageningen Academic Publishers, The Netherlands, pp. 499-506.

Paoli, J-N., Strauss, O., Tisseyre, B., Roger, J-M. and Guillaume, S. 2004. Fusion de données géoréférencées (Fusion of georeferenced data), Actes de la XIIe conférence sur la logique floue et ses applications. pp. 77-84.

Tisseyre, B., Mazzoni, C., Ardoin, N. and Clipet, C. 2001. Yield and harvest quality measurement in precision viticulture – application for a selective vintage. In: Third European Conference on Precision Agriculture, edited by G. Grenier and S. Blackmore, pp. 133-138.

Yager, R. R., Ovchinnikov, S., Tong, R. M. and Nguyen, H. T. 1987. Fuzzy Sets and Applications: Selected Papers by L.A. Zadeh. (Ed.) John Wiley & Sons, New-York.