# AROMA AND QUALITY ASSESSMENT FOR VERTICAL VINTAGES USING MACHINE LEARNING MODELLING BASED ON WEATHER AND MANAGEMENT INFORMATION

Authors:Sigfredo FUENTES<sup>1\*</sup>, Claudia GONZALEZ VIEJO<sup>1</sup>, Xiaoyi WANG<sup>1</sup> and Damir D. TORRICO<sup>1</sup>

<sup>1</sup> School of Agriculture and Food, Faculty of Veterinary and Agricultural Sciences, University of Melbourne, VIC 3010, Australia

## \*Corresponding author: sfuentes@unimelb.edu.au

## Abstract:

**Context and purpose of the study** - Wine quality traits are usually given by parameters such as aroma profile, total acidity, alcohol content, colour and phenolic content, among others. These are highly dependent on the weather conditions during the growing season and management strategies. Therefore, it is important to develop predictive models using machine learning (ML) algorithms to assess and predict wine quality traits before the winemaking process.

Material and methods - Samples in duplicates of Pinot Noir wines from vertical vintages (2008 to 2013) of the same winery located in Macedon Ranges, Victoria, Australia were used to assess different chemical analytics such as i) aromas using gas chromatography – mass spectrometry, ii) color density, iii) color hue, iv) degree of red pigmentation, v) total red pigments, vi) total phenolics, vii) pH, viii) total acidity (TA), and ix) alcohol content. Data from weather conditions from the specific vintages were obtained both from the bureau of meteorology (BoM) and the Australian Wine Availability Project (AWAP) climate databases. Such data consisted of: i) solar exposure from veraison to harvest (V-H), ii) solar exposure from September to harvest (S-H), iii) maximum January solar exposure, iv) degree days from S-H, v) maximum January evaporation, vi) mean maximum temperature from veraison to harvest, vii) mean minimum temperature from V-H, viii) water balance from S-H, ix) solar exposure from V-H, x) degree hour accumulation with base 25 - 30 °C, xi) degree hour accumulation with base 25 °C, xii) degree hour accumulation with base 30 °C, xiii) degree hour accumulation with base 35 °C, and xiv) total cumulative degree days accumulation with base 10 °C. All data were used to develop two machine learning (ML) regression models using Matlab® R2018b. The best models obtained were using artificial neural networks (ANN) with the Levenberg-Marquardt algorithm with 5 neurons for Model 1 and 9 neurons for Model 2. Model 1 was developed using the 14 parameters from the weather data as inputs to predict 21 aromas found in the wines from the six different vinatges. Model 2 was developed using the same 14 parameters from weather data and the eight chemical parameters as targets and outputs.

**Results** - Both models obtained presented very high accuracy to predict wine quality trait parameters. Model 1 had an overall correlation coefficient R = 0.99 with a high performance based on the mean squared error (MSE = 0.01), while Model 2 had an overall correlation coefficient R = 0.98 with a high performance (MSE = 0.03). These models would aid in the prediction of wine quality traits before its production, which would give anticipated information to winemakers about the product they would obtain to make early decisions on wine style variations.

Keywords: wine quality; machine learning; weather; aromas

## 1. Introduction

Red wine has a wide variety of aromatic volatile compounds, which, along with some chemical parameters such as total acidity, pH, alcohol content, color and phenolic content, among others are responsible for the specific wine quality. These components depend on many different factors such as cultivar, weather, type of soil, sun exposure, water status and canopy, among others (González-Barreiro, Rial-Otero, Cancho-Grande, & Simal-Gándara, 2015). According to Jackson and Lombard (1993), a lower TA and higher pH are developed under warmer temperatures, which provide a wine with lower quality traits. Likewise, at higher temperatures lower aromatic volatile compounds are produced, which also leads to lower quality wines. Regarding the effect of temperature in red grapes color, it has been reported that the best temperature range to develop anthocyanins is 17 - 26 °C as lower or higher temperatures lead to lower color pigmentation (Drappier, Thibon, Rabot, & Geny-Denis, 2019; Jackson & Lombard, 1993; Pirie, 1977).

Up to date, there are no studies involving the prediction of chemical and aromatic compounds in wine based on weather conditions to predict the final wine quality using machine learning. Therefore,

the aim of this study was to develop machine learning (ML) models using weather information from vinatges to predict wine quality traits before the winemaking process in vertical vintages (2008 – 2013) of Pinot Noir wines from the same winery. This was achieved by developing two artificial neural network (ANN) models using 14 weather-related parameters from the bureau of meteorology (BoM) and the Australian Wine Availability Project (AWAP) climate database to predict i) 21 aromatic volatile compounds and ii) eight chemical parameters.

### 2. Material and methods

#### Study area

All the wines and weather data used for this study were taken for Pinot Noir cultivar and vintages from 2008 – 2013 and located in Macedon Ranges in the sub-region of Romsey/Lancefield, south of the Great Dividing Range in Victoria, Australia. This site has an elevation of 540 m.a.s.l and is located at distance from the mitigating influence of the ocean.

#### Weather data acquisition

For all vintages (2008 – 2013), weather data conditions such as i) solar exposure from veraison to harvest (V-H), ii) solar exposure from September to harvest (S-H), iii) maximum January solar exposure, iv) degree days from S-H, v) maximum January evaporation, vi) mean maximum temperature from V-H and vii) mean minimum temperature from V-H were obtained from the BoM. While other parameters such as viii) water balance from S-H (WB-S-H), ix) solar exposure from V-H, x) accumulation of degree hours with base 25 - 30 °C, xi) degree hours, base 25 °C, xii) degree hours, base 30 °C, xiii) degree hours, base 35 °C, and xiv) total cumulative degree days with base 10 °C were obtained from the Australian Water Availability Project (AWAP) database.

#### Gas Chromatography – Mass Spectroscopy

Pinot Noir wine samples from all vintages (2008 – 2013) were analyzed for volatile compounds using a gas chromatograph 6850 series II with mass selective detector 5873 (GC-MSD; Agilent Technologies, Inc., Santa Clara, CA, USA) along with an autosampler PAL 120 (CTC Analytics AG, Zwingen, Switzerland). A J&W DB-WAX column (Agilent Technologies, Inc., Santa Clara, CA, USA) with a length of 30 m, inner diameter 0.25 mm and 0.25 µm film was used. Helium (ultrahigh purity, BOC Australia, North Ryde, NSW, Australia) was used as carrier gas at a constant flow of 2.0 ml/min. Samples were injected using the headspace method with a solid-phase microextraction (SPME) polydimethylsiloxane (PDMS) 100 µm fiber (Agilent Technologies, Inc., Santa Clara, CA, USA). A total of 1 mL of each sample was diluted in 9 mL of Milli-Q water along with 2 g of sodium chloride and 200  $\mu$ L of 4-octanol as internal standard (10 mg  $L^{-1}$ ) and ethyl nonanoate for quality control (10 mL  $L^{-1}$ ). This was poured into a 20 mL screw cap vial with an 18 mm magnetic screw cap with a polytetrafluoroethylene (PTFE) and silicone liner, this vial was agitated at 220 rpm at 35 °C for 10 min. A splitless mode was used for the inlet at 50:1 and opened after 30 s. The oven was set at 40 °C for 4 min and brought up to 220 °C for 20 min. The National Institute of Standards and Technology (NIST; National Institute of Standards and Technology, Gaithersburg, MD, United States) library was used to identify compounds. The peak area of the identified compounds was obtained.

#### Chemical measurements

The pH of all wine samples was measured with 50 mL of wine using a HI 5221 pH-meter (Hanna Instruments, Keysborough Vic, Australia), the instrument was previously calibrated with buffer solutions for pH 7, 4 and 10. Furthermore, titratable acidity (TA) was measured through alkaline titration with sodium hydroxide 0.1 N (Sigma-Aldrich Chemicals company, St. Louis, MO, USA) according to the Association of Official Agricultural Chemists (AOAC) Official Method 962.12. On the other hand, a M501 single beam scanning UV/visible spectrophotometer (Spectronic Camspec Ltd, Leeds, UK) was used to calculate wine color density, wine color hue, degree of red pigmentation, total red pigments and total phenolics. Wine color density (CD) was calculated using the absorbance values at 420, 520 and 620 nm using the equation  $CD = A_{420} + A_{520} + A_{620}$ , while for wine color hue (CH) used the equation  $CH = A_{420} / A_{520}$ . Total phenolics were calculated using the absorbance value at 740 nm and the diluted sample with gallic acid (Sigma-Aldrich Chemicals company, St. Louis, MO, USA) at a dilution factor of 6. Total red pigments and degree of red pigmentation were measured using the absorbance values at 520 nm using 1 *M* of hydrochloric acid (Sigma-Aldrich Chemicals company, St. Louis, MO, USA) to acidify the sample to pH 1. Alcohol was measured using an ebulliometer.

## Statistical analysis

Two ML models were obtained, having the best performance, using ANN, specifically the Levenberg-Marquardt training algorithm. Model 1 was developed to predict 21 aromatic volatile compounds using as inputs the 14 parameters from weather conditions described. On the other hand, Model 2 was developed with the same 14 weather-related parameters as inputs to predict eight chemical components. Both models were developed using random data division with 70% of the samples for training, 15% for validation with a mean squared error performance algorithm and 15% for testing using a default derivative function. A trimming was conducted using 3, 5, 9 and 10 neurons to find the best models with no over- or underfitting, finding that 5 neurons were best for Model 1 and 9 neurons for Model 2. Figure 1 shows the diagrams for both models. Statistical data obtained for the models consisted of correlation coefficient (R), slope (b) and performance as means squared error (MSE).

#### 3. Results and discussion

#### 3.1. Aromatic Volatile Compounds

Table 1 shows the aromatic volatile compounds found in the wine samples and the aromas associated to them. Overall, all wine samples measured, presented the highest peaks for fruity, grape, banana, cognac and mushroom aromas (ethyl acetate, ethyl nonanoate and isoamyl alcohol). It can be observed that the 2008 vintage had the highest peak area for the 1-hexanol, which is associated with pungent, alcoholic and fruity aromas, but was the lowest in ethyl octanoate (musty / creamy / mushroom) and ethyl decanoate (fruity / apple) when compared to the other samples. Vintage 2009 was the highest in fruity / grape (ethyl acetate), fruity / strawberry (Butanoic acid, 2-methyl-, ethyl ester), sweet apple / pineapple (ethyl isovalerate), banana / cognac (isoamyl alcohol and isoamyl acetate), green / citrus / waxy (1-octanol) and rose / floral and bready (phenethyl alcohol). On the other hand, vintage 2010 was the highest in diethyl succinate (cooked apple). The highest peak areas for ethyl butyrate, ethyl heptanoate and isobutanol were found in vintage 2011, while for vintage 2012 the highest peaks were ethyl hexanoate, ethyl octanoate, ethyl decanoate, hexyl acetate, ethyl lactate, and 1-heptanol. Vintage 2013 presented the highest peak area for cognac / apple / winey (ethyl nonanoate), floral / sweet (Octanoic acid, 2-methyl-, ethyl ester) and floral / rose (benzyl alcohol).

Volatile compound	Aroma*	oma* 2008 2009		2010	2011	2012	2013
Ethyl acetate	Fruity / Grape	1985 2159 1927 2477		2000	2069		
Ethyl butyrate	Sweet / Fruity / Apple	Sweet / Fruity / 103 175 60 190 Apple		151	138		
Ethyl hexanoate	Fruity / Pineapple	927 1492 796 1693		1693	1708	1191	
Ethyl heptanoate	Pineapple / Berry	19 19 19 20		17	17		
Ethyl octanoate	Musty / Creamy / Mushroom	1222 1570 1379 206		2061	2200	2061	
Ethyl nonanoate	Cognac / Apple / Winey	5399 5113 6		6090	6182	5932	6310
Ethyl decanoate	Fruity / Apple	85	93	140	143	194	183
Butanoic acid, 2-methyl-, ethyl ester	Fruity / Strawberry	112	161	59	87	72	80
Ethyl isovalerate	Sweet apple / Pineapple	102	208	75	109	64	87
Isobutanol	Winey	281	332	277	386	217	370
Isoamyl alcohol	Banana / Cognac	4446	5113	3604	4760	3471	4262
Hexyl acetate	Apple / Pear	.pple / Pear 12 14		16	7	17	11
Ethyl lactate	Fruity / Butter	200	173	202	221	246	191
Octanoic acid, 2-methyl-, ethyl ester	Floral / Sweet	139	32	48	47	49	162

Table 1. Peak area of the aromatic volatile compounds found in the wine samples and the aroma associated

Diethyl succinate	Cooked apple	273	283	293	166	269	234
Isoamyl acetate	Banana	450	542	364	451	327	499
1-Hexanol	Pungent / Fruity / Alcoholic	387	301	257	370	313	243
1-Heptanol	Pungent / Leafy green	27	26	20	38	43	19
1-Octanol	Green / Citrus / Waxy	25	28	15	22	20	16
Benzyl alcohol	Floral / Rose	17	14	27	23	25	28
Phenethyl alcohol	Rose / Floral / Bready	102	125	102	90	85	102

\*The aromas were obtained from The Good Scents Company (2018)

Abbreviations: W = wine, the number in the samples indicate the vintage

The aromatic compounds identified in the wine samples analyzed have been previously reported for Pinot Noir wines (Fang & Qian, 2005; Girard, Yuksel, Cliff, Delaquis, & Reynolds, 2001; Song et al., 2015). All compounds found belong to esters and alcohols groups, which have been reported as the largest groups of volatile compounds in red wines (Girard et al., 2001; Torrea et al., 2011). The esters group provide fruit-related aromas, while alcohols are more pungent, alcoholic and vegetal-like aromas such as floral and green (Dashko et al., 2015; The Good Scents Company, 2018; Ugliano, Travis, Francis, & Henschke, 2010). The 2011 vintage had the largest peak area of total esters associated lower temperatures (9 – 19 °C) and solar exposure and highest water balance, as it was a rainy year. This is in accordance with Jackson and Lombard (1993) who reported that higher fruit aroma intensities are found at mean temperatures between 9 and 20 °C. On the other hand, the wine sample from 2008 presented the largest peak are of total alcohols, regarding the weather conditions, this season presented the highest degree days S-H and highest maximum January evaporation, this coincides with the findings from Bonada et al. (2015), who reported higher floral-related aromas with higher temperatures. Wine from 2010 had the lowest peak area for total esters, that vintage season had the highest mean minimum temperature V-H. Wine from 2013 had the lowest peak area for total alcohols and that year presented the highest solar exposure V-H and highest degree hours with base 25 - 30 °C, degree hours base 25 °C and total cumulative degree days. Wines with high esters and low alcohols present more intensive aromas (Dashko et al., 2015; Jackson & Lombard, 1993).

## **3.2.** Chemical parameters

Table 2 shows the results of the chemical parameters measured for all samples. It can be observed that the 2011 vintage presented the lowest wine color density, highest color hue, lowest total red pigments and lowest phenolics. It also had slightly higher TA and pH, along with 2008, had the lowest alcohol percentage. On the other hand, wine from 2009 had the highest wine color density, while 2008 had the highest degree of red pigmentation and total phenolics. Regarding the total red pigments and alcohol content, 2012 was the highest, while 2013 was the lowest in wine color hue and degree of red pigmentation.

Tuble 2: Results of the enemical parameters measured for each sample							
Vintage	2008	2009	2010	2011	2012	2013	
Wine colour density	4.46	4.57	3.29	2.24	3.69	3.18	
Wine colour hue	1.18	1.22	1.21	1.26	1.15	1.05	
Degree of red pigment coloration (%)	62.56	56.01	51.52	45.30	39.50	36.01	
Total red pigments (AU)	3.27	3.67	2.89	2.18	4.35	4.32	
Total phenolics (AU)	50.02	45.91	33.66	23.74	28.69	30.09	
рН	3.54	3.51	3.46	3.56	3.49	3.51	
Titratable Acidity (g L <sup>-1</sup> )	6.1	5.9	5.9	6.15	6.05	5.8	
Alcohol (%)	12.6	13.9	13.8	12.7	14.2	13.7	

Table 2. Results of the chemical parameters measured for each sample

Wines from the 2009 and 2013 vintages had the lowest TA values and the highest mean maximum and minimum temperatures. These results also agree with Van Leeuwen and Darriet (2016), and Jackson and Lombard (1993) who reported that in warmer weather conditions acid production,

especially malic acid, is lower thus TA tends to be low (< 6 g L<sup>-1</sup>). It has been reported in previous studies that in Pinot Noir there was a more intense color at temperatures around 20 °C during daytime compared to those grown at higher temperatures close to 30 °C (Flamini, Mattivi, Rosso, Arapitsas, & Bavaresco, 2013; Kliewer & Torres, 1972; Teixeira, Eiras-Dias, Castellarin, & Gerós, 2013). However, in this study the highest color density, degree of red pigmentation and total phenolics were found in wines from 2008 and 2009, which presented mean maximum temperatures close to 30°C with 26 and 29 °C, respectively. This coincides with Teixeira et al. (2013) who reported that the in warmer climates with higher sun exposure, the phenolics and anthocyanins concentration is higher, producing higher quality wines. Jackson and Lombard (1993), mentioned that day temperatures close to 15 °C provided better pigmentation, this coincides with the findings in the present paper as 2008, 2009 and 2010 presented mean minimum temperatures close to 15 °C (10 – 13 °C). Regarding alcohol content, warmer climates produce higher alcohol wines compared to cooler climates (Becker, 1985), which coincides with the findings in the present study.

#### 3.3 Machine Learning Modelling

As shown in Table 3, Model 1 had a very high correlation coefficient (R = 0.99) for all three stages and the overall model with a high performance (MSE = 0.01) and b = 1.00. The overall model presented a low percentage of outliers (6.3%) based on the 95% confidence bounds (Figure 1A). Likewise, Model 2 had a high correlation coefficient R = 0.98 for the overall model, and R > 0.96 for the three stages with a high performance and slope close to unity (b ~ 1). The overall model (Model 2), also presented a low percentage of outliers (6.6%) based on the 95% confidence bounds (Figure 1B). In both models, the fact that the training performance was lower than the one for validation and testing and that the performance of validation and testing is the same, shows that there was no overfitting in the models. Another sign of no overfitting is that the validation and training R were very close in both models.

Stage	Samples	Observations	R	MSE	Slope				
Model 1									
Training	26	546	0.99	<0.01	1.00				
Validation	5	105	0.99	0.01	1.00				
Testing	5	105	0.99	0.01	0.99				
Overall	36	756	0.99	-	1.00				
Model 2									
Training	26	208	0.99	0.01	0.98				
Validation	5	40	0.97	0.03	0.95				
Testing	5	40	0.96	0.03	0.96				
Overall	36	288	0.98	-	0.98				

**Table 3.** Statistical results of the machine learning models showing the correlation coefficient (R), performance based on means squared error (MSE) and slope

The very high accuracy of both models showed that it is possible to predict the quality of the wines based on the aromatic volatile compounds (Model 1) and chemical compounds (Model 2) as a result of the weather data before harvest. This is due to the aforementioned relationship between the different parameters and the weather conditions, especially due to temperature and solar exposure. Therefore, these models could aid in the early decision-making for the wineries to foresee the final product they will obtain, which may modify their original production plans. This would allow to reduce any economical losses due to unexpected wine quality when assessed once the production is ready.

## 4. Conclusions

The link between weather conditions, quality of grapes and consequently the wine quality traits allowed to obtain highly accurate ML models (R > 0.98), which would aid in the prediction of wine quality based on aromas and chemical parameters before its production, which would give anticipated information to winemakers about the product they would obtain to make early decisions of wine style and avoid any economical losses.

#### 5. Acknowledgments

This research was supported by the Digital Viticulture program funded by the University of Melbourne's Networked Society Institute, Australia.

### 6. Literature cited

- Becker, N. (1985). *Site selection for viticulture in cooler climates using local climatic information*. Paper presented at the Proceedings of the International Symposium on Cool Climate Viticulture and Enology. DA Heatherbell et al.(eds.).
- Bonada, M., Jeffery, D., Petrie, P. R., Moran, M. A., & Sadras, V. O. (2015). Impact of elevated temperature and water deficit on the chemical and sensory profiles of B arossa S hiraz grapes and wines. *Australian journal of grape and wine research*, *21*(2), 240-253.
- Dashko, S., Zhou, N., Tinta, T., Sivilotti, P., Lemut, M. S., Trost, K., . . . Vrhovsek, U. (2015). Use of nonconventional yeast improves the wine aroma profile of Ribolla Gialla. *Journal of Industrial Microbiology & Biotechnology*, 42(7), 997-1010.
- Drappier, J., Thibon, C., Rabot, A., & Geny-Denis, L. (2019). Relationship between wine composition and temperature: Impact on Bordeaux wine typicity in the context of global warming. *Critical reviews in food science and nutrition*, *59*(1), 14-30.
- Fang, Y., & Qian, M. (2005). Aroma compounds in Oregon Pinot Noir wine determined by aroma extract dilution analysis (AEDA). *Flavour and Fragrance Journal, 20*(1), 22-29.
- Flamini, R., Mattivi, F., Rosso, M., Arapitsas, P., & Bavaresco, L. (2013). Advanced knowledge of three important classes of grape phenolics: anthocyanins, stilbenes and flavonols. *International journal of* molecular sciences, 14(10), 19651-19669.
- Girard, B., Yuksel, D., Cliff, M., Delaquis, P., & Reynolds, A. (2001). Vinification effects on the sensory, colour and GC profiles of Pinot noir wines from British Columbia. *Food Research International*, *34*(6), 483-499.
- González-Barreiro, C., Rial-Otero, R., Cancho-Grande, B., & Simal-Gándara, J. (2015). Wine aroma compounds in grapes: A critical review. *Critical reviews in food science and nutrition*, *55*(2), 202-218.
- Jackson, D., & Lombard, P. (1993). Environmental and management practices affecting grape composition and wine quality-a review. *American Journal of Enology and Viticulture*, 44(4), 409-430.
- Kliewer, W. M., & Torres, R. E. (1972). Effect of controlled day and night temperatures on grape coloration. *American Journal of Enology and Viticulture, 23*(2), 71-77.
- Pirie, A. (1977). Phenolics accumulation in red wine grapes. Vitis vinifera.
- Song, J., Smart, R., Wang, H., Dambergs, B., Sparrow, A., & Qian, M. C. (2015). Effect of grape bunch sunlight exposure and UV radiation on phenolics and volatile composition of Vitis vinifera L. cv. Pinot noir wine. *Food Chemistry*, 173, 424-431.
- Teixeira, A., Eiras-Dias, J., Castellarin, S., & Gerós, H. (2013). Berry phenolics of grapevine under challenging environments. *International journal of molecular sciences*, 14(9), 18711-18739.
- The Good Scents Company. (2018). The Good Scents Company Information System. Retrieved from http://www.thegoodscentscompany.com
- Torrea, D., Varela, C., Ugliano, M., Ancin-Azpilicueta, C., Francis, I. L., & Henschke, P. A. (2011). Comparison of inorganic and organic nitrogen supplementation of grape juice–Effect on volatile composition and aroma profile of a Chardonnay wine fermented with Saccharomyces cerevisiae yeast. *Food Chemistry*, 127(3), 1072-1083.
- Ugliano, M., Travis, B., Francis, I. L., & Henschke, P. A. (2010). Volatile composition and sensory properties of Shiraz wines as affected by nitrogen supplementation and yeast species: rationalizing nitrogen modulation of wine aroma. *Journal of Agricultural and Food Chemistry*, *58*(23), 12417-12425.
- van Leeuwen, C., & Darriet, P. (2016). The impact of climate change on viticulture and wine quality. *Journal of Wine Economics*, 11(1), 150-167.



**Figure 1.** Model diagrams showing a the two-layer feedforward networks with sigmoid functions using 14 weather-related parameters and A) 5 hidden neurons and 21 outputs / targets for Model 1, and B) 9 neurons and 8 outputs /

targets for Model 2.



**Figure 2.** Overall models showing the observed (x-axis) and estimated (y-axis) values for the output parameters for A) 21 aromatic volatile compounds (Model 1) and B) 8 chemical parameters (Model 2) using 14 weather-related variables. 95% confidence bounds are shown in both models.