Data Mining and χ2 Test Based Hybrid Approach to Modelling Climate Effects on Grape Crop in Varieties of Kumeu, New Zealand

S. Shanmuganathan, *Senior Member, IACSIT*, P. Sallis, and A. Narayanan

***Abstract*—The paper elaborates upon a hybrid approach consisting of data mining and statistical methods, to modelling seasonal climate effects, i.e., arising from year-to-year variability in weather conditions, on grape crop of three different varieties cultivated in northern New Zealand. Recent research using an iterative χ2 method based approach to modelling climate effects on "high" and "low" yearly yields (of perennial crops) with data at the regional (macro) and grape yield from different vineyards, with climate data at macro scale, are briefly outlined. The grape varieties studied are *Chardonnay*, *Pinot Noir* and *Pinot Gris*. The results show interesting patterns in the nexuses between extreme daily weather conditions and grape crop data in terms of daily maximum, temperature observed for "low" and "high" yields, and within the macro and meso scale data, covering a period of less than ten years.**

***Index Terms*—Year-to-year variability, seasonal patterns, extreme weather conditions.**

I. Introduction

Recent statistical method based approaches to modelling the year-to-year variability in daily extreme weather conditions on grape crop require data on yearly yield covering at least three decades. Even though New Zealand grapevine growing can be traced back to the late 19th century, grapevines have been planted at various stages but not always with success [1]. Hence, data available for a complete conventional statistical method based analysis, e.g.,χ2 method based (see section II), is considered to be inconsistent and inadequate.

On the other hand, grape crop prediction/ estimation has in recent years become vital for the success of viticulture and winemaking at all scales (vineyard and wine regions of a nation), and levels (vineyard operational, financial management and wine marketing) [2] [3]. In view of these facts, researchers at AUT's Geoinformatics Research Centre (GRC) investigated a few hybrid approaches to modelling the year-to-year variability in daily extreme weather conditions and its effects on grape crop using daily maximum, minimum and soil minimum temperatures, along with *Chardonnay* yearly grape yield data from a vineyard in northern New Zealand. In the next section, both, the iterative χ2 test based and GRC's hybrid method approaches to modelling climate effects on yield using daily extreme weather data at a meso scale are outlined.

Manuscript received May 30, 2012; revised July 30, 2012.

Subana Shanmuganathan and Philip Sallis are with Geoinformatics Research Centre, Auckland University of Technology, Private Bag 92006, Auckland 1142, New Zealand (e-mail: subana.shanmuganathan@aut.ac.nz; philip.sallis@aut.ac.nz).

Ajit.Narayanan is with School Computing and Mathematical Sciences, Auckland University of Technology, Private Bag 92006, Auckland 1142, New Zealand (e-mail: ajit.narayana@aut.ac.nz).

II. Recent Methods for Modelling Climate Effects on Grape Crop and Wine Quality

A. *Iterative χ2 Method Based Approach*

Using an iterative χ2 test approach as elaborated in [4] [5] Australian research [6] presented details on how the authors modelled the influences of daily extreme weather conditions on grapevine phenology and wine quality in four of the country's major wine regions. The Australian research concluded the approach as a useful way for establishing the nexuses between key weather variables and berry ripening/ wine quality processes as the authors described the available knowledge as "qualitative and fragmented".

The Australian study was carried out using data from four of Australia's major wine regions, namely, Hunter Valley, Margaret River, Coonawara and the Barossa Valley. The regional wine ratings were used in the study as surrogate for wine quality for comparing the frequency of defined weather conditions and the "high" (top 25%) and "poor" (bottom 25%) vintages at the regional scale. The results of this study produced the exact maximum (and minimum) temperatures associated with better quality wine in the different regions, such as 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 a quantitative assessment for establishing the timing and magnitude of weather influences on wine quality on a regional scale with data covering at least three decades.

B. *Data Mining and χ2 Method Based Approach*

The lack of sufficient long-term consistent data on grape crop i.e., covering a minimum of three decades to conduct statistical method/s based modelling led to the use of 1) an unsupervised artificial neural network based clustering and 2) an iterative χ2 test method approaches by GRC researchers. The results presented in [7] [8] show how such hybrid approaches could be used to extract useful information from data considered to be insufficient for any analysis purely based on conventional methodologies. The results produced the maximum and minimum (ambient air and soil) temperatures and frequencies (number of days) within the annual cycle of grapevine growth, linked to *Chardonnay* grapevine phenology and yearly "high"/"low" yields in the particular vineyard in northern New Zealand. However, there is a problem as to which method is the most appropriate when such long-baseline data is not available or, if available, not reliable or consistent.

The hybrid approach described in the next section focuses on between-variable differences rather than averages over a number of samples to produce results that may be useful for modellers and viticulturists. Between-variable differences are captured through the use of a range of data mining techniques commonly used by machine learning researchers.

III. THE HYBRID APPROACH

The section illustrates a different hybrid approach applied to modelling the effects of year-to-year variability in extreme daily temperature on yearly yields (in tons/area unit grapes) in three different grape varieties namely *Chardonnay*, *Pinot Noir* and *Pinot Gris*, as the yield data available is seen as insufficient for analysis with conventional/statistical methodologies. The daily maximum temperature data gathered and logged at Henderson River Pk, (36.85539S, 174.62383E), one of the National Institute of Water and Atmosphere's (NIWA's) metrological stations, and obtained from NIWA's web portal (http://cliflo.niwa.co.nz), is converted into a matrix of frequency/ number of days at 3^{o}C intervals during consecutive three week windows, for 45 weeks prior to harvest. The temperature data logged at Henderson River Pk is used as it is the closet station with data covering the whole time span (1998-2010). The hybrid approach used here to find knowledge on the links between weather conditions and the "high"/ "low" yields in the three varieties consists of the following:

- Data mining techniques, data clustering and CRT, C5 and QUEST tree classification tests for discovering patterns/ association rules within the pre-processed weather and grape yield data (three varieties separately and together).
- Iterative $\chi 2$ tests, to check on the significance of the patterns /rules observed (and interpretations arrived at) between the two sets of variables, yield and corresponding temperature frequencies, for each variety are run separately.
-

IV. DATA PREPARATION

A. Grape Crop Data

The grape crop data of the three different varieties analysed in this study was segregated into quartiles and the criteria for classifying "high (top 25%)" and "low (bottom 25%)" yield years are shown in (Fig. 1).

1) Chardonnay:

The 12 year (1998-2009) *Chardonnay* crop data describes yield in terms of tons/ha, Brix (Sugar content), pH, acid and harvest date. Of this data set yearly yields that fall within the top 25% quartile of yield and either top 25% quartile of Brix or acid are classified as "high" and similarly records that fall within the bottom 25% criteria are classified as "low" yield years respectively. Accordingly, 1999, 2002 and 2006 are classified as high and 2001 and 2008 as low yield years.

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum	Percentiles 25th	50th (Median)	75th
Yield	261	8.108	3.1945	2.2	13.1	6.100	8.200	10.100

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum	Percentiles 25th	50th (Median)	75th
Brix	96	21.695	1.0816	19.4	22.9	21.300	22.000	22.500

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum	Percentiles 25th	50th (Median)	75th
Acid	96	9.327	.8328	8.0	12.1	8.800	9.200	9.700

Fig. 1. Chardonnay yearly yield quartiles based on yield, Brix and acid

TABLE I: THE ARRANGEMENT OF CHANNELS

Variety	Yield Class	Yield ta/ha	TA	Years
Pinot Noir	Low	<4.74	>7.75	2005 & 2010
Pinot Noir	High	>8.24	>8.0	2004 & 2006
Pinot Gris	Low	<7.1	<7.2	2007 & 2009
Pinot Gris	High	>9.14	>8.0	2004 & 2006

B. Pinot Noir and Pinot Gris:

Similar to the way *Chardonnay* vineyard yearly yield production data was classified, that of *Pinot Noir* (2003-2010) and *Pinot Gris* (2004-2010) was also classified into "high" and "low" yield years based on yield and TA (titratable acidity or "total" acidity). The top and bottom quartiles (25%) of yield and TA, "high" and "low" yield years classified for *Pinot Noir* and *Pinot Gris* are listed in table I.

C. Weather Data

The NIWA's daily maximum weather is used to create matrices of numbers of days recorded in each of the continuous classes at 3^{o}C intervals (8.1-11^{o}C, 11.1¬14^{o}C, 14.1-17^{o}C, 17.1-20^{o}C, 20.1-23^{o}C, 23.1-26^{o}C, 26.1-29^{o}C and 29.1-32^{o}C) within moving 3 week windows, each window in succession adding a new week and dropping the first one as the window advanced. Time span of each matrix is 45 weeks prior to harvest and separate matrices have been created for daily maximum temperature data for all the three varieties based on [8].

V. THE RESULTS

The weather data converted into a matrix of week against temperature was analysed using Kohonen self-organising map (SOM) based clustering, CRT, C5 and QUEST based tree classification in Clementine and are discussed here.

A. Kohonen SOM Based Clustering

Kohonen SOM clustering produced all temperature ranges except for (32.1-35^{o}C) as significant in the clustering for all three varieties.

B. CRT Analysis

A set of rules generated by CRT analysis on the weather frequency data and "high"/ "low" yield years for *Pinot Noir* from this vineyard, is listed in Table II.

TABLE II: CRT RULES FOR *PINOT NOIR* "HIGH" AND "LOW" YIELD YEARS

Rules for PNH - contains 5 rule(s)
Rule 1 for PNH (22; 1.0)
if 11.1-14 <= 1.5 and week <= 17.5 and 20.1-23 <= 12.5 and 20.1-23 <= 7.5
and dates in ["(2003 5 14-6 3)" "(2003 5 21-6 10)" "(2003 5 28-6 17)" "(2003 6 11-7 1)" "(2003 6 4-6 24)" "(2003 7 23-8 12)" "(2003 7 30-8 20)" "(2005 4 19-5 9)" "(2005 4 26-5 16)" "(2005 5 10-5 30)" "(2005 5 17-6 6)" "(2005 5 24-6 13)" "(2005 5 3-5 23)" "(2005 5 31-6 20)" "(2005 6 14-7 4)" "(2005 6 21-7 11)" "(2005 6 28-7 18)" "(2005 6 7-6 27)" "(2005 7 12-8 1)" "(2005 7 19-8 8)" "(2005 7 26-8 15)" "(2005 7 5-7 25)"] then PNH
Rule 2 for PNH (7; 1.0)
if 11.1-14 <= 1.5 and week <= 17.5 and 20.1-23 <= 12.5 and 20.1-23 > 7.5
and dates in ["(2003 4 16-5 6)" "(2003 4 23-5 13)" "(2003 4 30-5 20)" "(2003 4 9-4 29)" "(2003 5 7-5 27)" "(2005 4 12-5 2)" "(2005 4 5-4 25)"] then PNH
Rule 3 for PNH (5; 1.0)
if 11.1-14 <= 1.5 and week > 17.5 and 20.1-23 <= 1.5
and dates in ["(2005 8 2-8 22)" "(2005 8 9-8 29)" "(2006 1 24-2 13)" "(2006 1 31-2 20)" "(2006 2 7-2 27)"] then PNH
Rule 4 for PNH (36; 0.833)
if 11.1-14 <= 1.5 and week > 17.5 and 20.1-23 > 1.5 and 17.1-20 <= 12.5 and 20.1-23 <= 6.5 then PNH
Rule 5 for PNH (8; 1.0)
if 11.1-14 > 1.5 and week > 10.5
and dates in ["(2003 6 18-7 8)" "(2003 6 25-7 15)" "(2003 7 16-8 5)" "(2003 7 2-7 22)" "(2003 7 9-7 29)" "(2003 8 13-9 3)" "(2003 8 21-9 10)" "(2003 8 6-8 27)"]
then PNH

Rules for PNL - contains 8 rule(s)
Rule 1 for PNL (3; 1.0)
if 11.1-14 <= 1.5 and week <= 17.5 and 20.1-23 <= 12.5 and 20.1-23 <= 7.5
and dates in ["(2004 5 27-6 16)" "(2004 6 3-6 23)" "(2009 7 30-8 19)"] then PNL
Rule 2 for PNL (4; 1.0)
if 11.1-14 <= 1.5 and week <= 17.5 and 20.1-23 <= 12.5and 20.1-23 > 7.5
and dates in ["(2004 4 29-5 19)" "(2004 5 6-5 26)" "(2009 4 16-5 6)" "(2009 4 9-4 29)"] then PNL
Rule 3 for PNL (2; 1.0)
if 11.1-14 <= 1.5 and week <= 17.5 and 20.1-23 > 12.5 then PNL
Rule 4 for PNL (16; 1.0)
if 11.1-14 <= 1.5 and week > 17.5 and 20.1-23 <= 1.5
and dates in ["(2004 9 16-10 6)" "(2004 9 2-9 22)" "(2004 9 23-10 13)" "(2004 9 9-9 29)" "(2005 1 20-2 9)" "(2005 1 27-2 16)" "(2005 2 10-3 2)" "(2005 2 17-3 9)" "(2005 2 3-2 23)" "(2009 8 13-9 2)" "(2009 8 20-9 9)" "(2009 8 6-8 26)" "(2009 9 24-10 14)" "(2010 1 28-2 17)" "(2010 2 11-3 3)" "(2010 2 4-2 24)"] then PNL
Rule 5 for PNL (40; 0.575)
if 11.1-14 <= 1.5 and week > 17.5 and 20.1-23 > 1.5 and 17.1-20 <= 12.5 and 20.1-23 > 6.5 then PNL
Rule 6 for PNL (9; 0.889)
if 11.1-14 <= 1.5 and week > 17.5 and 20.1-23 > 1.5and 17.1-20 > 12.5 then PNL
Rule 7 for PNL (12; 1.0)
if 11.1-14 > 1.5 and week <= 10.5 then PNL
Rule 8 for PNL (16; 1.0)
if 11.1-14 > 1.5 and week > 10.5
and dates in ["(2004 6 24-7 14)" "(2004 7 1-7 21)" "(2004 7 15-8 4)" "(2004 7 22-8 11)" "(2004 7 29-8 18)" "(2004 7 8-7 28)" "(2004 8 12-9 1)" "(2004 8 19-9 8)" "(2004 8 26-9 15)" "(2004 8 5-8 25)" "(2009 6 18-7 8)" "(2009 6 25-7 15)" "(2009 7 16-8 5)" "(2009 7 2-7 22)" "(2009 7 23-8 12)" "(2009 7 9-7 29)"] then PNL

Default: PNH

a. PNH: Pinot Noir *high* yield PNL: Pinot Noir *low* yield

The CRT results give precise indications of winter temperature that led to "high"/ "low" yields. For example, **Rule 5 for PNH (8; 1.0)** if 11.1-14 > 1.5 and week > 10.5 and dates in ["(2003 6 18-7 8)" "(2003 6 25-7 15)" "(2003 7 16-8 5)" "(2003 7 2-7 22)" "(2003 7 9-7 29)" "(2003 8 13-9 3)" "(2003 8 21-9 10)" "(2003 8 6-8 27)"] then PNH.

This rule 5 could be interpreted as that the daily maximum temperature between 11.1-14°C over a day and a half per year (dpy) in June-August (2003) led to "high" yield years for *Pinot Noir*.

The interpretations arrived at from the CRT rules produced for the Kumeu data set are presented in Table II.

C. Pinot Noir High Yield Years- 5 Rule(s)

Daily maximum Temperature between 11.1-14 (<= 1.5 days per year (dpy))

1. and 20.1-23 °C (< 7.5 dpy) during April-August
2. and 20.1-23 °C (7.5-12.5 dpy) during April-May
3. and 20.1-23 °C (<=1.5 dpy) during August and late January – February
4. and 20.1-23 °C (<=1.5 dpy) during August and late January - February

Daily maximum Temperature between 11.1-14 (>1.5 dpy)

5. during June - early September → high yield years

D. Pinot Noir Low Yield Years- 5 Rule(s)

Daily maximum Temperature between 11.1-14 (<= 1.5 dpy)

1. and 20.1-23 °C (<= 12.5 dpy) and 20.1-23 oC (<=7.5 dpy) in May-mid August)
2. and 20.1-23 °C (<= 12.5 dpy) and 20.1-23 (> 7.5 dpy) during April- May
3. and 20.1-23 (> 12.5 dpy) in April-May (week 17.5)
4. and 20.1-23 °C (<= 1.5 dpy) in August - October and February-March
5. and 20.1-23 °C (> 1.5 dpy) and 17.1-20 (<= 12.5 dpy) and 20.1-23 °C (dpy) > 6.5 in April-May (week 17.5)
6. and 20.1-23 °C (> 1.5 dpy) and 17.1-20 (> 12.5 dpy) during April-May (week 17.5)
7. during June-July (week <= 10.5)
8. during June-August --> low yield years

All of the above CRT rules relating to both "high" and "low" yield years for *Pinot Noir* (Table II) are concerned with winter cold and provide details on how daily maximum temperature affects the vine during its dormancy period. Just one rule (no.4) describes this variable over January-February flowering period for "high" yield years. None of the other rules provides any information on berry ripening period. This is the case with analyses on *Chardonnay*, *Pinot Noir* and *Pinot Gris* yields against daily maximum temperature data. Hence, to see the daily temperature effects during the ripening period, CRT analysis was ran only for frequencies within the five temperature ranges during week 30-45 for all the three grape varieties together and the rules provide details of November-February (flowering and berry ripening). Based on the results (Table III) for *Chardonnay*- as per

Rule 1: daily maximum temperature 14.1-17°C <= 0.5 dpy and 32.1-35 °C <= 0.5 dpy and 23.1-26 °C > 11.5 dpy in January –February led to "high" yield.

Rule 2:
if 14.1-17 > 0.5 dpy in November-December again led to "high" yield.

Similarly, rules relating to temperature, magnitude and timing that led to "high" and "low" yield years in the three varieties studied are listed in Table III.

TABLE III: CRT RULES FOR *CHARDONNAY*, *PINOT NOIR* AND *PINOT GRIS* DURING WEEK 30-45 PRIOR TO HARVEST

Rules for CH - contains 2 rule(s)
Rule 1 for CH
if 14.1-17 <= 0.5 and 32.1-35 <= 0.5 and 23.1-26 > 11.5
and dates in ["(1999 1 29-2 18)" "(2001 12 7-12 27)" "(2002 1 11-1 31)" "(2002 1 4-1 24)" "(2002 2 1-2 21)" "(2002 2 15-3 7)" "(2002 2 8-2 28)" "(2005 12 2-12 22)" "(2006 12 16-1 5)" "(2006 12 23-1 12)" "(2006 12 30-1 19)"] then CH
Rule 2 for CH
if 14.1-17 > 0.5
and dates in ["(1998 11 13-12 3)" "(1998 11 20-12 10)" "(1998 11 27-12 17)" "(2000 10 26-11 15)" "(2000 11 2-11 22)" "(2001 11 2-11 22)" "(2003 10 30-11 19)" "(2006 11 10-11 30)"]
and dates in ["(1998 11 13-12 3)" "(1998 11 20-12 10)" "(1998 11 27-12 17)" "(2001 11 2-11 22)"] then CH
Rules for CL - contains 3 rule(s)
Rule 1 for CL
if 14.1-17 <= 0.5 and 32.1-35 <= 0.5 and 23.1-26 <= 11.5 and 26.1-29 > 8.5
Rule 2 for CL
if 14.1-17 <= 0.5 and 32.1-35 <= 0.5 and 23.1-26 <= 11.5 and 26.1-29 > 14.5
and dates in ["(2001 2 8-2 28)" "(2007 2 16-3 8)" "(2008 1 10-1 30)" "(2008 1 24-2 13)"] then CL
Rule 3 for CL
if 14.1-17 > 0.5
and dates in ["(1998 11 13-12 3)" "(1998 11 20-12 10)" "(1998 11 27-12 17)" "(2000 10 26-11 15)" "(2000 11 2-11 22)" "(2001 11 2-11 22)" "(2003 10 30-11 19)" "(2006 11 10-11 30)"]
and dates in ["(2000 10 26-11 15)" "(2000 11 2-11 22)" "(2003 10 30-11 19)" "(2006 11 10-11 30)"] then CL

Rules for PGH - contains 2 rule(s)
Rule 1 for PGH
if 14.1-17 <= 0.5 and 32.1-35 <= 0.5 and 23.1-26 <= 11.5 and 26.1-29 <= 14.5 and 23.1-26 <= 8.5then PGH
Rule 2 for PGH
if 14.1-17 <= 0.5 and 32.1-35 <= 0.5 and 23.1-26 > 11.5
and dates in ["(2003 12 4-12 24)" "(2005 11 29-12 19)" "(2006 12 20-1 9)" "(2006 12 27-1 16)" "(2007 1 12-2 1)" "(2007 1 19-2 8)" "(2009 12 12-1 1)" "(2009 2 20-3 12)"]
and dates in ["(2003 12 4-12 24)" "(2005 11 29-12 19)" "(2006 12 20-1 9)" "(2006 12 27-1 16)"] then PGH
Rules for PGL - contains 2 rule(s)
Rule 1 for PGL
if 14.1-17 <= 0.5 and 32.1-35 <= 0.5 and 23.1-26 > 11.5
and dates in ["(2003 12 4-12 24)" "(2005 11 29-12 19)" "(2006 12 20-1 9)" "(2006 12 27-1 16)" "(2007 1 12-2 1)" "(2007 1 19-2 8)" "(2009 12 12-1 1)" "(2009 2 20-3 12)"]
and dates in ["(2007 1 12-2 1)" "(2007 1 19-2 8)" "(2009 12 12-1 1)" "(2009 2 20-3 12)"] then PGL
Rule 2 for PGL
if 14.1-17 <= 0.5 and 32.1-35 > 0.5 then PGL

Rules for PNL - contains 2 rule(s)
Rule 1 for PNL
if 14.1-17 <= 0.5 and 32.1-35 <= 0.5 and 23.1-26 <= 11.5 and 26.1-29 > 14.5
and dates in ["(2010 1 28-2 17)" "(2010 2 11-3 3)" "(2010 2 4-2 24)"] then PNL
Rule 2 for PNL
if 14.1-17 > 0.5
and dates in ["(2004 11 11-12 1)" "(2004 11 18-12 8)" "(2004 11 25-12 15)" "(2004 12 2-12 22)" "(2004 12 9-12 29)" "(2005 12 16-1 5)" "(2009 10 29-11 18)" "(2009 11 19-12 9)" "(2009 11 26-12 16)" "(2009 12 3-12 23)"] then PNL

Default: CH

CH/CL *Chardonnay high/low* PGH/PGL *Pinot Gris high/low* PNH/PNL *Pinot Noir high/low*

$\chi 2$ tests were ran on the total values of "high" and "low" yield year temperature frequencies to establish any correlations between the variables being analysed. The $\chi 2$ associations at *p-value* <=0.05 confidence interval could be interpreted as that there is a correlation between the two variables, the temperature at the respective frequency during the respective week and "high"/"low" yield for the respective grape variety. This enables the rejection of the null hypothesis, implying that the correlation established could be considered as true and occurring not by chance.

In the data set being studied in this research, even though there were eight 3°C daily maximum temperature intervals (8.1-11°C, 11.1¬14 °C, 14.1-17°C, 17.1-20°C, 20.1-23°C, 23.1-26°C, 26.1-29°C and 29.1-32°C) for 45 weeks prior to harvest, all of them were not included in the $\chi 2$ analysis as the test cannot be run with 0 frequencies. To avoid this issue all under 23°C temperature classes were combined to form a <23°C and similarly a >26°C class was formed adding all above 26°C for the berry ripening period i.e., 16 weeks prior to harvest (week 30-45), as performed in the earlier studies.

The $\chi 2$ results in Table IV give the ideal temperature range, week and magnitude linked to "high"/ "low" yields in the three grape varieties cultivated in Kumeu. New Zealand.

TABLE IV: $\chi 2$ RESULTS SHOWING DAILY MAXIMUM TEMPERATURE RANGE, MAGNITUDE AND TIMING RELATING TO DIFFERENT GRAPE VARIETIES GROWN IN KUMEU, NEW ZEALAND

PG	Wk	<23°C	23-26 °C	>26 °C	X2	p-value	h/l
13N-4De	30	13	7.5	0.5	3.86	0.050	low
4-24Dec	33	12	7.5	0.5	6.82	0.009	low
1-20Jan	37	6	7	8	4.00	0.046	low
12-22Feb	43	6	7.5	7.5	4.00	0.046	high
19F-11M	44	9.5	7.5	4	5.54	0.190	high
		3.5	7	10.5	5.83	0.016	low
26F-18M	45	11.5	8	1.5	4.24	0.400	high
				9.5	11.64	0.001	low
early Feb mid March (berry ripening) → < 23 °C produces high yield and >26 °C leads to low yield							
PN	**wk**	**<23 °C**	**23-26 °C**	**>26 °C**	**X2**	**p-value**	**h/l**
12N-1Dec	32	17	3.5	0.5	4.50	0.034	high
19N-8Dec	33	13	7.5	0.5	9.94	0.002	high
26N-15De	34	17.5	2.0	1.5	4.74	0.029	low
		9.5	8.5	3.0	8.05	0.005	high
3-23 Dec	35	16.0	4.0	1.0	12.90	0.000	low
	35	4.5	12.0	4.5	8.00	0.005	high
					4.46	0.035	high
17D-6Jan	37	11.5	8.0	1.5	4.24	0.400	low
		5.5	9.5	6.0	5.40	0.020	high
24D-13Ja	38	5.0	8.5	7.5	6.37	0.012	high
31D-20Ja	39	7	9	5	5.56	0.018	low
		2	8	11	4.50	0.034	high
4-24Feb	44	4	6.5	10.5	5.44	0.020	high
11F-3Mar	45	6	7.5	7.5	9.32	0.002	high
early Feb early March (berry ripening) → < 23 °C produces high yield							
Char	**wk**	**<23 °C**	**23-26 °C**	**>26 °C**	**X2**	**p-value**	**h/l**
11-31No	31	18.0	13.0	0.5	7.11	0.008	high
18N-7De	32	15.0	13.0	3.5	6.08	0.014	high
2-22Dec	34	16.0	10.5	5.0	4.59	0.032	high
9-29Dec	35	12.0	15.5	4.0	5.77	0.016	high
16D-5Ja	36	9.0	15.5	7.0	7.35	0.007	high
20J-9Feb	41	4.5	6.5	9.0	6.40	0.011	low
		0.5	11.0	20.0	8.35	0.004	high
27J-6Feb	42	4.5	9.0	6.5	6.40	0.011	low
		0.5	13.0	18.0	10.8	0.001	high
3-23 Feb	43	2.5	17.5	11.5	5.45	0.020	high
17F-9 Ma	45	3.0	6.5	13.5	4.12	0.042	low
		9.0	15.5	7.0	6.0	0.014	high
					7.35	0.007	high
early Feb early March (berry ripening) → < 26°C produces high yield							

Ja: January F/Fe: February N: Nov D/De/Dec: December

PN: *Pinot Noir*; Char: *Chardonnay*

Interestingly, both CRT rules and $\chi 2$ test results complement each other. For example, rule 2 for *Chardonnay* 14.1-17°C <= 0.5 dpy and 32.1-35°C <= 0.5 dpy and 23.1-26°C > 11.5 dpy in January-February led to "high" yield (Table III). The $\chi 2$ test results showed that daily maximum >26°C 20.0 dpy over this period as related to "high" yield (Table IV with $\chi 2$ value = 8.35 at *p-value* 0.004).

VI. Conclusions

The paper discussed of a hybrid approach of data mining (CRT analysis based tree classification method) and iterative $\chi 2$ test based methodology, investigated to modelling the effects of climate change in weather conditions (year-to-year variability) and on grape production in different varieties, namely, *Chardonnay*, *Pinot Noir* and *Pinot Gris*, from vineyards in northern New Zealand. The grapevine and winemaking industry of New Zealand lack continuous data on crop i.e., at least covering a period of three decades, and the situation impedes the application of any purely conventional data analysis based modelling methods. In order to overcome the issue, AUT's GRC researchers applied a hybrid approach that produced the temperature, magnitude and the timing that seemed to be linked to "high"/"low" yearly grape production in the three wine varieties analysed.

Based on the $\chi 2$ test results <26°C during early February-early March (berry ripening) favours "high" yield in *Chardonnay*. Similarly, <23°C favours *Pinot Noir* and *Pinot Gris* berry ripening in the vineyards studied in this research. It is also evident that during this time interval >26°C (over 9.5 dpy) over a three weeks time led to "low" yield in *Pinot Gris* but not on *Pinot Noir*, the latter is seem to be tolerant up to over 10.5 dpy. On the hand, for *Chardonnay* it is 13.5 dpy, more tolerant to high temperature (>26°C) during this season.

Finally, based on the results of data mining and iterative $\chi 2$ test methods it can be concluded that even though the grape crop data is insufficient i.e., less than 10 years, to carry out conventional/ rigorous data analysis methodologies, the hybrid approach investigated produced useful information on the temperatures over winter dormancy and berry ripening periods that favour "high" yield in the grape varieties chosen in the study, cultivated in northern New Zealand.

Acknowledgment

The authors wish to thank Winemaker Michael Brajkovich of Kumeu River Wines and Peter Sumich of Auckland University of Technology (AUT) for support given to conduct this research.

References

[1] M. Cooper. The Principal Wine Regions of New Zealand. Book Chanter in Wine Atlas of New Zealand (second edition). Hodder Moa, 2008 pp 50-51

[2] I. Dami., Methods of Crop Estimation in Grapes. 2010. [Online]. Available: www.oardc.ohio-state.edu/grapeweb/OGEN/07262006/CropEstimation06.pdf.

[3] M.Trought., Yield Management and Prediction. Science Report, Marlborough Wine Research Centre, Yield modelling of Sauvignon Blanc in Marlborough (extracted from Annual Report 2005-06) 2009. [Online]. Available: www.wineresearch.org.nz/projects/SR06-01aYieldManagmentPrediction.pdf.pdf

[4] J. M. Caprio and H. A. Quamme, Weather conditions associated with apple production in the Okanagan Valley of British Columbia. in *Canadian Journal of Plant Science 79*, 1999, pp. 129-137.

[5] J. M. Caprio and H A. Quamme, Weather conditions associated with grape production in the Okanagan Valley of British Columbia and potential impact of climate production in the Okanagan Valley of British Columbia and potential impact of climate change in *Canadian Journal of Plant Science 82*, 2002, pp.755-763.

[6] C. J. Soar, V. O. Sadras, and P. R. Petrie, Climate drivers of red wine quality in four contrasting Australian wine regions. *Australian Journal of Grape and Wine Research 14*, 2008. pp.78–90.

[7] S. Shanmuganthan, P. Sallis, and A Narayanan, Modelling the effects of daily extreme weather on grapevine and wine quality. David A. Swayne, Wanhong Yang, A. A. Voinov, A. (Eds.) in *proc. of the International Environmental Modelling and Software Society (iEMSs), 2010 International Congress on Environmental Modelling and Software Modelling for Environment's Sake*, Fifth Biennial Meeting, Ottawa, Canada. 2010 6pp. also available at http://aut.researchgateway.ac.nz/handle/10292/1723

[8] S Shanmuganathan, P Sallis, and A Narayanan, (2010) Data Mining Techniques for Modelling the Influence of Daily Extreme Weather Conditions on Grapevine Yield and Wine Quality in *proc. 2nd Int Conf. on Computational Intelligence, Communication Systems and Networks* (CICSyN 2010) Liverpool, UK, 2010. Published by IEEE 6pp.