DOI: 10.58895/ksp/1000044906-1 erschienen in:
OCM 2015 - 2nd International Conference on Optical Characterization of Materials, March 18th – 19th, 2015, Karlsruhe, Germany : Conference Proceedings
DOI: 10.58895/ksp/1000044906 | https://www.ksp.kit.edu/site/books/m/10.58895/ksp/1000044906/

Non-destructive determination of grape berry sugar concentration using visible/near infrared imaging and possible impact on wine quality

Magali Lafontaine,[1]* Zrinka Bockaj,[2]* Maximilian Freund,[1]
Kai-Uwe Vieth[2] and Christian Negara[2]

[1] Hochschule Geisenheim University,
Von Lade Str. 1, 65366 Geisenheim (Germany)
[2] Fraunhofer IOSB,
Fraunhoferstrasse 1, 76131 Karlsruhe (Germany)
* both authors contributed equally to this work

Abstract Reducing heterogeneity in harvest material may be beneficial for wine quality and this goal may be achieved through advanced berry sorting systems. The general aim is to assess if a relationship could be found between berry sugar concentration and hyperspectral images to determine the possible impact on wine quality. Grapes were picked at different stages of maturity in a one-year time interval and the berries were sorted according to their size and density. Hyperspectral images of the groups were obtained in the vis/NIR wavelength range with a complete spectrum from 400 nm to 2500 nm. Our results showed that vis/NIR images can be used as a tool to improve the segregation of berries from all tested grape varieties based on their sugar content. The PLSR algorithm is trained on all grape varieties together and later validated on each variety separately, proving the possibility of using a general regression model with constant parameters to predict sugar concentrations. Finally, the impact on quality was tested for red wines. Pinot noir berries with higher sugar concentrations presented more color since anthocyanin concentration was higher. Nevertheless, tannin concentration in skins and seeds tended to decrease with increasing sugar concentration. Groups with higher sugar concentration resulted in wines with higher anthocyanin and lower tannin concentration.

1 Introduction

The extension to which variability of quality traits for wine production is encountered in commercial vineyards has been highlighted by the development of new mapping and sorting techniques. In recent years, optical berry sorting machines have been developed permitting to eliminate material other than grape and rotten berries. Advanced berry sorting systems further allow for discrimination between berries of different size and color. This technology would therefore allow producers to create wines of different quality or style from a single production unit in a targeted way. The project ”GrapeSort” financed by a German ZIM research project of the BMWI involved two research institutions, Fraunhofer IOSB and Hochschule Geisenheim University with two industrial partners. The aim of this project was to determine suitable parameters correlated with grape and wine composition to be implemented as a quality criterion for improving sorting methods of the fruits [1]. One challenge would be to discriminate the berries according to their ripeness as a high variability, up to 50%, in their sugar concentration can be found. The aim of the study was to determine if a relationship could be found between berry sugar concentration and hyperspectral images of several varieties to implement this methods on a sorting machine. The second step was to assess the possible relevance of sorting berry according to their sugar concentration in relation to wine quality.

2 Berry size and density segregation

While focusing mainly on 3 varieties, in 2013 *Vitis vinifera* L. cv. Pinot noir, Pinot blanc and Riesling, the sortiment was broadend in 2014 with Pinot précoce, Dornfelder, Acolon, Pinot meunier, Lemberger, Trollinger, Müller Thurgau, Pinot gris and Gewürztraminer. Berries collected at different stages of maturity were classified according to their diameter with a sip column and then segregated according to their density in an interval of 5° Oe. All the berries were introduced in the less dense solution, the floating berries were collected, thereafter, the concentration of the solution was increased. The same process was repeated until all the berry population was sorted. Sorting according to density was successful as density solution concentration was correlated with berry sugar concentration ($R^2 = 0.98$). For each date, the berries were distributed according to a gaussian bell-shape showing the heterogeneity in sugar concentration up to 50% (Pinot noir 2013 : Figure 1.2A) and it seemed that the variability of the distribution was growing until harvest. Hyperspectral images were captured in the vis/NIR wavelength range with a complete spectrum from 400 nm to 2500 nm on 20-berries subsamples from each density group.

3 Hyperspectral imaging for sugar concentration prediction

The conducted measurements confirm the validity of involving NIR spectroscopy with the primary focus in determining the sugar concentration of grapes. In this paper, we continue the work of [2] and include more grape varieties in the general learning phase. Followed by the application of multivariate data analysis, we investigated the correlation of the prepared hyperspectral images with the laboratory data. At the end, the final validation of the results obtained with the regression analysis will be evaluated and compared for different grape varieties.

3.1 Regression analysis and data preparation

Although the main purpose of determining the sugar concentration is to sort berries into two different groups (with low and high sugar concentration), which generally implies the usage of classification algorithms, we found out that regression algorithms yield better results. Additionally, the possibility of changing the process of automatic berry sorting by only introducing a variable threshold for sugar concentration would significantly simplify the work of the winemaker. The regression problem can be transformed into a two-class classification problem by assigning the label ”high sugar concentration” (class A) or ”low sugar concentration” (class B) according to a defined threshold value, e.g. the median of the ground truth data. Another benefit in contrast to classification algorithm is that the complete information about the ground truth is taken into account, not only the class affiliation.

Moreover, it is not necessary to retrain the classifier every time when changing the threshold, but only to apply the new threshold value after the regression to define the classes for high and low sugar concentration.

Since the PLSR *(Partial Least Square Regression)* [3] has already been proven as an effective regression algorithm often used with hyperspectral imaging and also in the earlier phases of this project [2], we continued using it for the purpose of this study. PLSR attempts to determine the relationship between a dependent variable (also called *responses*) and one or more independent variables (also called *predictors*) by extracting from the predictor a set of orthogonal factors called latent variables that have the best predictive power.

In our case, the optimal number of latent variables proved to be 20. Moreover, the hyperspectral images of the berries will be used as predictors, and in laboratory measured sugar concentration of the berries as responses. As a result of regression analysis, the general regression model will be obtained as an vector of weights for each of the 99 features of the spectra. Finally, the sugar concentration values can be predicted from original hyperspectral images for each berry using a vector multiplication between a spectrum written as a vector and the computed regression coefficients. It has turned out, that the spectra from the SWIR wavelength range from 1000 nm to 2500 nm are not as useful as the vis/NIR spectra from the wavelength range from 400 nm to 1000 nm to assess the sugar concentration. Therefore the analysis is restricted to the latter.

During the training of the regression algorithm, additional adjustments to the original datasets have been applied, resulting in big improvements of the final results. First of all, the mean spectrum is computed for each berry (object) by averaging all spectra of every pixel of a given berry which also speeds up the whole process by reducing the data which has to be processed.

Furthermore, in this specific case, only the upper range of the vis/NIR wavelength range (708 - 1025 nm), represented by 99 features is extracted and used for regression because the difference in the results when using the whole and half wavelength-band is insignificant. They are even slightly better in the second case. One possible explanation of this phenomena could be that the information contained in the lower wavelength-band is not important for regression and could probably contain noise.

The dataset on which the regression is performed consists of two different datasets taken in 2013 and 2014 containing 12 different varieties with following properties shown in Table 1.1.

Table 1.1: Properties of each dataset used to train the regressor

Dataset	**Dates**	**Different grape varieties**	**Number of pixels**	**Number of objects (berries)**	**Amount of pixel used**
2013	07.10.2013 17.09.2013	3	1,218,687	1,248	100 %
2014	29.08.2014 01.09.2014 12.09.2014 15.09.2014 26.09.2014 30.09.2014	9	1,991,737	1,980	100 %

3.2 Validation and results

To ensure internal validity, the results were computed using one-fold cross-validation.

Further evidence that supports our basic premises about creating the regression model with one combined dataset are shown in Table 1.2. Including a time range of one year and 12 different grape varieties we still gained good results on only one specific grape variety. As a measurement for quality of the regression results, two different metrics will be used: RMSE *(root-mean-square-error)* [3] and PPMC *(Pearson Product Moment Correlation)* [3]. RMSE measures the typical distance of the data to the regression line and will be computed as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{1.1}$$

where y_i are the ground truths , $\hat{y}_i$ the predicted values and n the number of features.

The correlation coefficient measures the strength and the direction of a linear relationship between two variables (the ground truths and the predicted values).

Table 1.2: Regression results for each grape variety of dataset 2013 and 2014

Grape varieties	Correlation coefficients	RMSE
Pinot noir 2013	0.829	6.601
Pinot blanc 2013	0.918	5.553
Riesling 2013	0.930	5.855
Gewürztraminer	0.875	4.759
Pinot Meunier	0.868	6.358
Dornfelder	0.830	4.867
Pinot blanc	0.843	5.200
Pinot meunier	0.751	5.122
Acolon	0.832	4.819
Pinot précoce	0.892	5.627
Pinot noir	0.823	5.290
Müller Thurgau	0.823	6.353
Lemberger	0.807	6.260
Trollinger	0.854	4.621

The regression performance of the PLSR algorithm trained with the images of the combined datasets containing all grape varieties taken in 2013 and 2014 was validated on the images of Pinot blanc and Müller Thurgau from 2014 and shown in Figure 1.1. In addition to that, RGB images are also shown, but they differ only marginally and cannot be used in assessing the sugar concentration. These results give a significant insight in the predictive possibilities of our method.

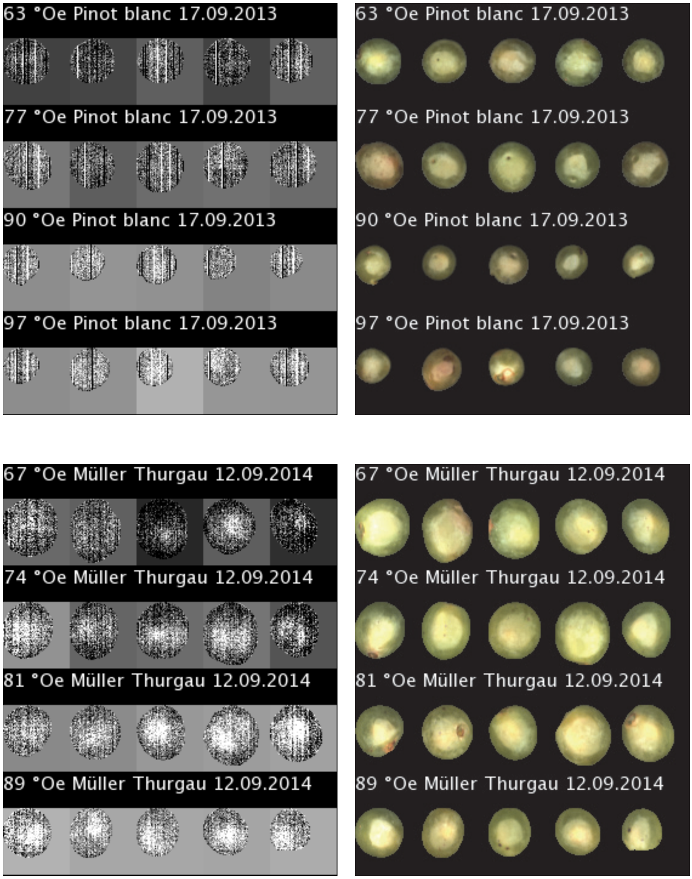


Figure 1.1: Predicted sugar concentration using the PLSR algorithm and RGB images in comparison. The algorithm was trained with 12 different grape varieties with a range of sugar concentrations: 37 - 116 °Oe. The background for grey-value images shows the mean predicted sugar concentration for each berry.

We indeed showed that it would be possible to discriminate the berries according to their sugar concentration by means of NIR spectrometry. The further question to be answered is if it would be relevant to sort berries according to their sugar concentration. One trial with 100kg Pinot Noir in 2013 [1] showed that the two sorted batches were different in average sugar concentration by 10%. In the trial in 2013 only an RGB camera and an NIR camera with a bandpass filter were used. Therefore, the segregation according to the sugar concentration was not as good as the achievable segregation using NIR spectrometry.

4 Possible impact on wine quality

4.1 Primary compounds

Results of Pinot noir picked on the 06.10.2013 are presented to illustrate the question. Variability in sugar concentration was important as the range of sugar concentration was between 70° Oe and 110° Oe for one single sampling date (Table1.3). With increasing sugar concentration, the total acidity (together with the malic acid concentration) were decreased, while pH increased. The alpha amino acids increased with increasing maturity meaning that an increasing concentration of nitrogen would be available for the yeast during fermentation. All the parameter indicate an increase in quality with increasing sugar concentration.

Table 1.3: Primary compounds analysis of the musts according to the sugar concentration of the different groups. Example for Pinot noir berry population picked on the 06.10.2013.

Sugar content Oe	Amino acids mg/L	Total acidity g/L	Malic acid g/L	pH
70	200	11.10	8.40	2.80
77	201	10.90	8.15	2.90
83	208	9.75	6.72	3.09
90	230	8.97	5.65	3.06
97	256	8.72	5.39	3.18
103	276	8.85	5.66	3.16
110	278	9.12	5.82	3.23

4.2 Secondary compounds

Berry phenolic compounds are crucial to red wine quality as they contribute to red wine color, color stability, structure and mouthfeel. The analysis of phenolics by wet chemistry was proceeded according to [4] for anthocyanins (A) and tannins (T).

Berry anthocyanins contribute to the color of a red wine, their accumulation starts in berry skin after the onset of ripening, veraison and is at maximum around harvest. Anthocyanin concentration and content increased with increasing sugar concentration in berries (Figure 1.2B). This is in accordance with previous work that berries from higher density showed an increase in anthocyanin concentration for several red varieties [5, 6]. Indeed anthocyanin concentration would be strongly correlated to sugar accumulation in berries [7]. Sucrose has been shown to induce anthocyanin production in grape skin [8] what was related to a modulation of the transcription of some key biosynthetic enzymes of the phenylpropanoid pathway [9].

Berry tannins participate mainly to the mouthfeel of a red wine and their accumulation occurs early in grape development, mainly before veraison in both berry skin and seeds. Tannin concentration in seeds and skins was highest in less ripen samples and their concentration in berries decreased with increasing density. But though the decrease in seed tannins seemed established, some studies reported no major changes for tannins of berry skin [10] what may depend on overall ripeness [5, 11]. However, it was pointed out that not only the amount but the extractability of the tannin compounds is changing due to the different interaction of tannins binding with cell wall proteins or polysaccharides.

4.3 Wine analysis

As it would be difficult to predict the actual extraction of phenolics into wine from berry analysis, a method of micro-scale winemaking was developed [12] to produce wine and the method for phenolic extraction was repeatable as the relative standard deviation (RSD) represented around 8%-11% (TP), 5-12% (A) and 8-12% (T).

Wine anthocyanins concentration increased with increasing density (Figure 1.2C). This was already observed for Cabernet sauvignon wine

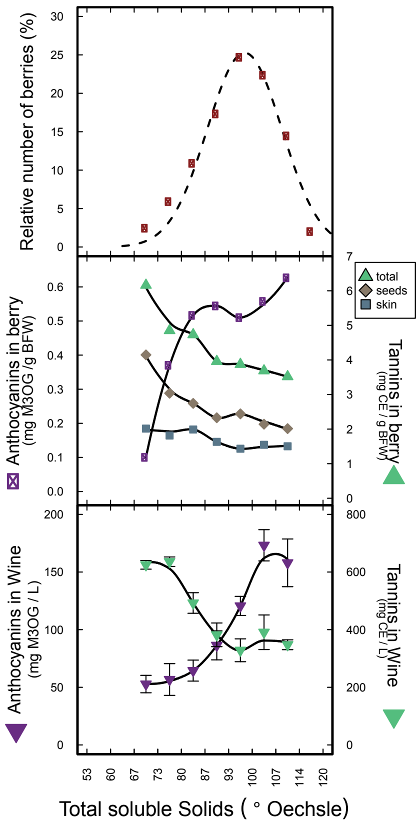


Figure 1.2: Results for Pinot noir bunch samples picked on the 16.10.2013. A: Distribution of the berries according to their density group as ° Oe. B: Berry phenolics compounds as mg per g berry fresh weight according to their density group, Anthocyanins (A) in berry skin, tannins (T) in berry skins and berry seeds together with total tannin concentration. C: Wine phenolics compounds as mg per L according to their density group, Anthocyanins (A) in wines, tannins (T) in wines.

as anthocyanin concentration and color intensity increased with higher density [13]. This may however be due to the higher concentration of anthocyanins in berry skin or higher ethanol concentration, with higher anthocyanin extraction [13]. A change in extractability of anthocyanins with higher density may also be an explanation as an increase of extractability through grape ripening was observed [14], as a consequence of the cellular wall degradation by pectolytic enzymes, making cell wall pectins permeable to the changes that occur during vinification. Therefore, not only the accumulation of the anthocyanins but also the ease of extraction would be the main factors affecting their extraction into wines [15].

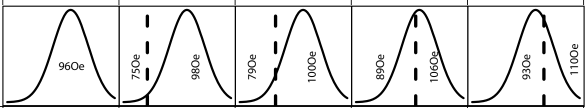
Wine tannins concentration decreased with increasing density. Reports on extraction of tannins are more confusing as for anthocyanins. A significant increase of tannin concentration of the wine with the density of the grapes was previously reported [13]. It was suggested that insufficiently ripened grapes would have a lower extractability of tannins from skins and a higher extractability of tannins from seeds [16] showing a higher contribution of seeds when less ripen [13, 17]. Discrepancies in results may be explained by the berry size, indeed, when sorting berries according to their density and size, wine tannin concentration decreased with increasing maturity [18] but berry size seemed to play larger role.

5 Setting the threshold

We showed that hyperspectral images in the vis/NIR wavelength range can be used for the segregation of berries based on their sugar content. The winemaker could easily transform the regression into the classification by simply giving the threshold value for the sugar concentration to separate berries below and above this threshold. However, deciding about the threshold would be another challenge to be solved depending on the type of wine targeted (Figure 1.3).

6 Conclusion and outputs

We showed that hyperspectral images in the vis/NIR wavelength range can be used as a tool to improve the segregation of berries from all tested grape varieties based on their sugar content. Indeed, discriminating berries according to their sugar concentration would allow to modify



	A	B		C		D		E	
	no sorting	threshold **77°Oe**		threshold **83°Oe**		threshold **90°Oe**		threshold **97°Oe**	
%	100%	2%	98%	19%	81%	61%	39%	83%	17%
TSS (°Oe)	96°Oe	75°Oe	98°Oe	79°Oe	100°Oe	89°Oe	106°Oe	93°Oe	110°Oe
T (mg/L)	398	553	376	472	361	408	373	399	348
A (mg/L)	121	60	127	72	135	113	167	120	158

Figure 1.3: Theoretical wine composition calculated based on the results of Pinot noir (Figure 1.2). Case study depending on the threshold separating the "low" and "high" sugar concentration groups: A: no sorting, B: threshold at 77° Oe, C: threshold at 83° Oe, D: threshold at 90° Oe, E: threshold at 97° Oe.

the wine style as an increase in sugar would lead to wines with more color since anthocyanin concentration in berries was higher. Nevertheless, tannin concentration in skins and seeds tended to decrease with increasing sugar concentration leading to wines with decreasing tannin concentration. In further steps, the results of the PLSR algorithm with vis/NIR spectra are used for designing the camera system of the sorting machine.

References

1. M. Lafontaine, M. Freund, K.-U. Vieth, and C. Negara, "Automatic fruit sorting by non-destructive determination of quality parameters using visible/near infrared to improve wine quality: I. Red wine production," *NIR news*, vol. 24, no. 8, pp. 6–8, 2013.
2. C. Negara, K.-U. Vieth, M. Lafontaine, and M. Freund, "Automatic fruit sorting by non-destructive determination of quality parameters using visible-near infrared to improve wine quality: II. Regression analysis," *NIR news*, vol. 25, no. 1, pp. 4–6, 2014.
3. W. Kessler, *Multivariate Datenanalyse: für die Pharma-, Bio- und Prozessanalytik*. Wiley-VCH Verlag GmbH and Co. KGaA Weinheim, 2006.
4. J. F. Harbertson, J. A. Kennedy, and D. O. Adams, "Tannin in Skins and

Seeds of Cabernet Sauvignon, Syrah, and Pinot noir Berries during Ripening," *American Journal of Enology and Viticulture*, vol. 53, no. 1, pp. 54–59, 2002.

5. F. Torchio, E. Cagnasso, V. Gerbi, and L. Rolle, "Mechanical properties, phenolic composition and extractability indices of Barbera grapes of different soluble solids contents from several growing areas," *Analytica Chimica Acta*, vol. 660, no. 1-2, pp. 183–9, 2010.
6. S. Río Segade, E. S. Vázquez, and E. D. Losada, "Influence of ripeness grade on accumulation and extractability of grape skin anthocyanins in different cultivars," *Journal of Food Composition and Analysis*, vol. 21, no. 8, pp. 599 – 607, 2008.
7. A. Pirie and M. G. Mullins, "Interrelationships of Sugars, Anthocyanins, Total Phenols and Dry Weight in the Skin of Grape Berries during Ripening," *American Journal of Enology and Viticulture*, vol. 28, no. 4, pp. 204–209, 1977.
8. A. J. G. Pirie and M. G. Mullins, "Changes in anthocyanin and phenolics content of grapevine leaf and fruit tissues treated with sucrose, nitrate and abscisic acid," *Plant Physiology*, vol. 58, pp. 468–472, 1976.
9. M. Ferri, L. Righetti, and A. Tassoni, "Increasing sucrose concentrations promote phenylpropanoid biosynthesis in grapevine cell cultures," *Journal of Plant Physiology*, vol. 168, no. 3, pp. 189–195, 2011.
10. D. Fournand, A. Vicens, L. Sidhoum, J. M. Souquet, M. Moutounet, and V. Cheynier, "Accumulation and Extractability of Grape Skin Tannins and Anthocyanins at Different Advanced Physiological Stages," *Journal of Agricultural and Food Chemistry*, vol. 54, no. 19, pp. 7331–7338, 2006.
11. L. Rolle, F. Torchio, S. Giacosa, S. R. Segade, E. Cagnasso, and V. Gerbi, "Assessment of Physicochemical Differences in Nebbiolo Grape Berries from Different Production Areas and Sorted by Flotation," *American Journal of Enology and Viticulture*, vol. 63, no. 2, pp. 195–204, June 2012.
12. M. Lafontaine, R. Scholz, M. Stoll, and H. R. Schultz, "Anthocyanin concentration in Pinot noir berries and their extraction into wine as affected by rootstock," *Proceedings 17th International Symposium GiESCO*, vol. Asti, Alba CN, Italy 29.08-02.09, 2011.
13. N. Kontoudakis, M. Esteruelas, F. Fort, J. M. Canals, V. D. Freitas, and F. Zamora, "Influence of the heterogeneity of grape phenolic maturity on wine composition and quality," *Food Chemistry*, vol. 124, no. 3, pp. 767 – 774, 2011.
14. N. Saint-Cricq de Gaulejac, N. Vivas, and Y. Glories, "Maturation phénolique des raisins rouges. relation avec la qualité des vins. compara-

ison des cépages merlot et tempranillo," *Progrès Agricole et Viticole*, vol. 115, no. 13-14, pp. 306–318, 1998.

15. I. Romero-Cascales, A. Ortega-Regules, J. M. Lopez-Roca, J. I. Fernandez-Fernandez, and E. Gomez-Plaza, "Differences in Anthocyanin Extractability from Grapes to Wines According to Variety," *American Journal of Enology and Viticulture*, vol. 56, no. 3, pp. 212–219, 2005.
16. C. Peyrot des Gachons and J. A. Kennedy, "Direct Method for Determining Seed and Skin Proanthocyanidin Extraction into Red Wine," *Journal of Agricultural and Food Chemistry*, vol. 51, no. 20, pp. 5877–5881, 2003.
17. J. L. Pastor del Rio and J. A. Kennedy, "Development of Proanthocyanidins in *Vitis vinifera* L. cv. Pinot noir Grapes and Extraction into Wine," *American Journal of Enology and Viticulture*, vol. 57, no. 2, pp. 125–132, 2006.
18. M. Lafontaine, M. Stoll, and H. R. Schultz, "Berry size and maturity affecting phenolic extraction in Pinot Noir wines," in *Proceedings 18th International Symposium GiESCO*, J. Queiroz and A. Carneiro, Eds., vol. 28. Porto (Portugal) 07.07-11.07: Ciência e Técnica Vitivinícola, 2013, pp. 396–400.