

Near-infrared spectroscopy and pattern-recognition processing for classifying wines of two Italian provinces

A.G. Mignani ^a, L. Ciaccheri ^{a*}, B. Gordillo ^b, A.A. Mencaglia ^a, M.L. González-Miret ^b,
F.J. Heredia ^b, A. Cichelli ^c

^a CNR-Istituto di Fisica Applicata “Nello Carrara”

Via Madonna del Piano, 10 – 50019 Sesto Fiorentino (FI), Italy

^b Lab. Color y Calidad de Alimentos, Univ. de Sevilla, Facultad de Farmacia – 41012 Sevilla, Spain

^c Università degli Studi “G. D’Annunzio”, Dip. Economia – 65127 Pescara, Italy

ABSTRACT

This paper presents an experiment making use of the near-infrared spectrum for distinguishing the wines produced in two close provinces of Abruzzo region of Italy. A collection of 32 wines was considered, 18 of which were produced in the province of Chieti, while the other 14 were from the province of Teramo. A conventional dual-beam spectrophotometer was used for absorption measurements in the 1300-1900 nm spectroscopic range. Principal Component Analysis was used for explorative analysis. Score maps in the PC1-PC2 or PC2-PC3 spaces were obtained, which successfully grouped the wine samples in two distinct clusters, corresponding to Chieti and Teramo provinces, respectively. A modelling of dual-band spectroscopy was also proposed, making use of two LEDs for illumination and a PIN detector instead of the spectrometer. These data were processed using Linear Discriminant Analysis which demonstrated satisfactory classification results.

Keywords: wine, classification, spectroscopy, NIR, geographic origin, multivariate data analysis

1. CLASSIFICATION OF WINES: WHY OPTICAL TECHNOLOGIES

The promotion of wines with a unique geographical connotation is considered a strategic factor for protecting and boosting the European share of the wine market. Labels bearing denomination of origin (PDO) and geographical indication (PGI) trademarks are often used to highlight the peculiarities of wines, for better visibility to consumers and differentiations with respect to similar products with lower price. The *terroir*, as the *specificity of place*, has a fundamental influence on the wine quality. The role of the *terroir* includes not only the soil type of that region, but also the climate, the weather, the vines and vineyards, and anything else that can possibly differentiate one piece of land from another, even close.

Investigating and classifying wine differences for authentication purposes has been widely accomplished using conventional analytical techniques such as high performance liquid chromatography ¹, gas chromatography ², liquid chromatography/mass spectrometry ³, and elemental analysis ⁴.

Beside analytical techniques suitable for laboratory use only, optical spectroscopy has emerged as a rapid and non-destructive tool for quick measurements, since the entire spectrum from the ultraviolet to the mid-infrared is capable to highlighting even minimal differences between wines ^{5, 6, 7, 8, 9, 10, 11, 12, 13}. In fact, tannins, phenolic compounds, other pigments, and the different content of water, sugars, and ethanol, greatly influence the optical spectrum and allow for distinguishing the different geographic areas of wines. Spectroscopic data are usually processed by means of multivariate data analysis, demonstrating that the combination of spectroscopy and chemometrics provides a modern and straightforward tool for wine classification and authentication. The near-infrared band, which is particularly informative for the concentration of water, sugars, and ethanol, demonstrated effectiveness for assessing the chemical composition and the aroma, and for monitoring the fermentation process ^{14, 15, 16, 17, 18, 19}.

This paper presents an experiment which was carried out in a small band of the near-infrared spectrum for distinguishing the wines produced in two close provinces of Abruzzo, which is a central region of Italy. A conventional dual-beam

* Email: l.ciaccheri@ifac.cnr.it – phone: +39 055 522 6322

spectrophotometer was used for absorption measurements in the near-infrared range, showing the most significant differences among the various samples in the 1300-1900 nm band. Principal Component Analysis was firstly used for explorative analysis. Score maps in the PC1-PC2 or PC2-PC3 spaces successfully grouped the wine samples in two distinct clusters, corresponding to Chieti and Teramo provinces, respectively. Then, an innovative model was tested, which simulates the use of two LEDs only for illumination, and a PIN detector instead of the spectrometer. Linear Discriminant Analysis was used for data processing, obtaining about 12% classification error.

2. THE WINE COLLECTION AND THE OPTICAL SPECTRA

Abruzzo is a region of central Italy. The Eastern part overlooks the Adriatic Sea, while the Western part includes the Gran Sasso mountains. The wine production of this region is an important economic resource. Most of the 4 million hectolitres annually produced are bearing PDO and PGI labels: this means that the production is distinctly oriented towards quality rather than quantity^{20, 21}. The export market is worth over 120 million euro, a figure that showed a growing trend and registered a record over the past five years. The most popular vine types of this region are Montepulciano d’Abruzzo, Trebbiano d’Abruzzo, Pecorino and Chardonnay, and other vines frequently cultivated are Sangiovese, Merlot, Malvasia and Cerasuolo.

This study considered a selection of wines from Teramo and Chieti provinces, respectively located in the North-East and South-East part of Abruzzo, as shown in Figure 1. Table I summarizes the characteristics of the collection, which represented a significant regional figure. It was made of 32 red, rosé and white wines of different vines, produced in 2007: 18 samples were from the Chieti province, while the other 14 samples were produced in the Teramo province.

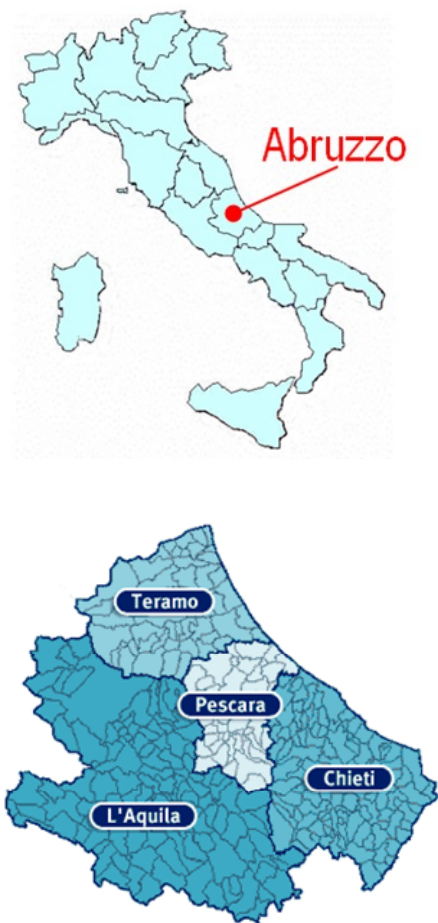


Figure 1. The location of the Abruzzo region in Italy (top), and of the Teramo and Chieti provinces (bottom).

Code	Province	Village	Brand	Variety	Wine type
1	CH	Orsogna	Orsogna	Sangiovese	white
2	CH	Orsogna	Orsogna	Sangiovese	white
3	CH	Orsogna	Orsogna	Montepulciano	red
4	CH	Orsogna	Orsogna	Montepulciano	red
5	CH	Orsogna	Orsogna	Malvasia	white
6	CH	Orsogna	Orsogna	Chardonnay	white
7	CH	Orsogna	Orsogna	Pecorino	white
8	CH	Orsogna	Orsogna	Sangiovese	red
9	CH	Orsogna	Orsogna	Trebbiano	white
10	CH	Orsogna	Orsogna	Montepulciano	red
11	CH	Ortona	Ortona	Chardonnay	white
12	CH	Ortona	Ortona	Pecorino	white
13	CH	Ortona	Ortona	Trebbiano	white
14	CH	Ortona	Ortona	Montepulciano	red
15	CH	Ortona	Ortona	Merlot	red
16	CH	Ortona	Ortona	Cerasuolo	rosé
17	TE	Silvi	Silvi	Montepulciano	red
18	TE	Silvi	Silvi	Chardonnay	white
19	TE	Silvi	Silvi	Trebbiano	white
20	TE	Silvi	Silvi	Pecorino	white
21	TE	Giulianova	GiovanPietro	Cerasuolo	rosé
22	TE	Giulianova	GiovanPietro	Montepulciano	red
23	TE	Giulianova	GiovanPietro	Trebbiano	white
24	TE	Giulianova	GiovanPietro	Pecorino	white
25	TE	Canosa	Nicola	Cerasuolo	rosé
26	TE	Castilenti	SanLorenzo	Cerasuolo	rosé
27	TE	Castilenti	SanLorenzo	Pecorino	white
28	TE	Castilenti	SanLorenzo	Montepulciano	red
29	TE	Castilenti	SanLorenzo	Trebbiano	white
30	TE	Castilenti	SanLorenzo	Chardonnay	white
31	CH	Canosa	Nicola	Chardonnay	white
32	CH	Torino di Sangro	Mucci	Cerasuolo	rosé

Table I. The wine collection.

A conventional dual-beam spectrophotometer was used for absorption spectroscopy measurements. Quartz cuvettes with 1 mm light path were used, with water in the reference channel. Figure 2-top shows the absorption spectra in the entire visible and near-infrared range. The visible spectra are clearly dominated by the color of wines, and therefore were disregarded. The spectroscopic differences in the near-infrared band are highlighted in Figure 2-bottom: the most significant differences among the various samples are in the 1300-1900 nm range. These differences are related to the first overtone of the OH stretch of water, and a combination of stretch and deformation of the OH group in water and ethanol^{22, 23}.

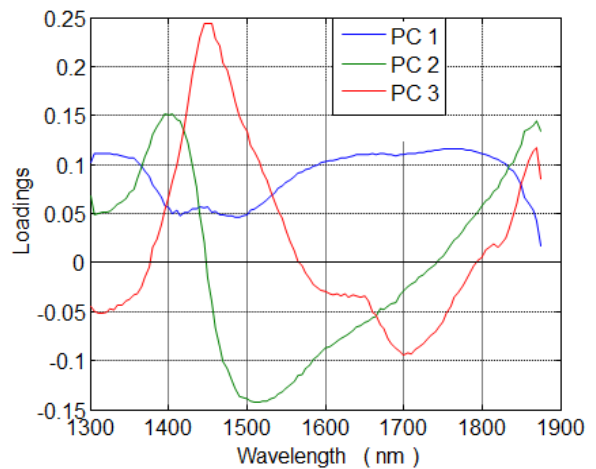
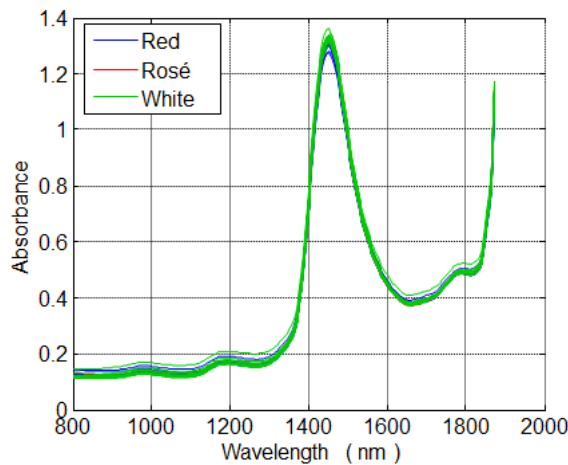
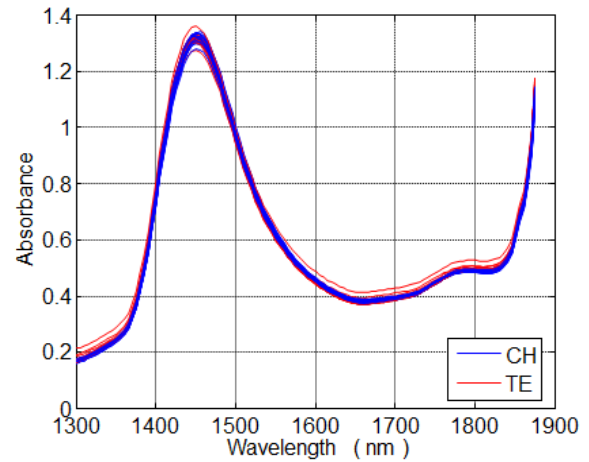
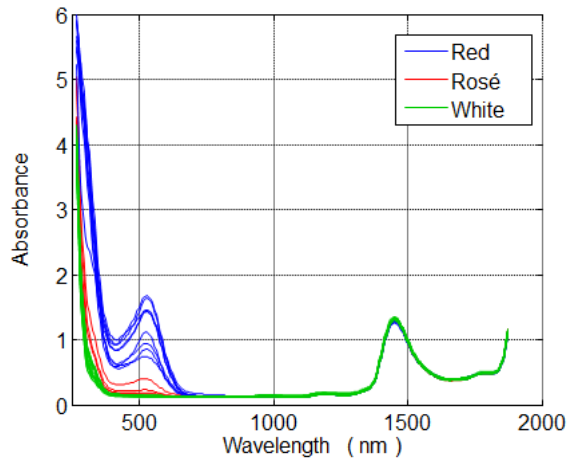


Figure 2. Wide band absorption spectra of all wines (top), and near-infrared band only (bottom).

Figure 3. Selected near-infrared band for data processing (top), and related PCA loadings (bottom).

3. GEOGRAPHIC CLASSIFICATION

The Principal Component Analysis (PCA), which is one of the most popular techniques for explorative analysis and data dimensionality reduction, was used for processing the spectroscopic data in the 1300-1900 nm range. PCA provides new variables and coordinates for identifying the wine samples in a 2D or 3D map. The coefficients giving the weight of each variable in the new space are called *loadings*. The new variables are called Principal Components (PCs), and have the following properties:

- PCs are mutually uncorrelated (orthogonality).
- 1st PC (PC1) has the largest variance among all possible linear combination of the starting variables.
- PCn has the largest variance among all linear combination of the starting variables that are orthogonal to PC1 ... PC(n - 1).

This means that high order PCs has little weight in distinguishing the samples, and can be disregarded with little loss of information. The loading plots are useful to interpret the score map: they show what variables are important for a given PC: variable with 0 loading has no importance, a variable with high (positive or negative) loading is important for differentiating the wines^{24, 25}.

Figure 3-bottom shows the loading of PCA processing of wine spectra in the 1300-1900 nm band:

- PC1 has a nearly constant behavior, and is related to the fluctuations of the baseline.
- PC2 expresses the difference between the absorbance at 1400 nm and 1500 nm. Note that these two wavelengths are positioned respectively on the ascending and descending slope of the peak at 1450 nm. Usually, this behavior is related to a shift in the central wavelength of the peak.
- PC3 has the maximum loading in the center of the peak, and is clearly related to the peak height.

Figure 4 shows the score maps in the PC1-PC2 and PC2-PC3 spaces which successfully grouped the wine samples in two distinct clusters, corresponding to the Chieti and Teramo provinces, respectively.

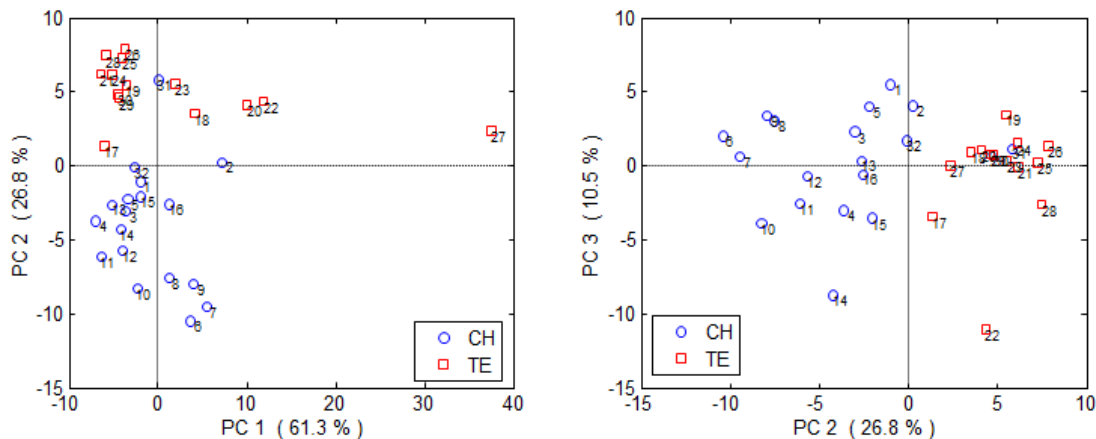


Figure 4. Results of PCA processing of spectroscopic data in the 1300-1900 nm band: clustering according to the Chieti and Teramo provinces.

4. PREDICTIVE MODEL FOR DUAL-BAND SPECTROSCOPY

Since the discriminating component PC2 is related to the differential absorbance between 1400 and 1500 nm, an attempt was made to classifying the wines from the two province simply using two LED for illumination, and a PIN detector instead of the spectrometer. This dual-band scheme could be interesting for implementing a low-cost device for consumers. Indeed, the growing interest of consumers to use cheap devices for self-assessment of food quality, was the motivation that inspired this work.

For this scope, the spectra of Figure 3-top were convoluted with Gaussian weighting functions, thus modelling the emission of commercially-available LEDs. Two bands were chosen, fitting the bands of the most evident spectroscopic differences. Then, by calculating the integral of the absorption spectra in these bands, each wine sample was identified by two coordinates. The characteristics of commercially available LEDs were considered in the simulation²⁶. Because LEDs emitting exactly at 1400 nm and 1500 nm were not readily available, we made a compromise by choosing LEDs centered at 1450 nm, and 1480 nm, respectively. Since the emission bands of these LEDs were wider than 100 nm, the use of notch filters was considered, so as to obtain FWHM=12 nm for both LEDs, as displayed in Figure 5-left.

The data set was processed using the Linear Discriminant Analysis (LDA), which is a robust and reliable technique for automatic object classification²⁷. Like PCA, LDA projects a high-dimensional pattern onto a subspace of smaller dimension, but the axes for projection are chosen using a different criterion. Actually, LDA is a tool that is specifically suited for identification, and looks for those variables that show a large spread among different clusters (inter-class variance), but limited variance within each cluster (intra-class variance). Given an N-class problem, the LDA extracts from the data matrix N-1 Discriminating Functions (DFs), which correspond to Principal Components in the PCA, but show a better resolution with regard to poorly-separated clusters. In our case, being a two-class problem, we extracted one DF only.

Figure 5-right shows the results of the LDA processing: the correct classification rate using all samples for calibration was 81.25%. The correctly classified wines were 15 from Teramo out of 18, and 11 from Chieti out of 14. The leave-one-out cross-validation procedure gave a success rate of 78.1%, quite close to the calibration value.

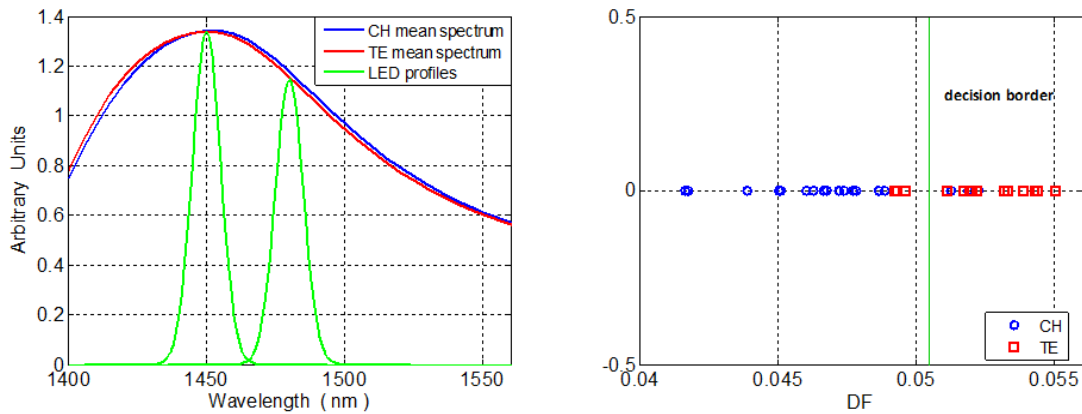


Figure 5. The model of dual-band spectroscopy – Wine absorption spectra in the limited 1400-1500 nm band, including the emission spectra of two LEDs (left). Results of LDA processing (right).

5. PERSPECTIVES

Absorption spectroscopy in the near-infrared band, combined to multivariate data processing, demonstrated effectiveness for wine authentication. Wines produced in the Abruzzo region of central Italy were successfully classified according to two provinces, Teramo and Chieti, respectively. Moreover, a model of dual-band spectroscopy was proposed, making use of two LEDs for illumination and a PIN detector instead of the spectrometer. Also this model showed a good classification according to provinces, demonstrating the potentials for implementing a low-cost device for consumer use.

ACKNOWLEDGEMENTS

The Consejería de Innovación, Ciencia y Empresa (Junta de Andalucía) is acknowledged for partial financial support (IAC07-I-1664).

REFERENCES

- ¹ S.A. Bellomario, X.A. Conlan, R.M. Parker, N.W. Barnett, M.J. Adams, “Geographical classification of some Australian wines by discriminant analysis using HPLC with UV and Chemiluminescence detection”, *Talanta* 80, 833-838 (2009).
- ² S.M. Rocha, E. Coelho, J. Zrosliková, I. Delgado, M.A. Coimbra, “Comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry of monoterpenoids as a powerful tool for grape origin traceability”, *Journal of Chromatography A*, 1161, 292-299 (2007).
- ³ L. Jaitz, K. Siegl, R. Eder, G. Rak, L. Abranco, G. Koellensperger, S. Hann, “LC-MS/MS analysis of phenols for classification of red wine according to geographic origin, grape variety and vintage”, *Food Chemistry* 122, 366-372 (2010).
- ⁴ J.M. Jurado, A. Alcàzar, A. Palacios-Morillo, F. De Pablos, “Classification of Spanish DO white wines according to their elemental profile by means of support vector machines”, *Food Chemistry*, 135, 898-903 (2012).
- ⁵ D. Cozzolino, W.U. Cynkar, P.A. Smith, “Can spectroscopy geographically classify Sauvignon Blanc wines from Australia and New Zealand?”, *Food Chemistry*, 126, 673-678 (2011).
- ⁶ D. Cozzolino, M. Haldstock, R.G. Damberg, W.U. Cynkar, P.A. Smith, “Mid infrared spectroscopy and multivariate analysis: a tool to discriminate between organic and non-organic wines grown in Australia”, *Food Chemistry*, 116, 761-765 (2009).
- ⁷ L. Liu, D. Cozzolino, W.U. Cynkar, R.G. Damberg, L. Janik, B.K. O’Neill, C.B. Colby, M. Gishen, “Preliminary study on the application of visible-near infrared spectroscopy and chemometrics to classify Riesling wines from different countries”, *Food Chemistry*, 106, 781-786 (2008).

-
- ⁸ C.J. Bevin, R.G. Damberg, A.J. Fergusson, D. Cozzolino, "Varietal discrimination of Australian wines by means of mid-infrared spectroscopy and multivariate analysis", *Analytica Chimica Acta*, 621, 19-23 (2008).
- ⁹ M. Urbano, M.D. Luque de Castro, P-M. Pérez, J. Garcia-Olmo, M.A. Gómez-Nieto, "Ultraviolet-visible spectroscopy and pattern recognition methods for differentiation and classification of wines", *Food Chemistry*, 97, 166-175 (2006).
- ¹⁰ L. Liu, D. Cozzolino, W.U. Cynkar, M. Gishen, C.B. Colby, "Geographic classification of Spanish and Australian Temprillo red wines by visible and near-infrared spectroscopy combined with multivariate analysis", *Journal of Agricultural and Food Chemistry*, 54, 6754-6759 (2006).
- ¹¹ C.J. Bevin, A.J. Fergusson, W.B. Perrey, L.J. Janik, D. Cozzolino, "Development of a rapid fingerprinting system for wine authenticity by mid-infrared spectroscopy", *Journal of Agricultural and Food Chemistry*, 54, 9713-9718 (2006).
- ¹² D. Cozzolino, H.E. Smyth, K.A. Lattey, W. Cynkar, L. Janik, R.G. Damberg, I.L. Francis, M. Gishen, "Relationship between sensory analysis and near infrared spectroscopy in Australian Riesling and Chardonnay wines", *Analytica Chimica Acta*, 539, 341-348 (2005).
- ¹³ D. Cozzolino, H.E. Smyth, M. Gishen, "Feasibility study on the use of visible and near-infrared spectroscopy together with chemometrics to discriminate between commercial white wines of different varietal origins", *Journal of Agricultural and Food Chemistry*, 51, 7703-7708 (2003).
- ¹⁴ W. Cynkar, D. Cozzolino, R.G. Damberg, "The effect of sample storage and homogenisation techniques on the chemical composition and near infrared spectra of white grapes", *Food Research International*, 42, 653-658 (2009).
- ¹⁵ H.E. Smyth, D. Cozzolino, W.U. Cynkar, R.G. Damberg, M. Sefton, M. Gishen, "Near infrared spectroscopy as a rapid tool to measure volatile aroma compounds in Riesling wine: possibilities and limits", *Analytical Bioanalytical Chemistry*, 390, 1911-1916 (2008).
- ¹⁶ D. Cozzolino, W.U. Cynkar, R.G. Damberg, M.D. Mercurio, P.A. Smith, "Measurement of condensed tannins and dry matter in red grape homogenates using near infrared spectroscopy and partial least squares", *Journal of Agricultural and Food Chemistry*, 56, 7631-7636 (2008).
- ¹⁷ D. Cozzolino, M.J. Kwiatowski, R.G. Damberg, W.U. Cynkar, L.J. Janik, G. Skouroumounis, M. Gishen, "Analysis of elements in wine using near infrared spectroscopy and partial least square regression", *Talanta*, 74, 711-716 (2008).
- ¹⁸ D. Cozzolino, R.G. Damberg, L. Janik, W.U. Cynkar, M. Gishen, "Analysis of grapes and wine by near infrared spectroscopy – a review", *Journal of Near Infrared Spectroscopy*, 14, 279-289 (2006).
- ¹⁹ J. Fernández-Navales, M.I. López, M.T. Sánchez, J.A. García, J. Morales, "A feasibility study on the use of a miniature fiber optic NIR spectrometer for the prediction of volumic mass and reducing sugars in white wine fermentations", *Journal of Food Engineering*, 89, 325-329 (2008).
- ²⁰ Web: <http://www.consorzio-viniabruzzo.it/>
- ²¹ Web: <http://www.italy-wine-tours.net/en/winetourabruzzo/winetourabruzzo/winetourabruzzo.php>
- ²² D. Cozzolino, M. Parker, R.G. Damberg, M. Herderich, M. Gishen, "Chemometrics and visible-near infrared spectroscopic monitoring of red wine fermentation in a pilot scale", *Biotechnology and Bioengineering*, 95, 1101-1107 (2006).
- ²³ D. Cozzolino, M.J. Kwiatkowski, M. Parker, W.U. Cynkar, R.G. Damberg, M. Gishen, M.J. Herderich, "Prediction of phenolic compounds in red wine fermentations by near infrared spectroscopy", *Analytica Chimica Acta*, 513, 73-80 (2004).
- ²⁴ M.J. Adams, [Chemometrics in Analytical Spectroscopy], The Royal Society of Chemistry, Cambridge, UK (1995).
- ²⁵ J.E. Jackson, [A User's Guide to Principal Components], J. Wiley & Sons Inc., Hoboken, NJ (2003).
- ²⁶ Web: http://www.roithner-laser.com/led_diverse.html
- ²⁷ B.G.M. Vandeginste, D.L. Massart, L.C.M. Buydens, S. De Jong, D.J. Lewi, J. Smeyers-Verbeke, [Handbook of Chemometrics and Qualimetrics], Elsevier Science BV, Amsterdam – Chapter 33 (1998).