2010 Second International Conference on Computational Intelligence, Communication Systems and Networks

Data Mining Techniques for Modelling the Influence of Daily Extreme Weather Conditions on Grapevine, Wine Quality and Perennial Crop Yield

Subana Shanmuganathan and Philip Sallis
Geoinformatics Research Centre
School of Computing and Mathematical Sciences
Auckland University of Technology, New Zealand.
e-mail: subana.shanmuganathan@aut.ac.nz

Ajit Narayanan
School of Computing and Mathematical Sciences
Auckland University of Technology, New Zealand
e-mail: philip.sallis@aut.ac.nz,
ajit.narayanan@aut.ac.nz

Abstract-*The influences of daily weather extremes, such as maximum/ minimum temperatures, humidity, and precipitation, are observable in perennial crop phenology that in turn determines the annual crop yield in quality and quantity. In viticulture, grapevine phenology determines the quality of vintage produced from the grapes apart from the best effects by winemaker. Following a brief review of current literature in this research domain, the paper describes a data mining approach being developed to data association modelling to depict dependency relationships between daily weather extremes, grapevine phenology and yield indicators using data from a vineyard in northern New Zealand and daily weather extremes logged at a nearby meteorology station. An artificial neural network algorithm was used to classify the data associations and the chi-square test was used to establish the degree of dependence between the related variable values. The initial results of the approach to daily maximum weather conditions show potential.*

Keywords; *self-organising maps, decision tree, χ2 test method*

I. INTRODUCTION

The variability in daily extreme weather conditions, such as maximum and minimum in temperature, wind velocity, humidity and precipitation, influences plant physiology and growth stages significantly and this could ultimately result in substantial shifts in crop phenology/ growth events, such as budburst, flowering, *Veraison (in grapevines)* and harvest. In the case of grapevine phonology, the influences in turn typically lead to changes in berry sugar content and the composition of other major components, such as the proteins (procompounds) responsible for wine colour, aroma and flavour phenols. These are the ingredients that determine the quality of vintage produced from the grapes despite the best efforts made by the winemaker. The quality of any vintage relating it its wine style depends fundamentally on the grapes used but they are in turn influenced particularly by the weather conditions that ripened the grapes. The final and very significant influence comes from the wine maker's experience and talent [1] [2].

The centuries-old traditional "Terroir x cultiva" concept is used by contemporary viticulturists who continue to adapt ways and means to compensate for seasonal daily weather variability. An example of this is irrigation control. This is in order to optimise grapevine growth, which is seen to be very common during the berry ripening period to ensure that the best kind of grapes for refining the vintage within the wine appellation are produced from the vineyard [3][4][5].

The next section of this paper reviews published work in modelling daily recordings of selected extreme weather conditions and their effects on different perennial crop yield in addition to grapevine yield and wine quality. The methodology section follows, which describes data sources, elements and pre-processing as well as data analysis methods used in this research, such as an unsupervised algorithmic artificial neural network called Kohonen self-organising map (SOM) [1] based data mining and the statistical methods applied to verifying the apparent results. The final section describes the results, showing the potential of the research methodology for gaining more insight in the intuitive proposition that grapevine yield and vintage quality are influenced by daily weather extremes more precisely than has been previously demonstrated.

II. LITERATURE ON DAILY WEATHER INFLUENCE ON GRAPEVINE AND OTHER PERENNIAL CROPS

Based on an innovative approach with χ2 test method by [6] [7], an Australian study [8] described research that modelled the varying influence of daily extreme weather conditions on grapevine phenology and wine quality in four of Australia's major wine regions. In the original study [6], the influence of daily temperature and precipitation on annual apple production was modelled with frequencies of daily weather variability. Each weather variable separately consisted of a matrix of days with values in each of the continuous classes (at 2^{o} F intervals for temperature) within a period of a moving 3 week window. Each window in succession added a new week and dropped the first week as the window advanced, for testing with an iterative χ2 test approach. For this, apple annual production data was initially separated into quartiles by level of production and then each upper and lower quartiles were tested with a combined mid two quartile values to find the critical (cardinal or turning) point in each climate variable. The χ2 tests were then run for high-low and low-high scans to see any deviation in test values generated for the extreme quartiles from those of the combined mid-quartiles, an increase followed by a decrease (or a turning point) in a

[1] Kohonan SOMs are feed forward artificial neural networks that are based on an unsupervised algorithmic processing. Approaches based on SOM techniques provide an excellent tool for data mining in areas where prior knowledge on problem domain is limited.

978-0-7695-4158-7/10 $26.00 © 2010 IEEE
DOI 10.1109/CICSyN.2010.15

given scan being referred to as the cardinal value of the respective climate variable. Based on the test results, the study concluded that a significant association existed between the climate variables analysed and annual apple production within the 72-year period 1920–1991. The results and associations established included, flower bud initiation in June (30°C) and flower bud development in August (26°C) with poor production in the following year. November, December and February (critical value range, –7°C to –29°C) were found to be the months during which the main climatic factor limiting the apple production occurred in the data set. These time periods correspond with the occurrence of historical winter injury events. Similarly, daytime temperatures that influenced apple production adversely as well as favourably with respective *bad* and *high* yield years were established along with some indications and actions, such as early irrigation, for avoiding a potential *bad* yield year. Production favoured by mild temperatures during bloom and adversely affected by both very low temperatures and unseasonably high temperatures were explained to be coinciding with the temperature requirements of pollination and pollen tube growth. *High* production years were also associated with a lack of low night time temperature in spring, explained to be associated with frost in low lying areas. Hot, dry weather during August of a harvest year was found to be having a negative impact on apple production possibly because of loss in net photosynthesis, lower fruit size or apple sunburn. Warm weather during harvest favoured production, probably because of improved conditions for harvest operations and low fruit losses from autumn frosts.

The Australian study [8] of course, looked at quantifying what was described by the authors as "qualitative and fragmented knowledge" on the links between key weather variables during berry ripening and wine quality using this χ^2 test method in four of Australia's major wine regions, the regions being the Hunter Valley, Margaret River, Coonawara and the Barossa Valley. The regional wine ratings were used in the study as surrogate for wine quality for comparison between the *high* (top 25%) and *poor* (bottom 25%) vintages in relation to the frequency of defined weather conditions. The results of this study showed that with maximum temperature, better quality was associated with temperatures above 34°C throughout most of ripening in the Hunter, below 28°C in early January in the Margaret River, 28–33.9°C towards harvest in Coonawarra, and below 21.9°C in late January and early February and 28–30.9°C towards harvest in the Barossa. It was concluded that the approach provided a means for quantitative assessment allowing for the timing and magnitude of weather influences on wine quality and that this was possible on a regional basis.

III. The Data Issues and Methodology

This section examines the formulation of data sets and analytical issues when applying the χ^2 test method to gain further insight in the data dependencies between daily extreme recordings in weather variables, grapevine growth and wine quality as they were described in the research outlined in Section 2.

A. Daily weather and wine quality data sets

Daily weather data, i.e., maximum, minimum and grass minimum temperatures, has been obtained from the National Institute of Water and Atmosphere (NIWA) gathered at the nearest meteorology station from April 1997 to March 2009 [9]. Each of these weather variables is converted into a matrix of occurrence frequency at continuous 3°C intervals between the maximum and minimum recorded for that variable during a moving 3 week window for 45 weeks prior to the harvest date of each yield year.

Grapevine yield data for the same period was obtained from the grape grower. This data consisted of Vintage, Grapes harvested (in Yield tons/hectare, Harvest Date, Brix (dissolved sugar-to-water mass ratio of a liquid), Acid and pH [must]. Of the 12 vintages, 1998-2000 are classified as *moderate*, 2001, 2003, 2005 as *low* and 2002, 2004 and 2006 as *high* by the winemaker. The two sets of data (climate and grapevine yield) are used in this research to see the associations between them (as well as the degree of any existing association) and are described in the next section.

B. The methodology

Daily extreme weather frequencies within a moving 3 week window over 45 weeks prior to harvest and grapevine yield from a vineyard spanning a 12 year period were analysed using data mining techniques. The χ^2 method used in the literature referenced in Section 2 was regarded as inadequate as a single test due to lack of sufficient data so alternative analytical techniques were used, particularly those connectionist methods using the Kohonen algorithms embodied in self-organising maps (SOMs) [10].

The 12 year grapes yield data consists of *low* and *high* production (3 years of each) and of the rest, years 1997-2000 are described as *moderate* as indicated by the winemaker. Thus, 2007-2009 can be used for testing the prediction capability of the models being investigated for this purpose. The *high* and *low* years are considered as *upper* and *lower* quartiles respectively for the χ^2 test. However, in the SOM based data mining, *moderate* years as well are included to see the correlations between the dependent (yield) and independent (weather) variables of this research by analysing the SOM cluster patterns and profiles.

IV. The Results

This section illustrates the results of the data mining approach using the Kohonen SOM algorithm and an analysis using the χ^2 test for this research domain.

A. SOM results

A SOM (figure 1) of 100 nodes with principal plane ratio 100:67 trained with 8 cycles, 0.5 tension under *normal, exact mode* [11], for 135 data points with 9 temperature intervals, clustering favoured by rate (priority 4) and week (priority 2), was created and the findings are discussed here.

Figure 1. SOM of week, yield class and occurrence frequencies at continuous 3oC intervals within the lowest and highest of maximum daily temperature recorded in a moving 3 week windows for 1-45 weeks prior to every harvest date during the 12 year period (1997-2009).

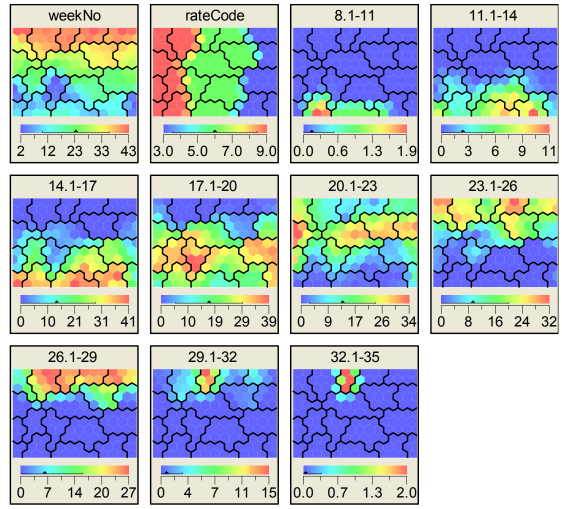


Figure 2. SOM components of week, yield class (rate code *low*(3), *moderate* (6) and *high*(9)) and the occurrence frequencies of temperature intervals (in total days for the three years) during moving 3 week windows (1-45 prior to harvest). The components show the patterns between different temperature range, week and yield class.

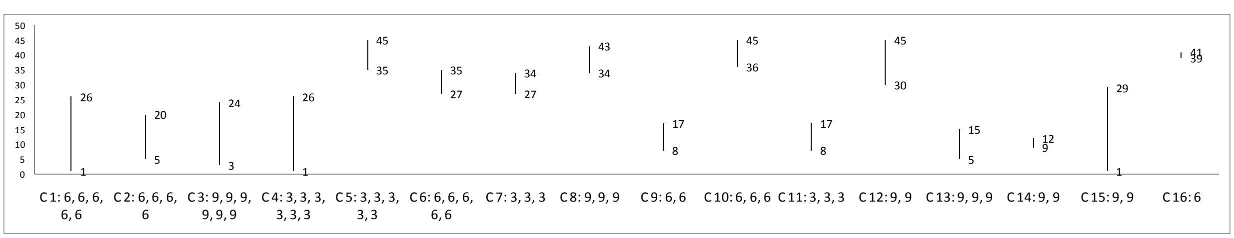
The SOM clustering of frequencies of daily maximum temperature at 3C° intervals (8.1-11, 11.114, 14.1-17, 17.1-20, 20.1-23, 23.1-26, 26.1 29 and 29.1-32) show the associations between the variable analysed i.e. week, and respective temperature frequencies for *low (3), moderate (6)* and *high (9)* yield years, during the 45 weeks prior to harvest. The map provides a means of visualising the associations between different temperature frequencies during week 1-45 (figures 1-3 Tables I and II).

The SOM cluster profiles listed in Tables 1 and II give details of the associations between different yield year classes and temperature frequencies/intervals. For example, C1, C4 and C15 with rates 3, 6 and 9 consist of frequencies for week. 1-26, 1-26 and 1-29 respectively, hence show the *low, moderate* and *high* yield year profiles, the variation among them in daily extreme weather data during this time period (within the 45 moving weeks prior to harvest).

TABLE I. THE 16 SOM CLUSTERS AND THEIR PROFILES. SOM CLUSTERING IS FAVOURED BY WEEK AND YIELD CLASS. CLUSTERS WITH SIMILAR WEEK SPAN BUT WITH DIFFERENT RATES ARE HIGHLIGHTED RATE CODE LOW 3, MODERATE 6 AND HIGH 9

SOM cluster No.	C 1: 6, 6, 6, 6, 6	C 2: 6, 6, 6, 6	C 3: 9, 9, 9, 9, 9, 9	C 4: 3, 3, 3, 3, 3, 3	C 5: 3, 3, 3, 3, 3	C 6: 6, 6, 6, 6, 6	C 7: 3, 3, 3	C 8: 9, 9, 9	C 9: 6, 6	C 10: 6, 6, 6	C 11: 3, 3, 3	C 12: 9, 9	C 13: 9, 9, 9	C 14: 9, 9	C 15: 9, 9	C 16: 6
Minimum	1	5	3	1	35	27	27	34	8	36	8	30	5	9	1	39
Maximum	26	20	24	26	45	35	34	43	17	45	17	45	15	12	29	41
08.1-11°C	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.30	0.00	1.50	0.00	0.00	0.00
11.1-14°C	0.60	6.71	2.08	1.19	0.00	0.00	0.00	0.00	8.11	0.00	8.60	0.00	3.50	8.00	0.00	0.00
14.1-17°C	7.50	27.00	16.92	17.38	0.18	0.67	3.88	0.00	37.56	0.00	36.50	0.17	31.67	41.50	2.29	0.00
17.1-20°C	32.00	25.57	30.58	32.06	1.55	10.89	18.50	0.30	15.22	0.43	17.00	5.67	24.50	13.00	21.14	0.00
20.1-23°C	19.40	3.71	11.67	10.75	13.45	27.44	28.38	13.40	1.11	8.57	0.60	24.17	1.67	0.50	32.57	6.33
23.1-26°C	3.40	0.00	1.75	1.56	23.27	18.33	10.00	24.90	0.00	28.14	0.00	26.00	0.17	0.00	7.00	17.67
26.1-29°C	0.10	0.00	0.00	0.06	22.82	5.22	2.25	21.80	0.00	23.71	0.00	6.83	0.00	0.00	0.00	23.67
29.1-32°C	0.00	0.00	0.00	0.00	1.73	0.44	0.00	2.60	0.00	2.14	0.00	0.17	0.00	0.00	0.00	13.33
32.1-35°C	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.00
rate Code	6.00	6.00	9.00	3.00	3.00	6.00	3.00	9.00	6.00	6.00	3.00	9.00	9.00	9.00	9.00	6.00

TABLE II. GRAPH SHOWING THE TIME SPAN (IN WEEK) FOR THE 16 SOM CLUSTERS IN FIGURE 1.



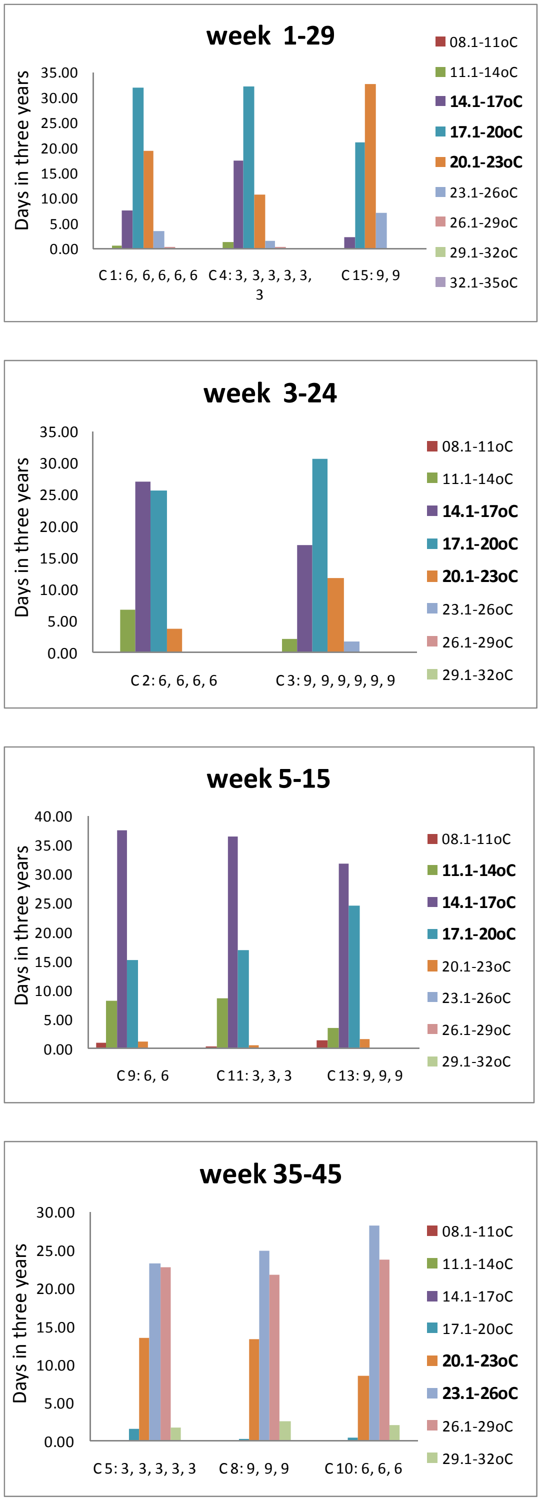


Figure 3. a - d (top to bottom) Graphs showing the associations between maximum daily temperature frequencies (in bold) and yield years rated as 3: *low*, 6: *moderate* and 9: *high*, during the specified growth stages (in week no. within the 45 moving weeks prior to harvest).

Based on the SOM clustering, week 1-29 (figure 3 a) could be seen as having associations with occurrence frequencies at maximum daily temperature ranges 14.1-17, 17.1-20 and 20.1-23 ^{o}C. Higher frequencies of 17.1-20 ^{o}C 30/3-> 10 days per year lead to *low/ moderate* annual yield. Meanwhile, higher frequencies of 14.1-17 ^{o}C relates to *low* yield. On the other hand, higher frequencies i.e., 30/3-> 10 days per year of 20.1-23 ^{o}C relates to *high* yield.

Similarly, during week 35-45 (figure 3 d) temperature ranges 20.1-23 and 23.1-26 ^{o}C show associations with different yield classes (figure 3 d). During this time, higher frequencies i.e., 27/3-> 9 days per year of 23.1-26 ^{o}C temperature lead to *moderate* yield. The difference between *low* and *high* seems to be very subtle, it could be either a day at 17.1-20 ^{o}C or under 25/3-> 8.33 days per year at 23.1-26 ^{o}C daily maximum temperature.

In order to verify the SOM results and to establish the degree of the associations between grapevine yield and weather variables iterative $\chi 2$ tests were carried out between the *low* Vs *high* yield classes and their respective occurrence frequencies at 3 ^{o}C daily maximum temperature intervals during moving 3 week windows (week 31-45), and the results are discussed in the next section.

B. *$\chi 2$ test results*

Initial results of χ^2 test conducted for moving week 31-45 show the associations between *high* and *low* yield years and at three sets of temperature intervals analysed in this research. The temperature intervals have to be reduced to only three classes as χ^2 test cannot be conducted with zero values for frequency hence temperature intervals 8.1-11, 11.1-14, 14.1-17, 17.1-20 and 20.1-23^{o}C were combined and a <23^{o}C temperature class was created. Similarly, 26.1-29 and 29.1-32^{o}C were added to create a >26.1^{o}C. For all three classes occurrence frequencies during moving 3 week windows between 31-45 prior to harvest were analysed to overcome the 0 frequencies that makes χ^2 test meaningless, which is an issue in this research arising due to insufficient yield data. Despite this drawback, the χ^2 test results give the precise week, occurrence frequency and the temperature interval critical to annual yield, in this case *high* and *low* yield year classes.

TABLE III. FREQUENCY DISTRIBUTION OF DAILY MAXIMUM TEMPERATURE AND WEEK NOS.

week No.	<23	23.1-26	>26	chi square rate	p-value
31	11.67	**8.00**	1.33	8.000	0.005
32	17.67	3.00	0.33	9.228	0.002
32	8.67	**9.00**	**3.33**	7.364	0.007
33	16.00	4.67	0.33	7.247	0.007
33	8.33	8.33	**4.33**	10.286	0.001
44	5.33	8.00	7.67	6.125	0.013
45	4.33	7.33	9.33	14.235	0.000
45	**9.00**	10.00	2.00	4.900	0.027

The table shows the daily maximum temperature and the week nos associated with *low* (grey) and high (brown **bold**) *yield* years with their $\chi 2$ rate and p-values in a vineyard from northern New Zealand.

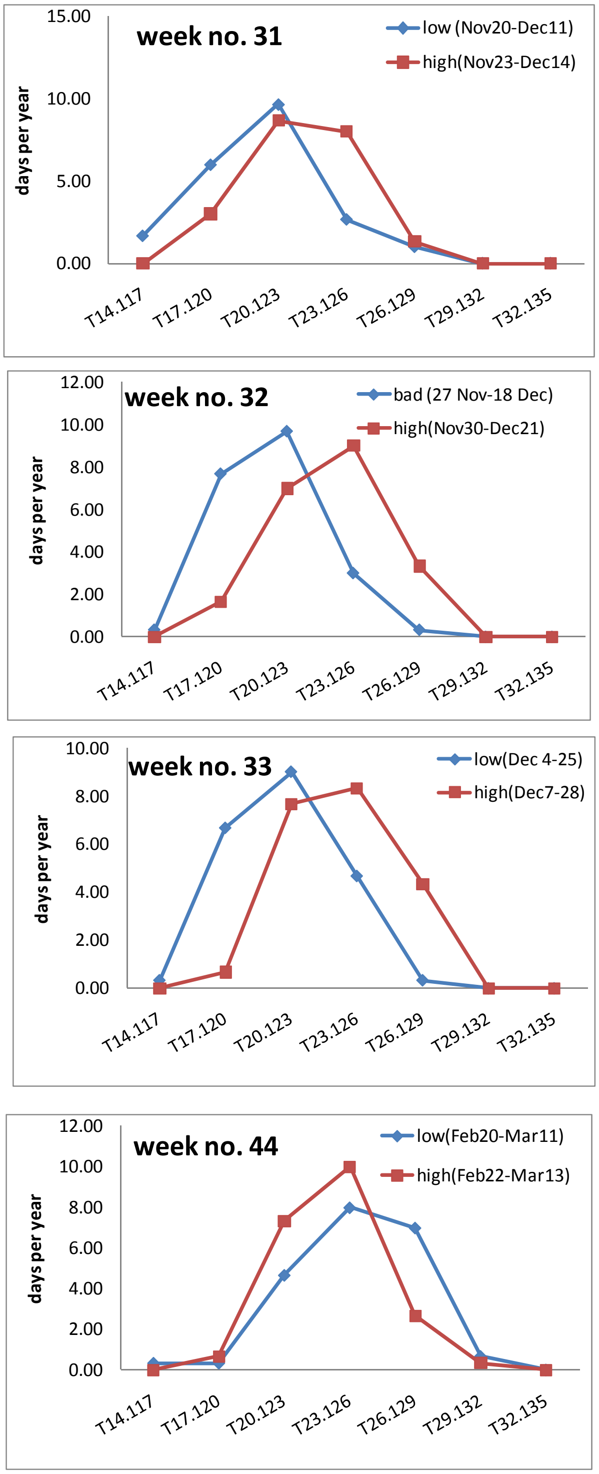


Figure 4. Graphs top to bottom: a:8 days (frequency) at temperature interval at 23.1-26 °C during week 31 (between Nov 20-Dec 11) relate to *high* yield years. b: during week 32 just over 17 days at <23 °C during (Nov 23-Dec14) relate to *low* yield years. Meanwhile in the same week 9 days at 23.1-26 °C and just over 3 days at >26.1 °C lead to *high* yield years c: during week 33, 16 days of <23 °C relate to *low* yield and just over 4 days at >26 °C relate to *high* yield years. d: during week 44 over 7 days > 26 °C relate to *low* yield years

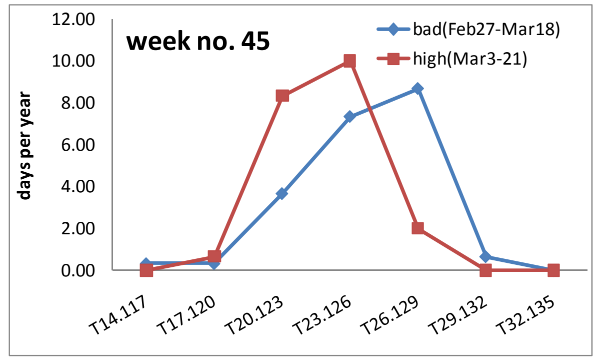


Figure 5. During week 45 just over 9 days/year of temperature (frequency) at >26°C relates to *low* yield years, during the same week 9 days at <23°C during relate to *high* yield years (Table III).

The χ^2 test results even thought narrowed down to week 31-45 prior to harvest and with occurrence frequencies redefined to three temperature intervals, still produced the degree of daily maximum weather associated with annual *low* and *high* yield year classes (figures 4 a-d and 5). These results give the specific occurrence of frequencies and the respective temperature ranges along with the week details.

Interestingly, the $\chi 2$ test results conform to the SOM results as well as local anecdotal knowledge. The SOM cluster profiles imply that occurrence frequencies at temperature intervals 14.1-17, 17.1-20 and 20.1-23 °C as determinants of yield year classes during week 1-29 (figure 3). Similarly, SOM results point to temperature intervals 20.1-23 and 23.1-26 °C determinants of yield during week 35-45.

Based on the local grapevine grower knowledge, high temperatures during late November-early December (or the flowering season) favour better yield and the $\chi 2$ test results not only reflect this but give the precise week details, temperature intervals and the occurrence frequencies as well. Likewise, late February to mid March is the berry ripening time during which higher frequencies >26.1 °C, are associated with *low* yield years, again well portrayed by SOM and $\chi 2$ test results. Higher frequencies of high temperature during berry ripening period could hasten sugar decomposition before colour, aroma and taste flavour phenols get to favourable levels required for fine wine production.

V. CONCLUSIONS

The paper illustrated a data mining method (Kohonen SOM based) used to depict the influence of daily extreme weather conditions on grapevine phenology, annual crop yield and wine quality. The initial SOM results of this work show associations between daily maximum temperature intervals 14.1-23C° during week 1-29 and >26C° during week 35-45 (ten weeks prior to harvest).

Despite a drawback arising due to lack of sufficient data, associations between daily maximum temperature occurrence frequencies in <23C° during week 32-33 and >26 C° during week 44-45 with *low* yield were confirmed by

χ^2 test, similarly respective frequencies of 23.1-26 C^o during 31-32 and <23 C^o during week 44-45 associated *low* and *high* yield years have been confirmed. This results show that more days of <23 C^o in mid November to early December and late February >26 C^o are linked with *low* years in terms of yield. Meanwhile, more days in mid November to early December and late February of temperature 23.1-26C^o and late February to early March with <23 C^o are associated with *high* yield years. Further research and results on the associations between daily extreme weather conditions i.e., minimum and soil (grass minimum), grapevine phenology and annual grapevine yield/wine quality using this approach appear in [12].

ACKNOWLEDGMENT

The authors thank the management, especially Michael Brajkovich of Kumeu River Wines in Henderson, Auckland, New Zealand and Peter Sumich for grapevine yield data and advice. Prof Jeffry Hunter and Mathangi Shanthakumar are acknowledged for help with statistical data analysis methodologies.

REFERENCES

[1] Jones, G V, and Davis, R E,. (2000) Climate Influences on Grapevine Phenology, Grape Composition, and Wine Production and Quality for Bordeaux, France. Vols. Am. J. Enol. Vitic., Vol. 51, No. 3, 2000 pp249-261.

[2] Sallis, P.J., Shanmuganathan, S., Pavesi, L., and Jarur, M. (2008) Kohonen Self-organising maps in the mining data mining of wine taster comments,:, 2008. Data Mining IX, Data Mining, Protection, Detection and other Security Technologies 2008. WIT press, Cadiz, Spain pp. 125-139.

[3] Shanmuganathan, S., Sallis, P., and Narayanan, A., 2009 Statistical data analysis incorporating web text mining to establish correlations between grape wine taster comments and wine ratings.. s.l. : Published by IEEE computer society, 2009. In proceeding of the 6th International Conference on Information Technology in Asia 2009 (CITA09) "Towards Human-Centered Computing" held in Hilton Kuching Hotel, Sarawak, 6 - 9th July 2009. pp 258-262.

[4] Shanmuganathan, S., Sallis, P., and Narayanan, A., Modelling climate change effects on wine quality based on expert opinions expressed in free-text format: The WEBSOM Approach. M. Kppen et al. (Eds.), Springer-Verlag Berlin Heidelberg 2009. pp. 915-925., 2009. ICONIP 2008, Part I, LNCS 5506.

[5] Shanmuganathan, S., Sallis, P., and Narayanan, A., Unsupervised artificial neural nets for modelling the effects of climate change on New Zealand grape wines. B. Anderssen et al. (eds), 2009. 18th IMACS World Congress - MODSIM09 International Congress 13-17 July 2009, Cairns, Australia. ISBN: 978-0-9758400-7-8. pp. 803-809.

[6] Caprio, J M, and Quamme H A,. (1998) Weather conditions associated with apple production in the Okanagan Valley of British Columbia., Agriculture and Agri-Food Canada Pacific Agri-Food Research Centre, Summerland, British Columbia, Canada, V0H 1Z0 1998, Vol. Contribution no.1075 129-137.

[7] Caprio, J M, and Quamme, H A,. (2002) Weather conditions associated with grape production in the Okanagan Valley of British Columbia and potential impact of climate change. Agriculture and Agri-Food Canada, Pacific Agri-Food Research Centre Summerland, British Columbia, Canada V0H 1Z0 : 2002, Contribution no. 2139. pp755-763.

[8] Soar, C J, Sadras, V O, and Petrie, P R,. Climate drivers of red wine quality in four contrasting Australian wine regions. 2008, Australian Journal of Grape and Wine Research 14, pp78–90, 2008.

[9] National Institute of Water & Atmospheric Research. The National Climate Database, National Institute of Water & Atmospheric Research. http://cliflo.niwa.co.nz/.

[10] Kohonen, T., Self-Organizing Maps (Paperback) pringer; 3rd ed. edition (Dec 28 2000) ISBN-10: 3540679219 ISBN-13: 978-3540679219. pages 520.

[11] Viscovery, 2010 www.viscovery.net/welcome last accessed 9 February 2010.

[12] Shanmuganathan, S., Sallis, P., and Narayanan, A., Modelling the effects of daily extreme weather on grapevine and wine quality in International Environmental Modelling and Software Society (iEMSs) 2010 in Proc. International Congress on Environmental Modelling and Software Modelling for Environment's Sake, Fifth Biennial Meeting, Ottawa, Canada, David A. Swayne, Wanhong Yang, A. A. Voinov, A. Rizzoli, T. Filatova (Eds.) http://www.iemss.org/iemss2010/index.php?n=Main.Proceedings (*in press*).