

# Modelling the seasonal climate variability and its effects on vintage wines from Marlborough, NZ

Subana Shanmuganathan and Ana Perez Kuroki  
Geoinformatics Research Centre,  
School of Computing and Mathematical Sciences,  
Auckland University of Technology  
New Zealand

Ajit Narayanan and Philip Sallis  
Geoinformatics Research Centre,  
School of Computing and Mathematical Sciences,  
Auckland University of Technology  
New Zealand

**Abstract**— The paper presents some interim results from an ongoing research on the application of data/text mining methodologies being investigated to modelling the seasonal climate variability and its effects on the world famous Marlborough vintage wines. The research in an extension to the investigations thus far conducted on modelling the effects of seasonal climate variability on Kumeu wines and all the sub-projects contribute to an overarching project which is aimed at developing a suitable set of schemes/ procedures for the identification and characterisation of wines produced from New Zealand's major wine regions. The distinctive New Zealand wine styles along with the regions from where the wines come from are initially elaborated. The major issues regarding the topic are; firstly, there is no single method that could be considered as the best way to establish the links between precise independent (climate/ weather) and the rather imprecise dependent (subjective wine quality) data sets. Secondly, the data on New Zealand wine quality is not sufficient enough to perform any conventional rigorous analytical approaches as data on wine quality spans only a decade, hence we look at data/text mining methods and a combination of explorative and statistical data analysis methodologies to resolve the issues. Following a brief outline on the methods investigated and results achieved in the Kumeu wine case study, the paper presents the new methods and approaches explored with Marlborough wines produced from 1996 to 2007. Finally, wine descriptors that are found to be linked with wine quality and therefore considered as correlated to the regional climatic conditions experienced in different wine regions of New Zealand, are discussed.

**Keywords-component; formatting; style; styling; insert (key words)**

## I. INTRODUCTION

The quality of a vintage wine depends on the climate, the seasonal weather conditions that ripened the grapes as much as the winemaker experience and talent. This has been well-documented in literature as seen in (1) (2). But knowledge on "how" and the "exact" weather conditions that affect the grape composition and the ripening process is fragmented (3). The major reason for this being lack of consistent data on wine quality to scientifically validate the anecdotal evidence, by establishing the correlation/s between two vital sets of factors, namely, the wine quality which is arguably a subjective issue when compared with the more precise and readily available weather data. This is especially the case with wines produced from most of the new wine producing countries, such as New Zealand, Australia, Chile and South Africa. However, in recent

years, the interest in scientifically understanding the effects of long- and short-term effects of climate change has been significant in both *old* as well as *new* world wine producing countries to improve vineyard and grape wine production. Nonetheless lack of consistent data makes the modelling of climate change effects of grapevine growth and wine quality difficult using conventional methodologies.

In view of the above facts, researchers at AUT's Geoinformatics Research Centre (GRC) began investigating into experimenting with approaches that are generally referred to as other than the rigorous statistical data analysis methodologies to modelling the influences of weather conditions on the quality of wine vintages of different regions within New Zealand. The ultimate aim of this effect is to find a set of ideal features that can be used for the identification and characterisation of wines produced from New Zealand's regions under different seasonal conditions (4). Section II gives an outline on the results obtained with approaches already investigated using wine ratings (and sommelier comments (5) and grape yield (6) from Kumeu River Wines with data spanning only a decade or so. Section III looks at Marlborough wine and climate data being used in the research, following which future direction of this research is briefly outlined.

## II. CLIMATE EFFECTS ON WINES OF DIFFERENT VINTAGES

Recent research in modelling the seasonal variability in weather conditions and its effects on grapevine phenology and wine produced from the grapes of that season showed that, of the many factors, weather as the main influencing factor at this temporal scale (7).

### A. Year- to year-variability in climate and its effects on wine

In viticulture it is believed through cultivation practices that vintage-to-vintage variability effects, especially on fruit quality is imposed by two broader categories, namely *terroir* (climate, soil) and *cultiva* (or variety). Of these two, the influence from *terroir* factors is described to be major at this temporal scale with their impacts established scientifically as (50%) for climate, (25%) for soil and (10%) for *cultivar* by van Leeuwen (8). In this research climate is concluded to be the major influencing factor, its affects described as realised specifically through vine water balance status. The study looked at the influence of all variables representing *terroir* and

*cultiva* simultaneously on vintage-to-vintage vine development and berry composition of non irrigated *vitis vinifera* on gravelly soil (with heavy clay subsoil and sandy soil as well as water table within the reach of roots) with *Merlot*, *Cabernet franc* and *Cabernet Sauvignon*. The climate variables used for the study consisted of maximum and minimum temperatures, degree days (base of 10°C), sunshine hours, ET<sub>0</sub>, rainfall, and water balance for a four year period from 1996 to 2000. Similarly, (9) (10) studied the seasonal climate effects on grapevine phenology and wine quality and concluded temperature and weather related factors during berry ripening period as the main deterministic factor/s in wine quality.

### B. Year-to-year (vintage) variability and its effects on Kumeu wines

The initial results of the analysis carried out on Kumeu (New Zealand) wine descriptor frequencies of 12 grouped descriptors in each year from 1997-2006, for 30 Kumeu wines discussed in (8) showed the correlations between yearly total of 12 std deviation in monthly average temperature against wine descriptor frequencies. The wine descriptors were grouped using the WEBSOM approach to reduce the 51 descriptors to 12 groups. The WEBSOM<sup>1</sup> of 45 nodes was created with 51 descriptors and commercial software package viscovery.

An interesting observation made from this study was that year 1998, with the highest *ssd/meant* within the period analysed consisted of high descriptor frequencies for clusters C2, C3, C6 and C10 descriptors. Meanwhile, year 2002 with the lowest *ssd/mean* temperature consisted of higher frequencies for C5, C8 and C11 descriptors. Discriminant analysis ran on the data set produced 11 words (underlined) as contributing factors in determining the variable vintage (or year considered as a dependent variable on the 11 descriptors). C1 descriptors are present in all years. *ssd/meanT*: standard deviation of daily mean temperature for the growing season (Sep.-April). Source for temperature data: National Institute for Water and Atmospheric Research (NIWA). The model correctly classified 76.7% of original grouped cases and 50.0% of cross-validated grouped cases (30 cases in total).

The results show that there are correlations between temperature variations and some descriptors

### III. THE METHODOLOGY

The research elaborated upon here looked at different methods initially for modelling the correlations between wine descriptors of 778 Marlborough vintages and their ratings for studying the wine descriptor/ ratings correlations with seasonal weather conditions. The following are the steps of the methodology adopted here:

1) a matrix of 195 descriptors selected as useful and relevant after discarding common and rare words from the corpus of 2351 words extracted from the original wine comments given for 778 wines used in the analysis (for further details on the text mining method (vector space model see 9).

2) ascertain the correlations between the descriptor groups and different wine styles/ ratings to select the high and low wine rate years for wine styles produced from the Marlborough region. The methods investigated in this research are: websom for grouping the co-occurring words (describing a feature) relating to wine style perhaps indicating high or low ratings. The Marlborough main wine styles being analysed in the paper are: Chardonnay, Pinot Noir, Riesling, Sauvignon Blanc, Pinot Gris and Gewürztraminer.

3) establish the correlations between style descriptors and ratings using statistical methods

4) study the correlations between style descriptors and ratings using data mining (DM) methods. The DM methods used in the research are JRip rules and J48 tree based classification of WEKA software.

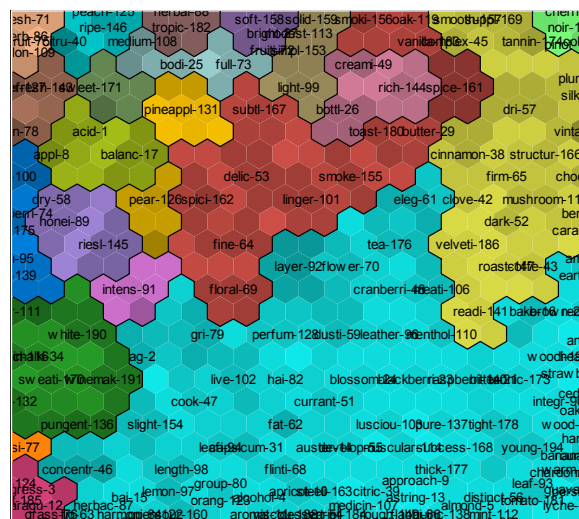
5) 254 Chardonnay vintages of New Zealand produced between 1996-2007 are studied to establish the wine descriptors related to the regions using statistical methods and then with data mining techniques.

### IV. RESULTS

Results obtained from different statistical and DM investigated methods to establish correlations between wine descriptors and vintages of different NZ wine styles are discussed in this section

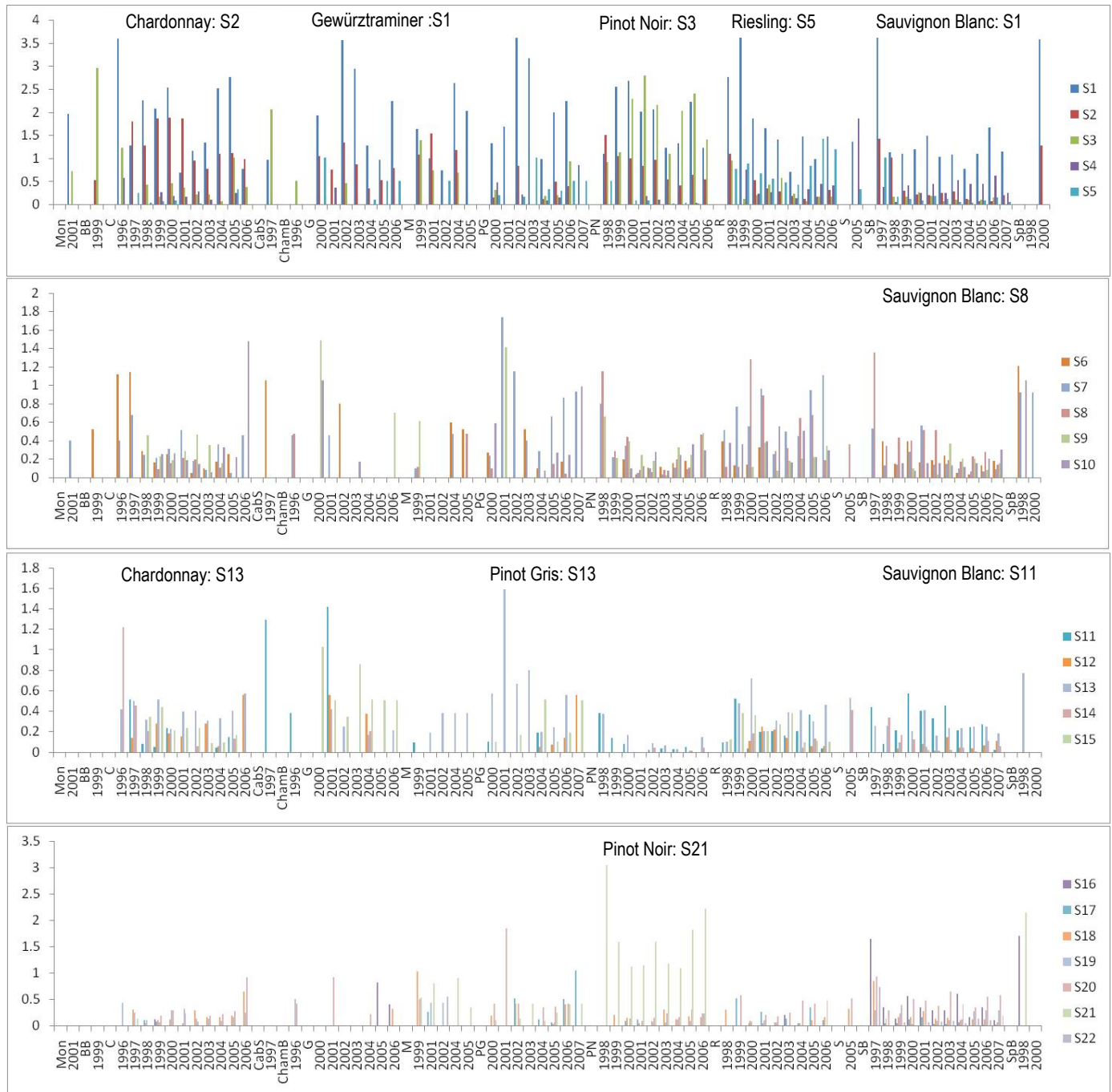
#### 1) The WEBSOM approach to Marlborough wine styles

Using the WEBSOM approach word segments (features) relating to Marlborough vintages of different wine styles were analysed to establish the words that correlate to the wine styles.



**Figure 1.** WEBSOM of 195 descriptors extracted from sommelier comments provided for 253 Marlborough vintages of styles produced from this famous wine region. The different segments in the SOM show the descriptors used to state the features (S1-S22) of the vintages by sommeliers.

<sup>1</sup> WEBSOM: The WEBSOM method provides an approach to organise a textual document collection onto a graphical map display with an facilitates to browse the document collection. The approach allows for content-directed search within the collection. Initially each document is encoded into a histogram of word categories that are used by the self-organizing map (SOM) algorithm to create a map based on the similarities in the contexts of the words. The encoded documents can be organized on several hierarchical SOMs, in which nearby locations contain similar documents (9)



**Figure 2.** S1 (Sauvignon Blanc): eleg-61, tea-176, layer-92, flower-70, ampl-6 earthi-60, cranberri-48, meati-106, bake-16, brown-28, red-142, gri-79, perfum-128, dusti-59, leather-96, menthol-110, ag-2, woodi-193, anis-7 heavi-85 strawberri-164, live-102, hai-82, blossom-24, blackberri-23, raspberri-140, bitter-21, tannic-173, cook-47, currant-51, integr-90, cedar-33 oaki-120, slight-154, fat-62, luscious-103, pure-137, tight-178, wood-192, leafi-94, capsicum-31, auster-14, develop-55, muscular-114, success-168, young-194, hard-83 round-150, concentr-46, length-98, flinti-68, thick-177, warm-189, lemon-97, group-80 orang-123, apricot-10, steeli-163, citric-39, leaf-93 tomato-181, persist-129, fig-63, herbac-87 power-135, bai-15 harmoni-84 sharp-151, open-122, sour-160, alcohol-4 aromat-11 strong-165,

viscou-188, dessert-54, variet-184, medicin-107 syrapi-172, rough-149, approach-9 astring-13 flabbi-66 fleshi-67 zesti-195, quinc-138, almond-5 butterscotch-30, distinct-56 mint-112 nutti-118, banana-1 S2 (Chardonnay): smoki-156, oak-119, vanilla-183, spice-161, subtl-167, toast-180, butter-29, delic-53, smoke-155, spici-162, linger-101, fine-64, floral-69 (chardonnay) S3: smooth-157, suppl-169, complex-45, tannin-174, plum-134 silki-152, dri-57, vintag-187, cinnamon-38, structur-166, firm-65, chocol-37, clove-42, mushroom-115, dark-52, berri-20 caramel-32, velveti-186, roast-147, coffe-43, readi-141 (Pinot Noir) S19: herbal-88 tropic-182 S20: clean-41 crisp-50 fresh-71 herb-86 melon-109, grapefruit-75 S21: black-22 cherri-36 noir-117 pinot-133, cola-44 S22: grassi-77

The 22 word segments (figure 2) generated from the WEBSOM (figure 1) illustrate the correlations between the Marlborough wines styles and the wine descriptors (figure 2). Correlations were found between Chardonnay: S2, Pinot Noir: S3, Riesling: S5, Sauvignon Blanc: S1 Pinot Gris: S13 and Gewürztraminer: S1 descriptors. The 195 wine descriptors used were extracted from sommelier comments from a web magazine (<http://buyingguide.winemag.com/regions/new-zealand>) for the vintages 1996-2007.

Based the DM results some descriptors can be attributed to certain styles. For example, either a) higher weight of lychee or b) higher pear and spice along with low values of butter and ripe relate to *Gewürztraminer* style of Marlborough. Similarly, the descriptors and their respective weights relating to Marlborough's Pinot Noir, Pinot Gris, Riesling and Chardonnay are listed in Table 1. Anything that does not meet the conditions in the Table 1 is stated as Sauvignon Blanc (SB). Hence, to reveal the exact descriptors to SB, comments of SB vintages are analysed alone. Similarly, Chardonnay from New Zealand regions produced during 1996 and 2006 as well are studied as individual collections and the results are discussed in the next section.

TABLE I. MARLBOROUGH VINTAGES (1997-2007) AND STYLES

(lyche104 >= 0.97) => style2=G (9.0/1.0)	JRip rules
(pear126 >= 0.51) and (spice161 >= 0.53) and (butter29 <= 0) and (ripe146 <= 0) => style2=G (4.0/0.0)	
(gri79 >= 0.57) => style2=PG (15.0/1.0)	
(appl8 >= 0.23) and (bai15 >= 1.4) => style2=PG (4.0/1.0)	
(almond5 >= 0.73) => style2=PG (4.0/1.0)	
(riesl145 >= 0.46) => style2=R (33.0/1.0)	
(lime100 >= 0.31) and (sweet171 >= 0.35) => style2=R (10.0/3.0)	
(apricot10 >= 0.87) => style2=R (9.0/4.0)	
(dry58 >= 0.34) and (bodi25 >= 0.3) => style2=R (4.0/1.0)	
(oak119 >= 0.35) and (cherri36 <= 0) and (pinot133 <= 0) => style2=C (46.0/11.0)	
(pear126 >= 0.34) and (butter29 >= 0.75) => style2=C (8.0/0.0)	
(toast180 >= 0.43) and (cherri36 <= 0) => style2=C (13.0/4.0)	
(chardonnai35 >= 0.52) => style2=C (13.0/2.0)	
(cherri36 >= 0.24) => style2=PN (148.0/8.0)	
(pinot133 >= 0.42) => style2=PN (19.0/1.0)	
(plum134 >= 0.71) => style2=PN (6.0/2.0)	
(tannin174 >= 0.43) => style2=PN (3.0/1.0)	
(cola44 >= 0.38) => style2=PN (2.0/0.0)	
=> style2=SB (428.0/64.0)	

(WEKA) show the correlations between the descriptors and Marlborough wine styles

TABLE II. MARLBOROUGH VINTAGE (1996-2006) DESCRIPTORS & RATINGS

(veget-111 >= 0.37) and (fruit-37 <= 0) and (fresh-36 <= 0.26) => rate scale=low (11.0/3.0)
(asparagu-8 >= 0.6) and (fruit-37 <= 0) => rate scale=low (7.0/2.0)
(sour-99 >= 0.94) => rate scale=low (3.0/0.0)
(heavi-50 >= 0.9) => rate scale=low (6.0/2.0)
(group-45 >= 0.84) => rate scale=low (5.0/2.0)
(complex-22 >= 0.4) => rate scale=high (24.0/10.0)
=> rate scale=med (325.0/71.0)

JRip rules show the correlations between Marlborough SB vintages and descriptors

## 2) Marlborough Sauvignon Blanc vintages and descriptors

The collection of comments given for 381 Marlborough vintages was converted into matrix of 118 wine descriptors and their rates transformed into "low <80" "medium (med) >79 and <90" and "high >89" based on their rating (100 point) for analysing with DM techniques using WEKA software. The JRip rules listed in Table 2 produced (at training 76% and cross validation 70%) for this indicate that high values of the descriptor *veget* (meaning vegetable 6 vegetal 26) and low values of fruit and fresh as related to "low" (<80) rate in Marlborough SB vintages. Similarly, either higher weights of a) asparagus (>= 0.6) and fruit (<= 0) or b) high sour-99 (0.94) or c) heavy (>= 0.9) group-45 (>= 0.84) are related to low rated SB vintages. On the other hand, high value of complex (>= 0.4) if related to higher ratings with an accuracy of 24.0 correct and 10.0 wrong.

TABLE III. MARLBOROUGH SB VINTAGES (1997-2007) & RATINGS

complex-22 <= 0
asparagu-8 <= 0.4
rich-88 <= 0.36
creami-25 <= 0
group-45 <= 0
bean-12 <= 0
honei-54 <= 0.49: med (278.0/57.0)
honei-54 > 0.49
finish-34 <= 0.1: med (4.0)
finish-34 > 0.1: high (6.0/1.0)
bean-12 > 0
bean-12 <= 0.75
fresh-36 <= 0: med (3.0)
fresh-36 > 0: low (2.0)
bean-12 > 0.75: low (2.0)
group-45 > 0
lime-63 <= 0: low (5.0)
lime-63 > 0: med (2.0)
creami-25 > 0
melon-68 <= 0: med (8.0/1.0)
melon-68 > 0: high (2.0)
rich-88 > 0.36
veget-111 <= 0
melon-68 <= 0
grassi-43 <= 0
sweet-104 <= 0.52
lime-63 <= 0
tropic-109 <= 0: med (10.0)
tropic-109 > 0: high (3.0/1.0)
lime-63 > 0: high (3.0/1.0)
sweet-104 > 0.52: high (2.0)
grassi-43 > 0: high (2.0)
melon-68 > 0: high (3.0)
veget-111 > 0: low (2.0)
asparagu-8 > 0.4
fruit-37 <= 0.05: low (9.0/2.0)
fruit-37 > 0.05: med (11.0/2.0)
complex-22 > 0
linger-64 <= 0
herbal-53 <= 0.36
fruit-37 <= 0.17
appl-5 <= 0
eleg-30 <= 0
nectarin-72 <= 0: med (8.0)
nectarin-72 > 0: high (2.0)
eleg-30 > 0: high (2.0)
appl-5 > 0: high (2.0)
fruit-37 > 0.17: high (5.0)
herbal-53 > 0.36: high (3.0)
linger-64 > 0: med (2.0)



Meanwhile, classifier model created J48 pruned trees (Table III, with WEKA, (at training 82% and cross validation 67%)) produced rules with additional (to the JRip) words relating to rate classes (low, medium and high). The additional words being, *creami (creamy)*, *bean*, *honei (honey)*, *lime*, *melon*, *grassi (grassy)*, *sweet*, *tropic*, *nectarine*, *eleg (elegant)*, *apple*, *fruit*, *herbal*, and *linger*.

### 3) New Zealand's Chardonnay vintages and descriptors

Finally, Chardonnay vintages produced from regions in New Zealand between 1996 and 2007 are analysed separately to identify the region related words and their correlations between ratings as analysis conducted all the vintages together using either WEBSOM descriptor groupings or JRip/ J48 model classifiers did not produce any interesting rules. However, rules obtained analysing the chardonnay vintages separated based on the regions produced interesting rules to establish the correlations between “low”, “med” and “high” rate classes and are presented in Table IV.

TABLE IV. HAWK'S BAY VINTAGES (1996-2006) AND DESCRIPTORS

<p><u>Gisborne</u>  sweet-19 &lt;= 0    spice-18 &lt;= 0      appl-1 &lt;= 0.27: med (28.0/7.0)      appl-1 &gt; 0.27: high (2.0)    spice-18 &gt; 0: high (3.0/2.0)  sweet-19 &gt; 0    vanilla-23 &lt;= 0: med (3.0)    vanilla-23 &gt; 0: low (3.0)</p> <p>Training: 76%  Cross validation : 46%</p> <p><u>Waipara</u>  toast-8 &lt;= 0.26    citru-3 &lt;= 0: med (8.0/2.0)    citru-3 &gt; 0: high (2.0/1.0)  toast-8 &gt; 0.26: high (3.0)</p> <p>Training: 76%  Cross validation : 38%</p> <p><u>Hawke's Bay</u>  lime-19 &lt;= 0    ripe-28 &lt;= 0.23      orang-23 &lt;= 0        creami-9 &lt;= 0          honei-17 &lt;= 0            intens-18 &lt;= 0: med (19.0/3.0)            intens-18 &gt; 0: high (5.0/1.0)            honei-17 &gt; 0: high (2.0)          creami-9 &gt; 0: high (2.0)        orang-23 &gt; 0: high (3.0)    ripe-28 &gt; 0.23: med (8.0/1.0)  lime-19 &gt; 0: med (6.0/1.0)</p> <p>Training: 86%  Cross validation : 48%</p>
--

## V. CONCLUSIONS

The paper illustrated statistical and data mining approaches being investigated to establishing the correlations between wine descriptors (embedded in sommelier comments) and wine ratings in order to modelling the seasonal weather influences on New Zealand wine vintages. The results show potential for establishing the variability in wine styles and regions. The approaches can be further developed to identify the special attributes i.e., wine descriptors, that could be used to relate wine style as well as quality i.e., ratings and then to model the seasonal weather variability on New Zealand vintage wines. The methods can as well provide a means to develop a viticulture zoning scheme as well as geographical indicators for New Zealand vintage wines.

## VI. FUTURE WORK

The research is on going and it is anticipated that once the specific descriptors for different wine styles and regions are identified and verified they could be incorporated into an GIS to model the climate as well as other local influences, such as environmental (soil, aspect, slope) at more finer scales, i.e., within vineyards on grapevine phenology and wine quality.

## ACKNOWLEDGMENT

The authors wish to acknowledge the support extended by Master Winemaker Michael B of Kumeu River Wines and Peter Sumich of AUT.

## REFERENCES

- [1] Ashenfelter, O, Ashmore, D and Lalonde, R. *Bordeaux Wine Vintage Quality and the Weather*. Chance vol 8 No. 41995: 7-14., 1995.
- [2] Shanmuganathan, S and Sallis, P. *Modelling climate change effects on wine quality based on expert opinions expressed in free-text format: The WEBSOM Approach*. ICONIP 2008, Part I, LNCS 5506, 2009. Springer-Verlag Berlin Heidelberg 2009. pp. 915-925. , 2009.
- [3] Soar, C J, Sadras, V O and Petrie, P R. (2008) *Climate drivers of red wine quality in four contrasting Australian wine regions*. s.l. : Australian Journal of Grape and Wine Research 14, 78–90., 2008.
- [4] Shanmuganathan, S. *Viticultural zoning for the identification and characterisation of New Zealand "Terroirs" using cartographic data*. in proceedings of GeoCart'2010 and ICA Symposium on Cartography © New Zealand Cartographic Society Inc, 2010 [Eds, 2010.
- [5] S, Shanmuganathan and Sallis, P. *Shanmuganathan, S. and Sallis, P. (2010). Determining the relationship between climate variations and wine quality: the WEBSOM approach*. In Simard, S. (Ed), Climate Change and Variability, Sciyo, pp. 389-408, 2010.
- [6] Shanmuganathan, S, Sallis, P and Narayanan, A. *Data mining and X2 test based hybrid approach to modelling climate effects on grape crop in varieties of Kumeu, New Zealand* in proceedings of [eds] Thatcher, S., 2011 3rd International Conference Conference on Machine

- Learning and Computing (ICMLC 2011) Singapore, 26-28 February 2011, 978-1-4244-925 3-4 /11 ©2011 IEEE Vol 1 pp 355-359, 2011.
- [7] Shanmuganathan, S, Sallis, P and Narayanan, A. *Data mining techniques for modelling seasonal climate effects on grapevine yield and wine quality* in Second International Conference on Computational Intelligence, Communication Systems and Networks Networks (CICSyN), Liverpool, UK, 28 – 30 July, 84-89, 2010.
- [8] van Leeuwen, Cornelis, Friant, Philippe, Chone, Xavier, Tregoat, Olivier, Koundouras, Stephanos, Dubourdieu, Denis. *Influence of Climate, Soil, and Cultivar on Terroir*. s.l. : 2004, Am. J. Enol. Vitic 2004 55:207-217., 2004.
- [9] Jones, G V. *Climate change: observations, projections, and general implications for viticulture and wine production*. Practical Winery & Vineyard: 44-64., 2007.
- [10] Jones, G V and Davis, R E. *Climate Influences on Grapevine Phenology, Grape Composition, and Wine Production and Quality for Bordeaux, France*. Am. J. Enol. Vitic., Vol. 51, No.3, 2000 pp249-261. 2000.
- [11] Shanmuganathan, S, Sallis, P and Narayanan, A. *Unsupervised artificial neural nets for modelling the effects of climate change on New Zealand grape wines*. In B. Anderssen et al. (eds) /18th IMACS World Congress - MODSIM09 International Congress. s.l. : In B. Anderssen et al. (eds) /18th IMACS World Congress - MODSIM09 International Congress on Modelling and Simulation/, 13-17 July 2009, Cairns, Australia. ISBN: 978-0-9758400-7-8. pp. 803-809., 2009.
- [12] Honkela, T, et al. *Self-organizing maps of very large document collections: Justification for the WEBSOM method*. In Balderjahn, I., Mathar, R., and Schader, M., editors, *Classification, Data Analysis, and Data Highways*,. Springer, Berlin. pp 245-252.: <http://lib.tkk.fi/Diss/2000/isbn9512252600/article4.pdf>, 1998.
- [13] Best, S, León, L and Claret, M. *Use of Precision Viticulture Tools to Optimize the Harvest of High Quality Grapes* in proceedings of Information and Technology for Sustainable Fruit and Vegetable Production FRUTIC 05, 12-16 September 2005, Montpellier France, 2005. <http://www.symposcience.org/exl-doc/colloque/ART-00001647.pdf>.