Delineating Vineyard Zones by Fuzzy K-Means Algorithm Based on Grape Sampling Variables

Ana Belén González-Fernández[1], José Ramón Rodríguez-Pérez[1], Enoc Sanz-Ablanedo[1], José Benito Valenciano[3], Victoriano Marcelo[2*]

[1]Grupo de Investigación en Geomática e Ingeniería Cartográfica (GEOINCA), Escuela Superior y Técnica de Ingeniería Agraria, Universidad de León, Avenida de Astorga, s/n, 24401, Ponferrada, León, Spain. Phone: +34 987442022.

[2]Departamento de Ingeniería y Ciencias Agrarias, Escuela Superior y Técnica de Ingeniería Agraria, Universidad de León, Avenida de Astorga, s/n, 24001, Ponferrada, León, Spain. Phone: +34 987442029.

[3]Departamento de Ingeniería y Ciencias Agrarias, Escuela Superior y Técnica de Ingeniería Agraria, Universidad de León, Avenida de Portugal, 41, 24071, León, Spain. Phone: +34 987295242.

***Corresponding author:** Email: v.marcelo@unileon.es Phone: +34 987442029. Fax: +34 987442070

ABSTRACT:

This study describes a method for delineating management zones using interpolated maps of grape characteristics recorded in 2013 and 2014 in a Godello vineyard located in the Bierzo Denomination of Origin (León, Northwest Spain). Ten variables were analyzed and recorded for the sampled vines (50 vines/ha). Interpolated maps reflecting each variable and year were created by spatial interpolation (kriging) from the sampled points. Principal component analysis was used to detect relationships between variables and to select the variables to be used to create the cluster classification. Using the fuzzy k-means classification algorithm implemented in the Management Zone Analyst (MZA v.1.0.0) software, several zones were delineated by combining the studied variables. The results delineated 2 different management areas composed of 3 zones each based on winery objectives: (1) to increase grape production (combining the yield for 2013 and 2014); and (2) to improve grape composition (combining the pH for 2013 and 2014).

Key words: cluster classification, Godello, grape characteristics, management zones, precision viticulture, *Vitis vinifera* L.

Highlights

Delineate management zones improves the management of vineyards.

Management zones were delineated using interpolated maps of grape variables.

PCA was used to select variables as inputs for the fuzzy k-means classification.

Zone location and number were optimized using MZA software and the fuzzy k-means algorithm.

The results were 2 new zones: one to raise grape production and other to improve grape composition.

Abbreviations

TA: titratable acidity ($kg \cdot m^{-3}$)

BW: weigh of 100 berries ($kg \cdot 10^{-3}$)

MI maturity index ($TSS \cdot TA^{-1}$)

CW: cluster weight ($kg \cdot 10^{-3}$)

PW: shoot pruning weight ($kg \cdot 10^{-3} m^{-1}$)

RI: Ravaz index ($Y \cdot PW^{-1}$)

TSS: TSS: total soluble solids (ºBrix)

WS: weight of shoots ($kg \cdot 10^{-3}$)

Y: yield ($kg \cdot 10^{-3} m^{-1}$)

1. Introduction

Grape production and quality are not even within the same parcel. Moreover, the spatial variability of the variables that define grape production and quality are not stable between campaigns, although the distribution pattern tends to remain the same (Arnó et al., 2011). One approach to optimize production goals is to delineate homogeneous blocks for separate management (Proffitt et al., 2006). The use of different management zones means better management of cropping practices, including segregated harvests and the establishment of monitoring points (Santesteban et al., 2010). The management zoning approach is usually used in vineyards in Australia, Chile, New Zealand, South Africa and the USA. For Australian vineyards, Bramley and Hamilton (2004) identified zones with similar yield characteristics, and Bramley (2005) identified zones based on maturation and quality variables. For Chilean vineyards, Esser and Ortega (2002) delineated homogeneous zones according to soil characteristics and related these with grape composition, production and vigor variables. In USA, California, Johnson et al. (2001) delineated management zones according to vine vigor. In Spain, Arnó et al. (2005) related yield maps with leaf petiole composition, and González-Fernández et al. (2016) identified homogeneous zones on the basis of a combination of grape composition, production and vigor variables.

Delineating zones with similar characteristics requires an understanding of the spatial stability of the variables that define final grape characteristics (Keller, 2015), such as grape composition and production (Cortell et al., 2005). Grape composition is usually defined by total soluble solids (TSS), which estimates the probable alcohol content of the wine (Alburquerque et al., 2007). pH and titratable acidity (TA) define the acidity of grapes and the organoleptic characteristics of a wine (Blouin and Guimberteau, 2003). Production variables like yield (Y), Mean weight of one cluster calculated from the ratio Y/number of clusters (CW) and the weight of a sample of 100 berries (BW) depict the productive potential of a vineyard. The inverse relationship between grape composition and production variables has been widely documented (González-Fernández et al., 2012). At maturity, grape size increases due to water accumulation in the berry; however, if water accumulation is excessive, the concentrations of the components that determine quality drop, although the quantities remain the same (Walker et al., 2005). Production variables are directly related to vine vigor variables like mean shoot weight (WS) and total pruning weight (PW). Production variables are directly related to vine vigor variables like mean shoot weight (WS) and total pruning weight (PW). To produce quality wine, some authors (Cortell et al., 2005) recommend ensuring vine balance, estimated using the Ravaz index (RI), which is the ratio between Y and PW (Ravaz, 1911). The recommended RI range is between 4 and 10 (Champagnol, 1984).

For financial and practical reasons, determination all variables related to production and quality for all vines in a vineyard is not possible. One solution is to select a sample of vines from a plot and interpolate the data to the whole area. Goovaerts (1999) indicated that the value of a variable studied in 2 different locations will be more similar for nearer locations. Of several interpolation techniques available (inverse distance, triangulation, etc.), kriging is widely used because it considers data variability from the variogram and usually results in lower error than other techniques (Brooker, 1986).

Several multivariate analysis methods are available to delineate management zones. A useful tool for delineating management zones is cluster classification analysis, which classifies individual data into different classes (clusters). The individuals belonging to each homogeneous cluster are grouped according to proximity criteria defined by a distance function (Urretavizcaya et al., 2014). One of the most widely used cluster algorithms is fuzzy k-means, which groups data into k classes whose centroid reflects the minimum Euclidean distance from each data point (Tagarakis et al., 2013). In a study of management zone delineation according to yield, Arnó et al. (2011) compared the k-means and fuzzy k-means algorithms, concluding that fuzzy k-means led to more compact and balanced zones over time.

Baluja et al. (2013) used fuzzy k-means classification to identify management zones for the production of different types of Spanish wines. For a study of Greek vineyards, with yield and grape composition as the reference variables, Tagarakis et al. (2013) used fuzzy k-means to delineate homogeneous blocks according to the vegetation index and soil characteristics. Urretavizcaya et al. (2014) used fuzzy k-means to investigate the importance of early berry sampling in defining management zones.

Our aim was to optimize the management of the plot by delineating management zones according to grape composition, production and vigor variables using the fuzzy k-means algorithm. The novelty of this study in relation to previous studies is that input are grape composition and production variables using variables measured by winegrowers that do not require specific additional variables for delineation.

2. Materials and Methods

2.1. Study site and experimental layout

The study was conducted in a vineyard cv. Godello (rootstock SO4: *Vitis berlandieri* and *Vitis riparia*) measuring 22704.3 m^2, located in the Bierzo Denomination of Origin (León, Northwest Spain; 42.606 N, 6.692W (WGS84)). The vineyard, planted in 1992 in a 1.1 m x 3.0 m pattern, is formed of bilateral cordons and vertical shoots positioned with 2 pairs of wires. Cultivation practices (weed control, fertilization, tillage, phytosanitary treatments, etc.) are the same for the whole plot.

To create the management zones, 50 vines/ha were sampled by choosing 1 line in 5 and 1 vine in 10 in each line. Each sampled vine was geo-referenced using a centimeter-precision Topcon Hiper+ GPS receiver (Topcon Corporation, Tokyo, Japan) with real-time kinematic correction. The vineyards were sampled in 2013 and 2014 in order to characterize grape composition, production and vine vigor variables.

2.2. Workflow

The methodology involved 3 main steps: (1) data collection (physical and chemical analyses of grape composition, production and vine vigor variables); (2) statistical characterization; and (3) mapping (spatial interpolation, clustering and delineation of zones) and selection of suitable cluster classifications (Fig. 1). Grape composition and production variables were measured immediately after data collection. Principal components analysis (PCA) was implemented to study the relationship between variables and to select cluster combinations. Spatial autocorrelation was depicted in a semivariogram aimed at fitting the best model to interpolated maps created for each variable and year. Spatially interpolated maps of grape composition, production and vine vigor variables were included in a cluster classification and the optimal number of zones and the most suitable cluster classification were selected.

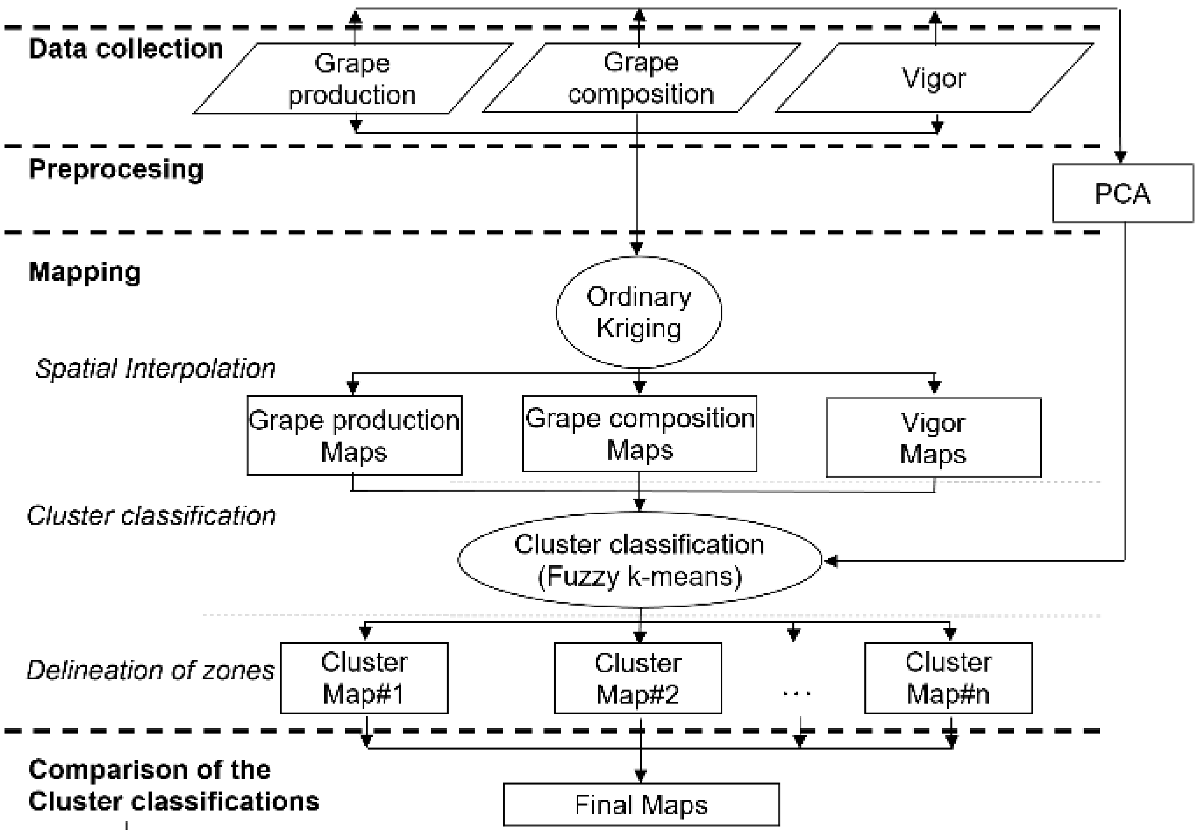

Fig. 1: Flowchart of the most suitable variables for delineating zones identified by the fuzzy k-means algorithm.

2.3. Data collection

2.3.1. Grape composition

In September of each year (2013 and 2014), 30 berries were picked from each of the 2 cordons of the selected vine, and another 30 were picked from the nearest cordon of the 2 adjacent vines (total 120 berries). Immediately after picking the berries were placed in hermetically sealed plastic bags and stored in a cooler to preserve their characteristics. They were then crushed and the juice filtered through cheesecloth to determine TSS, pH and TA.

The wet chemical analyses were performed following standard methods (European Commission Regulation (EC) No 2676/90). TSS (ºBrix) was determined at 20°C using an Atago PR1 digital refractometer (Atago Co., Tokyo, Japan); pH was measured using a Crison GLP21 electronic pH-meter (Crison Instruments, S.A., Alella, Barcelona, Spain); and TA was determined by acid-base titration using sodium hydroxide (0.1 N) to an endpoint pH of 8, with values expressed as tartaric acid (g/L). Grape maturity was defined by the maturity index (MI) as the ratio between TSS and TA (Bisson, 2001).

2.3.2. Production

Picked grapes were weighed to calculate the weight of 100 berries (BW; $kg \cdot 10^{-3}$). At harvest time, following the same criterion as applied to the sampled grapes, all the clusters were weighed and counted. Distance between the nearest cordon of 2 adjacent vines was measured, and total grape production was expressed per linear meter so as to obtain Y ($kg \cdot 10^{-3}\ m^{-1}$). The number of clusters was used to calculate CW ($kg \cdot 10^{-3}$).

2.3.3. Vigor

The vigor variables were determined during dormant periods in December in 2013 and 2014. All sampled vines were pruned. Shoots were weighed and counted in order to calculate PW ($kg \cdot 10^{-3} m^{-1}$) and WS ($kg \cdot 10^{-3}$). RI was calculated as the ratio between Y and PW (Ravaz, 1911).

2.4. Statistical preprocessing

The collected data were first analyzed to detect and delete potential outliers and then PCA without rotation was applied to the grape composition, production and vine vigor values. These statistical analyses were carried out using the software package SPSS v.21.0 (SPSS Inc., Chicago, Illinois, USA).

2.5. Mapping

2.5.1. Variogram analysis and spatial interpolation

Experimental variograms were calculated for each variable measured in 2013 and 2014 and for the mathematical average for both years. The variogram estimator used was the classical variogram (Panatier, 1996). No restrictions for distance were applied as we used a lag distance greater than the size of the work area. A variogram grid of 200 radial divisions was used and 180 angular divisions were also considered in order to allow subsequent anisotropic analysis. The experimental variograms were modeled using 2 components; the first was a nugget effect (Cressie, 1991) and the second was either a linear or spherical component (Panatier, 1996), depending on whether or not model length was greater than the maximum lag distance. Anisotropy with 2 perpendicular main directions and a tolerance of 60 degrees was considered. The parameters of the model components were fitted manually with the help of interactive graphics that considered different numbers of lags and different anisotropy directions. The experimental variograms were used to obtain continuous grids or maps for each vineyard variable. Point kriging was the interpolation method used (Isaaks and Srivastava, 1989). Variogram analyses and interpolations were performed using Surfer 11.0.642 software (Golden Software Inc, Golden, Colorado, USA).

2.5.2. Cluster and zone delineation

Mapped grape composition, production and vine vigor variables were input to the cluster classification by fuzzy k-means algorithm using Management Zone Analyst (MZA) 1.0.0 software (Agriculture Research Service, University of Missouri-Columbia, USA), which creates potential management zones using unsupervised fuzzy classification. MZA software provides concurrent output for a range of numbers of clusters, so that the user can identify the optimal number of management zones (Fridgen et al., 2004). MZA calculates statistical data, including the variance-covariance matrix values used to select the method of similarity in the zone delineation process, depending on the case: one classification variable (Euclidean), equal variances and covariances nearest to 0 (Euclidean), unequal variances and covariances nearest to 0 (Diagonal), or unequal variances and covariances different from 0 (Mahalonobis). The optimal number of management zones is defined by normalized classification entropy (NCE) and the fuzziness performance index (FPI). NCE indicates the disorganization resulting from dividing the dataset into different classes. FPI (values between 0 and 1) denotes the degree of separation between created classes, with values closer to 0 indicating greater separation between classes (Fridgen et al., 2004). In both NCE and FPI, the optimal number of zones is indicated by a minimum value.

PCA was used to plot the scores and to identify variable clusters as a criterion to select variables to be used as inputs for the MZA software. In order to optimize the selection of variables for zoning delineation, we sought the minimum number of variables that contained the maximum of information about the vineyard. In this research, the delineation zones corresponded to cluster classifications created by using (1) different input combinations for each variable for both study years, (2a) different input combinations in function of PCA groups for both study years, and (2b) different input combinations in function of PCA groups for each study year separately.

Visualizing the spatial distribution of site-specific management units resulting from the delineated zones is crucial. Moreover, given that the study was conducted in a commercial vineyard, the selected zones had to be based on both statistical results (lowest FPI and NCE) and implementation feasibility (shape of the delineation zones). The management zones defined by the MZA were thus mapped using ArcGIS v10.2 software, with the spatial location of the zones used as a criterion to compare the different zones.

3. Results

3.1. Principal components analysis

PCA was conducted on the correlation matrix produced from the 10 variables for the Godello grape variety over 2 years (Fig. 2). The PCA plots provided a visual overview of how different variables were influenced by years. Principal components (PC) 1, 2 and 3 had eigenvalues that were greater than 1.0. The first 3 PCs explained 73.88 % of the total variance (PC 1: 30.37%; PC 2: 24.29%; PC 3: 19.22 %) (Fig. 2). Furthermore, loading for PC 1 and PC 2 indicated that some variables described the same variation among samples. The underlying dimension for PC 1 was grape composition and vigor variables, with positive loading for WS (0.77), PW (0.76) and TA (0.68) but negative loading for MI (-0.79), TSS (-0.61) and RI (-0.61) on the left side of the plot. PC 2, which referred to production variables, was loaded positively for CW (0.82), Y (0.73) and BW (0.53). PC 3 was loaded positively for pH (0.65).

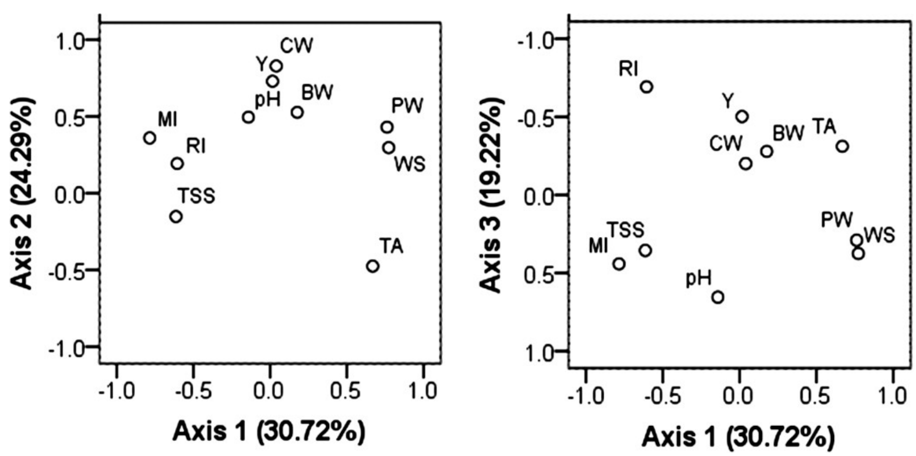

Fig. 2: Principal components extracted by factorial analysis of all the studied parameters. The main components were calculated for grape, production and vigor parameters measured in 117 vines for the 2 studied harvests (A 2013 and B 2014).

<u>Input variables:</u> Composition variables: TSS: total soluble solids (ºBrix); TA: titratable acidity ($g \cdot L^{-1}$); MI maturity index. Production variables: BW: weigh of 100 berries ($kg \cdot 10^{-3}$); CW: cluster weight ($kg \cdot 10^{-3}$); Y: yield ($kg \cdot 10^{-3} m^{-1}$). Vigor variables: PW: shoot pruning weight ($kg \cdot 10^{-3} m^{-1}$); WS: weight of shoots ($kg \cdot 10^{-3}$); RI: Ravaz index.

Since the relationships between variables had to be studied while minimizing the influence of the environment, all variables for the 2 years were included in the PCA. The above results indicate that the grape composition and vigor variables were related to PC 1, while the production variables were related to PC 2.

3.2. Fuzzy k-means

A total of 22 cluster classifications were created following the criteria cited in Materials and Methods (Table 1), 10 obtained by combining data for each variable for both study years (1), 4 obtained as different input combinations in function of PCA groups for both study years (2a), and 8 obtained as different input combinations in function of PCA groups for each study year separately (2b). Note that CC21 cluster classifications remained once pH for the 2 study years was combined, following the criterion of combining variables for both study years (1) and for combining variables detected by the PCA for both study years (2a).

Looking at the combined data for the 2 study years for each variable (1), FPI and NCE values were lowest for the cluster classification created with Y for 3 zones (CC5; FPI=0.0327 and NCE=0.0157); next lowest were those for CC2 (FPI=0.0359; NCE=0.0195) and CC3 (FPI=0.0449; NCE=0.0215). The highest FPI and NCE values were obtained for CC8 (0.0750 and 0.0372, respectively). All cluster classifications were created using the Mahalanobis method to measure similarity. The results indicate that 3 was the optimal number of zones for most of the cluster classifications created by combining data for each variable for the 2 study years.

Looking at the different input combinations in function of PCA groups for both study years (2a), CC3 obtained the lowest PFI and NCE (0.0449 and 0.0215, respectively), while CC11 obtained the highest

PFI and NCE (0.176 and 0.0948, respectively). The Mahalanobis method was again the measure of similarity used and the optimal number of zones was 3.

Table 1: Parameters of the cluster classifications (Fuzzy K-Means) used to the study

Classification type	Cluster classification	Input variables	Method of similarity	N	FPI	NCE
	CC1	TSS 13; TSS 14	Mahalanobis	3	0.0529	0.0258
	CC2	TA 13; TA 14	Mahalanobis	4	0.0359	0.0195
	CC3	pH 13; pH 14	Mahalanobis	3	0.0449	0.0215
	CC4	MI 13; MI 14	Mahalanobis	3	0.0574	0.028
1	CC5	Y 13; Y 14	Mahalanobis	3	0.0327	0.0157
	CC6	CW 13; CW 14	Mahalanobis	3	0.0488	0.024
	CC7	BW 13; BW 14	Mahalanobis	3	0.0537	0.0261
	CC8	PW 13; PW 14	Mahalanobis	3	0.0759	0.0372
	CC9	WS13; WS 14	Mahalanobis	3	0.0516	0.0254
	CC10	RI 13; RI 14	Mahalanobis	3	0.0469	0.0226
	CC11	TA 13; TA 14; WS 13; WS 14; PW 13; PW 14	Mahalanobis	3	0.176	0.0948
2a	CC12	TSS 13; TSS 14; RI 13; RI 14; MI 13; MI 14	Mahalanobis	5	0.1107	0.0735
	CC13	CW 13; CW 14; BW 13; BW 14; Y 13; Y 14	Mahalanobis	3	0.1481	0.0815
	CC3	pH 13; pH 14	Mahalanobis	3	0.0449	0.0215
	CC14	TA 13; WS 13; PW 13	Mahalanobis	3	0.0963	0.0496
	CC15	TSS 13; RI 13; MI 13	Mahalanobis	3	0.0907	0.0467
	CC16	CW 13; BW 13; Y 13	Mahalanobis	4	0.0711	0.0399
2b	CC17	pH 13	Euclidean	2	0.0199	0.0071
	CC18	TA 14; WS 14; PW 14	Mahalanobis	4	0.0863	0.049
	CC19	TSS 14; RI 14; MI 14	Mahalanobis	4	0.0665	0.0378
	CC20	CW 14; BW 14; Y 14	Mahalanobis	2	0.0729	0.0411
	CC21	pH 14	Euclidean	2	0.0308	0.0109

Type of classification: 1: using different input combinations of each variables combining the 2 years of the study. 2a: different input combinations in function of PCA groups combining the 2 years of the study. 2b different input combinations in function of PCA groups for each year of the study.

Input variables: Composition variables**:** TSS: total soluble solids (ºBrix); TA: titratable acidity (g/L); MI maturity index. Production variables: Y: yield (kg·10^{-3}m^{-1}); CW: cluster weight (kg·10^{-3}); BW: weigh of 100 berries (kg·10^{-3}); Vigor variables: PW: shoot pruning weight (kg·10^{-3}m^{-1}); WS: weight of shoots (kg·10^{-3}); RI: Ravaz index. Year: 13: sampled in 2013; 14: sampled in 2014.

N: optimal number of management zones; FPI: fuzziness performance index; NCE: normalized classification entropy.

For different input combinations in function of PCA groups for each study year separately (2b), CC17 and CC21 obtained the lowest FPI (0.0199 in 2013 and 0.0308 in 2014) and NCE (0.0071 in 2013 and 0.0109 in 2014). These were the only cluster classifications for which the Euclidean measure of similarity method was used. CC14 had the highest FPI and NCE (0.0963 and 0.0496, respectively).

In general, the lowest FPI and NCE values were obtained for CC17 and CC21, followed by combinations of CC5, CC2 and CC3. The highest FPI and NCE values were obtained for CC11 (0.176 and 0.0948, respectively). As mentioned above, 3 was detected as optimal number of management zones. Although the CC8 cluster obtained the highest FPI and NCE values (combining data for the 2 years of the study) for each variable (1), these values were lower than when the classification according to PCA groups (2a and 2b) was used, except for the clusters that used pH (CC3, CC17 and CC21).

3.3. Management zone maps

Maps were created with the optimal number of zones for each cluster classification.

(1) Input combinations for each variable for both study years

Maps created using the optimal zones based on the combinations for each variable for both study years are shown in Fig. 3. It can be observed that vigor variables showed no clear spatial distribution, as the zones were mixed within the plot. CC10 was the only cluster classification that input vigor variables with defined zones in the plot, denoting zone 2 in the center, zone 3 in the northeast and zone 2 in the northwest and southeast.

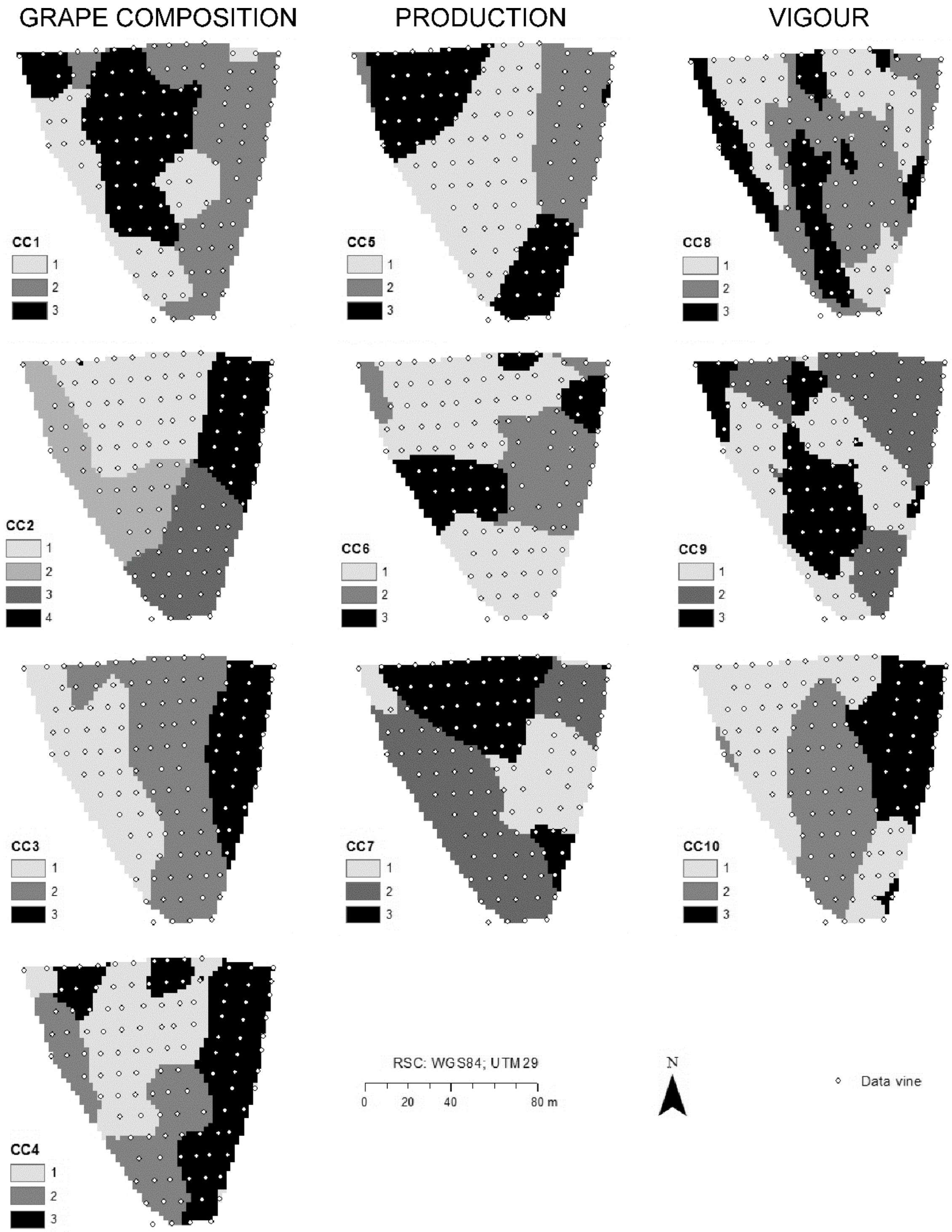

Fig. 3: Management zone maps based on using different input combinations of each variable combining the 2 years of the study.

Composition variables: CC1: Combination of total soluble solids (ºBrix) measured in 2013 and 2014; CC2: Combination of titratable acidity (g·L^{-1}) measured in 2013 and 2014; CC3 Combination of pH measured in 2013 and 2014; CC4:

Combination of maturity index measured in 2013 and 2014. Production variables: CC5: Combination of yield ($kg \cdot 10^{-3} m^{-1}$) measured in 2013 and 2014; CC6: Combination of: cluster weight ($kg \cdot 10^{-3}$) measured in 2013 and 2014; CC7: Combination of the weight of 100 berries ($kg \cdot 10^{-3}$) measured in 2013 and 2014; Vigor variables: CC8: Combination of shoot pruning weight ($kg \cdot 10^{-3} m^{-1}$) measured in 2013 and 2014; CC9: Combination of weight of shoots ($kg \cdot 10^{-3}$) measured in 2013 and 2014; CC10: Combination of Ravaz index measured in 2013 and 2014.

Combinations using the production variables point to the cluster classification created with CC5, with zone 1 located in the central part of the plot, zone 2 in the northeast and zone 3 in the parts of the plot with the lowest elevation (northwest and southeast).

Maps created using grape composition variables as input showed that combination of CC2 and CC3 were the most spatially defined zones. For CC2, 4 defined zoned were identified (zone 1 in the north-center, zone 2 in the northwest, zone 3 in the south and zone 4 in the northeast). CC3 showed 3 clear zoned distributed as zone 1 in the western part of the plot, zone 2 in the center and zone 3 in the east.

(2a) Input combinations in function of PCA groups for both study years

Maps created using the optimal zones based on combinations of variables in function of PCA groups for both study years are shown in Fig. 4. Again, CC3 was the cluster classification which best defined zones.

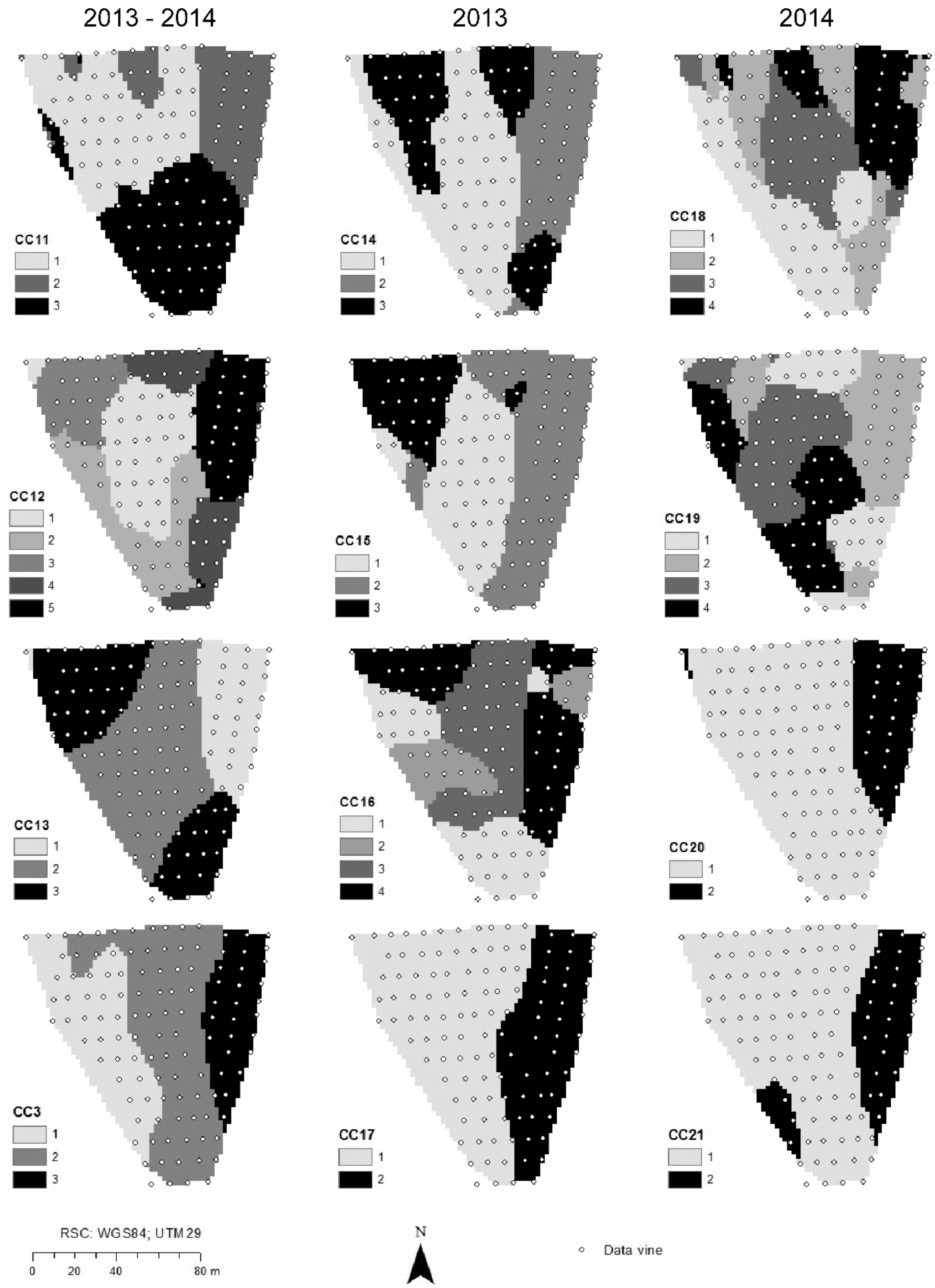

Fig. 4: Management zone maps based on using different input combinations in function of PCA groups combining the 2 years of the study (2013-2014) and for each year of the study (2013 or 2014).

<u>2013-2014:</u> CC11: Combination of titratable acidity ($g \cdot L^{-1}$), weight of shoots ($kg \cdot 10^{-3}$) and shoot pruning weight ($kg \cdot 10^{-3} m^{-1}$) measured in 2013 and 2014; CC12: Combination of total soluble solids (ºBrix), Ravaz index and maturity index measured in

2013 and 2014; CC13: Combination of yield ($kg\cdot10^{-3}m^{-1}$), cluster weight ($kg\cdot10^{-3}$) and weight of 100 berries ($kg\cdot10^{-3}$) measured in 2013 and 2014; CC3 Combination of pH measured in 2013 and 2014. 2013: CC14: Combination of titratable acidity ($g\cdot L^{-1}$), weight of shoots ($kg\cdot10^{-3}$) and shoot pruning weight ($kg\cdot10^{-3}m^{-1}$) measured in 2013; CC15: Combination of total soluble solids (ºBrix), Ravaz index and maturity index measured in 2013; CC16: Combination of yield ($kg\cdot10^{-3}m^{-1}$), cluster weight ($kg\cdot10^{-3}$) and weigh of 100 berries ($kg\cdot10^{-3}$) measured in 2013; CC17 Combination of pH measured in 2013. 2014: CC18: Combination of titratable acidity (g/L), weight of shoots ($kg\cdot10^{-3}$) and shoot pruning weight (kg·10-3m^{-1}) measured in 2014; CC19: Combination of total soluble solids (ºBrix), Ravaz index and maturity index measured in 2014; CC20: Combination of yield ($kg\cdot10^{-3}m^{-1}$), cluster weight ($kg\cdot10^{-3}$) and weigh of 100 berries ($kg\cdot10^{-3}$) measured in 2014; CC21 Combination of pH measured in 2014.

The group created with the production variables, CC13, had a very similar distribution as the group created using the combination of CC5, with zone 2 in the central part of the plot, zone 1 in the northeast and zone 3 in the lower elevation zones (northwest and southeast).

Using CC11 as input, zone 2 was located in the northeast of the plot and zone 3 in the south; zone 1 was in the north-center and northwest, but mixed with parts of zones 2 and 3.

The map created by CC12 had 5 zones mixed in the plot, so was not considered suitable for the creation of management zones.

(2b) Input combinations in function of PCA groups for each study year separately

Maps created using the optimal zones based on combinations of variables in function of PCA groups for each study year are shown in Fig. 4, which indicates that CC17 and CC21, both with 2 zones, were the cluster classifications with the best defined spatial distributions: zone 1 in the center and western part of the plot and zone 2 in the east. The same trend was evident for CC20 but the combination of these variables for 2013, that is CC16, had 4 optimal zones that were mixed within the plot.

The rest of the combinations showed no defined spatial distribution, as management zones were intermixed in the plot.

4. Discussion

The studied vineyard is currently managed as one block. However, more suitable zone-based management needs to take into account other variables, including grape characteristics. Kontoudakis et al. (2011), for instance, indicated that a low percentage of weak-quality grapes reduces the organoleptic quality of wines. To achieve the winery objectives, it is necessary to study the spatial distributions of variables based on winery specifications (Tagarakis et al., 2014). The methodology described above is a useful approach to improving management zone delineation by wineries.

4.1. Selection of suitable input variables

Obtaining grape composition, production and vigor variables, as well as requiring hard work over a short period of time, is an impractical approach if measurements have to be repeated during the growing season. To simplify fieldwork, it is desirable to define management zones with the lowest possible number of variables. We combined and studied data on each variable referring to 2 years.

However, in studies by Urretavizcaya et al. (2014) and Tagarakis et al. (2014), the differentiated management zones created by combining more than 1 variable (TSS, TA, pH, total phenolic content, malic acid concentration, tartaric acid concentration, yeast assimilable nitrogen, total anthocyanins and extractable anthocyanins in the case of the former, and TSS, TA, BW and total anthocyanin content in the case of the latter) resulted in greater vine potential than resulted from using a single variable. Keller (2015) concluded that grape characteristics were defined by berry composition and other variables were defined as production variables. The PCA was conducted to study possible relationships between variables and to select suitable combination of variables to input into the cluster classification. The PCA showed that production variables were related, as they were all located in the same group (Jolliffe, 2002). Grape composition and vigor variables were mixed in 2 groups. pH was located in an isolated group, probably due to the inverse relationship between pH and the other grape composition variables (Kodur, 2011; Saxton et al., 2009).

4.2. Cluster classifications

The 22 cluster classifications with 1 or more variables were studied and compared. Following Tagarakis et al. (2013), in order to obtain the cluster classifications we used the Euclidean measure of similarity for cluster classifications with a single variable and otherwise the Mahalanobis method. Lower FPI and NCE values were obtained using combinations of a single variable. The lowest values were obtained using pH for each year (CC17 and CC21), with 2 as the optimal number of zones. The next lowest NCE and FPI values were obtained by CC5, CC2 and CC3. The optimal number of zones was 3 for CC5 and CC3 and 4 for CC2.

CC3, which combined pH for the 2 study years, was the only PCA group with a low FPI and NCE. Although the location of management zones in the map combining the production variables, CC13, was similar to that for CC5, CC13 obtained one among the highest FPI and NCE values. Obviously, using a single value to delineate management zones makes sampling easier. So, in contrast with the conclusions of Tagarakis et al. (2014) and Urretavizcaya et al. (2014), we found that the most suitable cluster classifications were obtained using a single variable. Bearing in mind the conclusions of Arnó et al., (2012) and Urretavizcaya et al. (2014) that the most suitable zones for top quality wines had more favorable grape composition variables compared to production variables, we studied individual management zones in function of each group of variables, that is, grape composition, production and vigor. Delineating the same vineyard in more than one type of management zone in order to achieve winery objectives was proposed by Tagarakis et al. (2013), who suggested 2 different delineations in function of production yield and grape quality.

The most suitable delineations in function of grape composition were CC2 and CC3. Using pH or TA to delineate management zones may improve the quality of the wine. pH and TA determine the acidity of grapes at harvest time, which also affects wine color, microbiological stability and organoleptic characteristics (Blouin and Guimberteau, 2003). Both those cluster classifications defined management zones with low FPI and NCE values, but for an optimal number of 3 zones in CC3 and 4 zones in CC2. Studies by Taylor et al. (2003) and Arnó et al. (2011) concluded that a system based on a large number of zones may be complex to implement. Tagarakis et al. (2013) also indicated that delineation in more than 3 zones in small vineyards would be impractical. Therefore, classification into 3 zones was considered most appropriate for this study, with CC3 as the most suitable cluster classification to delineate management zones in function of grape composition. Moreover, CC3 was detected by the PCA as homogeneous and, since the zones were located longitudinally in the plot, their management may be easier.

Production variables provide an estimate of the productive potential of a vineyard. Looking at maps created with production variables as input, CC5 was found to be the most suitable, given the low FPI and NCE values and defined management zones.

Of the maps created with vigor variables, CC10 reflecting RI for 2 years obtained the lowest FPI and NCE values. However, zones 1 and 2 for CC10 were intermixed in the plot, so this delineation of management zones was not suitable in practical terms.

Bramley (2005), Bramley et al. (2011) and Cortell et al. (2005) observed a direct relationship between production and vigor variables. In this research, this relationship was reflected in the fact that the zones delineated by CC10 and CC5 were similar. Thus, both production and vigor variables for the studied vineyard could be improved by implementation of CC5 management zones, which would ensure balance in the vineyard.

5. Conclusion

We delineated vineyard management zones based on grape composition, production and vine vigor variables using the fuzzy k-means algorithm in a case study referring to the Bierzo Denomination of Origin (northwest Spain). The results indicate that cluster classifications using the same variables for 2 studied years were more accurate than those created using PCA-created groups.

On the basis of fuzziness performance index (FPI) and normalized classification entropy (NCE) values, a fuzzy k-means classification based on must pH for the 2 study years (pH 2013 and pH 2014) was considered the most suitable delineation of vineyard management zones. Combined data for titratable acidity (TA) for the 2 years also resulted in feasible zone delineations.

Considering production variables, the delineation based on yield (Y) for 2013 and 2014 resulted in 3 zones. Delineated zones based on vigor variables were not suitable for implementation in the studied vineyard.

According to our findings, the proposed protocol is a suitable method for vineyard management zoning delineation. The contribution of this research is that it takes advantage of data used by winegrowers to assess grape composition and production variables and does not require any specific additional variables for delineation.

Acknowledgments

This work was supported by the Universidad de León, Spain [grant number 2016/00145/001-T102]. The authors acknowledge the assistance of the Bodegas y Viñedos Bergidenses, SAT. support.

References

Alburquerque, M., Arranz, C., Barajas, E., Yuste, J., 2007. Modification of productive and qualitative behavior of Tempranillo variety grown with deficitary irrigation through the use of different kinds of buds on the winter pruning. Acta Hortic. 754, 183-190. doi:10.17660/ActaHortic.2007.754.23.

Arnó, J., Bordes, X., Ribes-Dasi, M., Blanco, R., Rosell, J.R., Esteve, J., 2005. Obtaining grape yield maps and analysis of within-field variability in Raimat (Spain). In J.V. Stafford (Eds.), Precision Agriculture '05, (pp. 899-906). Wageningen Academic Publishers, Wageningen. doi:10.3920/978-90-8686-549-9.

Arnó, J., Martínez-Casasnovas, J.A., Ribes-Dasi, M., Rosell, J., 2011. Clustering of grape yield maps to delineate site-specific management zones. Span. J. Agric. Res. 9, 721-729. doi: 10.5424/sjar/20110903-456-10.

Arnó, J., Rosell, J., Blanco, R., Ramos, M., Martínez-Casasnovas, J. 2012. Spatial variability in grape yield and quality influenced by soil and crop nutrition characteristics. Precis. Agric. 13, 393-410. doi: 10.1007/s11119-011-9254-1.

Bisson, L., 2001. In search of optimal grape maturity. Practical Winery & Vineyard Journal, July/August, 32-43.

Baluja, J., Tardaguila, J., Ayestaran, B., Diago, M.P., 2013. Spatial variability of grape composition in a Tempranillo (Vitis vinifera L.) vineyard over a 3-year survey. Precis. Agric. 14, 40-58. doi: 10.1007/s11119-012-9282-5.

Blouin, J. Guimberteau, G., 2003. Maduración y Madurez de la Uva. Mundiprensa, Madrid [in Spanish].

Bramley, R. 2005. Understanding variability in winegrape production systems 2. Within vineyard variation in quality over several vintages. Aust. J. Grape Wine R. 11, 33-42. doi: 10. 1111/j.1755-0238.2005.tb00277.x.

Bramley, R., Hamilton, R., 2004. Understanding variability in winegrape production systems. Aust. J. Grape Wine R. 10, 32-45. doi: 10.1111/j.1755-0238.2004.tb00006.x.

Bramley, R., Ouzman, J., Thornton, C., 2011. Selective harvesting is a feasible and profitable strategy even when grape and wine production is geared towards large fermentation volumes. Aust. J. Grape Wine R. 17, 298-305. doi:10.1111/j.1755-0238.2011.00151.x.

Brooker, P., 1986. A parametric study of robustness of kriging variance as a function of range and relative nugget effect for a spherical semivariogram. Mat. Geology 18, 477-488.

Champagnol, F., 1984. Elements de physiologie pe la vigne et de viticulture générale. Imprimerie Déhan, Montpellier.

Cortell, J.M., Halbleib, M., Gallagher, A.V., Righetti, T.L., Kennedy, J.A., 2005. Influence of vine vigor on grape (*Vitis vinifera L*. Cv. Pinot Noir) and wine proanthocyanidins. J. Agr. Food Chem. 53, 5798-5808. doi: 10.1021/jf0504770.

Cressie, N., 1991. Statistics for Spatial Data. John Wiley and Sons, New York.

Esser, A., Ortega, R., 2002. Aplicaciones de la viticultura de precisión en Chile: Estudio de casos. Revista Agronomía y Forestal UC 5, 17-21 [in Spanish].

Fridgen, J.J., Kitchen, N.R., Sudduth, K.A., Drummond, S.T., Wiebold, W.J., Fraisse, C.W., 2004, Management Zone Analyst (MZA): Software for subfield management zone delineation. Agron. J. 96, 100-108. doi: 10.2134/agronj2004.1000.

González-Fernández, A.B., Marcelo, V., Valenciano, J.B., Rodríguez-Pérez, J.R., 2012. Relationship between physical and chemical parameters for four commercial grape varieties from the Bierzo region (Spain). Acta Hortic. 147, 111–117. doi: 10.1016/j.scienta.2012.09.009.

González-Fernández, A.B., Rodríguez-Pérez, J.R., Sanz, E., Ordoñez, C., 2016. Vineyard zone delineation by cluster classification based on annual grape and vine characteristics. Precis. Agric. 1, 49. doi. 10.1007/s11119-016-9475-4.

Goovaerts, P., 1999. Geostatistics in soil science: state-of-the-art and perspectives. Geoderma 89, 1-45. doi: 10.1016/S0016-7061(98)00078-0.

Isaaks, E.H., Srivastava, R.M., 1989. An Introduction to Applied Geostatistics. Oxford University Press, New York.

Johnson, L.F., Bosch, D., Williams, D., Lobitz, B., 2001. Remote sensing of vineyard management zones: Implications for wine quality. Appl. Eng. Agric. 17, 557-560. doi: 10.13031/2013.6454.

Jolliffe, I.T., 2002. Principal Components Analysis. (2nd ed.). Springer, New York.

Keller, M., 2015. The Science of Grapevines: Anatomy and Physiology (2nd ed.). Academic Press, London.

Kodur, S., 2011. Effects of juice pH and potassium on juice and wine quality, and regulation of potassium in grapevines through rootstocks (Vitis): A short review. Vitis 50, 1-6.

Kontoudakis, N., Esteruelas, M., Fort, F., Canals, J.M., De Freitas, V., Zamora, F., 2011. Influence of the heterogeneity of grape phenolic maturity on wine composition and quality. Food Chem. 124, 767-774. doi: 10.1016/j.foodchem.2010.06.093.

Pannatier, Y., 1996. VarioWin - Software for Spatial Data Analysis in 2D. Springer-Verlag, New York.

Proffitt, T., Bramley, R., Lamb, D., Winter, E. 2006. Precision Viticulture: A New Era in Vineyard Management and Wine Production. Ashford S., Adelaide.

Ravaz, L., 1911. L'Effeuillage De La Vigne. Annales d'Ecole Nationale d'Agriculture Montpellier 11, 216-244.

Santesteban, L., Miranda, C., Jiménez, C., Fuentemilla, M., Urretavizcaya, I., Tisseyre, B., Guillaume, S., Royo, J., 2010. Evaluación del interés del índice NDVI para la delimitación de unidades de manejo diferenciado en una explotación vitícola. Revista de Teledetección 33, 11-16 [in Spanish].

Saxton, V.P., Creasy, G.L., Paterson, A.M., Trought, M.C.T., 2009. Behavioral responses of European blackbirds and Australasian silvereyes to varying acid and sugar levels in artificial grapes. Am. J. Enol. Viticult. 60, 82-86.

Tagarakis, A., Liakos, V., Fountas, S., Koundouras, S., Gemtos, T.A., 2013. Management zones delineation using fuzzy clustering techniques in grapevines. Precis. Agricul. 14, 18-39. doi: 10.1007/s11119-012-9275-4.

Tagarakis, A., Koundouras, S., Papageorgiou, E., Dikopoulou, Z., Fountas, S., Gemtos, T., 2014. A fuzzy inference system to model grape quality in vineyards. Precis. Agricul. 15, 1-24. doi: 10.1007/s11119-014-9354-9.

Taylor, J.C., Wood, G.A., Earl, R., Godwin, R.J., 2003. Soil factors and their influence on within-field crop variability, part II: spatial analysis and determination of management zones. Biosyst. Eng. 84, 441-453. doi: 10.1016/S1537-5110(03)00005-9.

Urretavizcaya, I., Santesteban, L.G., Tisseyre, B., Guillaume, S., Miranda, C., Royo, J.B., 2014. Oenological significance of vineyard management zones delineated using early grape sampling. Precis. Agricul. 15, 111-129. doi: 10.1007/s11119-013-9328-3.

Walker, R.R., Blackmore, D.H., Clingeleffer, P.R., Kerridge, G.H., Rühl, E.H., Nicholas, P.R., 2005. Shiraz berry size in relation to seed number and implications for juice and wine composition. Aust. J. Grape Wine Res. 11, 2-8. doi: 10.1111/j.1755-0238.2005.tb00273.x.

Tables

Table 1: Parameters of the cluster classifications (Fuzzy K-Means) used to the study

Classification type	Cluster classification	Input variables	Method of similarity	N	FPI	NCE
	CC1	TSS 13; TSS 14	Mahalanobis	3	0.0529	0.0258
	CC2	TA 13; TA 14	Mahalanobis	4	0.0359	0.0195
	CC3	pH 13; pH 14	Mahalanobis	3	0.0449	0.0215
	CC4	MI 13; MI 14	Mahalanobis	3	0.0574	0.028
1	CC5	Y 13; Y 14	Mahalanobis	3	0.0327	0.0157
	CC6	CW 13; CW 14	Mahalanobis	3	0.0488	0.024
	CC7	BW 13; BW 14	Mahalanobis	3	0.0537	0.0261
	CC8	PW 13; PW 14	Mahalanobis	3	0.0759	0.0372
	CC9	WS13; WS 14	Mahalanobis	3	0.0516	0.0254
	CC10	RI 13; RI 14	Mahalanobis	3	0.0469	0.0226
	CC11	TA 13; TA 14; WS 13; WS 14; PW 13; PW 14	Mahalanobis	3	0.176	0.0948
2a	CC12	TSS 13; TSS 14; RI 13; RI 14; MI 13; MI 14	Mahalanobis	5	0.1107	0.0735
	CC13	CW 13; CW 14; BW 13; BW 14; Y 13; Y 14	Mahalanobis	3	0.1481	0.0815
	CC3	pH 13; pH 14	Mahalanobis	3	0.0449	0.0215
	CC14	TA 13; WS 13; PW 13	Mahalanobis	3	0.0963	0.0496
	CC15	TSS 13; RI 13; MI 13	Mahalanobis	3	0.0907	0.0467
	CC16	CW 13; BW 13; Y 13	Mahalanobis	4	0.0711	0.0399
2b	CC17	pH 13	Euclidean	2	0.0199	0.0071
	CC18	TA 14; WS 14; PW 14	Mahalanobis	4	0.0863	0.049
	CC19	TSS 14; RI 14; MI 14	Mahalanobis	4	0.0665	0.0378
	CC20	CW 14; BW 14; Y 14	Mahalanobis	2	0.0729	0.0411
	CC21	pH 14	Euclidean	2	0.0308	0.0109

Type of classification: 1: using different input combinations of each variables combining the 2 years of the study. 2a: different input combinations in function of PCA groups combining the 2 years of the study. 2b different input combinations in function of PCA groups for each year of the study.

Input variables: Composition variables**:** TSS: total soluble solids (ºBrix); TA: titratable acidity (g/L); MI maturity index. Production variables: Y: yield ($kg \cdot 10^{-3} m^{-1}$); CW: cluster weight ($kg \cdot 10^{-3}$); BW: weigh of 100 berries ($kg \cdot 10^{-3}$); Vigor variables: PW: shoot pruning weight ($kg \cdot 10^{-3} m^{-1}$); WS: weight of shoots ($kg \cdot 10^{-3}$); RI: Ravaz index. Year: 13: sampled in 2013; 14: sampled in 2014.

N: optimal number of management zones; FPI: fuzziness performance index; NCE: normalized classification entropy.

Figure 1

Click here to download high resolution image

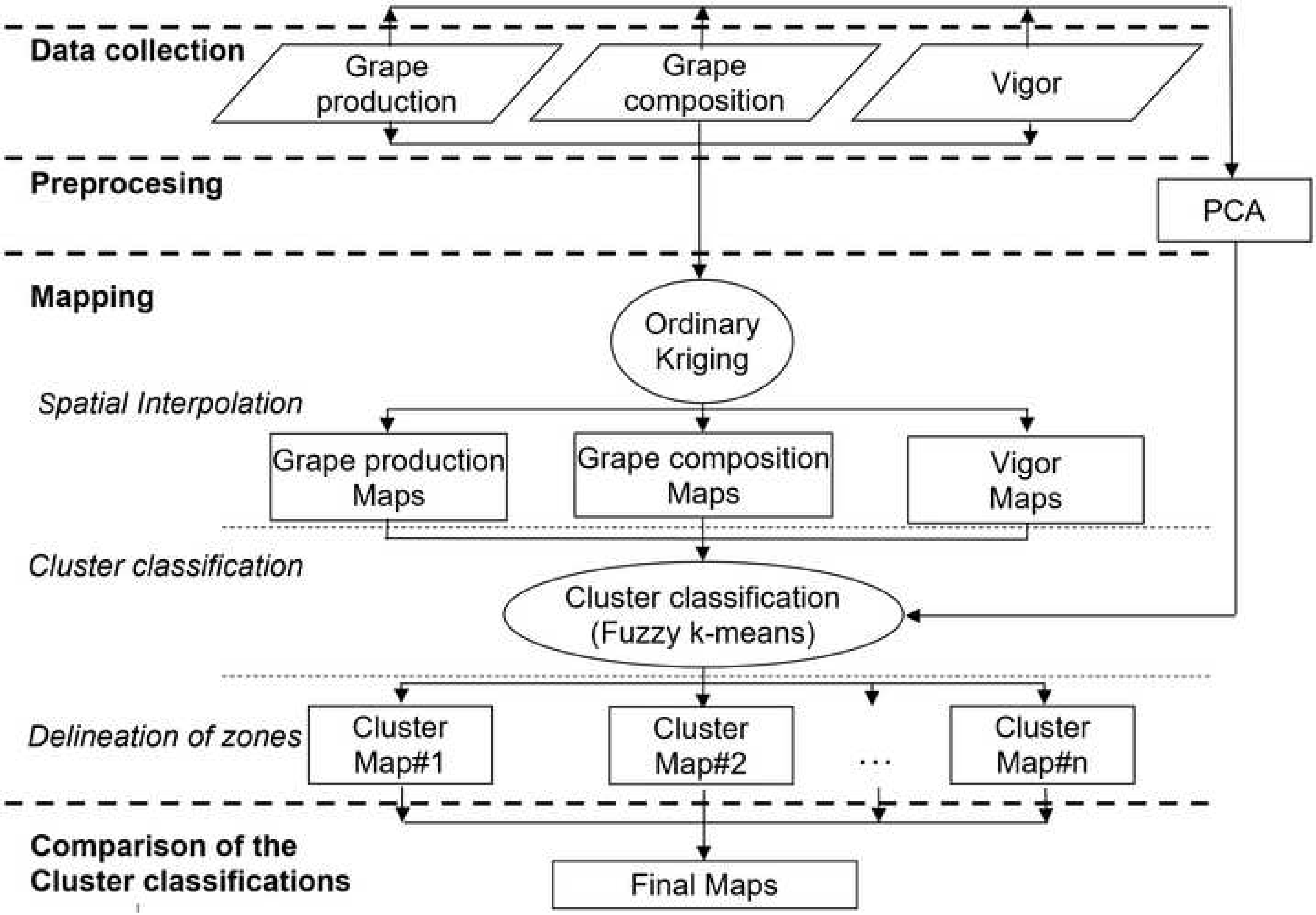

Figure 2

Click here to download high resolution image

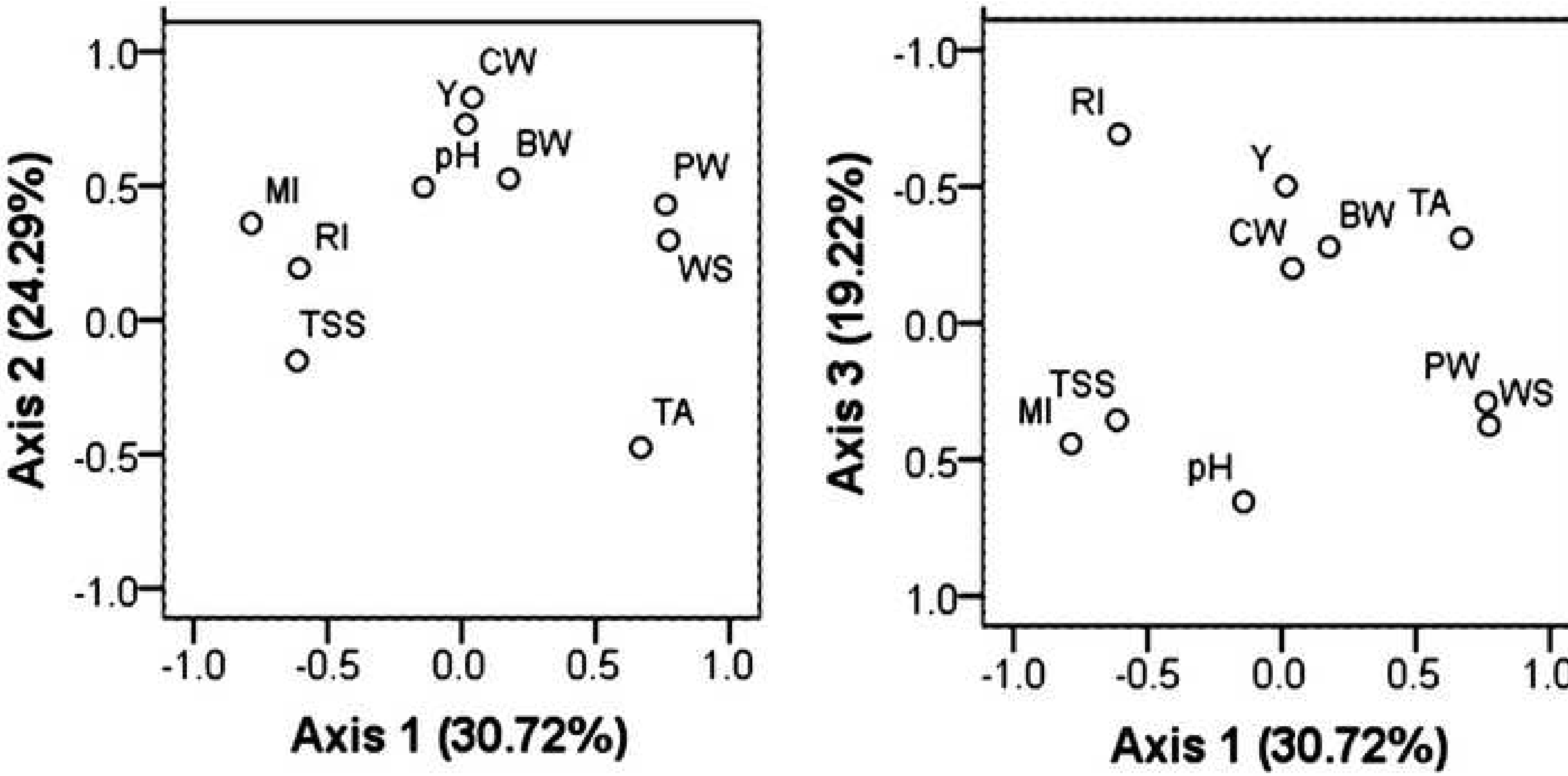

Figure 3
Click here to download high resolution image

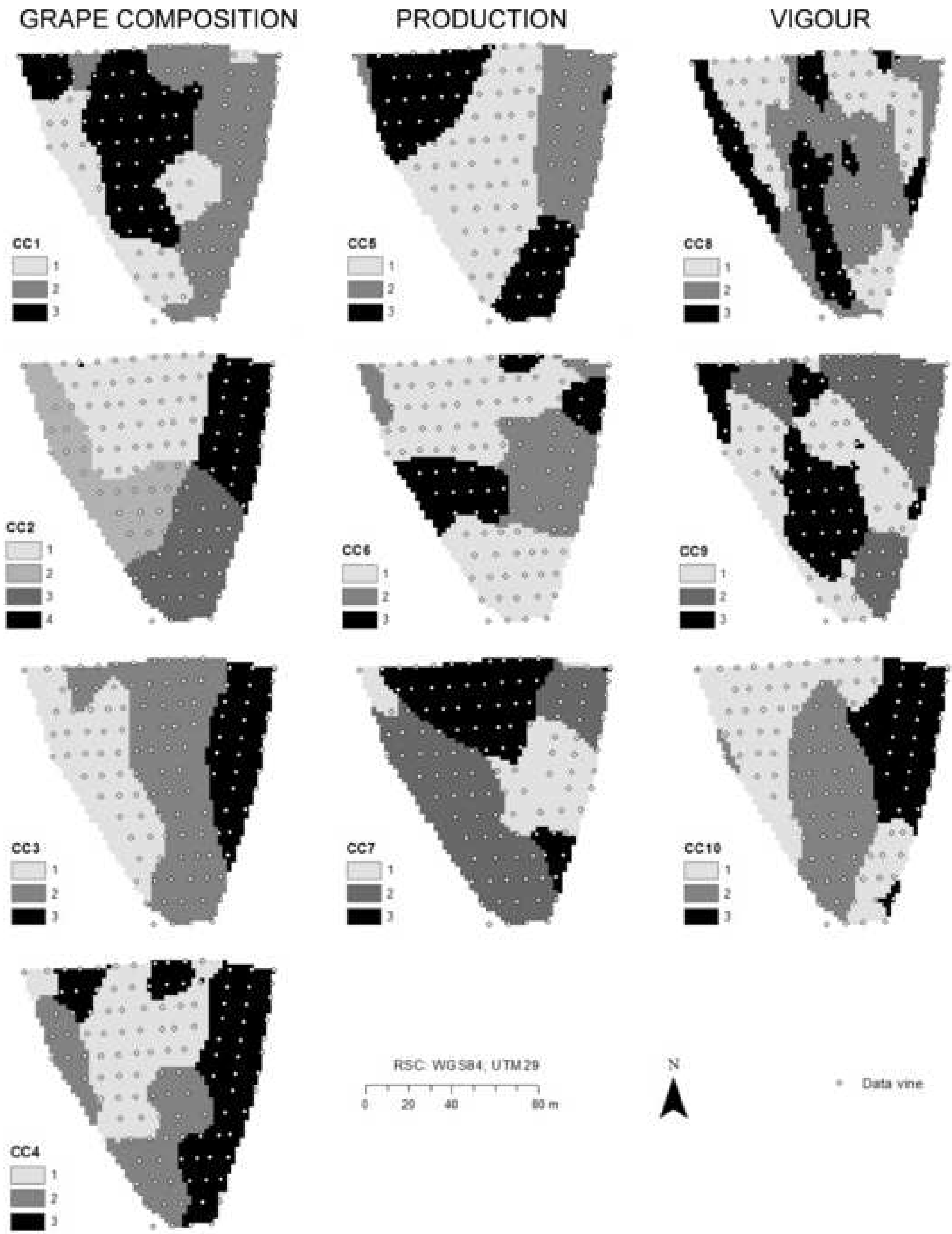

Figure 4

Click here to download high resolution image