Pixel clustering in spatial data mining; an example study with Kumeu wine region in New Zealand

S Shanmuganathan[a] and J Whalley[a,b]

[a]Geoinformatics Research Centre (GRC), [b]School of Computing and Mathematical Sciences
Auckland University of Technology (AUT), New Zealand
Email: subana.shanmuganathan@aut.ac.nz

Abstract: This paper describes an approach to pixel clustering using self-organising map (SOM) techniques in order to identify environmental factors that influence grape quality. The study area is the Kumeu grape wine region of northern New Zealand (NZ). SOM methods first introduced by Kohonen in the late 1980s, are based on two layered feed forward artificial neural networks (ANNs) with an unsupervised training algorithm. They are useful in projecting multidimensional input data onto low dimensional displays while preserving the intrinsic properties in the raw data by which the detection of previously unknown knowledge in the form of patterns, structures and relationships is enhanced.
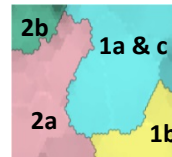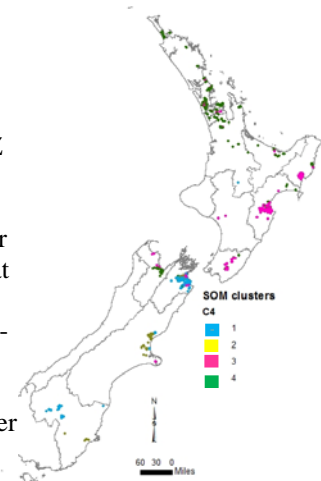
Figure 1a: SOM and NZ maps showing the SOM clusters of 437,888 NZ vineyard pixels, the major distinguishing attributes at this macro scale are: elevation, annual average- and minimum-temperatures, induration, drainage and monthly water ratio balance (figure 1b)

In modern day viticultural zoning approaches, factors that contribute to grape quality are typically categorised into three classes; *terrior* (climate, soil type, topography of a location), *cultiva* (the variety of the vine) and dependent factors such as berry quality indicators (e.g.: Brix and pH) and wine quality/market price.

Many modern viticulturists rely on expert knowledge and intuition to establish viticultural zones in conjunction with Geographic Information Systems (GIS) to further subdivide a wine region and vineyards into zones. The most common scale for such zoning has been the "*meso*" scale and the factors used for the characterisation of vineyards, varies extensively. The most adopted factors used for zoning are grapevine growth phenology (growing degree days (GDD), frost days/timing, berry ripening temperature range) for which comprehensive knowledge on local viticulture and wine quality is essential. Hence, for characterising vineyards from the *new world* or wine regions with insufficient knowledge for zoning is considered as a challenging task. For such instances, the SOM approach discussed in this paper provides a means to resolving a lack of extensive historical knowledge especially, when establishing zoning systems. The case study presented demonstrates the advantages of the SOM approach to identifying the ideal discerning attributes for zoning between and within vineyard/s using available geocoded digital data. The results of the SOM based clustering and data mining approach show that water deficit, elevation (along with hill shade and aspect) and annual average/minimum temperatures, are the main contributory factors for zoning vineyards in the Kumeu wine region at the *meso* scale. Interestingly, the elevation, annual average- and minimum-temperatures, induration, drainage and monthly water ratio balance are found to be the discerning factors at the macro conforming some of the currently used factors in NZ.

Cluster	pixel count	Ele vation	Ave Temp	A min Temp	A sol Radiation	Indu ratin	Exch Catio n	Acid sol P	Che limitat on	Age	Slop e	Drai nag e	Wat BR	Water deficit
1a&c	177191	**128.59**	**12.04**	**1.57**	14.92	**3.11**	1.97	3.79	1.00	**1.87**	0.06	**4.34**	1.62	**219.95**
1b	93607	**62.37**	**11.62**	**1.09**	14.07	**3.31**	2.01	3.86	1.00	**1.16**	0.03	**4.88**	1.70	**208.26**
2a	127694	**36.85**	**13.35**	**3.20**	14.72	**1.23**	**2.21**	2.46	1.07	**1.37**	0.04	3.28	**1.76**	**179.55**
2b	39396	**93.84**	**13.74**	**4.59**	14.89	**2.28**	**1.42**	1.62	0.94	**1.71**	0.06	3.74	**2.67**	**54.10**
Total	**437888**													

Figure 1b: SOM cluster profiles, WatBR: monthly water balance ratio.

Keywords: *Self-organising map clustering, viticulture, infield variability*

1.   INTRODUCTION

The traditional approaches that are still in use for characterising (or zoning) the wine regions using either simple or complex indices, were originally developed based on extensive knowledge relating to viticulture and wine quality gained over decades and in some cases centuries of wine making, such as "*terroir x cultiva[1]*" (Shanmuganathan, 2010). This makes the zoning of vineyards from the *new world* or new wine regions a challenging task.

Over the recent years the collective analysis of spatial and aspatial attribute data from disparate sources, incorporated into a GIS has become a useful and popular approach to studying the spatial patterns, such as correlations and trends within multi-sourced data sets in many application domains. For example, historic census (Chi & Zhu, 2008), healthcare (Bissonnette, Wilson, Bel, & Shah, 2012) (Wei, Tedders, & Tian, 2012) and socio-economics (Xiaonian, Yi, Zhang, & Liu, 2011) are some of the domains that have used this approach. This research clearly demonstrates the usefulness of such an approach when developing an understanding of issues involving multiple complex factors in a spatial context.

This paper outlines the main approaches to the integrated analysis of multiple attribute data in a spatial context using GIS. Consequently, the approach investigated is described in detail. The results of a new approach, using SOM and TDIDT (Top-Down Induction of Decision Tree) techniques to identify meaningful independent factors for zoning a wine region at the *meso* scale are presented. The results show that the approach can be applied successfully to analyse spatial and aspatial attributes describing a land area at different levels of detail, especially in less known problem domains. This approach also has wider implications in that it can be applied to temporal change of the attributes as clustered zones to understand the change and its effects, for example, the effects of potential climate change for decision making that would otherwise require expensive high resolution satellite/aerial imagery and analysis.

2.   SPATIAL AND NON-SPATIAL ATTRIBUTE DATA ANALYSIS

Both simple and complex spatial data analysis methods are efficient and useful when there is sufficient knowledge in the problem domain. The most commonly used methods that are applied to the analysis of integrated spatial attribute data can be grouped into four main categories, namely; retrieval/classification/measurement, overlay, neighbourhood and connectivity of network functions.

In addition, topographic functions, i.e., spatial attributes, can be computed, from elevation information usually in raster format, either as a digital elevation model (DEM) or a digital terrain model (DTM). Using the eight orthogonal (O) and diagonal (D) neighbours of a cell, spatial attributes such as slope, aspect, and topographic position (ridge, valley, and knoll), of a given land area can be ascertained. These topographic parameters are often highly correlated with the distribution of plant and animal species hence, are frequently used in remote sensing applications to distinguish spectrally similar habitats. For example, it is often spectrally difficult to separate coastal dunes and sandy flats but they can be separated using the slope and the topographic position of the two similar habitats ().

3.   CLUSTERING IN SPATIAL DATA MINING

Increasingly, new algorithms are being investigated for clustering spatial data aimed at improving the efficiency of the clustering process (Chauhan, Kaur, & Alam, 2010). Recently there has been considerable research reported that attempts to refine certain aspects of the clustering, aspects such as; improving cluster quality in large volumes of high dimensional data sets (Qian & Zhang, 2004), noise removal (Ester, Kriegel, Jörg S,, & Xu , 1996), uncertainty (Li, Shi, & Liu, 2010), data pre-processing and reduction of running time consumed for clustering (Qian and Zhang 2004).

4.   THE METHODOLOGY

Modern GIS have functions that enable the integration, manipulation, visualisation and analysis of geo-coded data. They enable analysts to pre-process digital map layers that consist of attribute data on various landscape features, observations and measurements.

---

[1] *"Terroir"* is a concept, has been recently defined as an interactive ecosystem, in a given place, including its climate, soil. The vine is the "*cultivar*". The term is frequently used to explain the hierarchy of high-quality wines. It relates the sensory attributes of a wine to the environmental conditions in which the grapes were grown (Leeuwen & Seguin, 2006, Journal of Wine Research, 2006, Vol. 17, No.1,10)
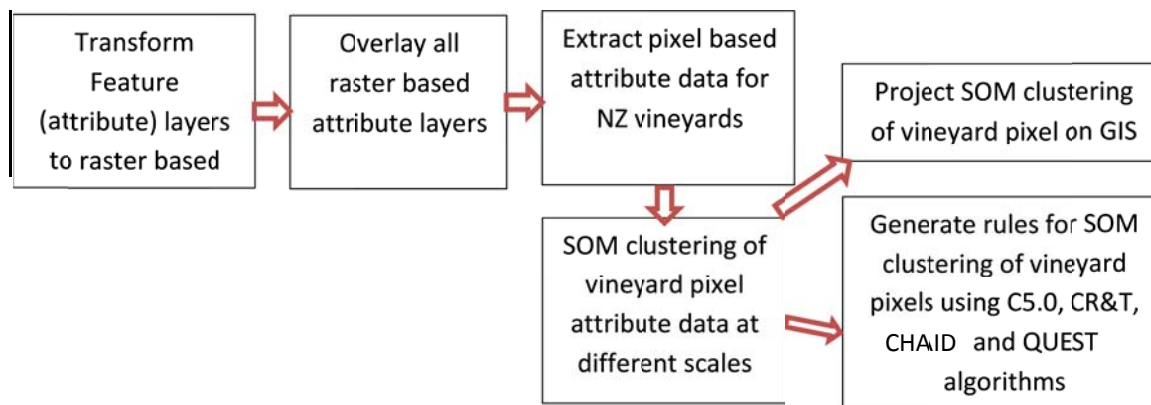
Figure 2. The processes used for data pre-processing and SOM clustering

Table 1. *Terrior* attributes used as features for the pixel clustering to identify zones in the Kumeu vineyards

Climate variables	Land form variables	Soil variables
1. Mean annual temperature: strongly influences plant productivity. 2. Mean minimum winter Temperature: influences plant survival. 3. Mean annual solar radiation: determines potential productivity. 4. Monthly water balance ratio: indicates average site "wetness". 5. Annual water deficit: gives an indication of soil dryness, it is calculated using mean of daily temperature, daily solar radiation and rainfall (Leathwick, Morgan, Wilson, Rutledge, McLeod, & Johnston, 2002)	1. Elevation 2. Slope: Major driver of drainage, soil rejuvenation and microclimate 3. Aspect 4. Hill shade	1. Drainage: influences the oxygen availability in upper soil layers. 2. Acid soluble phosphorous: indicates a key soil nutrient 3. Exchange calcium: both a nutrient and a determinant of soil weathering. 4. Induration (hardness): determines soil resistance to weathering. 5. Age: separates recent, fertile soils from older less fertile soils. 6. Chemical limitation of plant growth: indicates the presence of salinity of ultramafic substances.

In New Zealand, map layers and scientific datasets (e.g.: soil variables, landforms and climate variables that are considered as contributory to the classification of *Terrior* into viticultural zones, are freely available from the Landcare Research's Land Resource Information Systems (LRIS) Portal (http://lris.scinfo.org.nz/). The polygon map layers obtained from LRIS and employed in this study are detailed in Table 1. These layers were pre-processed, using procedures available in ArcGIS10.1 (www.esri.com), in order to transform the dataset into an appropriate format for clustering (Figure 2). The first pre-processing step transformed the polygon data sets into raster format and in the second they were projected into one co-ordinate system.

Finally, point attribute data for the pixels relating to NZ vineyards was extracted from all the raster layers into one table. The SOM clustering was then performed on a data set of 7,858 pixels (Figure 1a) relating to Kumeu wine region vineyards alone. This Kume pixel data is a subset of the original 437,888 points generated relating to all the wine growing regions of New Zealand (Figure 1b). Viscovery (www.viscovery.net), a commercial software package, was used to perform the SOM clustering and rule induction was performed using TDIDT algorithms (RT, CHAID and QUEST) available in SPSS Clementine (http://www.spss.com/clementine/).

5. SOM CLUSTERING AND RULES GENERATED AT THE MESO SCALE

When establishing the clusters the number of clusters was progressively increased from 2 clusters to 18 clusters in order to study the clustering and the cluster profiles in the NZ wide research. NZ maps were over laid with these clusters to visualise the spatial distribution of the clusters (figures 1a and b show 4 clusters generated using SOM).

In the research, the data subset relating to the Kumeu wine region alone is studied to study the use of the approach at this scale. The ten SOM clusters generated for the Kumeu vineyards and the pixels and their profiles (Figures 3-6) show that the variability can be observed in the values of attributes among and even within vineyards that could help in the vineyard management decision making relating to selective spraying /harvesting. However no significant variation was observed in the slope and chemical limitation of plant growth within these vineyards. This may be due to the fact that the slope resolution (50M) used was not sufficient enough in details for clustering. The other useful observations made from the SOM are:

- Annual solar radiation, annual average and minimum temperatures, acid soluble phosphorous, drainage, elevation, cation exchange, induration, monthly water balance and annual water deficit show similarity in corresponding high and low areas in the clustering and can be used for zoning of the vineyards.
- Aspect and hill shade show variability that can be used for zoning purposes.
- Age (soil) has one cluster that is 1 year (new fertile) and rest of the clusters are 2 years old (less fertile).
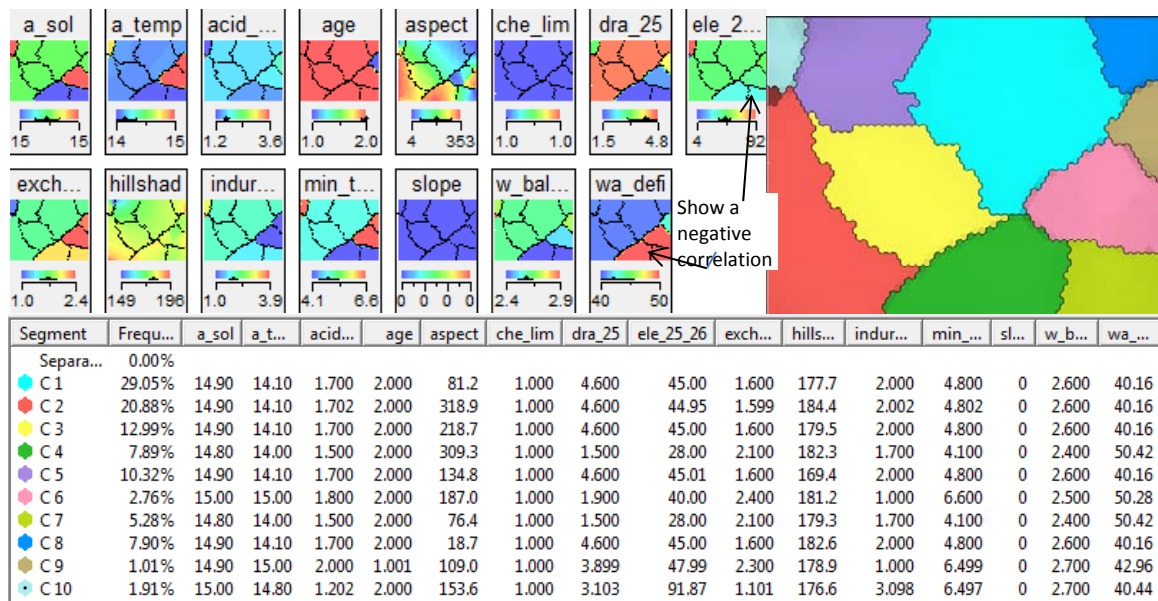

Segment	Frequ...	a_sol	a_t...	acid...	age	aspect	che_lim	dra_25	ele_25_26	exch...	hills...	indur...	min_...	sl...	w_b...	wa_...
Separa...	0.00%															
C 1	29.05%	14.90	14.10	1.700	2.000	81.2	1.000	4.600	45.00	1.600	177.7	2.000	4.800	0	2.600	40.16
C 2	20.88%	14.90	14.10	1.702	2.000	318.9	1.000	4.600	44.95	1.599	184.4	2.002	4.802	0	2.600	40.16
C 3	12.99%	14.90	14.10	1.700	2.000	218.7	1.000	4.600	45.00	1.600	179.5	2.000	4.800	0	2.600	40.16
C 4	7.89%	14.80	14.00	1.500	2.000	309.3	1.000	1.500	28.00	2.100	182.3	1.700	4.100	0	2.400	50.42
C 5	10.32%	14.90	14.10	1.700	2.000	134.8	1.000	4.600	45.01	1.600	169.4	2.000	4.800	0	2.600	40.16
C 6	2.76%	15.00	15.00	1.800	2.000	187.0	1.000	1.900	40.00	2.400	181.2	1.000	6.600	0	2.500	50.28
C 7	5.28%	14.80	14.00	1.500	2.000	76.4	1.000	1.500	28.00	2.100	179.3	1.700	4.100	0	2.400	50.42
C 8	7.90%	14.90	14.10	1.700	2.000	18.7	1.000	4.600	45.00	1.600	182.6	2.000	4.800	0	2.600	40.16
C 9	1.01%	14.90	15.00	2.000	1.001	109.0	1.000	3.899	47.99	2.300	178.9	1.000	6.499	0	2.700	42.96
C 10	1.91%	15.00	14.80	1.202	2.000	153.6	1.000	3.103	91.87	1.101	176.6	3.098	6.497	0	2.700	40.44

Figure 3. SOM map (top left) created with 7,858 Kumeu sub region pixels alone, SOM components (top right) and SOM cluster profiles (bottom) show the patterns in factors used in the pixel clustering. Aspect and hill shade vary in a similar manner and throughout the SOM and they both are related to elevation.

Of the vineyard attributes analysed (table 1), water deficit and elevation (along with aspect and hill shade) were found to be the main contributing factors to the variability observed a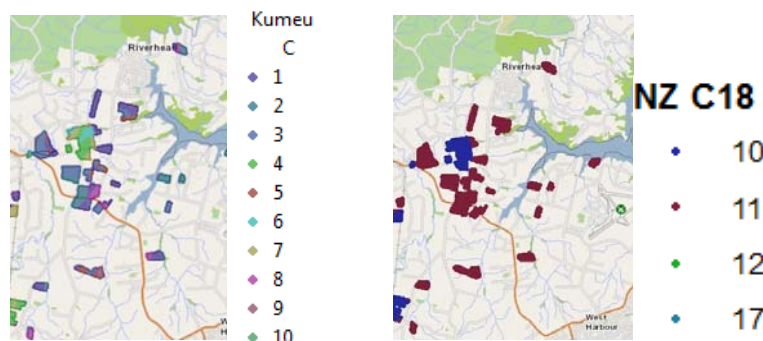mong and within vineyards in the Kumeu wine region (figure 3). These two factors are negatively correlated to each other, the higher the elevation the lower the water deficit. Meanwhile, the C5 and CRT rules (in figures 5 and 6) generated using SOM cluster as classes show the conditions and patterns relating to the SOM clusters (similarities and the dissimilarities between potential viticulture zones in Kumeu vineyards).

Figure 4. The geographical distribution of the ten SOM clusters (left) of figure 2 and (right) SOM clusters 10, 11, 12 and 17 of the 18SOM created using the original 437,888 points generated for all vineyards in New Zealand. Both the ten Kumeu only and 18 all NZ pixel clustering show the variability even within vineyards however, in the former, the pixels give more zones within the Kumeu vineyards.

Rule no	Instance; confidence	Rule asp; aspect, hs; hill shade, wd; water deficit, ele; elevation 25m resolution	SOM cluster
1	46; 1.0	if wd<= 40.16 and asp <= 106.56 and > 29.15 and hs <= 173 and > 172	one
2	309; 1.0	if wd<= 40.16 and asp <= 136.38 and > 29.15 and hs <= 175 and > 173	one
3	5; 1.0	if wd<= 40.16 and asp <= 145.11 and > 136.38 and hs <= 174 and > 173	one
4	1,916; 1.0	if wd<= 40.16 and asp <= 151.34 and > 29.15 and hs <= 180 and > 175	one
5	4; 1.0	if wd<= 40.16 and asp <= 156.37 and > 151.34 and hs <= 176 and > 175	one
6	2; 1.0	if wd<= 40.16 and asp <= 154.45 and > 151.34 and ele_25 in [45] and hs <= 177 and > 176	one
1	2; 1.0	if wd<= 40.16 and asp <= 277.27 and > 151.34 and ele_25 in [0] and hs <= 182 and > 176	two
2	7; 1.0	if wd<= 40.16 and asp <= 277.27 and > 264.29 and ele_25 in [45] and hs <= 182 and > 181	two
3	126; 1.0	if wd<= 40.16 and asp <= 277.27 and > 176 and hs > 182	two
4	63; 0.984	if wd<= 40.16 and asp <= 284.39 and > 277.27 and hs > 180	two
5	1,425; 1.0	if wd<= 40.16 and asp > 284.39	two
1	13; 0.923	if wd<= 40.16 and asp <= 154.45 and > 151.34 and ele_25 in [45] and hs <= 181 and > 177	three
2	958; 1.0	if wd<= 40.16 and asp <= 277.27 and > 154.45 and ele_25 in [45] and hs <= 181 and > 176	three
3	39; 1.0	if wd<= 40.16 and asp <= 264.29 and > 151.34 and ele_25 in [45] and hs <= 182 and > 181	three
4	8; 1.0	if wd<= 40.16 and asp <= 284.39 and > 277.27 and hs <= 180	three
1	**620; 1.0**	**if wd> 40.16 and asp > 190.35 and min_temp <= 4.8**	**four**
1	484; 1.0	if wd<= 40.16 and asp <= 151.34 and hs <= 172	five
2	22; 1.0	if wd<= 40.16 and asp <= 136.38 and > 106.56 and hs <= 173 and > 172	five
3	11; 1.0	if wd<= 40.16 and asp <= 151.34 and > 136.38 and hs <= 173 and > 172	five
4	9; 1.0	if wd<= 40.16 and asp <= 151.34 and > 145.11 and hs <= 174 > 173	five
5	10; 1.0	if wd<= 40.16 and asp <= 151.34 and > 136.38 and hs <= 175 and > 174	five
6	22; 1.0	if wd<= 40.16 and asp <= 156.37 and > 151.34 and hs <= 175	five
7	253; 1.0	if wd<= 40.16 and asp <= 277.27 and > 156.37 and hs <= 176	five
1	**217; 1.0**	**if wd> 40.16 and min_temp > 6.5**	**six**
1	**415; 1.0**	**if wd> 40.16 and asp <= 190.35 and min_temp <= 4.8**	**seven**
1	52; 1.0	if wd<= 40.16 and asp <= 29.15 and hs <= 180 and hs > 172	eight
2	591; 1.0	if wd<= 40.16 and asp <= 151.34 and hs > 180	eight
1	**79; 1.0**	**if wd> 40.16 and a_temp > 14.8 and min_temp <= 6.5 and > 4.8**	**nine**
1	**150; 1.0**	**if wd> 40.16 and a_temp <= 14.8 and min_temp <= 6.5 and > 4.8**	**Ten**

Figure 5. C5.0 tree rules created with 7,858 Kumeu pixels alone, water deficit (wd> or <= 40.16) is seen as the major discerning attribute then followed by aspect (asp), hill shade / elevation or both.

Rule No	Instances; confidence	Rule asp; aspect, hs; hill shade, wd; water deficit, ele; elevation 25m resolution	SOM Cluster
1	2,383; 0.956	if asp <= 151.99 and hs <= 180.5 and > 172.5 and wd<= 40.3	one
1	88; 1.0	if asp > 151.99 and <= 268.825 and ele_25 in [45] and hs > 176.5 and > 182.5	two
2	1,579; 0.97	if asp > 151.99 and > 268.825 and dra_25 > 4.25	two
1	973; 0.997	if asp > 151.99 and <= 268.825 and ele in [45] and hs <= 182.5 and > 176.5	three
1	186; 0.323	if asp > 151.99 and <= 268.825 and ele in [0 28 40 48 92] and hs > 176.5	four
2	560; 1.0	if asp > 151.99 and > 268.825 and drainage <= 4.25 and ele_25 in [28]	four
1	505; 0.958	if asp <= 151.99 and hs <= 172.5 and <= 180.5	five
2	283; 0.968	if asp > 151.99 and <= 268.825 and hs <= 176.5	five
1	119; 0.824	if asp > 151.99 and > 268.825 and drainage <= 4.25 and ele_25 in [40 48 92]	six
1	269; 1.0	if asp <= 151.99 and ele in [28] and hs <= 180.5 and > 172.5 and wd> 40.3	seven
2	174; 0.569	if asp <= 151.99 and ele in [28 40 48 92] and hs > 180.5	seven
1	591; 1.0	if asp <= 151.99 and ele in [45] and hs > 180.5	eight
1	148; 0.419	if asp <= 151.99 and ele in [40 48 92] and hs <= 180.5 and > 172.5 and wd> 40.3	nine

Figure 6: CRT tree rules created with Kumeu pixels alone show aspect (asp>/< 151.99) as major discerning factor followed by hill shade/elevation and then water deficit > 40.3. Drainage has been used in 2 rules.

asp; aspect, hs; hill shade, wd; water deficit, ele; elevation 25m resolution

```
ele = 0 or ele = 45 [ Mode: one ] (6,377)                          sp <= 0.06 [ Mode: three ] => three (156; 0.987)
  asp <= 31.7900 [ Mode: eight ] (585)                             sp > 0.06 [ Mode: three ] => three (81; 0.815)
      hs <= 180 [ Mode: eight ] => eight (63; 0.825)             hs > 178 [ Mode: three ] => three (105; 0.81)
      hs > 180 [ Mode: eight ] => eight (522; 1.0)         asp > 203.9300 and asp <= 262.1000 [ Mode: three ] (686
  asp > 31.7900 and asp <= 57.9900 [ Mode: one ] (684)          hs <= 182 [ Mode: three ] (633)
      hs <= 180 [ Mode: one ] => one (616; 1.0)                  a_temp <= 14.1 [ Mode: three ] (632)
      hs > 180 [ Mode: eight ] => eight (68; 1.0)                  hs <= 176 [ Mode: five ] => five (13; 1.0)
  asp > 57.9900 and asp <= 84.9900 [ Mode: one ] (715)             hs > 176 [ Mode: three ] => three (619; 1.0)
      hs <= 180 [ Mode: one ] (714)                              a_temp > 14.1 [ Mode: two ] => two (1; 1.0)
        hs <= 173 [ Mode: five ] => five (50; 0.7)             hs > 182 [ Mode: two ] => two (53; 1.0)
        hs > 173 [ Mode: one ] => one (664; 1.0)          asp > 262.1000 and asp <= 307.7300 [ Mode: two ] (553)
      hs > 180 [ Mode: eight ] => eight (1; 1.0)                 hs <= 181 [ Mode: three ] => three (154; 0.604)
  asp > 84.9900 and asp <= 107.4800 [ Mode: one ] (680)          hs > 181 [ Mode: two ] => two (399; 0.997)
      hs <= 173 [ Mode: five ] => five (209; 0.852)       asp > 307.7300 [ Mode: two ] => two (1,109; 1.0)
      hs > 173 [ Mode: one ] => one (471; 1.0)            ele = 28 [ Mode: four ] (1,035)
  asp > 107.4800 and asp <= 139.5900 [ Mode: one ] (675)        asp <= 203.9300 [ Mode: seven ] (418)
      hs <= 173 [ Mode: five ] => five (255; 1.0)                  asp <= 139.5900 [ Mode: seven ] => seven (353; 1.0)
      hs > 173 and hs <= 176 [ Mode: one ] => one (177; 0.977)    asp > 139.5900 [ Mode: seven ] => seven (65; 0.954)
      hs > 176 [ Mode: one ] => one (243; 1.0)                  asp > 203.9300 [ Mode: four ] => four (617; 1.0)
  asp > 139.5900 and asp <= 203.9300 [ Mode: five ] (690)    ele = 40 [ Mode: six ] => six (217; 1.0)
      hs <= 173 [ Mode: five ] => five (198; 1.0)           ele = 48 [ Mode: nine ] => nine (79; 1.0)
      hs > 173 and hs <= 176 [ Mode: five ] => five (150; 0.853)  ele = 92 [ Mode: ten ] => ten (150; 1.0)
      hs > 176 and hs <= 178 [ Mode: three ] (237)
```

Figure 7. Based on CHAID tree and rules (created with 7,858 Kumeu pixels alone) elevation is split into 5 classes (=0/=40, =28, =40, =48 and =92) as CHAID algorithm is a multi-node decision tree. Aspect and hill shade as well are used in the rules. In addition, for clusters three, five and two annual average temperature is used (in *italics*). SOM clusters six, nine and ten are defined purely on elevation with 217, 79 with 150 instances respectively all at 100% confidence. Clusters seven and four vary in elevation and aspect.
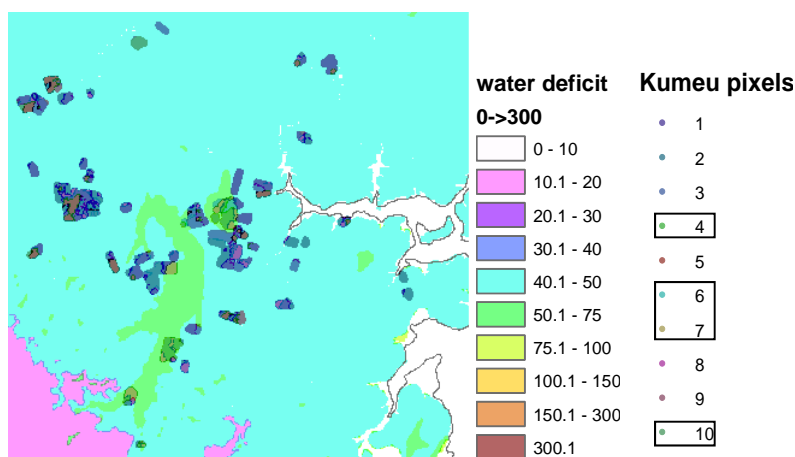

Figure 8. SOM clustering displayed over water deficit map of Kumeu sub region shows the pixels that are in the > or < than 40.16 water deficit areas, the clusters > 40.16 being 4, 6, 7 and 10.

In addition, annual average and minimum temperatures also show some variability in the CAHID and QUEST trees and rules (figures 8 and 9) even though the resolution of the two attributes are not sufficient enough for the meso scale characterisation by other methods.

6. CONCLUSIONS

Traditional approaches to characterising/ zoning land areas of interest using spatial thematic digital mappings requires extensive knowl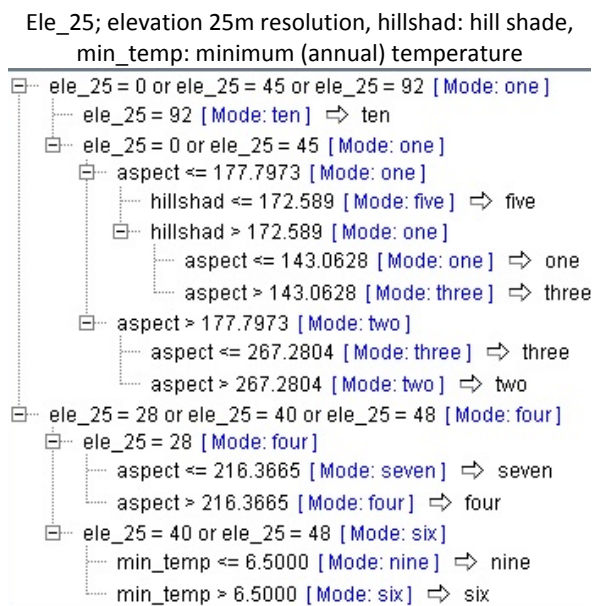edge of local environmental and crop related factors. This makes zoning practically impossible for areas where extensive knowledge does not exist. The SOM based clustering and TDIDT data mining approach gives a useful means to identifying the contributory attributes and areas for potential zones in new "*terroirs*". For the Kumeu wine region, it has been shown that water deficit, elevation (along with aspect and hill shade) as well as, to a lesser extent, annual minimum and average temperatures seem to be contributing to the variability at the *meso*-scale. This is interesting because in New Zealand at the regional/ macro-scale GDD, annual average and minimum temperatures are still used as the major deterministic factors when choosing a grape vine variety for planting (Shanmuganathan, 2010).

Ele_25; elevation 25m resolution, hillshad: hill shade,
min_temp: minimum (annual) temperature

Figure 9. Quest tree rules with elevation split into two main modes (=0, =45 /= 92) and (=28, =40/ =48) and then further into two classes each. All elevation classes are then divided into binary nodes based on aspect, hill shade and min temperature <= 6.5/6.5.
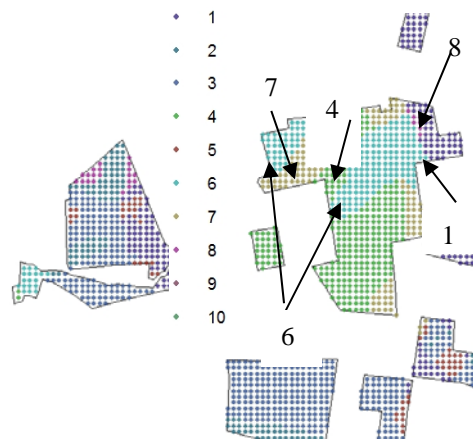

Figure 10. SOM clustering of Kumeu pixels showing the variability within and between vineyards. The Vineyard in the top right mainly consists of SOM clusters 4, 6, and 7 at elevations 28, 40 and 28 respectively and all with water deficit >40.16. The major difference between cluster 4 and 7 is aspect. The same vineyard also has areas from clusters 1 and 8 with water deficit <40.16 (C5 rule fig 5) and at elevation 48 m (CRT rule fig 6).

It could be concluded based on this approach that using relevant coarse digital attribute data suitable attributes for zoning at the *meso*-scale could be identified for wine regional /vineyard management decision making. Regression test performed show water deficit, age, hill shade, slope, aspect, min temp, acid sol phosphorous, induration as predictors with .407 adjusted R^2. More research is planned for the future to fine tune the approach with more *meso* scale data sets.

REFERENCES

Bissonnette, L., Wilson, K., Bel, S., & Shah, T. I. (2012). Neighbourhoods and potential access to health care: The role of spatial and aspatial factors. *Health & Place,* Volume 18, Issue 4, July 2012, 841-853.

Chauhan, R., Kaur, H., & Alam, M. (2010). Data Clustering Method for Discovering Clusters in Spatial Cancer Databases. *Inter' Journal of Computer Applications (0975 – 8887)* Vol 10– No.6, Nov 2010, 9-14.

Chi, G., & Zhu, J. (2008). Spatial Regression Models for Demographic Analysis. Popul. Res. Policy Rev (2008) 27, 17–42 DOI 10.1007/s11113-007-9051-8.

Ester, M., Kriegel, H.-p., Jörg S,, & Xu , X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Published in Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96) Simoudis E, Han J, Fayyad U M (eds.) 169 - 194.

Li, B., Shi, L., & Liu, J. (2010). Research on Spatial Data Mining Based on Uncertainty in Government GIS. 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2010), 10-12 August 2010 Yantai, China: 978-1-4244-5934-6/10 ©2010 IEEE. 2905-2908

Qian, Y., & Zhang, K. (2004). GraphZip: A Fast and Automatic Compression Method for Spatial Data Clustering Spatial Data Clustering. SAC '04, March 14-17, 2004, Nicosia, Cyprus (p. 5). Nicosia, Cyprus: 2004 ACM 1-58113-812-1/03/04.

Shanmuganathan, S. (2010). Viticultural Zoning for the Identification and Characterisation of New Zealand "Terroirs" Using Cartographic Data. *GeoCart'2010 and ICA Symposium on Cartography – Proceedings* Auckland: New Zealand Cartographic Society Inc. 53-64.

Wei, T., Tedders, S., & Tian, J. (2012). An exploratory spatial data analysis of low birth weight prevalence in Georgia. Applied Geography, Volume 32, Issue 2, March 2012, 195-207.

Xiaonian, L., Yi, Z., Zhang, F., & Liu, X. (2011). The Geographic Information Platform of New Socialist Countryside Comprehensive Services. *Procedia Environmental Sciences 11 (2011),* 3 – 10.