Modelling relationships between visible winegrape berries and bunch maturity

J. TANG[1], P.R. PETRIE[2,3] and M. WHITTY[1]

[1] School of Mechanical and Manufacturing Engineering, University of New South Wales, Sydney, NSW 2052, Australia
[2] South Australian Research and Development Institute, Waite Research Precinct, Urrbrae, SA 5064, Australia; [3] The Australian Wine Research Institute, Waite Research Precinct, Urrbrae, SA 5064, Australia
Corresponding author: Ms Julie Tang, email julie.tang@unsw.edu.au

Abstract

Background and Aims: Rapid and non-destructive techniques for measuring grape composition have the potential to improve harvest decisions and to allow the batching of fruit into parcels of similar composition to improve processing efficiency. If such techniques are to be based on the non-destructive assessment of visible berries, then relationships between the composition of visible berries and that of the whole bunch will need to be described.

Methods and Results: Total soluble solids, pH, TA and the concentration of anthocyanin, phenolic substances and tannin of visible berries and associated entire bunches of Shiraz and Chardonnay were measured over three seasons. Good correlations were obtained between the composition of visible berries and that of the entire bunch. Most relationships were not 1:1 but were consistent and stable across seasons.

Conclusions: The accuracy and stability across the seasons of relationships between the composition of visible berries and that of the entire bunch for both Chardonnay and Shiraz were evaluated.

Significance of the Study: The approaches used in this study to predict bunch composition from visible berry composition could aid the development of non-destructive techniques based on sensing visible berries to assess grape composition during ripening.

Keywords: *fruit composition, non-destructive methods, ripening, sampling, sugar concentration*

Introduction

Vineyard fruit is assessed regularly during ripening to estimate harvesting dates. Compositional parameters typically assessed include TSS, TA and pH (Olarte Mantilla et al. 2012). Composition varies within and between bunches and the magnitude of this variation can change as fruit ripens (Kliewer and Lider 1968, Pagay and Cheng 2010). At a bunch level, variation in sunlight exposure has been linked to variation in composition. For example, both high and low levels of fruit sunlight exposure can have a negative effect on anthocyanin accumulation in Merlot (Mabrouk and Sinoquet 1998); and a difference in phenolic substances was observed between the shaded and sunlight-exposed sides of Pinot Noir bunches (Price et al. 1995). The TSS of Thompson Seedless berries sampled from sunlight-exposed bunches on the exterior of the canopy was higher than that for berries sampled from the interior of the canopy, but berries sampled from the front (exposed) side of the sunlight-exposed bunches had a lower TSS than the berries from the rear (shaded) side (Kliewer and Lider 1968).

Fruit composition is assessed by sampling grape bunches from the vineyard and then transporting them back to the winery for analysis as recommended by Krstic (2003) [where bunches are randomly selected across vines and vineyards to account for spatial variation across vineyards and bunch-to-bunch variation within vines (Trought and Bramley 2011)]. Variation in composition between bunches within a vine is generally higher than that for the average composition of bunches between vines (Krstic et al. 2011), with greater variation of TA and TSS seen in less mature bunches (Rankine et al. 1962). When individual berries are sampled, berries from the apical location of a bunch do not consistently predict average TSS for the whole bunch (Wolpert and Howell 1984). Sampling from the surface of bunches tended to overestimate bunch TSS (Kasimatis and Vilas 1985) suggesting that whole bunches, as opposed to individual berries, should be preferred for assessment of composition. Sensory assessment of grapes in the vineyard or laboratory can also be used to assist harvest decisions and match fruit characteristics to wine style (Olarte Mantilla et al. 2012).

In the last decade, there has been a focus on the analysis of whole grapes by spectroscopy in the VIS (Cao et al. 2010), in the VIS-NIR (Fernandes et al. 2011, Giovenzana et al. 2015, Fadock et al. 2016) and in the NIR (Larra'ın et al. 2008, Ferna'ndez-Novales et al. 2009, Guidetti et al. 2010, Gonza'lez-Caballero et al. 2011) wavelengths and using active proximal sensors such as the Multiplex by Force-A (Ghozlen et al. 2010, Baluja et al. 2012). Irrespective of the wavelength used, all of these studies have relied on identifying a correlation between individual grape berries and their composition (Larra'ın et al. 2008, Ferna'ndez-Novales et al. 2009, Cao et al. 2010, Guidetti et al. 2010, Fernandes et al. 2011, Gonza'lez-Caballero et al. 2011, Giovenzana et al. 2015), or between visible grape berries and the composition of whole bunches (Baluja et al. 2012, Giovenzana et al. 2015, Fadock et al. 2016).

As there is a large variation in composition from berry-to-berry within bunches, the accuracy of non-destructive

doi: 10.1111/ajgw.12374

analysis methods may be compromised if all of the berries are not assessed and those sampled are not representative of all the berries. In order to develop rapid and feasible sensors, relationships between the composition of visible berries and that of the entire bunch need to be described. Thus, the aim of this research was to investigate the strength and stability of relationships between visible berries and average bunch maturity for several compositional parameters such as TSS, pH, TA and the concentration of anthocyanin, phenolic substances and tannin.

Materials and methods

Grape bunch sampling

Grape bunches were sampled from Chardonnay and Shiraz grown on own roots and planted in 2000 and 1999, respectively. Both cultivars were trained to a height of 1.1 m with vertically shoot positioned canopies in a vineyard located in Orange, NSW, Australia. This region has an average growing season temperature of 17.6°C (Hall and Jones 2010). Rows were oriented approximately north–south with a row spacing of 3 m and vine spacing of 2.1 m (Chardonnay) and 1.8 m (Shiraz). During the 2015/16, 2016/17 and 2017/18 seasons, the blocks received supplementary drip irrigation of approximately 0.52, 0.75 and 0.66 ML/ha to complement, respectively, the 331.4, 589.4 and 352 mm of rainfall during the season.

In 2016, a bunch was sampled from each side of the canopy (east and west facing) on 40 vines randomly distributed throughout the block for each cultivar. Bunches that were visible from the vine row were chosen and photographed perpendicular to the row (Figure 1) before being harvested. Chardonnay and Shiraz were sampled 3 and 2 weeks prior to harvest, respectively (Table 3). Each bunch was packed into a container with ice before being transported to the laboratory. Samples were stored at 4°C for approximately 5 days until they were analysed in a laboratory at the University of New South Wales. In 2017, bunches were sampled 1 week apart on three dates for each cultivar from post-veraison to harvest (Table 3). Ninety bunches were sampled from Chardonnay and 74 from Shiraz. Bunches were analysed in a laboratory at the vineyard site. In 2018, sampling was the same as in 2017 except 60 Chardonnay bunches were sampled on two dates and 69 Shiraz bunches were sampled on three dates approximately 1 week apart. The samples on the second date for Chardonnay were taken from a different block on the same vineyard but with equivalent row orientation and vine training due to an earlier than expected harvest of the test block.

Measurement of fruit composition

For each bunch, the corresponding reference image was used to identify the visible berries (defined as berries with 40% or more of a spherical face visible in the reference image). These berries were removed and weighed and then the remaining berries were removed and weighed. In 2016, prior to analysis, the number of non-visible Shiraz berries in the sample was reduced, so the sample size was similar to total berry number in the visible berry sample, and the sample berries weighed again.

Each set of berries was crushed and the juice decanted through a sieve to remove large particles. The TSS was measured with a temperature compensated digital refractometer (DBR-1, Starr Instruments, Dandenong South, Vic., Australia), pH with a pH meter (pH Cube, TPS, Brendale, Qld, Australia) and TA by manual titration to pH 8.2 (Iland et al. 2013) using 5 and 3 mL of juice for Chardonnay and Shiraz, respectively.

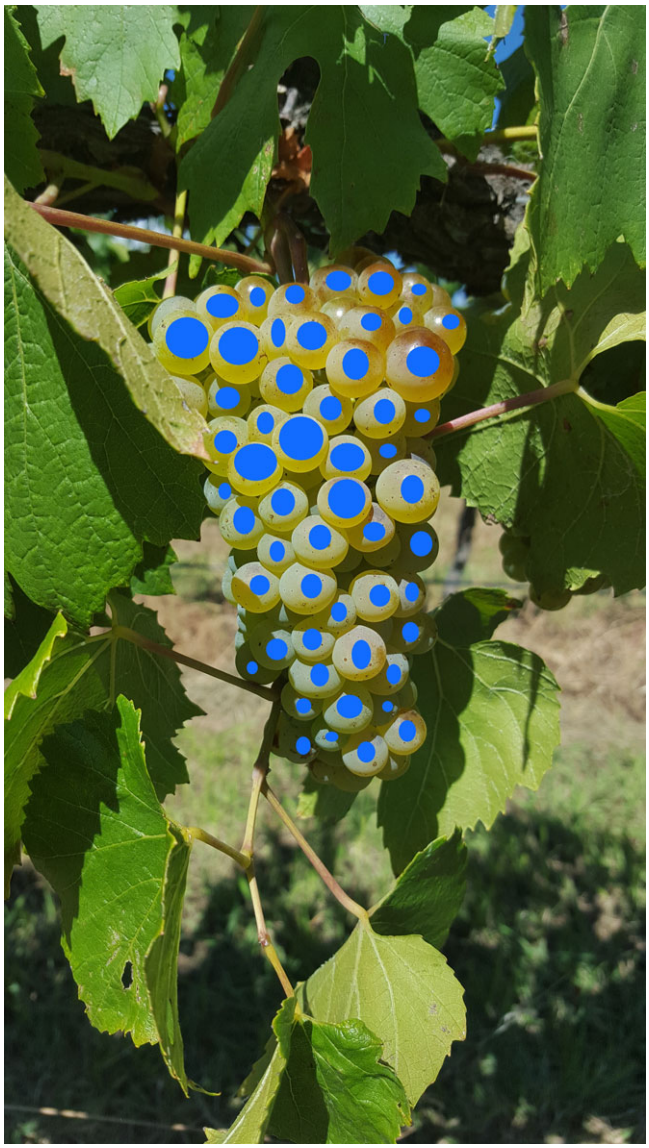

Figure 1. Reference image taken in field to assist in identifying visible berries on the grapevine. Berries marked in blue are considered 'visible'.

Prior to analysis of the concentration of anthocyanin, phenolic substances and tannin in the Shiraz, the juice was removed for the measurement of pH and TA was replaced with de-ionised water to maintain the concentration of anthocyanin, phenolic substances and tannin. The Shiraz sample was blended with a Rocket blender (MBR-1701, Homeland Housewares, Sherman Oaks, CA, USA) for 20 s to a uniform consistency. A 1 mL subsample was mixed with 10 mL of 50% v/v aqueous ethanol solution for 1 h on a rotary mixer before being centrifuged at 1688 × *g* for 5 min. The supernatant (herein termed 'the extract') was then stored in amber vials at −20°C for 7–14 days until further analysis. Previous research has shown that there is no significant loss of colour when samples are stored at this stage of processing for up to 3 months (Cynkar et al. 2004).

To measure anthocyanin and phenolic substances, 1 mL of the extract was mixed with 10 mL of 1 mol HCl and left to stand for 3 h before the absorbance was measured at 520 and 280 nm. To measure tannin, a Control and a

treatment sample were prepared following the methyl cellulose precipitable tannin (MCPT) assay (Iland et al. 2013) and the absorbance was read at 280 nm, with values reported as epicatechin equivalents in mg/g berry mass.

In 2017 and 2018, only TSS, pH and TA were measured; pH and TA were measured with a mini-titrator (Hanna Instruments HI84502-01, Woonsocket, RI, USA).

Measurement of bunch compactness

Bunch compactness was estimated as the berry number/cm of rachis (Sabbatini and Howell 2010). Unfortunately, rachis length and berry number were not recorded when bunches were processed. Therefore, bunch compactness was calculated based on the berry count, rachis mass and the relationship between rachis mass and rachis length in the previous season (2015). In the years of 2015–2017, 80 bunches per block were tagged at 20 locations in the same blocks used in this study. In 2015, rachis length and corresponding rachis mass at harvest for each tagged bunch were recorded. From 2016 and 2017 (the years of this study), rachis mass, bunch mass and manual berry counts of each remaining bunch were measured at harvest. Using the average rachis mass to length ratio derived from 2015 data, the following formula (Equation 1) was used to calculate average bunch compactness for 2016 and 2017.

$$c = \frac{\overline{bc_m}}{\overline{rw} \times \frac{\overline{rl_{v2015}}}{\overline{rw_{v2015}}}} \quad (1)$$

where c is the compactness (berry count to rachis length ratio), bc_m is the average manual berry count for specific year, rw is the average rachis mass for specific year, $rl_{v2015}$is the average rachis length in 2015 and. rw_{v2015} is the average rachis mass in 2015.

Analysis of results

The average proportion of visible berries in relation to the whole bunch for Chardonnay (range 0.15–0.61, $\mu = 0.4$, $\sigma = 0.08$) and Shiraz (range 0.21–1, $\mu = 0.46$, $\sigma = 0.11$) differed from 50% (represented as $\mu = 0.5$). Thus, a weighted average was taken to represent the bunch composition using the following formulae (Equations 2,3):

$$q_{\text{bunch}} = \frac{W_v \times q_v}{W_v \times W_n} + \frac{W_n \times q_n}{W_v \times W_n} \quad (2)$$

$$\text{pH}_{\text{bunch}} = log_{10}\left(\frac{W_v \times 10^{kv}}{W_v + W_n} + \frac{W_n \times 10^{kn}}{W_v + W_n}\right) \quad (3)$$

where q is the compositional parameter (°Brix, TA, anthocyanin, phenolic substances or tannin), w_v is the mass of visible berries, w_n is the mass of non-visible berries, q_v is the compositional parameter value of visible berries, q_n is the compositional parameter value of non-visible berries, k_v is the pH of visible berries and k_n is the pH of non-visible berries.

The relationship between visible berries and bunch composition was validated using a separate test set, by first splitting the data from each of the six data sets into six sets (for a total of 36 sets). The k-fold cross validation ($k = 5$) was used rather than leave-one-out cross-validation (LOOCV), which tends to be more unreliable because of overestimating variance in cross-validation error (Rao et al. 2008). Standard error of estimate (SE) and mean squared errors (MSE) were calculated for training and test sets.

The relationships were also compared with a 1:1 relationship. This was done using analysis of covariance (ANCOVA), which provides the MSE of each regression, and a P-value for the overall test. The R^2 value was also calculated for the 1:1 relationship. The above relationships were also derived between visible and non-visible berries; with results showing equivalent relationships to those presented here, that is a significant difference between visible berries and non-visible berries for all compositional parameters with the exception of TSS in Shiraz only. As the focus of this study is to compare visible berries to the whole bunch, the relationship between visible and non-visible berries is not discussed here.

Results

Total soluble solids

There was a strong linear relationship ($r^2 \geq 0.82$) between TSS of visible berries and bunch TSS for both cultivars in all three seasons (Table 1). Shiraz exhibited the highest standard error of validation test set (SERV) in 2017, which was likely due to larger variation in the data set and slightly lower r^2 value than in 2016 and 2018. For Chardonnay, there was a stronger relationship between visible berry TSS and bunch TSS in 2017 and 2018 (r^2 0.97–0.99) compared to that in 2016 (0.82). This was likely due to the larger sample size and smaller TSS range in 2017 and 2018 when compared to that in 2016 (Table 3).

In 2016, the MSE and SERV for the test set were smaller than the average training errors for both data sets, which may be attributed to the small size of the test set with respect to the data set range. In 2017 and 2018, test set errors were larger than training errors as is generally expected.

In terms of prediction ability, the standard error of the regression (SER) can be used as an estimate where 95% of predictions fall within $\pm 2 \times$ SER and errors associated with the trained regression models are summarised in Table 3. In 2016, the regression lines had prediction errors (95% prediction intervals) of ± 1.59 and ± 0.78°Brix for Chardonnay and Shiraz, respectively. In 2017, the prediction error was on average ± 0.66°Brix for both cultivars and in 2018 prediction errors lower again, averaging ± 0.43°Brix. This reduction in prediction error may be attributed to a larger concentration of samples for the TSS ranges measured in 2018 and 2017 compared to 2016.

When comparing the fitted regression slopes with a 1:1 slope, for Shiraz there was no significant difference between 2016 and 2018 (P-value > 0.05 and P-value > 0.1, respectively). In 2017, however, the deviation from a 1:1 slope was significant (P-value < 0.01). For Chardonnay, the deviation from the 1:1 model was significant in all years (P-value < 0.001). By combining the data points over all years, the slope of the TSS relationship generated for Shiraz did not differ significantly from a 1:1 (P-value > 0.1). For Chardonnay, the slope of the TSS relationship differed significantly from a 1:1 model (P-value < 0.001) with an SER of 0.55 in the generated model compared with SER of 0.62 in the 1:1 model. The r^2 values were both 0.98.

pH

The linear relationship between visible berry pH and entire bunch pH in 2016 (Figure 2c,d) was similar for Chardonnay

Table 1. Average errors of prediction for the relationships between the composition of visible berries and that of the whole bunch from training and the mean squared error and standard error of estimate of the validation test set when using the regression fitted on all training data.

Year	Cultivar	Parameter	Average training error				Test error	
			r^2	s	SER	MSE	MSE	SERV
2016	Chardonnay	TSS (°Brix)	0.82	0.78	0.81	0.74	0.39	0.62
		pH	0.92	0.034	0.029	0.0010	0.0020	0.045
		TA (g/L)	0.89	0.55	0.64	0.45	0.17	0.42
2016	Shiraz	TSS (°Brix)	0.98	0.39	0.41	0.18	0.12	0.35
		pH	0.89	0.033	0.032	0.0012	0.0006	0.026
		TA (g/L)	0.92	0.24	0.24	0.058	0.069	0.26
		Anthocyanin (mg/g)	0.75	0.30	0.34	0.15	0.046	0.21
		Phenolic substances (mg/g)	0.76	0.11	0.11	0.012	0.014	0.12
		Tannin (mg/g)	0.81	0.42	0.41	0.18	0.16	0.40
2017	Chardonnay	TSS (°Brix)	0.97	0.33	0.31	0.11	0.15	0.39
		pH	0.95	0.024	0.025	0.0006	0.0004	0.019
		TA (g/L)	0.96	0.26	0.27	0.078	0.046	0.214
2017	Shiraz	TSS (°Brix)	0.97	0.34	0.29	0.09	0.22	0.47
		pH	0.99	0.017	0.017	0.0003	0.0002	0.012
		TA (g/L)	0.97	0.13	0.12	0.015	0.028	0.17
2018	Chardonnay	TSS (°Brix)	0.99	0.21	0.19	0.038	0.077	0.28
		pH	0.96	0.016	0.017	0.0003	0.0002	0.014
		TA (g/L)	0.96	0.34	0.33	0.12	0.17	0.41
2018	Shiraz	TSS (°Brix)	0.98	0.24	0.23	0.054	0.073	0.27
		pH	0.98	0.017	0.017	0.0003	0.0003	0.019
		TA (g/L)	0.95	0.22	0.25	0.064	0.027	0.17
All	Chardonnay	TSS > 16 (°Brix)	0.98	0.29	0.29	0.084	0.095	0.31
		pH	0.98	0.027	0.026	0.0007	0.0009	0.030
		TA (g/L)	0.96	0.50	0.43	0.19	0.54	0.74
All	Shiraz	TSS ≥ 17 (°Brix)	0.97	0.34	0.34	0.12	0.14	0.38
		pH	0.97	0.026	0.024	0.0006	0.0008	0.028
		TA (g/L)	0.96	0.20	0.21	0.044	0.026	0.16

k = 5. MSE, mean squared error; SER, standard error of regression; SERV, standard error of estimate of validation test set.

(r^2 = 0.92) and Shiraz (r^2 = 0.89), with average training errors of approximately 0.001 MSE. In 2017 and 2018 (Figure 2c,d), similar fits are also seen for both cultivars (r^2 of 0.95/0.96 for Chardonnay and 0.99/0.98 for Shiraz).

Tables 1 and 2 present similar errors for the two data sets (MSE ±0.002). The predictive ability of the Shiraz relationships was improved in all years except 2018 when compared to the Chardonnay relationships (SERV in Table 1). This is also apparent for the trained regression model errors across the entire data set for each year (Table 2), except in 2016 when both trained regression models had an equivalent SER (0.033).

All the pH relationships between visible berries and entire bunches differed significantly from the 1:1 model (P-value < 0.001).

Titratable acidity

The r^2 value in 2016 for the relationship between visible berry TA and entire bunch TA was 0.89 for Chardonnay and 0.92 for Shiraz and similar to TSS and pH; in 2017 the r^2 value was higher being 0.96 for Chardonnay and 0.97 for Shiraz (Table 2). The standard error of the test set was lower than the SER of the training set for Chardonnay in 2016/17 and for Shiraz in 2018, indicating a large variation in the data sets or a smaller data set (2018 Shiraz). All other results had an SERV larger than SER of the training set, as expected. The SER for Chardonnay relationships in 2016 was larger than in 2017 and 2018 (0.56 compared to 0.26 and 0.34), which can be linked to a larger number of data points for a given range in 2017 and 2018 (Figure 2e). A similar phenomenon was seen for Shiraz from 2016 to 2017 (SER 0.24 in 2016 compared to 0.13 in 2017). For Shiraz in 2018, however, and perhaps due to the limited number of data points, the SER was 0.2. The MSE in Chardonnay was also much larger than that for Shiraz (0.306 compared to 0.057 in 2016, 0.067 compared to 0.016 in 2017 and 0.113 compared to 0.039 in 2018). The MSE was four times smaller in 2017 for both cultivars, indicating a much tighter fit to the regression model.

All the relationships between visible berry TA and entire bunch TA differed significantly from 1:1 (P-value < 0.001, 2018 Shiraz P-value < 0.01). The departure from 1:1 was more pronounced for Chardonnay (Figure 2e). The gradient of the regression line differed substantially for Chardonnay in 2016 (0.82 compared to an average of 1.04 in 2017/18). In 2017, the variation from 1:1 was due to a 270% increase in the MSE (0.183 compared to 0.067) when using the 1:1 model as opposed to the fitted regression. For Shiraz, the gradient stayed constant and the reason for variation from the 1:1 model was related to differences in the intercept (0.41–0.59), which was equivalent to at least 12% of the data set range (Figure 2f).

Anthocyanin, phenolic substances and tannin

These compositional parameters were measured only in 2016. In 2016, the predictive ability of the relationship between visible berries and entire bunch anthocyanin, phenolic substances and tannin was ±0.6, ±0.24 and ±0.8 mg/g berry mass, respectively.

For anthocyanin, the average r^2 value in training the model (Table 1) was 0.75, indicating a linear fit. There was some noise in the data set as well as outliers, which contributed to the lower r^2 value when compared to the models for TSS (Figure 3a). The larger average SERV of the test set

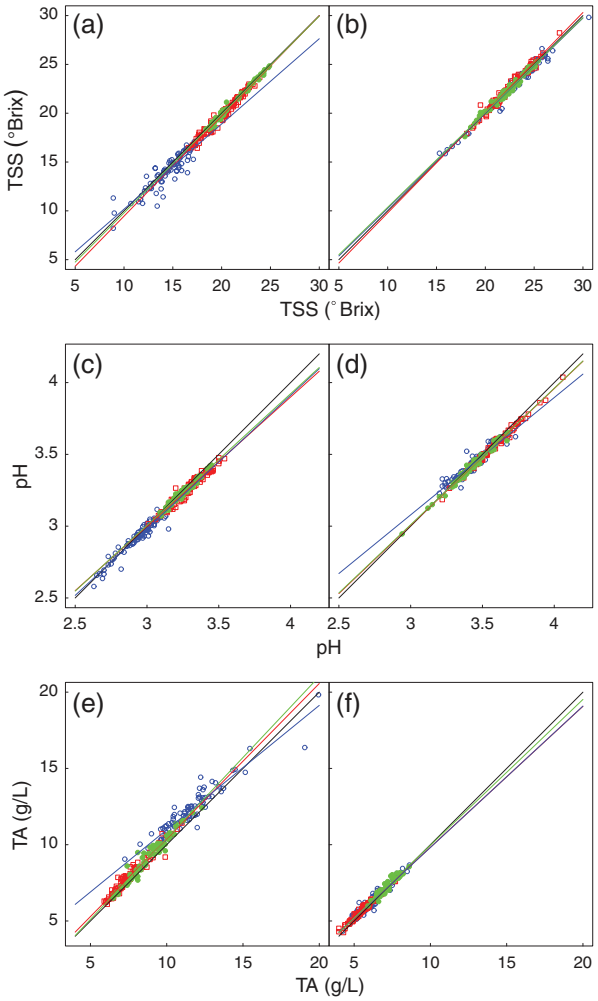

Figure 2. Relationship between (a, b) TSS, (c, d) pH and (e, f) TA of the visible berries in a bunch and that of the whole bunch for (a, c, e) Chardonnay and (b, d, f) Shiraz for 2016 data (○), 2017 data (□) and 2018 data (●). The 1:1 line (—), 2016 linear model (——), 2017 linear model (——) and 2018 linear model (——) are shown.

compared to the smaller SER of the calibration set in Table 1 can be attributed to a larger number of data points, difficult to predict points in the training set compared to the test set or to a relatively small test set. The SER associated with a model trained overall data was 0.21 and this corresponds to approximately 10% of the data range. The deviation from 1:1 for anthocyanin was significant (P-value < 0.001). By comparing the SER and r^2 of the linear model over the whole data set (0.18, 0.75) with that of the 1:1 relationship (0.20, 0.68), we can see that they are similar.

The linear relationship for phenolic substances had an r^2 value of 0.76, which was similar to that for anthocyanin (Table 2). The error associated with the test sets was similar to the model training errors and this indicated that the model was stable. The MSE is 0.017 for the test set, which is approximately 10% of the range. The relationship generated differed significantly from 1:1 (P-value < 0.01); however, the SER was similar for both models.

For tannin, the relationship between visible berry tannin and entire bunch tannin was strong (r^2 = 0.81). The MSE and SERV of the test set for this model were smaller than that of the average MSE and SER of the training sets, which can be due to a small data set (μ = 2.52, σ = 0.95) for the range of values covered (0.12–4.69). The relationship generated differed significantly from the 1:1 model (P-value < 0.001).

Discussion

Total soluble solids

Both the strength and the stability of the relationships between visible berry composition and whole bunch composition were evaluated, along with their adherence to a 1:1 relationship. Rankine (1995) suggested an accuracy of 1°Brix between the sample and entire crop as a target for estimation, which indicates that TSS errors of greater than 1°Brix are unsatisfactory. Hence, this figure was used to define whether relationships were sufficiently accurate. Krstic et al. (2011) identified sampling errors as a proportion of the data set mean for between vine variability, which was used here as a benchmark for bunch sampling error. These are 4–5% for TSS, 3–4% pH, 10–12% TA, 13–18% anthocyanin and 13–17% for phenolic substances.

The slopes of most trained regression relationships were close to 1 with r^2 values differing from 1:1 by 0.03 at most (Table 2, Figure 2a,b). The 2016 Chardonnay samples, however, covered a lower TSS range than the remainder of the samples, and hence the slope of that relationship (Table 2, Figure 2a) differed most from the 1:1 model, with higher visible berry TSS than bunch TSS. For winegrape bunches sampled at low TSS (<16°Brix), Kasimatis and Vilas (1985) found significantly higher TSS of approximately 0.9°Brix from samples of surface berries when compared to that of the remaining bunch. Visible berries in our analysis form a proportion of surface berries, and with the remaining bunch consisting mostly of non-surface berries, our result of visible berry TSS higher than total bunch TSS at low bunch TSS is consistent with surface berries tending to have TSS higher than that of the entire bunch (Kasimatis and Vilas 1985) and increased sunlight exposure being related to higher TSS (Reynolds et al. 1986). As grapes ripen TSS eventually plateaus as sugar importation slows. Any further increase in TSS is the result of evaporation of water from the surface of the berry (Coombe and McCarthy 2000). Trought et al. (2017) reported a reduction in the variation of berry TSS across a population of individually sampled berries as berries ripened, so it would be expected that the difference between visible and non-visible berries would also decrease with ripening, resulting in relationships that are closer to 1:1.

Using only data where TSS < 16°Brix from the 2016 Chardonnay data set, the trained linear model achieved accuracy of ±1.65°Brix (data not shown), which is greater than the target accuracy of 1°Brix (Rankine 1995). This low level of accuracy associated with the Chardonnay linear model at TSS < 16°Brix is likely due to the higher intra-bunch berry variation that is apparent at lower TSS (Trought et al. 2017).

A linear model using all the data with visible berry TSS ranging from 16.5 to 24.5°Brix achieved an accuracy of ±0.58°Brix for estimating the bunch TSS for Chardonnay (Table 2). A 1:1 model across the same visible berry TSS range achieved a prediction accuracy of ±0.70°Brix. Given the aforementioned target accuracy of 1°Brix, both models would be acceptable for estimating TSS for winemaking. If an additional threshold of 10% error, however, is added to the trained regression, then the 1:1 model is not appropriate for estimating bunch TSS from visible berry TSS for Chardonnay and, hence, it is recommended that a linear regression be developed (Table 3).

A TSS range of 16.5–24.5°Brix and a validation accuracy of 0.62°Brix across the three seasons compare favourably

Table 2. Evaluation of the trained regression models for Chardonnay and Shiraz compared to errors associated with a 1:1 assumption.

Year	Cultivar	Parameter	r^2	Trained regression model				r^2	1:1 model	
				Slope	Intercept	SER	MSE		SER	MSE
2016	Chardonnay	TSS (°Brix)	0.82	0.87	1.44	0.80	0.62	0.76	0.93	0.83***
		pH	0.92	0.93	0.19	0.033	0.0011	0.90	0.037	0.0013***
		TA (g/L)	0.89	0.82	2.83	0.56	0.31	0.68	0.98	0.93***
2016	Shiraz.	TSS (°Brix)	0.98	0.97	0.51	0.39	0.15	0.98	0.41	0.16
		pH	0.89	0.82	0.63	0.033	0.0011	0.85	0.039	0.0015***
		TA (g/L)	0.92	0.93	0.55	0.24	0.057	0.90	0.27	0.072***
		Anthocyanin (mg/g)	0.75	0.77	0.41	0.178	0.031	0.68	0.20	0.039***
		Phenolic substances (mg/g)	0.76	0.81	0.23	0.11	0.012	0.72	0.12	0.014**
		Tannin (mg/g)	0.81	0.80	0.49	0.42	0.17	0.76	0.47	0.21***
2017	Chardonnay	TSS (°Brix)	0.97	1.03	−0.81	0.33	0.11	0.95	0.44	0.19***
		pH	0.95	0.90	0.30	0.024	0.0006	0.86	0.039	0.0015***
		TA (g/L)	0.96	1.02	0.20	0.26	0.067	0.88	0.43	0.18***
2017	Shiraz	TSS (°Brix)	0.97	1.03	−0.47	0.33	0.11	0.97	0.36	0.12**
		pH	0.99	0.95	0.15	0.017	0.0003	0.96	0.028	0.0008***
		TA (g/L)	0.97	0.92	0.59	0.13	0.016	0.91	0.22	0.046***
2018	Chardonnay	TSS (°Brix)	0.99	1.01	−0.29	0.21	0.042	0.99	0.24	0.057***
		pH	0.96	0.92	0.26	0.016	0.000	0.92	0.023	0.0005***
		TA (g/L)	0.95	1.06	−0.20	0.34	0.11	0.89	0.49	0.23***
2018	Shiraz	TSS (°Brix)	0.98	0.97	0.66	0.24	0.055	0.98	0.25	0.059
		pH	0.98	0.95	0.16	0.017	0.000	0.97	0.021	0.000***
		TA (g/L)	0.88	0.96	0.41	0.20	0.039	0.85	0.23	0.050**
All	Chardonnay	TSS (°Brix)	0.98	1.01	−0.50	0.55	0.30	0.97	0.62	0.38***
		pH	0.98	0.95	0.13	0.026	0.000	0.97	0.034	0.001***
		TA (g/L)	0.96	1.01	0.35	0.49	0.24	0.92	0.68	0.46***
All	Shiraz	TSS (°Brix)	0.98	0.98	0.37	0.34	0.11	0.98	0.34	0.12
		pH	0.97	0.91	0.31	0.025	0.001	0.95	0.031	0.001***
		TA (g/L)	0.96	0.94	0.50	0.20	0.038	0.93	0.24	0.057***

Statistical significance of the 1:1 assumption is given by *, $P < 0.1$; **, $P < 0.01$; ***, $P < 0.001$; non-significant otherwise. MSE, mean squared error; SER, standard error of regression.

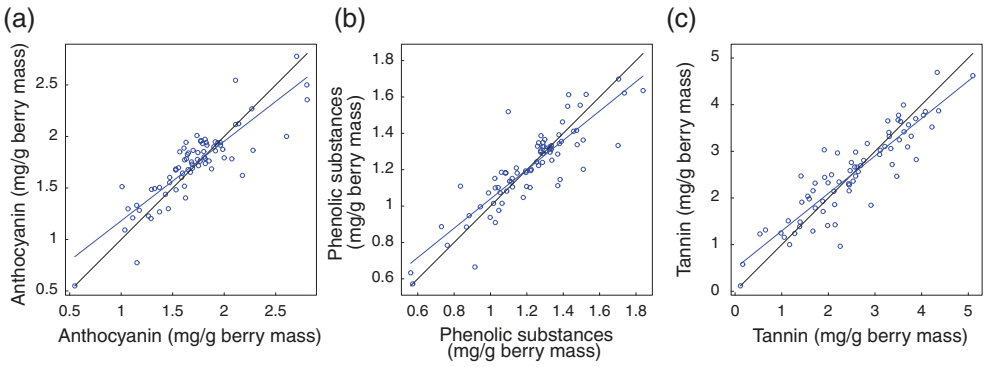

Figure 3. Relationship between the concentration of (a) anthocyanin, (b) phenolic substances and (c) tannin of visible berries in a bunch and that of the whole bunch for Shiraz. The 1:1 line (—) and the 2016 linear model (——) are shown.

with existing TSS estimation studies (Cao et al. 2010, Guidetti et al. 2010, Giovenzana et al. 2013). These, at best, were accurate to ±0.94°Brix for grape must [described as 0.47 SECV (Fernández-Novales et al. 2009)].

A linear regression using 2017 data estimated the 2018 data with an accuracy of ±0.49°Brix (data not shown) compared to the best case of the 2018 linear model estimating the 2018 data with an accuracy of ±0.42°Brix (Table 2), which shows that the relationship is stable across similar bunch TSS. Similarly, a performance comparison of 2017 data tested on a linear regression using 2018 data and the linear regression using 2017 data yielded an accuracy of ±0.72 (data not shown) and ±0.66°Brix, respectively. The small changes in accuracy across models for the same data indicates that both 2017 and 2018 Chardonnay models can be considered stable in the range 16.5–24.5°Brix and could, therefore, be used to predict bunch TSS from visible berry TSS for Chardonnay for that range.

The data collected across all years for Shiraz were across a similar range (Figure 2b, Table 3) with the majority falling between 17 and 28°Brix. Within this range, the 1:1 relationship and combined linear regression across the three seasons achieved an accuracy of ±0.76 and ±0.75°Brix, respectively. Hence, a 1:1 relationship could be used for predicting bunch TSS. Errors in cross-validation for the trained model (SERV in Table 1) resulted in an accuracy of ±0.76°Brix, which is better than the accuracy of ±0.94°Brix achieved by Fernández-Novales et al. (2009) on an NIR estimation on grape must and the accuracy of ±1.3°Brix achieved by Fadock et al. (2016) on a non-destructive VIS-NIR estimation on composite grape berry samples.

Shiraz models were considered stable across seasons given that 1:1 relationship performed well in each year with the accuracy deviating by at most 10% from the trained models (Table 3). Given that Spayd et al. (2002) found minimal effect (at most 0.8% TSS) on TSS at harvest across treatments of sunlight exposure, shading and artificial alteration of bunch temperature, it is possible that the TSS models presented here will hold for a warmer climate, however, this would need to be tested. Overall, and for

Table 3. Summary and evaluation of results and absolute errors (95% confidence) associated with derived models relating the composition of visible berries to that of the entire bunch.

Year	Cultivar	Parameter	Data		SD	Linear	Trained model		1:1 model	
			Range	Mean			± Error	Good	± Error	Good
2016	Chardonnay	TSS (°Brix)	7.7–17.7	14.07	1.87	Y	1.60	N	1.85	N
		pH	2.56–3.15	2.92	0.12	Y	0.066	Y	0.072	Y
		TA (g/L)	7.25–19.95	12.21	1.71	Y	1.12	N	1.95	N
2016	Shiraz	TSS (°Brix)	7.71–29.82	22.87	3.11	Y	0.78	Y	0.81	Y
		pH	3.23–3.63	3.42	0.10	Y	0.066	Y	0.078	N
		TA (g/L)	4.1–8.67	6.19	0.96	Y	0.48	N	0.54	N
		Anthocyanin (mg/g)	0.55–2.8	1.79	0.57	Y	0.36	N	0.40	N
		Phenolic substances (mg/g)	0.57–1.7	1.25	0.39	Y	0.22	N	0.24	N
		Tannin (mg/g)	0.12–4.69	2.52	0.95	Y	0.84	N	0.94	N
2017	Chardonnay	TSS (°Brix)	16.42–23.67	19.81	1.87	Y	0.66	Y	0.87	N
		pH	3.02–3.48	3.25	0.10	Y	0.048	Y	0.078	N
		TA (g/L)	6.1–11.42	8.03	1.25	Y	0.52	Y	0.86	N
2017	Shiraz	TSS (°Brix)	17.92–28.23	22.67	2.07	Y	0.66	Y	0.71	Y
		pH	3.19–4.04	3.56	0.14	Y	0.033	Y	0.056	N
		TA (g/L)	4.24–7.6	5.74	0.71	Y	0.25	Y	0.86	N
2018	Chardonnay	TSS (°Brix)	18.37–24.84	20.99	2.00	Y	0.42	Y	0.49	N
		pH	3.07–3.43	3.24	0.08	Y	0.031	Y	0.046	N
		TA (g/L)	6.5–12.42	9.145	1.47	Y	0.68	Y	0.99	N
2018	Shiraz	TSS (°Brix)	17.61–25.35	22.54	1.77	Y	0.48	Y	0.49	Y
		pH	2.95–3.65	3.406	0.12	Y	0.034	Y	0.042	N
		TA (g/L)	5.99–8.55	7.15	0.57	Y	0.40	Y	0.45	N
All	Chardonnay	TSS > 16 (°Brix)	16.01–24.84	19.96	2.19	Y	0.58	Y	0.71	N
		pH	2.56–3.48	3.13	0.19	Y	0.053	Y	0.069	N
		TA (g/L)	6.1–19.95	9.77	2.35	Y	0.98	N	1.36	N
All	Shiraz	TSS (°Brix)	17.01–29.82	22.89	2.01	Y	0.68	Y	0.68	Y
		pH	2.95–4.04	3.47	0.14	Y	0.051	Y	0.061	N
		TA (g/L)	4.24–8.67	6.38	1.03	Y	0.39	Y	0.48	N

Y, Yes; N, No.

simplicity, a 1:1 model could be used for Shiraz in the data range 17–28°Brix with an expected accuracy of approximately ±0.8°Brix. For Chardonnay, a linear regression model is appropriate only for the data range 16.5–24°Brix and would deliver an accuracy of approximately ±0.6°Brix.

pH

The slope for all the models for Chardonnay (Figure 2c) ranged between 0.9 and 0.95, and was consistent between seasons. With a slope less than 1 and the majority of data points lying below the 1:1 line, this indicates that visible berry pH is generally higher than total bunch pH, which is similar to the results reported by Kliewer and Lider (1968) that showed increased sunlight exposure of bunches resulted in higher pH.

The slope in 2016 for Shiraz (Figure 2d) was less steep and differed significantly from that in 2017 and 2018. The inconsistency of the 2016 slope is due to data being collected at only one date during the season, resulting in a small data range and an increased impact of outliers on the slope (Table 3) when compared to the 2017 and 2018 data sets. As the 2017 and 2018 data sets covered a larger range of pH values, the model fit was more constrained, resulting in a decreased effect of outliers and more consistent results across the 2 years. The approximately 1:1 fit in 2017 and 2018 characterised by the high r^2 values of 0.96 and 0.97 (Table 2) as well as the approximately 50% split in data points above and below the 1:1 model line (Figure 2d) can be explained by the minimal effect of sunlight exposure on the pH of fruit at harvest (Price et al. 1995, Intrigliolo et al. 2014).

The relationship between the pH of the visible berries and the whole bunch in the 3 years of the study was significantly different from 1:1 (P-value <0.001). In contrast to the slight but significant variation to the 1:1 model presented here, the statistically insignificant variation reported by Zoecklein et al. (1992) for berry samples with different levels of sunlight exposure in Chardonnay and Riesling at harvest can be explained by the small number of samples analysed (ten per replicate). Similarly, the insignificant variation in pH of shaded and sunlight-exposed bunches of Pinot Noir reported by Price et al. (1995) can be explained by the small number of samples (five) in the study resulting in a difference of 0.04 pH being reported as insignificant. All pH models exhibited <0.02 pH variation from a 1:1 relationship but were significantly different as a larger number of samples was used in our study. Therefore, we conclude that the effect of fruit sunlight exposure on pH is small but statistically significant. Across the 3 years of the study, using the 1:1 relationship for Chardonnay led to an absolute error of ±0.069 whereas using the linear model led to an error of ±0.053. For Shiraz, the error associated with using the 1:1 model assumption was ±0.062 compared with ±0.051 for the linear regression across the 3 years. With models for individual years, the 2016 models were more similar to 1:1 (difference in SER < 0.006), whereas for the other years of the study they were associated with an almost doubling of the errors of the linear regressions. Therefore, the 1:1 relationship to estimate pH should be used only if an accuracy of ±0.08 pH is acceptable.

Although Cao et al. (2010) observed an RMSEP of 0.1487, which they suggested was accurate enough for predicting pH (compared with RMSEP of 0.045 at most here), Iland et al. (2012) recommended a target accuracy of 95% for vineyard sampling, which is equivalent to a pH variation of 0.18 at most when using the mean of our data sets. All

pH models generated here have sufficient accuracy for vineyard sampling (±0.08 pH). In this work, we suggest that a 1:1 relationship for prediction is sufficient only if the error of the 1:1 model is not more than 10% greater than the trained linear model (Table 3). Therefore, we would not recommend that a 1:1 assumption be used to estimate pH for Chardonnay and Shiraz in the data range 2.56–3.48 and 2.95–4.04, respectively. Trained linear models represent a significant improvement.

All the models for Chardonnay were stable as they performed well across data from different years, with all predictions within ±0.02 pH (data not shown) from the best case estimates (estimates made using a model trained on the data in question). This resulted in predictions that were always within ±0.08 pH for the 2016 data set across all the models. The 2017 and 2018 predictions were always within ±0.05 pH across the different models. The 2017 and 2018 models for Shiraz performed well on the 2017 and 2018 data sets with predictions within ±0.005 pH (data not shown) from the best case estimates. The 2017 and 2018 models are able to predict the 2016 data set with prediction within ±0.03 pH of the best case estimates. Similar performance was seen for the 2016 model on the 2018 data set with slightly better performance on the 2017 data set (±0.02 pH).

It is interesting to note that even though there was variation in the range of pH values over which the data were sampled for both Chardonnay and Shiraz, relationships had a similar slope across years, provided data were sampled at more than one time interval during the season. Thus, we would suggest that linear regression models need to be developed for estimating pH using a minimum of two subsets of data sampled at least 1 week apart to maintain an accuracy of approximately ±0.05 pH. All of the developed models for Chardonnay were consistent across seasons and could be used interchangeably to estimate bunch pH from combined visible berry pH in the data range 2.56–3.48. For Shiraz, we would suggest either the 2017 or 2018 model could be used for estimating bunch pH in the range of 2.95–4.04. Temperature increases generally result in higher berry pH (Kliewer and Lider 1968, Spayd et al. 2002) and, therefore, it is unlikely that the regression models presented here will be usable in regions with a different average growing season temperature.

Titratable acidity

The relationship between the TA of the visible berries and that of the entire bunch was strong, with an r^2 value of 0.88 or higher. The 2016 data set for Chardonnay exhibited a significantly different slope of 0.82 when compared to that of 2017 and 2018 (Table 2). The 2016 Chardonnay slope suggested that visible berries had a TA value higher than that of non-visible berries. The large intercept value, however, means that this was true only at TA values greater than 16 g/L, which cannot be verified from our data set due to the small number of data points obtained in this range. Most of the Chardonnay data lay above the 1:1 line (Figure 2e), indicating that TA is higher in non-visible berries. As many of these are likely to be shaded, this result can be explained by the degradation of TA and malate at higher berry temperature (Reynolds et al. 1986) generally associated with increased sunlight exposure. The model slope for Shiraz was consistent (0.92–0.96) across the 3 years of the study.

Berry TA declines with ripening and most of the malic acid is metabolised (Iland et al. 2011) and this can be observed in our results as a closer fit to a 1:1 relationship (solid line) as TA drops in the Shiraz and Chardonnay. The higher bunch compactness of Chardonnay, and associated shade-related variation in berry temperature compared to Shiraz, may have contributed to the higher variability of berry TA > 8.5 g/L.

All regressions between visible berry TA and entire bunch TA differed significantly from 1:1 (P-values < 0.001 for all models except Shiraz in 2018, which had a P-value <0.01). For Shiraz, the maximum error associated with using a 1:1 relationship was ±0.544 g/L compared with an error of ±0.486 g/L for trained linear regression models. Therefore, a 1:1 assumption can be used only if an accuracy of around ±0.55 g/L is acceptable. For Chardonnay, the best accuracy achieved assuming a 1:1 relationship was ±0.864 g/L from the 2017 data set. Based on the 95% accuracy suggested by Iland et al. (2012), this translates to a TA variation of 0.42 g/L at most for Shiraz and 0.72 g/L for Chardonnay. Therefore, it is concluded that a 1:1 assumption is not suitable for estimation of bunch TA from visible berry TA in Chardonnay for the data range 6.1–19.95 g/L and Shiraz for the data range 4.24–8.67.

Regarding stability, the 2016 Chardonnay model performed poorly in 2017 and 2018, with an error of up to ±1.8 g/L. Much more stability was evident in the 2017 and 2018 Chardonnay models, which were able to predict both the 2017 and 2018 data sets within an accuracy of ±0.02 g/L. The 2017 and 2018 Chardonnay regression relationships were also able to predict the 2016 data set with an additional error of ±0.4 g/L, which is significantly better than the performance of the 2016 regression tested on the 2017 and 2018 data sets. For Shiraz, all models were quite stable with the poorest performance being within ±0.24 g/L, when predicting visible berry TA for the 2016 data set. Better performance was seen when predicting 2017 and 2018 data sets by all models with additional errors up to ±0.06 g/L. Therefore, all Shiraz models could be considered to be stable with the exception of the 2016. The Chardonnay 2016 model was unstable with an accuracy of ±1.12 g/L. Improved results, however, were evident for 2017 and 2018 (Table 2). Even though a combined model of 2017 and 2018 data presented a suitable level of error (±0.6 g/L), combining the 2016 data points in this data range resulted in an unsuitable level of error. Therefore, only the data range presented by 2017 data was suitable with a model error of ±0.78 g/L for the range 6.1–11.4 g/L across the 2-year data set. For Shiraz, a much lower error of ±0.39 g/L was present in the combined regression model over the 2 years, which we consider adequate for estimating TA in Shiraz within the data range 4.1–8.7 g/L. As increased temperature reduces TA in grape berries (Spayd et al. 2002), the TA models presented are expected to be suitable only for vineyards with similar growing season climate. For different climates, it is expected that the temperature gradient between the visible berries and the remainder of the bunch would differ.

Anthocyanin, phenolic substances and tannin

The concentration of anthocyanin, phenolic substances and tannin was expressed on a berry fresh mass basis. It is possible that if the concentration in skins was determined as distinct from the concentration in whole berries, a stronger fit (between skin concentration of visible berries and that of the rest of the bunch for each of the three maturity parameters above) may be obtained. This is due to the above compounds being largely expressed in the skin combined with

minimal changes to skin surface area when compared to the inner mesocarp (Roby et al. 2004). Nevertheless, as red wine is made from whole berries these are the most relevant measurement unit as their analysis will reflect the impact of fruit exposure and berry dehydration.

With an r^2 value of 0.75, the linear regression model used to estimate bunch anthocyanin from visible berry anthocyanin can be considered moderately strong. Although overexposure of bunches to sunlight can reduce anthocyanin concentration in berry skins (Bergqvist et al. 2001), our results (Table 2) suggest that visible berries exhibited a higher concentration of anthocyanin on average in the vineyards studied. Price et al. (1995) found that there was no significant difference in anthocyanin concentration for sunlight-exposed berry skin when compared to that of shaded berries, which contrasts with our results for visible and non-visible berries (P-value <0.001 in Table 2). Downey et al. (2004) found little effect of shading on anthocyanin concentration at harvest for bunches shaded from 9 weeks pre-veraison or earlier when compared to sunlight-exposed bunches. Leading up to harvest, however, differing trends between sunlight-exposed and shaded berries were seen when anthocyanin was measured as mg/g (as measured here) as opposed to mg/berry. The difference between the concentration of anthocyanin in shaded and sunlight-exposed berries appeared negligible up until 2 weeks after veraison (Downey et al. 2004) and then differed thereafter (Downey et al. 2004, Edo-Roca et al. 2013). The lack of consistency across years for anthocyanin accumulation (Downey et al. 2004) may explain some of the variance seen in our results (average r^2 of 0.75 and Figure 3a). Some bunches exhibited higher anthocyanin concentration in visible berries, whereas in other bunches non-visible berries exhibited a higher concentration of anthocyanin.

Spayd et al. (2002) found that some heat is required for anthocyanin synthesis, but that excessive hours above 35°C were detrimental for anthocyanin accumulation. These authors also showed that light was a factor in increasing anthocyanin concentration. Thus, it would be expected that differences between the sunlight-exposed berries and the average for the bunch will be more pronounced in warmer climates.

We also note that there were 6 days with high daily maximum temperature (>35°C) from post-veraison to harvest, which may have inhibited anthocyanin accumulation (Mori et al. 2007) in the visible berries, resulting in the anthocyanin concentration of bunches exceeding that of visible berries. This is more likely in instances where bunch temperature was higher, such as in bunches collected from the western side of the canopy.

Therefore, we conclude that complex relationships between anthocyanin accumulation in a bunch and factors such as temperature (Spayd et al. 2002, Tarara et al. 2008), TSS level (Dai et al. 2014) and light (Spayd et al. 2002, Downey et al. 2004, Tarara et al. 2008) cannot be estimated using a simple linear regression model between the anthocyanin concentration of visible berries and the bunch.

Using the suggested 95% accuracy as a benchmark (Iland et al. 2012), an accuracy only within ±0.15 mg/g is considered suitable. Our linear regression model was able to predict bunch anthocyanin to an accuracy of ±0.356 mg/g in the range of 1–2.5 mg/g berry mass with 95% confidence (P-value < 0.001), which is much higher than the desired accuracy for vineyard sampling. Although the prediction ability of the model is unsatisfactory, the relationship between the anthocyanin concentration of visible berries and the bunch appears to be linear and we conclude that the relationship obtained is promising.

For phenolic substances, consistent training and test errors (Table 1) indicate that a fitted regression model on all regression data is sufficient. The slope of 0.81 along with the number of points lying below the solid 1:1 line (Figure 3b) suggests that phenolic substances increase with sunlight exposure (Dokoozlian and Kliewer 1996, Bergqvist et al. 2001) as visible berries are likely to have a higher portion of the bunch sunlight exposure. A 1:1 model may overestimate bunch phenolic substances for a higher concentration of visible berry phenolic substances (Figure 3b), and this is seen in our results where there was a significant variation from the 1:1 model (P-value < 0.01). At a similar level of midday sunlight exposure, bunches could still vary in phenolic substances depending on location (north or south side of the canopy) (Bergqvist et al. 2001). This, however, was based in the Central San Joaquin Valley, which has a Mediterranean climate and higher temperature (average maximum of 25°C), which can inhibit accumulation of phenolic substances. This is not the case in our study, which was based in a location with an average maximum temperature of 18.1°C. Overall, the linear regression model presented here has an accuracy of prediction of ±0.22 mg/g berry mass compared to ±0.24 mg/g berry mass for the 1:1 assumption. Given that a prediction with 95% accuracy requires errors to be within ±0.1 mg/g berry mass, we cannot recommend the linear regression model or a 1:1 assumption for predicting phenolic substances for this data set.

For predicting bunch tannin from visible berry tannin, the slope of 0.8 suggests a slightly higher tannin concentration in visible berries, which may be related to increased sunlight exposure (Ristic et al. 2007). Downey et al. (2004) found that bunch shading had little effect on the tannin concentration (when measured in mg/g berry mass) from post-veraison to harvest in Shiraz. This appears in agreement with our results (Figure 3c). There is a statistically significant variation; however, from the 1:1 assumption (P-value < 0.001 in Table 2), with accuracy improving from ±0.94 to ±0.84 mg/g if a linear regression model is used. As there is approximately 10% variation between the two approaches, we consider the 1:1 assumption suitable for estimating the level of tannin in bunches from visible berry data if accuracy of ±0.94 mg/g is sufficient. A significant difference in bitterness was found for a tannin variation of 30 mg/L when comparing skin to seed bitterness (Brossaud et al. 2001). Given the 95% accuracy target suggested by Iland et al. (2012), we would not recommend the tannin regression model developed in this study. It is likely additional sampling would be required for model improvement.

Overall, it appears that there is a linear, and approximately 1:1, relationship between the tannin concentration of visible berries and that of bunches. When sunlight is excluded, temperature increases have been shown to result in higher seed tannin and decreased skin tannin (Ristic et al. 2007). This means that varying bunch temperature gradients for different climates may alter the slope of derived regressions.

Conclusion

Strong linear relationships between the visible berries and the entire bunch for the compositional parameters TSS

(r^2 = 0.82–0.99), pH (r^2 = 0.89–0.99) and TA (r^2 = 0.88–0.97) were observed for Chardonnay and Shiraz. Relationships between the visible berries and entire bunches had predictive power for anthocyanin and phenolic substances of r^2 = 0.75 and 0.76, respectively, and predictive power of r^2 = 0.8 for tannin, however, relationships were based on a limited data set. Although linear relationships were described, they were not always 1:1. The exception was for TSS in Shiraz, where a 1:1 relationship had a prediction error of ±0.76°Brix for data ranging from 17 to 28°Brix. Relationships predicting tannin also approximated a 1:1 relationship. Combined regression models across years were appropriate for TSS, pH and TA in Shiraz. A combined regression model was not appropriate for TSS and TA in Chardonnay as the range of values in 2016 was different to 2017 and 2018, leading to errors of ±1.1°Brix for TSS, and ±0.98 g/L for TA. Within the TSS range 16.5–24°Brix, however, a linear regression model, with an error of ±0.56°Brix, was appropriate for predicting bunch TSS from visible berry TSS for Chardonnay. A combined regression model for predicting bunch pH from visible berry pH was satisfactory, with an error margin of ±0.053 pH, which was similar to that for the individual season models (±0.066, ±0.048 and ±0.031 in 2016, 2017 and 2018, respectively).

In general, the regression models for Shiraz had lower prediction errors when compared to those for Chardonnay, which may be due to lower bunch compactness. We have also shown that across three seasons, and as long as the range of data collected is similar, linear models for TSS, pH and TA are sufficient for Shiraz and for pH of Chardonnay berries in the same vineyard. There is a need to develop predictive relationships for entire bunch compositional characters based on the composition of visible berries as a first step toward developing non-destructive methods for measuring grape composition. Our results suggest that variation in bunch compactness may contribute to variation in predictive relationships among cultivars, although we acknowledge that the data set is limited. Furthermore, relationships may need to be described for vineyards with different climates and training systems as the temperature differences between visible berries and non-visible berries are likely to differ as non-visible berries tend to be shaded.

This article investigated the strength and stability of linear relationships between visible berries and average bunch composition as a necessary step toward developing non-destructive, sensing-based systems in the vineyard. The relationships described were not always 1:1 and therefore models will need to be developed for different cultivars and different compositional parameters. For non-destructive sensing techniques such as the multiplex sensor, prediction errors (RMSEP%) of 5–9% for TSS (Herrera et al. 2003, Urraca et al. 2015), 3% for pH (Giovenzana et al. 2013), 18% for TA (Giovenzana et al. 2015) and 9–13% for anthocyanin (Ghozlen et al. 2010) have been reported. Comparing these with the results in this article for TSS and pH, the difference between using a 1:1 relationship and a trained regression model was insignificant in relation to errors associated with the sensors. For TA, there was a significant but small improvement when using a trained regression model over a 1:1 relationship. For anthocyanin, there was little difference between using a trained regression model and a 1:1 relationship.

References

Baluja, J., Diago, M.P., Goovaerts, P. and Tardaguila, J. (2012) Spatio-temporal dynamics of grape anthocyanin accumulation in a Tempranillo vineyard monitored by proximal sensing. Australian Journal of Grape and Wine Research **18**, 173–182.

Bergqvist, J., Dokoozlian, N. and Ebisuda, N. (2001) Sunlight exposure and temperature effects on berry growth and composition of Cabernet Sauvignon and Grenache in the Central San Joaquin Valley of California. American Journal of Enology and Viticulture **52**, 1–7.

Brossaud, F., Cheynier, V. and Noble, A.C. (2001) Bitterness and astringency of grape and wine polyphenols. Australian Journal of Grape and Wine Research **7**, 33–39.

Cao, F., Wu, D. and He, Y. (2010) Soluble solids content and pH prediction and varieties discrimination of grapes based on visible-near infrared spectroscopy. Computers and Electronics in Agriculture **71**, S15–S18.

Coombe, B.G. and McCarthy, M. (2000) Dynamics of grape berry growth and physiology of ripening. Australian Journal of Grape and Wine Research **6**, 131–135.

Cynkar, W., Cozzolino, D., Dambergs, R., Janik, L. and Gishen, M. (2004) The effects of homogenisation method and freezing on the determination of quality parameters in red grape berries of *Vitis vinifera*. Australian Journal of Grape and Wine Research **10**, 236–242.

Dai, Z.W., Meddar, M., Renaud, C., Merlin, I., Hilbert, G., Delrot, S. and Gome's, E. (2014) Long-term in vitro culture of grape berries and its application to assess the effects of sugar supply on anthocyanin accumulation. Journal of Experimental Botany **65**, 4665–4677.

Dokoozlian, N. and Kliewer, W. (1996) Influence of light on grape berry growth and composition varies during fruit development. Journal of the American Society for Horticultural Science **121**, 869–874.

Downey, M.O., Harvey, J.S. and Robinson, S.P. (2004) The effect of bunch shading on berry development and flavonoid accumulation in Shiraz grapes. Australian Journal of Grape and Wine Research **10**, 55–73.

Edo-Roca, M., Nadal, M. and Lampreave, M. (2013) How terroir affects bunch uniformity, ripening and berry composition in *Vitis vinifera* cvs. Carignan and Grenache. Journal International des Sciences de la Vigne et du Vin **47**, 1–20.

Fadock, M., Brown, R.B. and Reynolds, A.G. (2016) Visible-near infrared reflectance spectroscopy for nondestructive analysis of red winegrapes. American Journal of Enology and Viticulture **67**, 38–45.

Fernandes, A.M., Oliveira, P., Moura, J.P., Oliveira, A.A., Falco, V., Correia, M.J. and Melo-Pinto, P. (2011) Determination of anthocyanin concentration in whole grape skins using hyperspectral imaging and adaptive boosting neural networks. Journal of Food Engineering **105**, 216–226.

Fernández-Novales, J., López, M.-I., Sánchez, M.-T., Garćia-Mesa, J.-A. and González-Caballero, V. (2009) Assessment of quality parameters in grapes during ripening using a miniature fiber-optic near-infrared spectrometer. International Journal of Food Sciences and Nutrition **60**, 265–277.

Ghozlen, N.B., Cerovic, Z.G., Germain, C., Toutain, S. and Latouche, G. (2010) Non-destructive optical monitoring of grape maturation by proximal sensing. Sensors **10**, 10040–10068.

Giovenzana, V., Beghi, R., Malegori, C., Civelli, R. and Guidetti, R. (2013) Wavelength selection with a view to a simplified handheld optical system to estimate grape ripeness. American Journal of Enology and Viticulture **65**, 117–123.

Giovenzana, V., Civelli, R., Beghi, R., Oberti, R. and Guidetti, R. (2015) Testing of a simplified LED based vis/NIR system for rapid ripeness evaluation of white grape (*Vitis vinifera* L.) for Franciacorta wine. Talanta **144**, 584–591.

González-Caballero, V., Pérez-Maŕın, D., López, M.-I. and Sánchez, M.-T. (2011) Optimization of NIR spectral data management for quality control of grape bunches during on-vine ripening. Sensors **11**, 6109–6124.

Guidetti, R., Beghi, R. and Bodria, L. (2010) Evaluation of grape quality parameters by a simple Vis/NIR system. Transactions of the American Society of Agricultural and Biological Engineers **53**, 477–484.

Hall, A. and Jones, G.V. (2010) Spatial analysis of climate in winegrape-growing regions in Australia. Australian Journal of Grape and Wine Research **16**, 389–404.

Herrera, J., Guesalaga, A. and Agosin, E. (2003) Shortwave-near infrared spectroscopy for non-destructive determination of maturity of wine grapes. Measurement Science and Technology **14**, 689–697.
Iland, P., Dry, P., Proffitt, T. and Tyerman, S. (2011) The grapevine: from the science to the practice of growing vines for wine (Patrick Iland Wine Promotions: Adelaide, SA, Australia).
Iland, P., Bruer, N., Ewart, A., Markides, A. and Sitters, J. (2012) Monitoring the winemaking process from grapes to wine: techniques and concepts, 2d ed (Patrick Iland Wine Promotions: Adelaide, SA, Australia).
Iland, P., Bruer, N., Edwards, G., Caloghiris, S. and Wilkes, E. (2013) Chemical analysis of grapes and wine: techniques and concepts, 2nd edn (Patrick Iland Wine Promotions: Adelaide, SA, Australia).
Intrigliolo, D.S., Llacer, E., Revert, J., Esteve, M.D., Climent, M.D., Palau, D. and Gomez, I. (2014) Early defoliation reduces cluster compactness and improves grape composition in Mando, an autochthonous cultivar of *Vitis vinifera* from southeastern Spain. Scientia Horticulturae **167**, 71–75.
Kasimatis, A.N. and Vilas, E.P. (1985) Sampling for degrees Brix in vineyard plots. American Journal of Enology and Viticulture **36**, 207–213.
Kliewer, W. and Lider, L. (1968) Influence of cluster exposure to the sun on the composition of Thompson Seedless fruit. American Journal of Enology and Viticulture **19**, 175–184.
Krstic, M. (2003) Growing quality grapes to winery specifications: quality measurements and management options for grape growers (Winetitles: Adelaide, SA, Australia).
Krstic, M.P., Leamon, K., DeGaris, K., Whiting, J., McCarthy, M. and Clingeleffer, P. (2011) Sampling for wine grape quality parameters in the vineyard: variability and post-harvest issues. Blair, R.J., Williams, P.J. and Hoj, P.B., eds. Proceedings of the eleventh Australian wine industry technical conference; 7–11 October 2001; Adelaide, SA, Australia (Australian Wine Industry Technical Conference: Urrbrae, SA, Australia) pp. 87–90.
Larráın, M., Guesalaga, A.R. and Agosín, E. (2008) A multipurpose portable instrument for determining ripeness in wine grapes using NIR spectroscopy. IEEE Transactions on Instrumentation and Measurement **57**, 294–302.
Mabrouk, H. and Sinoquet, H. (1998) Indices of light microclimate and canopy structure of grapevines determined by 3D digitising and image analysis, and their relationship to grape quality. Australian Journal of Grape and Wine Research **4**, 2–13.
Mori, K., Goto-Yamamoto, N., Kitayama, M. and Hashizume, K. (2007) Loss of anthocyanins in red-wine grape under high temperature. Journal of Experimental Botany **58**, 1935–1945.
Olarte Mantilla, S.M., Collins, C., Iland, P.G., Johnson, T.E. and Bastian, S.E.P. (2012) Review: berry sensory assessment: concepts and practices for assessing winegrapes' sensory attributes. Australian Journal of Grape and Wine Research **18**, 245–255.
Pagay, V. and Cheng, L. (2010) Variability in berry maturation of Concord and Cabernet Franc in a cool climate. American Journal of Enology and Viticulture **61**, 61–67.
Price, S., Breen, P., Valladao, M. and Watson, B. (1995) Cluster sun exposure and quercetin in Pinot Noir grapes and wine. American Journal of Enology and Viticulture **46**, 187–194.
Rankine, B.C. (1995) Making good wine: a manual of winemaking practice for Australia and New Zealand (Pan Macmillan Australia: Sydney, NSW, Australia).
Rankine, B.C., Cellier, K. and Boehm, E.W. (1962) Studies on grape variability and field sampling. American Journal of Enology and Viticulture **13**, 58–72.
Rao, R.B., Fung, G. and Rosales, R. (2008) On the dangers of cross-validation. An experimental evaluation. Apte, C., Park, H., Wang, K. and Zaki, M.J., eds. Proceedings of the eighth SIAM international conference on data mining; 24–26 April 2008; Atlanta, GA, USA (Society for Industrial and Applied Mathematics: Philadelphia, PA, USA) pp. 588–596.
Reynolds, A., Pool, R. and Mattick, L. (1986) Influence of cluster exposure on fruit composition and wine quality of Seyval Blanc grapes. Vitis **25**, 85–95.
Ristic, R., Downey, M.O., Iland, P.G., Bindon, K., Francis, I.L., Herderich, M. and Robinson, S.P. (2007) Exclusion of sunlight from Shiraz grapes alters wine colour, tannin and sensory properties. Australian Journal of Grape and Wine Research **13**, 53–65.
Roby, G., Harbertson, J.F., Adams, D.A. and Matthews, M.A. (2004) Berry size and vine water deficits as factors in winegrape composition: anthocyanins and tannins. Australian Journal of Grape and Wine Research **10**, 100–107.
Sabbatini, P. and Howell, G.S. (2010) Effects of early defoliation on yield, fruit composition, and harvest season cluster rot complex of grapevines. HortScience **45**, 1804–1808.
Spayd, S.E., Tarara, J.M., Mee, D.L. and Ferguson, J.C. (2002) Separation of sunlight and temperature effects on the composition of *Vitis vinifera* cv. Merlot berries. American Journal of Enology and Viticulture **53**, 171–182.
Tarara, J.M., Lee, J., Spayd, S.E. and Scagel, C.F. (2008) Berry temperature and solar radiation alter acylation, proportion, and concentration of anthocyanin in Merlot grapes. American Journal of Enology and Viticulture **59**, 235–247.
Trought, M.C.T. and Bramley, R.G.V. (2011) Vineyard variability in Marlborough, New Zealand: characterising spatial and temporal changes in fruit composition and juice quality in the vineyard. Australian Journal of Grape and Wine Research **17**, 79–89.
Trought, M.C.T., Naylor, A.P. and Frampton, C. (2017) Effect of row orientation, trellis type, shoot and bunch position on the variability of Sauvignon Blanc (*Vitis vinifera* L.) juice composition. Australian Journal of Grape and Wine Research **23**, 240–250.
Urraca, R., Sanz-Garcia, A., Tardaguila, J. and Diago, M.P. (2015) Estimation of total soluble solids in grape berries using a hand-held NIR spectrometer under field conditions. Journal of the Science of Food and Agriculture **96**, 3007–3016.
Wolpert, J.A. and Howell, G.S. (1984) Sampling Vidal Blanc grapes. II. Sampling for precise estimates of soluble solids and titratable acidity of juice. American Journal of Enology and Viticulture **35**, 242–246.
Zoecklein, B., Wolf, T., Duncan, N., Judge, J. and Cook, M. (1992) Effects of fruit zone leaf removal on yield, fruit composition, and fruit rot incidence of Chardonnay and white Riesling (*Vitis vinifera* L.) grapes. American Journal of Enology and Viticulture **43**, 139–148.

Manuscript received: 30 April 2018
Revised manuscript received: 2 September 2018
Accepted: 10 September 2018