

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE ESCUELA DE INGENIERIA

INSTRUMENTO PORTÁTIL PARA MEDICIÓN DE COMPUESTOS FENÓLICOS EN VINO TINTO USANDO ESPECTROSCOPÍA RAMAN

ÁLVARO LEONARDO GALLEGO GUTIÉRREZ

Tesis para optar al grado de Magíster en Ciencias de Ingeniería

Profesor Supervisor:

ANDRÉS GUESALAGA M.

Santiago de Chile, Septiembre, 2009 © 2009, Álvaro Gallego



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE ESCUELA DE INGENIERIA

INSTRUMENTO PORTÁTIL PARA MEDICIÓN DE COMPUESTOS FENÓLICOS EN VINO TINTO USANDO ESPECTROSCOPÍA RAMAN

ÁLVARO LEONARDO GALLEGO GUTIÉRREZ

Tesis presentada a la Comisión integrada por los profesores:

ANDRÉS GUESALAGA M.

ALDO CIPRIANO Z.

EDMUNDO BORDEU S.

EDUARDO AGOSIN T.

Para completar las exigencias del grado de Magíster en Ciencias de Ingeniería

Santiago de Chile, Septiembre, 2009

(A mi familia y amigos, que siempre me han apoyado en todo)

AGRADECIMIENTOS

Agradezco a la gente del laboratorio de Enología de la Universidad, por su ayuda entregada para el muestreo de los vinos, en especial a Alejandra Zúñiga, Álvaro González y el profesor Edmundo Bordeu. Agradezco también a mi familia, amigos y compañeros que me incentivaron a trabajar, y terminar este trabajo. Finalmente agradezco a mi profesor supervisor Andrés Guesalaga por su apoyo y confianza para desarrollar esta investigación.

INDICE GENERAL

			Pág.
DED	ICA7	ΓORIA	ii
AGR	RADE	CIMIENTOS	iii
IND	ICE I	DE TABLAS	vi
IND	ICE I	DE FIGURAS	vii
RES	UME	N	x
ABS	TRA	CT	xi
1.	INT	RODUCCIÓN	1
	1.1.	Motivación	1
	1.2.	Objetivos	2
	1.3.	Experiencias anteriores	3
2.	MA	RCO TEÓRICO	5
	2.1	Compuestos fenólicos en el vino tinto	5
	2.2	Espectroscopía Raman	7
		2.2.1 Historia	7
		2.2.2 Teoría	8
		2.2.3 Espectroscopía Raman	11
	2.3	Regresión multivariable, Mínimos cuadrados parciales	15
		2.3.1 Introducción a la calibración multivariable	15
		2.3.2 PLS, Mínimos cuadrados parciales	16
3.	MA	TERIALES Y MÉTODOS	21
	3.1	Espectrómetro Raman	21
	3.2	Muestras	29
	3.3	Métodos analíticos tradicionales	30
	3.4	Adquisición de espectros	31
	3.5	Preprocesamiento de espectros	34
	3.6	Calibración	38

RESULTADOS	44
4.1 Calibración de espectros	44
4.2 Polifenoles totales	52
4.3 Antocianas	54
4.4 Taninos	57
CONCLUSIONES	60
BIBLIOGRAFIA	62
	4.1 Calibración de espectros 4.2 Polifenoles totales 4.3 Antocianas 4.4 Taninos CONCLUSIONES

INDICE DE TABLAS

	Pág.
Tabla 3-1: Especificaciones técnicas del espectrómetro BTR111 MiniRam	28
Tabla 4-1: Resultados calibración PLS utilizando diferentes métodos	44
Tabla 4-2: Resultados calibración PLS utilizando diferentes segmentaciones del	
espectro	51
Tabla 4-3: Resultados modelo de calibración PLS para polifenoles totales	53
Tabla 4-4: Resultados modelo de calibración PLS para antocianas	55
Tabla 4-5: Resultados modelo de calibración PLS para taninos	58

INDICE DE FIGURAS

Pá	ág.
Figura 1-1: Composición compuestos fenólicos totales en vino tinto	1
Figura 2-1: Estructura básica de los flavonoides	6
Figura 2-2: Diagrama del proceso de dispersión Rayleigh y Raman	. 10
Figura 3-1: Diagrama esquemático de un espectrómetro Raman	. 22
Figura 3-2: Configuración Czerny-Turner con grilla de difracción	. 24
Figura 3-3: Comparación del espectro Raman de una muestra de Tylenol con distintas	S
longitudes de onda de excitación. a) 473 nm. b) 785 nm	
Figura 3-4: Comparación de una zona de un espectro Raman con distinta resolución óptica	
ориса	. 20
Figura 3-5: Espectrómetro Raman B&W TEK BTR111 MiniRam	. 28
Figura 3-6: Software utilizado para la captura de espectros	. 32
Figura 3-7: Portacubetas B&W TEK BCH-100	. 33
Figura 3-8: Método de eliminación de fluorescencia con círculo	. 37
Figura 3-9: Interfaz gráfica de análisis del PLS_Toolbox 5.0.3 para MATLAB	. 39
Figura 3-10: Gráfico RMSECV vs LV para selección del número de variables latentes	3
para el modelo PLS	. 40
Figura 3-11: Gráfico de dispersión de valores predichos vs medidos	. 41
Figura 3-12: Gráfico de residuos de predicción	42

Figura 3-13: Gráfico del vector de regresión b	42
Figura 3-14: Gráfico de las VIP (Variable Importance in Projection)	43
Figura 4-1: Calibración modelo PLS de espectros en bruto, usando 9 LV	45
Figura 4-2: Calibración modelo PLS de espectros con MSC, usando 6 LV.	45
Figura 4-3: Calibración modelo PLS de espectros con eliminación de faplicando método del círculo con R=10, usando 7 LV	
Figura 4-4: Calibración modelo PLS de espectros con eliminación de faplicando método del círculo con R=20, usando 8 LV	
Figura 4-5: Calibración modelo PLS de espectros con eliminación de faplicando método del círculo con R=50, usando 6 LV	
Figura 4-6: Calibración modelo PLS de espectros con eliminación de faplicando método del círculo con R=35, usando 8 LV	
Figura 4-7: Calibración modelo PLS de espectros con eliminación de faplicando método del círculo con R=35 y Raman shift, usando	
Figura 4-8: Selección del radio óptimo para método del círculo	49
Figura 4-9: Comparación del método de eliminación de fluorescencia con una zona del espectro, para distintos radios a) R=10, b) R=20, R=50	c) R=35, d)
Figura 4-10: Calibración modelo PLS con solo la fluorescencia extre espectros, usando 8 LV	
Figura 4-11: Gráfico de residuos de predicción en modelo para polifenoles	totales 52

Figura 4-12: Gráfico de dispersión de valores predichos vs medidos para polifenoles
totales
Figura 4-13: Gráfico de las VIP para: a) espectro completo, b) zona 100-900 cm ⁻¹ 54
Figura 4-14: Gráfico de residuos de predicción en modelo para antocianas
Figura 4-15: Gráfico de dispersión de valores predichos vs medidos para antocianas 56
Figura 4-16: Gráfico de las VIP para: a) espectro completo, b) zona 100-900 cm ⁻¹ 56
Figura 4-17: Gráfico de residuos de predicción en modelo para taninos
Figura 4-18: Gráfico de dispersión de valores predichos vs medidos para taninos 58
Figura 4-19: Gráfico de las VIP para: a) espectro completo, b) zona 100-900 cm ⁻¹ 59

RESUMEN

A lo largo de este trabajo, se presenta el desarrollo de un nuevo método para medir los principales compuestos fenólicos en vinos tintos, específicamente los de la cepa Cabernet Sauvignon. Para esto se usó un espectrómetro Raman B&W TEK BTR-111 MiniRam, el cual utiliza el *scattering* Raman generado por las moléculas del vino para medir cualitativa y cuantitativamente la composición de cualquier tipo de material.

Se realizaron mediciones para muestras de vino tinto de diferentes años, viñas y cosechas, a las cuales se les midió su espectro Raman y además se les aplicó la técnica analítica tradicional de laboratorio, para obtener los valores de referencia necesarios en la construcción de los modelos de calibración.

A estos datos se les aplicó diferentes métodos de pre-procesamiento a través del software MATLAB, y luego con el PLS_Toolbox 5.0.3, se realizó la correlación entre los datos. De esta forma se obtuvieron los modelos para polifenoles totales, antocianas y taninos, los cuales fueron evaluados a través de los índices R^2 de validación y RMSEV.

Se obtuvieron excelentes correlaciones expresadas en valores de \mathbb{R}^2 de validación y RMSEV, lo cual permite considerar su uso en terreno y extender la técnica a otras cepas, así como también al análisis de otros compuestos en el vino.

Palabras Claves: espectroscopía, espectrómetro, Raman, compuestos fenólicos, vino, Cabernet Sauvignon, polifenoles, antocianas, taninos, fluorescencia, PLS.

ABSTRACT

This paper presents the development of a new method to measure the main phenolic compounds in red wine specifically those found in of Cabernet Sauvignon.

A B&W TEK MiniRam BTR-111 Raman spectrometer was used for this purpose, which is an instrument that uses the Raman scattering from the sample to analyze it qualitatively and quantify its molecular composition.

Measurements were made for samples of red wine from different seasons, vineyards and harvests, which were analyzed by the Raman instrument and also with standard analytical techniques in order to obtain the reference values needed for the model calibration.

Several pre-processing methods were tested for the data using MATLABTM, and the correlation between the sets of data was performed using the PLS_Toolbox 5.0.3. Models were adjusted for polyphenols, anthocyanins and tannins by optimizing the R^2 and RMSEV indices.

Excellent correlations were obtained, expressed in values for R^2 of validation and RMSEV, showing the potential use of this technique in field measurements and its extension to other varieties of wines and additional compounds in wine as well.

Keywords: spectroscopy, spectrometer, Raman, phenolics compounds, wine, Cabernet Sauvignon, polyphenols, anthocyanins, tannins, fluorescence, PLS.

1. INTRODUCCIÓN

1.1. Motivación

Los compuestos fenólicos en la uva tinta y en el vino, son componentes determinantes para la calidad de éste. Se encuentran naturalmente en la uva y también son desarrollados durante el proceso de vinificación. La cantidad de estos componentes dentro del vino depende tanto de la variedad vinífera, como de la cosecha y de las técnicas aplicadas al viñedo (Lampreave et al., 2000).

Existen dos grupos generales de estos compuestos: los no-flavonoides y los flavonoides, siendo este último, el grupo fenólico más importante en el vino. Dentro de este grupo se pueden encontrar diferentes familias de fenoles: flavonoles, antocianas, taninos, entre otros, los cuales se distribuyen en el vino como se ve en la figura 1-1.

Los Flavonoles se encuentran en el hollejo de la uva y se caracterizan por darle un efecto antioxidante al vino. Las antocianas también se encuentran en el hollejo de la uva y son las responsables de darle el color rojo a éstas y al vino. Mientras que los taninos se ubican en la semilla y en el hollejo, y son los que le dan la amargura y la astringencia al vino (Argyri et al., 2005).

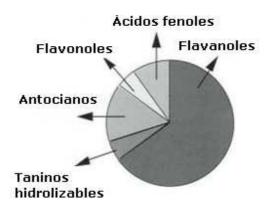


Figura 1-1: Composición compuestos fenólicos totales en vino tinto

En consecuencia, para analizar la calidad del vino, es necesario medir la presencia de estos compuestos, los cuales generalmente se miden con métodos de laboratorio utilizando procedimientos lentos y complejos en su ejecución. (Bordeu y Scarpa, 1998)

Debido a esto se han buscado otras alternativas más rápidas y eficientes para realizar estas mediciones, por ejemplo la espectroscopía, la cual en pocos segundos proporciona información sobre la composición molecular de la muestra analizada. En particular, la espectroscopía Raman se basa en el análisis de la luz que es dispersada por la muestra examinada al hacer incidir sobre ella un haz de luz monocromático. Sólo una pequeña parte de esta luz que se dispersa, presenta cambios de frecuencia respecto a la luz incidente, los cuales son producidos por las vibraciones y rotaciones internas que se dan dentro de las moléculas y son características de la composición molecular de cada material.

1.2. Objetivos

El objetivo general de esta tesis consiste en desarrollar un método para estimar en vinos tintos, específicamente la cepa Cabernet Sauvignon, los compuestos fenólicos más importantes que determinan la calidad de éste: antocianas, taninos y polifenoles totales. Para ello se propone el uso de un espectrómetro portable de bajo costo NIR-Raman como el B&W Tek BTR-111 MiniRam.

Los objetivos específicos son los siguientes:

- 1. Implementar un protocolo de preparación y medición de las muestras.
- 2. Identificar y aislar la información útil (Raman) dentro de los espectros obtenidos con el instrumento.
- 3. Determinar un modelo matemático para la estimación de los parámetros, utilizando la información obtenida de los espectros.
- 4. Comparar los resultados obtenidos con las mediciones entregadas por la técnica analítica tradicional.

1.3. Experiencias anteriores

Se encontraron trabajos similares a éste, pero utilizando otras técnicas de espectroscopía y aplicaciones relacionadas. Una de las técnicas más comunes en el área de la industria de alimentos es la espectroscopía IR (InfraRed spectroscopy), la cual debido a su versatilidad, ha sido utilizada en un sinnúmero de aplicaciones (Li-Chan, 1996). En particular, en el caso del vino esta técnica ha sido empleada para clasificar y cuantificar diferentes muestras de acuerdo a su origen (Picque et al., 2001), para la medición de la interacción entre proteínas y taninos (Edelmann y Lendl, 2002), y para la medición de otros compuestos en la etapa de la fermentación (Urtubia et al., 2004). También se ha intentado utilizar espectroscopía IR para la medición de la copigmentación en vinos, obteniéndose resultados con errores de predicción importantes debido a la interferencia de otros compuestos en la señal espectrofotométrica (Versari et al., 2004). Para el caso de la medición de color, específicamente del contenido de antocianas, se han desarrollado experiencias utilizando espectroscopía NIR (Near InfraRed spectroscopy) con resultados satisfactorios (Ayala et al., 1997; Rivas-Gonzalo et al., 1992). La espectroscopía NIR se ha utilizado además para medir madurez en uvas a través de su pH, °Brix y antocianas (Larraín et al., 2008).

La espectroscopía Raman se ha utilizado asimismo en el área de alimentos. Particularmente se han hecho estudios de la medición del contenido alcohólico de bebidas destiladas (Nordon et al., 2005; Nose y Hojo, 2006), así como también mediciones de acidez total, aminoácidos y contenido fenólico total (Nose et al., 2005). Para el caso del vino, sólo se han realizado mediciones preliminares de concentración de agua, etanol y sacarosa, exclusivamente en vino blanco, como una aplicación de una fibra óptica con núcleo líquido (Meneghini et al., 2008). Finalmente, se encontró una aplicación para el análisis, medición y determinación de aminoácidos y antocianas en vinos, utilizando espectroscopía NMR (*Nuclear magnetic resonance spectroscopy*) (Kosir y Kydric, 2002).

Cabe mencionar, que estas experiencias, aunque son similares, fueron realizadas utilizando espectrómetros científicos con diseños ópticos de laboratorio, complejos y de alto costo, por lo que no son aplicables a los objetivos de esta tesis, la cual es realizar la aplicación con un equipo portable y de bajo costo.

2. MARCO TEÓRICO

2.1 Compuestos fenólicos en el vino tinto

Los compuestos fenólicos tienen una gran importancia en enología debido al papel que juegan directa o indirectamente sobre la calidad de los vinos. Ésta se ve influenciada fuertemente por su estructura fenólica, compuesta entre otros, por taninos y antocianas, que son responsables de características sensoriales tan importantes como el color, cuerpo, astringencia y amargor, además de otros intereses nutricionales y farmacológicos.

Estos compuestos tienen su origen principalmente en la baya de la uva, a los cuales se agregan otros fenoles que aparecen durante el proceso de vinificación. Estos constituyentes específicos del vino son esencialmente los elementos procedentes de las etapas fermentativas, pero también aparecen numerosos productos resultantes de la evolución de los polifenoles nativos de la baya. Éstos, durante la vinificación y el envejecimiento, participan en diversos tipos de reacciones, dando lugar a una multiplicidad de estructuras nuevas. Además, en los vinos que han permanecido en barrica, se puede detectar la presencia de constituyentes extraídos de la madera utilizada durante la crianza. Así, la composición fenólica del vino depende tanto de la materia prima como de los métodos de vinificación utilizados, los cuales influyen sobre fenómenos físicos (como la difusión de compuestos desde las partes sólidas hacia el mosto) y sobre los fenómenos químicos en los cuales están implicados los polifenoles (oxidación, degradación y condensación).

Desde el punto de vista químico, los compuestos fenólicos son caracterizados por un núcleo benzénico que lleva uno o varios grupos hidroxilos. Estos se pueden clasificar en dos grandes grupos: los compuestos no-flavonoides y los flavonoides, siendo este último, el grupo fenólico más importante presente en el vino (Peña, 2003).

El grupo de los compuestos no flavonoides comprende a los ácidos fenoles (divididos en ácidos benzoicos (C6-C1) y ácidos cinnámicos (C6-C3)), y a otros derivados fenólicos como los estilbenos.

El grupo de los flavonoides, se caracteriza por presentar dos anillos de 6 carbonos unidos por un heterociclo central de 3 carbonos (C6-C3-C6). Su estructura molecular básica se puede ver en la figura 2-1.

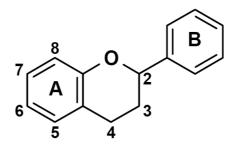


Figura 2-1: Estructura básica de los flavonoides

En esta familia se distinguen varias subclases según el grado de oxidación de su núcleo pirano (C_5H_6O). En primer lugar, se tienen los Flavonoles (como la quercetina, miricetina y el kaemferol y sus glicósidos) los cuales, se encuentran presentes en los hollejos, importantes debido a que participan en el color amarillo de los vinos blancos y por sus efectos antioxidantes benéficos para la salud.

Luego, están las antocianas que, contribuyen de manera preponderante al color de las uvas tintas y se encuentran principalmente, en el hollejo de éstas. El contenido y composición de las antocianas en la uva varían en función de la especie y de la variedad (Mazza y Miniati, 1993).

Finalmente, están los flavanoles o taninos condensados (taninos de la uva), que tienen como compuesto base los monómeros catequina y su isómero, la epicatequina. Éstos, se pueden encontrar tanto en las semillas como en el hollejo de la uva. Los taninos presentan una relación inversa en cuanto al amargor y astringencia a medida que aumentan de tamaño (aumenta el número de unidades de catequina o epicatequina en su estructura), disminuyendo el amargor y aumentando su astringencia, en los taninos más

grandes. Algunos alcanzan un tamaño en que no son solubles, por lo que no pueden reaccionar con las proteínas de la saliva, lo que hace que éstas precipiten, y por lo tanto, dejan de producir la sensación de astringencia.

Junto a los taninos condensados (provenientes de la uva), los taninos hidrolizables (extraídos de la madera durante el envejecimiento en barrica) también son responsables de la astringencia en los vinos tintos.

A medida que transcurre el tiempo, es posible apreciar la aparición de nuevos compuestos resultantes de la fermentación alcohólica (piruvato, vinil guayacol, vinil fenol), así como algunos pigmentos originados por la unión de las antocianas y taninos. Estos nuevos pigmentos, más estables que las originales antocianas, pueden presentar una coloración púrpura o anaranjada (más frecuente) que poco a poco contribuye a la modificación del color del vino en el tiempo.

En cuanto a los taninos en el vino, también presentan cambios en el tiempo, ya sea por precipitación, por ruptura, o por formación de otros de tamaño medio (polimerización), los cuales modifican las propiedades sensoriales de astringencia y amargor (Peña, 2003).

2.2 Espectroscopía Raman

2.2.1 Historia

Una de las técnicas analíticas que ha despertado mayor interés en los últimos años es la espectroscopía Raman. Esta técnica surgió a finales de los años 20 y está basada en el efecto Raman, fenómeno descubierto por el físico hindú Chandrasekhara Venkata Raman. Éste es un fenómeno físico de dispersión inelástica de la luz, producido por las vibraciones y rotaciones internas que se dan dentro de una molécula. Fue postulado por primera vez por Smekel en 1923 y observado en 1928 por Raman y Krishnan. En el experimento original, se enfocó el sol por un telescopio a unas muestras de agua y alcohol purificadas. A través de un lente se recolectó la radiación dispersada por la

muestra, la cual pasó a través de un sistema de filtros para demostrar la existencia de dispersión con una frecuencia distinta de la luz original. Por este descubrimiento, en 1930 C. V. Raman fue galardonado con el premio Nobel de Física.

2.2.2 Teoría

Cuando la luz interactúa con la materia, pueden ocurrir 3 fenómenos con los fotones de la luz incidente: Pueden ser absorbidos, dispersados o pueden no interactuar y pasar a través de ella. Si la energía del fotón que incide corresponde a la diferencia de energía entre el estado base y el estado excitado de una molécula, el fotón puede ser absorbido y la molécula es llevada a un estado de excitación de mayor energía. Éste es el caso que es aprovechado por la espectroscopía de absorción, en que se mide la pérdida de energía de la radiación de la luz incidente.

También es posible que el fotón interactúe con la materia y se disperse. En este caso, no es necesario que la energía del fotón sea igual a la diferencia de energía entre los dos estados energéticos de la molécula. Un ejemplo de este efecto es el color azul del cielo, ya que la luz azul (con mayor energía) es dispersada por las partículas de la atmósfera con mayor eficiencia que la luz roja (con menor energía). A diferencia de la espectroscopía NIR, la Raman utiliza una fuente de radiación de una sola frecuencia para iluminar la muestra, y es la radiación dispersada de la molécula (de distinta energía que la incidente) la que es detectada y medida. Es por esto, que la dispersión Raman no necesita que la radiación incidente coincida con la diferencia de energía entre el estado base y excitado. Desde el punto de vista cuántico, las variaciones de frecuencia observadas en el fenómeno de dispersión Raman son equivalentes a las variaciones de energía.

En la dispersión Raman, la luz interactúa con la molécula y distorsiona la nube de electrones que rodean al núcleo para llevarlo momentáneamente a un estado energético inestable llamado "estado virtual", el cual abandona inmediatamente para pasar a uno de

los niveles de energía estables, emitiendo así un fotón de energía igual a la diferencia de estos estados energéticos.

Este tipo de dispersión es debido a colisiones inelásticas. En este caso el resultado de la interacción fotón–molécula es una transferencia de energía entre ambos. Si la interacción con la molécula se produce solo en los electrones alrededor del núcleo sin afectar directamente a éste, se produce un choque elástico, en el cual ni el fotón ni la molécula sufre variaciones en su estado energético. Por lo tanto, la molécula vuelve al mismo nivel energético que tenía antes del choque y el fotón dispersado conserva la misma frecuencia que el incidente. Este proceso es conocido como dispersión Rayleigh y es el fenómeno de dispersión más intenso, ya que la mayoría de los fotones se dispersan de esta manera.

Si durante el proceso de dispersión ocurre interacción con el núcleo de la molécula, la energía será transferida ya sea, del fotón incidente a la molécula o de la molécula al fotón dispersado. En este caso, el proceso es inelástico y la energía emitida es distinta del fotón incidente. Éste es un proceso normalmente débil, en el cual uno de cada 10^6 - 10^8 fotones que se dispersan, es dispersado de esta forma.

En la dispersión Raman pueden darse dos casos: Raman Stokes y Raman Anti-Stokes. En el primer caso, la molécula después de salir del estado virtual, vuelve a un estado energético mayor que el inicial, es decir, se produce una transferencia de energía desde el fotón a la molécula, librando así un fotón dispersado de menor energía y frecuencia.

Para el segundo caso, la molécula vuelve del estado virtual a un estado energético menor que el inicial, produciéndose una transferencia de energía de la molécula al fotón. Por lo tanto, el fotón dispersado corresponde a uno de mayor energía y frecuencia que el fotón incidente. El fenómeno de la dispersión Rayleigh, Raman Stokes y Anti-Stokes puede esquematizarse como muestra la figura 2-2.

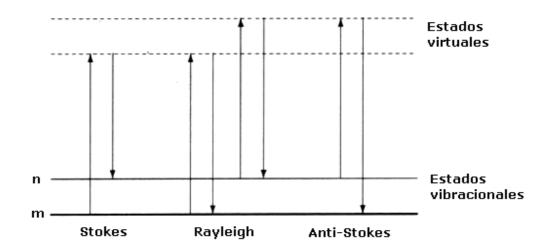


Figura 2-2: Diagrama del proceso de dispersión Rayleigh y Raman

Las frecuencias de las dispersiones Raman Stokes y Anti-Stokes van a depender, de la diferencia entre la frecuencia de la luz incidente y las frecuencias vibracionales moleculares permitidas. Para una molécula, sus diferentes enlaces poseen frecuencias de vibración, que dependen de la masa de los átomos involucrados y de la fuerza de unión entre ellos. Por lo tanto, cada material tendrá un conjunto de frecuencias diferentes que son características de su composición molecular.

La intensidad de la luz dispersada, tanto Raman como Rayleigh, va a depender del número de moléculas que interactúan con los fotones incidentes, siendo ésta mayor, cuando se iluminan mayor cantidad de moléculas (Pelletier, 1999).

A temperatura ambiente, la dispersión Stokes tiene una intensidad mayor que la dispersión Anti-Stokes, debido a que existen más moléculas con niveles energéticos vibracionales bajos, que moléculas con niveles altos y, por lo tanto, es más probable que se produzcan dispersiones Raman Stokes que dispersiones Raman Anti-Stokes.

2.2.3 Espectroscopía Raman

La espectroscopía Raman es una técnica de análisis molecular, no destructiva y altamente específica basada en el efecto Raman. Puede ser usada para mediciones cualitativas y cuantitativas, tanto para materiales orgánicos como inorgánicos.

Esta técnica analiza la luz dispersada por la muestra que se está analizando al hacer incidir sobre ella un haz de luz monocromático. Las dispersiones son características de cada material, ya que se deben a las vibraciones internas que ocurren en las moléculas. El conjunto de dispersiones producido por la muestra formará el espectro Raman propio de ésta.

Cada espectro Raman está formado por una banda que posee la misma frecuencia de la radiación incidente (correspondientes a la dispersión Rayleigh), y otras bandas de menor intensidad desplazadas respecto a la anterior. Estas bandas corresponden al efecto Raman y forman dos conjuntos diferentes: las bandas Stokes, de frecuencias inferiores a la incidente, y las bandas Anti-Stokes, de frecuencias superiores a la incidente. Estas bandas Stokes y Anti-Stokes aparecen en pares y de forma simétrica con respecto a la frecuencia central. Sin embargo, las bandas Stokes son considerablemente más intensas que las Anti-Stokes, por lo que generalmente son omitidas ya que proporcionan la misma información que las bandas Stokes.

Generalmente, la representación de estas bandas de dispersión se hace en función del número de onda normalizado $\bar{\nu}$, en vez de la longitud de onda o la frecuencia. El número de onda se define como la cantidad de longitudes de ondas que caben en un centímetro y se expresa como

$$\bar{\nu} = \frac{1}{\lambda} [cm^{-1}] \tag{2.1}$$

El número de onda $\bar{\nu}$, la longitud de onda λ , y la frecuencia ν están relacionados por:

$$\bar{\nu} = \frac{1}{\lambda} = \frac{\nu}{c} \tag{2.2}$$

Donde c corresponde a la velocidad de la luz.

Si se utiliza una representación de los espectros en función del número de onda desplazado según el número de onda del láser: $\bar{\nu} - \bar{\nu}_0$, se obtendrán los espectros con las bandas de dispersión independientes de la frecuencia de la fuente de excitación utilizada.

Generalmente, los espectros Raman que se obtienen, se componen de una señal útil y otra de ruido que perturba la información que existe en las bandas Raman. Dentro de los componentes que perturban la señal se puede encontrar el ruido aleatorio, el cual aparece debido a la naturaleza aleatoria de la luz y se presenta en forma de pequeñas fluctuaciones alrededor del valor medio de la intensidad de la señal.

Otro fenómeno que aparece como una señal de ruido es la fluorescencia. Es una emisión independiente de la dispersión Raman que se presenta como una línea de base que va cambiando lentamente y que puede llegar a cubrir completamente las bandas Raman menos intensas.

El espectro Raman, a través de sus bandas características de los componentes de la muestra iluminada por el láser, se puede utilizar como una huella digital para realizar análisis cuantitativos o cualitativos (Smith y Dent, 2005).

El análisis cuantitativo está basado en la medida de la intensidad de las bandas Raman, que forman el espectro. El análisis cualitativo en cambio, se basa en la comparación entre un espectro obtenido experimentalmente y otros espectros que forman parte de una base de datos de referencia. Gracias a este tipo de análisis es posible identificar materiales o diferenciarlos de otros.

El conjunto de bandas Raman es característico y particular de cada material y permite la identificación unívoca del mismo. Es por ello, que es de gran importancia obtener valores exactos de las posiciones de estas bandas, ya que serán éstas, las que se utilizarán para determinar finalmente la identificación del material. El problema es que

la obtención de esta información Raman no es siempre fácil ya que el espectro medido contiene, además de la señal Raman, ruido inherente tanto al proceso de medida como a la naturaleza de la muestra analizada, que puede llegar a enmascarar la señal Raman.

La calidad de la información que se puede extraer del espectro Raman no depende únicamente de la señal Raman, sino también de las distintas fuentes de ruido que intervienen en su adquisición. Solo una pequeña fracción de los fotones es dispersada inelásticamente, así que, por lo general, las líneas Raman son muy débiles y pueden ser fácilmente cubiertas por ruido. En la adquisición de la señal Raman intervienen distintas fuentes de ruido, las más importantes se nombran a continuación:

- Ruido generado debido a la naturaleza cuántica de la luz. Se trata de un ruido aleatorio, distinto para cada espectro capturada.
- Ruido generado por la instrumentación que depende del diseño del instrumento utilizado para el análisis. Este ruido incluye el ruido introducido por el CCD como el ruido térmico, el ruido de lectura o la respuesta en frecuencia del detector.
- Ruido generado por fuentes externas, producido fuera del instrumento y la muestra que se está analizando. Generalmente está causado por alguna fuente de luz externa que contamina la señal en algún punto del instrumento. Por ejemplo, el ruido que produce la luz de los tubos fluorescentes o la luz solar. Otro caso de ruido externo es el generado por partículas de alta energía, como lo son los rayos cósmicos que llegan al detector del instrumento Raman. Los rayos cósmicos liberan un gran número de electrones que son indistinguibles de los fotoelectrones. Los electrones generados por los rayos cósmicos se concentran en uno, o máximo dos elementos del detector. El resultado, es un peak muy estrecho y de gran intensidad en el espectro Raman. Estos peaks ocurren

- infrecuentemente, a tiempos aleatorios y en posiciones también aleatorias del espectro.
- Ruido generado por la muestra que incluye emisiones ópticas no deseadas como es el caso de la fluorescencia. Este fenómeno se debe a la absorción de fotones de una determinada longitud de onda seguida por la emisión de fotones de una longitud de onda mayor. El fotón incidente, es absorbido por la molécula y hace que ésta pase a un estado electrónico excitado. Después, la molécula salta a otro estado excitado de menor energía emitiendo un fotón de mayor longitud de onda que el que absorbió. Esta emisión se produce en un intervalo de tiempo muy corto después de la absorción, millonésimas de segundo, por lo que este fenómeno sólo tiene lugar mientras dura la iluminación. La fluorescencia, es una de las principales fuentes de ruido en espectroscopía Raman sobretodo, en muestras no transparentes. La fluorescencia tiene un espectro mucho más ancho que las bandas Raman y suele presentarse como una suave curvatura de la línea de base del espectro y puede alcanzar una intensidad que llegue a cubrir por completo las bandas Raman.

Como se puede ver, existen muchas fuentes de ruido que pueden llegar a cubrir a la señal Raman. Para poder lograr la identificación correcta de los componentes de la muestra, es imprescindible tener una buena razón señal a ruido en el espectro. Por una parte, se utilizan técnicas ópticas para mejorar esta razón, como por ejemplo, una excitación de elevada energía y altamente monocromática (para obtener el mayor nivel de señal posible), y por otra parte, es necesaria la reducción de ruido con procesamientos de las señales de los espectros obtenidos a través de software, utilizando diversos métodos ya existentes.

La espectroscopía Raman al ser un método no invasivo que ofrece información muy detallada basada en la vibración de las moléculas de la muestra analizada, permite analizar compuestos de diversos tipos que se encuentren en cualquier estado: líquido, sólido o gaseoso. En general, esta técnica es aplicable a diversos ámbitos ya que sirve

para identificar una gran cantidad de compuestos y materiales. Fue usada inicialmente para examinar materiales inorgánicos, pero después su uso se amplió considerablemente para el análisis de compuestos orgánicos. Recientemente, esta técnica ha tenido un importante crecimiento en aplicaciones biomédicas y alimenticias, así como también en áreas como geología, astronomía, semiconductores, arte, arqueología y biotecnología. También ha habido avances en la aplicación de esta técnica en la industria petroquímica, industria farmacéutica, de medio ambiente, en el área forense y de análisis de procesos (Smith y Dent, 2005).

2.3 Regresión multivariable, Mínimos cuadrados parciales

2.3.1 Introducción a la calibración multivariable

Para entregar información cuantitativa sobre algún compuesto en una muestra a partir de la medición de un espectro de ésta, es necesario crear una relación entre las variables medidas y la concentración del componente. Existen casos en que la correlación se hace observando una sola variable (una longitud de onda) y se compara con los cambios en la concentración del compuesto. Ésta es conocida como calibración simple o univariada, donde el modelo es construido a partir de una simple regresión lineal, lo cual conlleva a que el método tenga muchas limitaciones. Este modelo solo puede entregar información correcta cuando no existan contribuciones a la variable medida desde otras fuentes y solo el componente objetivo sea el que contribuya a la variable.

La calibración multivariada, en cambio, toma en cuenta todas las variables (espectro completo) entregando así una mejor visión del sistema entero, lo cual entrega varias ventajas, por ejemplo, una razón señal a ruido reducido debido a la información redundante involucrada en el proceso de calibración. Además, este tipo de modelos puede realizar una muy buena cuantificación incluso en la presencia de otros compuestos y variables físicas o químicas que pudieran interferir en la medición. Al

tener una visión del espectro completo es posible cuantificar más de un compuesto, cuando se encuentran presentes mezclados en alguna solución.

Por lo tanto, la calibración multivariada puede construir un modelo que compense las especies que interfieren excepto por aquellas que tienen alguna influencia en alguna longitud de onda que incluya otra información relevante.

Para construir el modelo es necesario un conjunto de mediciones representativas (set de calibración) con concentraciones conocidas del compuesto de interés (valores de referencia).

Existen varias técnicas de calibración para construir modelos multivariables como por ejemplo, CLS (*Classical Least Squares*), ILS (*Inverse Least Squares*) y MLR (*Multiple Linear Regression*). Sin embargo, hoy en día se ha hecho muy popular el uso de Modelos de Regresión Lineal Multivariables más complejos como son PLS (*Partial Least Squares*), PCR (*Principal Component Regression*) y RR (*Ridge Regression*).

2.3.2 PLS, Mínimos cuadrados parciales

La calibración por mínimos cuadrados parciales (PLS, por sus siglas en inglés), fue introducida por Herman Wold (1975) para ser aplicada en ciencias económicas y sociales. Sin embargo, gracias a las contribuciones de su hijo Svante Wold, ha ganado popularidad en el área de la química conocida como Quimiometría o *Chemometrics*, en donde se analizan datos que se caracterizan por muchas variables, con problemas de multicolinealidad, y pocas unidades experimentales en estudio, como es el caso de la espectroscopía.

La idea motivadora de PLS fue heurística, por este motivo, algunas de sus propiedades son todavía desconocidas, a pesar de los progresos alcanzados por Helland (1988), Hoskuldson (1988), Stone y Brooks (1990), entre otros. La metodología PLS generaliza y combina características del Análisis de Componentes Principales (PCA) y Análisis de Regresión Múltiple (MRA). La demanda por esta metodología y la evidencia

de sus resultados, han ido en aumento, convirtiéndose así en uno de los métodos de calibración más utilizados en la actualidad para el análisis de espectros.

Los métodos de calibración multivariable que utilizan modelos de regresión lineal parten desde el supuesto, de que existe una relación lineal entre el espectro y alguna característica de la muestra analizada que pueda ser predicha. Básicamente, los métodos de regresión lineal establecen un modelo como un polinomio lineal de la forma:

$$y = b_0 + b_1 x_1 + \dots + b_n x_n + e \tag{2.3}$$

Donde y es la variable predicha, b_0 es el offset, b_n los coeficientes de regresión, x_n la respuesta medida de la longitud de onda n, y e es el residuo. Otra forma de representarlo es matricialmente:

$$\mathbf{y} = \mathbf{X} \cdot \mathbf{b} + \mathbf{e} \tag{2.4}$$

Donde \mathbf{y} es un vector columna con las mediciones de calibración, \mathbf{X} es la matriz de espectros (m mediciones x n longitudes de onda) \mathbf{y} \mathbf{b} el vector de regresión. Para obtener este vector, es necesario calcular la inversa $(\mathbf{X}^T\mathbf{X})^{-1}$ por lo que ésta debe existir, es decir, el determinante no puede ser cero. Esto agrega la restricción de que el número de mediciones debe ser igual o mayor que el número de longitudes de onda involucrados en la calibración, lo cual muchas veces es un impedimento. En estos casos es conveniente utilizar métodos como PLS, PCR o RR en donde este requerimiento no se hace necesario.

El PLS realiza una regresión lineal luego de descomponer la matriz de espectros **X** en matrices de *scores* y *PLS factors* o variables latentes. Esto puede ser interpretado como una proyección en un nuevo sistema de coordenadas con menores dimensiones que el espacio original de variables independientes. Las variables latentes son determinadas por la máxima varianza de las variables independientes que son relevantes para predecir las variables dependientes. Existen varios algoritmos para calcular un modelo PLS, siendo

los más comunes SIMPLS y NIPALS. Este último es el más intuitivo y su aproximación matemática para el caso de una variable de predicción, se puede representar como:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{e} \tag{2.5}$$

$$\mathbf{y} = \mathbf{U}\mathbf{q}^T + \mathbf{f} \tag{2.6}$$

Donde \mathbf{e} y \mathbf{f} son las matrices residuales, \mathbf{T} y \mathbf{U} matrices de *scores*, y \mathbf{P}^T y \mathbf{q}^T los *loadings* o matrices de factores. El objetivo del PLS es modelar la relación entre los *scores* de la matriz \mathbf{X} y los *scores* del vector \mathbf{y} , de tal manera de encontrar factores capaces de capturar varianza y además lograr correlación. Un conjunto de vectores conocidos como *weights* \mathbf{W} , es el que establece las relación entre los *scores* de \mathbf{X} e \mathbf{y} .

$$\mathbf{U} = \mathbf{TW} \tag{2.7}$$

Todas estas matrices son calculadas secuencialmente en un proceso iterativo, en donde se busca la convergencia para calcular el vector de relaciones para cada variable latente. Cuando el modelo completo es calculado, el vector de regresión resultante queda como:

$$\hat{\mathbf{b}} = \mathbf{P}(\mathbf{P}^T \mathbf{P})^{-1} \mathbf{W} \mathbf{q}^T \tag{2.8}$$

$$\mathbf{y} = \mathbf{X}\hat{\mathbf{b}} \tag{2.9}$$

SIMPLS entrega exactamente el mismo resultado que NIPALS para regresiones simples y una mínima diferencia en el resultado para regresiones multivariables. El objetivo explícito de SIMPLS es maximizar la covarianza, además este algoritmo es más rápido y con mejores resultados, ya que está basado en el criterio de la covarianza mientras que NIPALS no.

Generalmente, los parámetros estimados para los modelos de calibración son elegidos usando criterios de error. Existen tres parámetros típicos para determinar el nivel de error de predicción de un modelo:

• RMSEC (*Root Mean Square Error of Calibration*): Mide el nivel de error entre la respuesta del modelo de calibración \hat{y}_i y su correspondiente valor de referencia y_i para las m muestras usadas en la calibración del modelo. Su fórmula es:

$$RMSEC = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2}$$
 (2.10)

• RMSEV o RMSEP (Root Mean Square Error of Validation or Prediction): Mide el nivel de error entre los valores predichos \hat{y}_i y sus respectivos valores de referencia y_i para el conjunto de p muestras de validación. Su valor es calculado como:

$$RMSEV = \sqrt{\frac{1}{p} \sum_{i=1}^{p} (y_i - \hat{y}_i)^2}$$
 (2.11)

• RMSECV (Root Mean Square Error of Cross-Validation): Al igual que el anterior, mide el error de predicción, pero utilizando el método de validación cruzada para las muestras de calibración. Este método separa el conjunto de muestras en varios subconjuntos y realiza iterativamente una calibración dejando uno de estos conjuntos fuera para usarlo como validación.

El método de validación cruzada es uno de los métodos más conocidos para evaluar el número de variables latentes que se usarán en el modelo.

• R^2 (Índice de Correlación Múltiple): Indica el porcentaje de variación de la variable dependiente explicado por el modelo. Existe un índice de correlación para el conjunto de calibración R_C^2 y uno para el conjunto de validación R_V^2 . Estos son calculados como:

$$R^2 = \left(\frac{\sigma_{y\hat{y}}}{\sigma_y \sigma_{\hat{y}}}\right)^2 \tag{2.12}$$

Donde σ_y corresponde a la desviación estándar de los valores de referencia, $\sigma_{\hat{y}}$ la desviación estándar de los valores predichos, y $\sigma_{y\hat{y}}$ la covarianza entre los valores de referencia y predichos.

3. MATERIALES Y MÉTODOS

3.1 Espectrómetro Raman

Para poder utilizar la técnica de espectroscopía Raman como herramienta de análisis, es necesario un equipo especialmente diseñado para esto. Como se mencionó anteriormente, esta técnica presenta muchas limitaciones debido a la débil intensidad del efecto Raman, por lo que es muy difícil observar sin el equipamiento adecuado.

Luego de ser descubierto este efecto en los años 30, comenzó rápidamente el interés científico para observar y utilizar este fenómeno, pero dado el poco desarrollo de las tecnologías necesarias para visualizar este efecto, y el importante adelanto de la espectroscopía infrarroja, es que este interés se dejó de lado. El descubrimiento del láser en 1960 significó el resurgimiento de la espectroscopía Raman, aunque fue necesario esperar el desarrollo de sistemas de detección más eficientes para poder observar satisfactoriamente el efecto. La tecnología actual de detectores y sistemas ópticos permiten hoy en día obtener el espectro Raman de casi cualquier muestra. Además la incorporación de la fibra óptica como medio de transmisión de la luz aumenta las ventajas del análisis a través de espectroscopía Raman, ya que con la fibra se aíslan todos los campos electromagnéticos externos que pueden afectar a la luz lo que lleva a aumentar la calidad de los espectros obtenidos. Al mismo tiempo la fibra permite una mayor flexibilidad al momento de tomar espectros, permitiendo así tener muestras alejadas del propio instrumento. Esto se traduce en una reducción de costos y amplía considerablemente las aplicaciones de uso de esta técnica.

Así como actualmente existe una extensa gama de aplicaciones para la espectroscopía Raman, existe una gran variedad de instrumentos capaces de captar este efecto. El funcionamiento básico de estos, es el mismo para todos, diferenciándose entre si, por la complejidad del diseño óptico, y por la calidad y capacidades de los componentes interiores.

En la gran mayoría de los casos, la muestra es iluminada usando un rayo láser, la luz dispersada se recoge con una lente y se envía a un detector a través de un elemento

dispersor. Las longitudes de onda cercanas a la del láser, dispersadas debido al efecto Rayleigh, son filtradas, mientras que el resto de la luz se dispersa en el detector. Un esquema generalizado de un espectrómetro Raman se muestra en la figura 3-1.

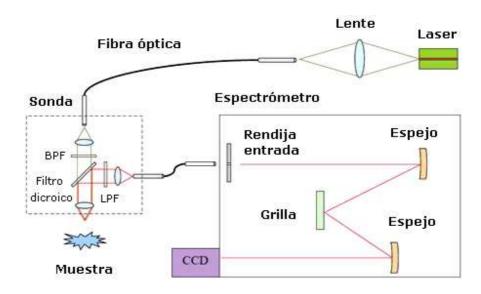


Figura 3-1: Diagrama esquemático de un espectrómetro Raman

Un espectrómetro Raman está compuesto básicamente por los siguientes componentes:

• Fuente de excitación: Se usa generalmente un láser como fuente de excitación en la espectroscopía Raman moderna debido a la alta energía que se requiere para producir dispersión Raman con la suficiente intensidad para poder ser medida correctamente. Los parámetros clave que determinan el desempeño de un láser en la espectroscopía Raman son su longitud de onda, el ancho del haz y la potencia. La longitud de onda debe ser elegida de acuerdo a las propiedades de la muestra y a la respuesta espectral del detector, pero se debe tener muy en cuenta que la elección de ésta afectará directamente a la intensidad del efecto Raman así

como también a la intensidad de fluorescencia en la muestra. Para mejorar la resolución espectral del sistema, el láser debe tener idealmente un ancho de haz tan estrecho como sea posible, y además, la potencia debe ser lo suficientemente grande para producir una señal detectable de la dispersión Raman, pero no tanto como para llegar a quemar la muestra.

• Sonda: La función principal de la sonda en el espectrómetro es dirigir el haz del láser hacia la muestra, recoger la señal, y separarla de la intensa dispersión Rayleigh producida por el láser, la cual es una de las principales dificultades de este tipo de espectroscopía. Teniendo en cuenta que la dispersión Raman es 10⁻⁶ - 10⁻⁸ veces menos intensa que la dispersión Rayleigh, son necesarias algunas técnicas de filtrado especiales. Una de ellas, la más usada, es utilizar filtros rechaza-banda (*Notch*), los cuales dejan pasar todo el espectro, excepto una pequeña banda, la cual se centra en la longitud de onda del láser para así eliminar la dispersión Rayleigh. Otra técnica utilizada cada vez menos, es incorporando dos o tres monocromadores los cuales se configuran de tal forma que separen la dispersión Raman de las otras radiaciones y además aumenten la dispersión y separación de los peaks Raman entre sí.

Ambas técnicas son eficientes presentando cada una tanto ventajas como desventajas. Las sondas basadas en filtros resultan más simples y menos costosas que las otras ya que aprovechan el avance y la efectividad de los filtro *notch*. Las sondas que usan monocromador presentan la ventaja de permitir ajustar en un rango la frecuencia del láser utilizado, así como observar el espectro cercano de la línea de la excitación láser.

En la sonda además, se encuentran otros filtros los cuales cumplen la función de mejorar la calidad de la señal Raman eliminando cualquier otro tipo de interferencia.

• Elemento de dispersión: Los espectrómetros Raman convencionales utilizan una configuración basada en la dispersión *Czerny-Turner* (ver figura 3-2). En ésta, la radiación incidente pasa primero por una rendija de entrada para luego ser colimada en un espejo convexo. La radiación, una vez colimada, es dispersada por una grilla de difracción, y luego es recogida por otro espejo convexo, formándose así múltiples imágenes de la radiación incidente a diferentes longitudes de onda en un detector CCD. Las intensidades relativas de esos componentes de longitud de onda son recibidas por el detector, para formar así el espectro de la muestra examinada. La resolución espectral del instrumento viene determinada por una serie de parámetros, siendo los más importantes el número de pixeles del CCD, la densidad de la grilla, el ancho de la rendija de entrada y la aberración espectral de los espejos.

Otras configuraciones más complejas usadas en algunos espectrómetros pueden ser del tipo *Echelle* o con interferómetro de transformada de Fourier.

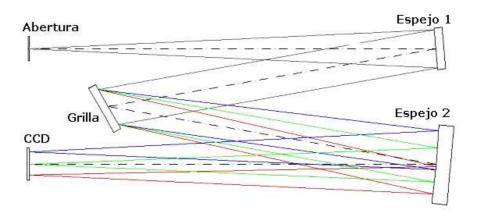


Figura 3-2: Configuración Czerny-Turner con grilla de difracción

• **Detector:** Hoy en día casi la totalidad de los espectrómetros Raman de dispersión utilizan un CCD (*Charge-Coupled Device*) como detector. Éste

cumple la función de convertir la radiación recibida en señales eléctricas, las cuales son enviadas a través de la conexión del espectrómetro a un computador u otro equipo para ser procesadas y así realizarle el análisis correspondiente. Los principales parámetros de un detector CCD son el número de pixeles, el cual afectará la resolución del espectro; y la respuesta en frecuencia del detector, la cual definirá el rango espectral del instrumento.

Todos estos componentes determinan las características y el rendimiento del espectrómetro. Los principales parámetros de rendimiento de estos instrumentos son:

• Longitud de onda de excitación: Influye directamente en la intensidad de la dispersión Raman y en la fluorescencia, según:

Dispersión Raman
$$\propto \frac{1}{\lambda^4}$$

Fluorescencia $\propto \frac{1}{\lambda}$

Su efecto en el espectro de una muestra de Tylenol se puede ver en la figura 3-3.

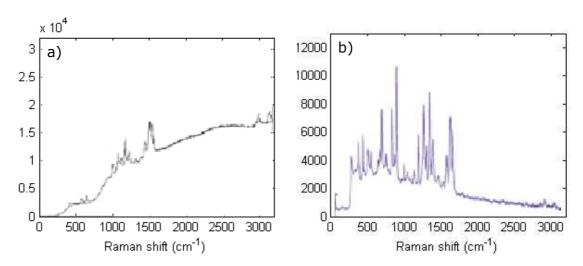


Figura 3-3: Comparación del espectro Raman de una muestra de Tylenol con distintas longitudes de onda de excitación. a) 473 nm. b) 785 nm.

• **Resolución óptica**: Depende del ancho del láser, largo focal, número de pixeles del CCD, aberraciones en la imagen y rango espectral. Las diferencias entre un mismo espectro con distintas resoluciones se aprecia en la figura 3-4.

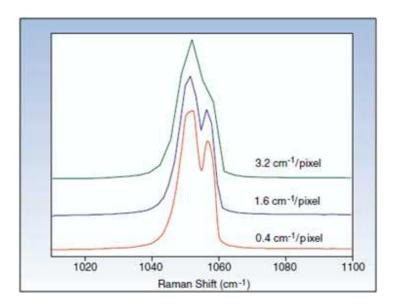


Figura 3-4: Comparación de una zona de un espectro Raman con distinta resolución óptica

- Potencia del láser: A mayor potencia se tiene una señal de dispersión Raman más fuerte, pero podría llegar a deteriorar y quemar la muestra.
- Rango Espectral: Depende de la banda de interés y está directamente relacionado con la longitud de onda del láser y CCD.
- **Diseño Óptico, Tamaño y Peso**: Se refiere a la complejidad del diseño interior del espectrómetro, así como del tamaño y peso que esto implica. Dependiendo de la aplicación que se necesite y el entorno de trabajo, es la importancia de estos parámetros.

A partir de estos parámetros, es posible dividir a los espectrómetros en tres categorías. En primer lugar se tienen los espectrómetros científicos, los cuales están diseñados para análisis y estudios científicos, entregando espectros de alta calidad y precisión, con resoluciones espectrales del orden de los 0,4 cm⁻¹ y rangos que alcanzan hasta 8000 cm⁻¹. Tienen un costo muy elevado y en general son fabricados con diseños ópticos muy complejos y de gran tamaño. En segundo lugar se tienen los espectrómetros industriales, los cuales se destacan por su fabricación especial para ser instalados dentro del mismo proceso, para mediciones tanto de monitoreo como de automatización, integrando las características necesarias para ser incorporado a un lazo de control.

Finalmente se tienen los espectrómetros portátiles, los cuales se caracterizan por su reducido tamaño y peso. Están diseñados para ser trasladados fácilmente y realizar mediciones en terreno. En general tienen menor resolución, rango espectral y potencia que los otros espectrómetros.

En el trabajo desarrollado en esta tesis, se utilizó un espectrómetro Raman B&W TEK BTR111 MiniRam (ver figura 3-5). Éste es un espectrómetro portátil, rápido y de fácil uso, pero con características técnicas bastante limitadas. Cuenta con una excitación láser de infrarrojo cercano (NIR), y funciona con una sonda especialmente diseñada (BAC100-785) la cual transmite la luz entre el instrumento y la muestra a través de dos fibras ópticas. Esta sonda incorpora internamente los respectivos filtros para la eliminación del Rayleigh y otras interferencias. Sus principales características técnicas se resumen en la tabla 3-1.

Tabla 3-1: Especificaciones técnicas del espectrómetro BTR111 MiniRam

Rango espectral	175 – 3100 cm ⁻¹ (<i>Raman shift</i>)
Resolución óptica	10 cm ⁻¹ FWHM
Resolución digital	16 bit, 250 kHz
Dark Noise	< 60 RMS a 9ms
Long. Onda láser	785 nm (± 1 nm)
Ancho haz láser	0,2 nm FWHM
Potencia láser	> 300 mW
Tiempo integración	9 – 65.535 msec.
Detector	CCD, 2048 pixeles
Dimensiones	22,6 x 16,2 x 8,2 cm.
Peso	2 kg.
Interfaz PC	USB



Figura 3-5: Espectrómetro Raman B&W TEK BTR111 MiniRam

3.2 Muestras

Para realizar la calibración del instrumento y validar la técnica Raman en la medición de compuestos fenólicos en el vino, se determinó inicialmente utilizar muestras de una sola cepa, y con esto evitar muestras con valores que se escaparan del grupo promedio. Se decidió utilizar vinos del tipo Cabernet Sauvignon debido a la importancia de esta cepa en el país.

Se buscó tener vinos de distintas viñas, distintas cosechas y años, para así obtener un rango amplio y homogéneo de muestras. Estas muestras en su mayoría fueron entregadas por el Laboratorio de Servicios de la Facultad de Agronomía e Ingeniería Forestal de la P. Universidad Católica de Chile. Estas botellas correspondían a vino de los años 2004 hasta 2007, de las viñas Anakena, Lomas de Cauquenes, Los Robles, San Pedro, Montes y Valdivieso, entre otras. Para incorporar al modelo un rango mayor con vinos más nuevos, se consiguieron algunas botellas de muestra del año 2008, en visitas realizadas a las viñas Aquitania, Terraustral, Martínez de Salinas, Tinajas del Maule y Terramater.

En total se utilizaron 114 muestras de 38 botellas distintas de Cabernet Sauvignon de los años 2004 a 2008. De éstas, un tercio se utilizó para validación y el resto para calibración. Estos conjuntos se seleccionaron de tal manera de abarcar el rango completo de valores, tanto para calibración, como validación.

Todas las muestras fueron previamente pasadas por un filtro de membrana de 0,6 μm. y luego almacenadas en tubos de plástico de 5 ml., cada una con su respectiva muestra de respaldo.

Se trabajó en general con grupos de 8 muestras las cuales se almacenaban en frio (aprox. 5°C) entre cada medición, cuidando que no pasen más de 3 días entre el muestreo Raman y la analítica de laboratorio, para así no alterar sus características.

3.3 Métodos analíticos tradicionales

Como se mencionó anteriormente, para realizar una calibración utilizando el método de PLS, se necesita para cada muestra un espectro junto a su respectivo valor de referencia para hacer la correlación. Estos valores fueron obtenidos realizando las metodologías utilizadas en el laboratorio de Enología de la Facultad de Agronomía e Ingeniería Forestal de la P. U. C. de Chile. Estos corresponden a un análisis espectrofotométrico en el cual se mide la absorbancia de una solución que contiene a la muestra.

Los análisis se hicieron generalmente en grupos de 4 u 8 muestras, luego de ser filtradas y almacenadas como se señaló en el punto anterior. Para cada muestra se hicieron 3 repeticiones, para así reducir lo más posible los errores, dejando fuera el valor medido más lejano y tomando como valor final el promedio de las otras dos. Además para aumentar la exactitud de los resultados es que se hicieron todas las mediciones de volumen utilizando micropipetas y repitiendo siempre el mismo procedimiento.

Los distintos análisis que se realizaron fueron:

• Polifenoles totales (DO280): (Bordeu y Scarpa, 1998) Se prepara una dilución 1:100 de la muestra en agua destilada, es decir, se mezclan 10 ml. de agua destilada con 0,1 ml. de la muestra de vino en un tubo de ensayo. Luego se mide esta solución en el espectrofotómetro a 280 nm., en cubetas de cuarzo de 10 mm. de paso óptico, y usando un blanco de agua destilada. El índice se obtiene como:

$$D0280 = Absorbancia \cdot 101 \tag{3.1}$$

• Antocianas totales: (Puissant y Leon, 1967) Se prepara una dilución 1:100 de la muestra en HCl 1% p/v, es decir, se mezclan 10 ml de HCl 1% p/v con 0,1 ml de la muestra de vino en un tubo de ensayo. Luego se mide esta solución en el

espectrofotómetro a 520 nm., en cubetas de poliestireno de 10 mm. de paso óptico, y usando un blanco de HCl 1% p/v. El índice se obtiene como:

Antocianas
$$(mg/lt) = Absorbancia \cdot 2298,76$$
 (3.2)

• Taninos totales por precipitación con metilcelulosa: (Sarneckis et al., 2006) Para este análisis se preparan 2 soluciones. En un tubo se mezclan 0,250 ml de vino con 2 ml de metilcelulosa (tubo metilcelulosa), mientras que en el otro se coloca sólo 0,250 ml de vino (tubo testigo). Luego a ambos tubos se le agregan 2 ml. de solución saturada de Sulfato de Amonio y se completan con agua destilada cada uno hasta los 10 ml. Se debe dejar reposar ambos tubos por 10 minutos, y luego llevarlos a centrifugar por 6 minutos a 4000 rpm. Después nuevamente se dejan reposar por 10 minutos para luego sacarle el sobrenadante a cada tubo y medirlo en el espectrofotómetro a 280 nm. en cubetas de cuarzo de 10 mm. de paso óptico y usando un blanco de agua destilada. El cálculo de la concentración de taninos se realiza como:

$$Taninos(mg/lt) = (Abs. testigo - Abs. metilcel.) \cdot 83,03 - 3,87$$
 (3.3)

3.4 Adquisición de espectros

Para la toma de los espectros Raman se utilizó un software creado especialmente para esta aplicación (ver figura 3-6), el cual fue diseñado utilizando el kit de desarrollo de software (SDK) incluido con el espectrómetro en un ambiente Visual Studio .NET en lenguaje C#. Este programa se diseñó para facilitar la toma de espectros cuando se quiere obtener grandes cantidades de estos y sobretodo de manera continua. El funcionamiento es bastante simple y sólo requiere introducir un identificador para la muestra y presionar el botón "Guardar" las veces que sea necesario para adquirir los

espectros. El programa automáticamente guarda en un archivo de texto el identificador, el número de espectro, y los datos capturados. Toda esta información, como también la del tiempo de integración y tipo de espectro, es almacenada dentro del archivo para poder ser leída y posteriormente utilizada para la agrupación de los espectros.

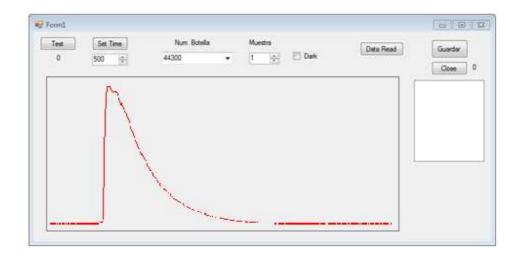


Figura 3-6: Software utilizado para la captura de espectros

Para la adquisición de los espectros Raman de las muestras se realizó el siguiente procedimiento:

- Dejar reposar las muestras, previamente almacenadas en frío, por 10 minutos a temperatura ambiente.
- Vaciar la muestra desde el tubo plástico a la cubeta de vidrio e introducirla en el porta cubetas BCH100 (ver figura 3-7) especialmente diseñado para toma de muestras en espectroscopía.



Figura 3-7: Portacubetas B&W TEK BCH-100

- Abrir el obturador (*shutter*) de la sonda para iluminar la muestra con el láser durante 30 segundos, para estabilizar los espectros obtenidos.
- Capturar 10 espectros de manera continua, cada uno con un tiempo de integración de 500 ms.
- Cerrar el shutter y capturar un espectro de oscuridad (Dark), el cual es sustraído a los otros espectros ya que sólo contiene interferencia para las mediciones.
- Retirar la cubeta, agitar la muestra, y volver a repetir el mismo procedimiento dos veces más.

De esta forma se tendrán $n \times 10 \times 3$ espectros para cada muestra, y $n \times 3$ espectros Dark. Todos estos espectros serán posteriormente agrupados y promediados para quedar finalmente con un espectro por muestra.

3.5 Preprocesamiento de espectros

Los modelos de regresión multivariable, como se mencionó anteriormente, no buscan una relación absoluta entre las longitudes de onda y la concentración del analito, sino que calculan el cambio relativo en el espectro e intenta correlacionarla con el respectivo cambio en la concentración del compuesto buscado en las muestras. Es por esto que los modelos tienden a ser tan robustos y permiten ser calibrados en presencia de otras interferencias.

Debido a este hecho es que existen algunas técnicas matemáticas para preparar los espectros, los cuales se aplican a los datos que serán usados en el modelo multivariable. El objetivo de estos algoritmos es eliminar la información redundante y aumentar las diferencias relevantes entre cada una de las muestras.

Para encontrar el mejor modelo de calibración se probaron y aplicaron diferentes tipos de procesamiento de señal. Algunos para acondicionar el espectro para la aplicación del método PLS, otros para eliminar fluorescencia y otros simplemente para filtrar fuentes de ruido que pudieran interferir. Cada uno de estos será explicado a continuación:

a) **Promedio y sustracción del** *Dark*: Antes de realizar cualquier otro procesamiento, se agrupan los espectros tomados para cada muestra. Recordando que cada una tiene 10 espectros, estos fueron promediados obteniendo así un espectro para cada muestra. Al utilizar un espectro promedio en vez de un solo espectro individual, se reduce el ruido aleatorio producido en el CCD del espectrómetro. A cada uno de estos espectros promedios se le restó además, el correspondiente espectro promedio *Dark*, para así dejar un espectro sólo con información de la muestra, eliminando cualquier otra información que pudiera interferir desde la sonda o del instrumento. Su algoritmo matemático es:

$$X_{i} = \sum_{j=1}^{10} X_{i,j} - \sum_{k=1}^{3} X_{Dark_{i,k}}$$
(3.4)

- b) **Filtro Savitzky Golay**: El filtrado a través del método de Savitzky Golay es uno de los más utilizados en señales espectroscópicas para la eliminación de ruido aleatorio y suavización de la señal (Člupek et al., 2007). La gran ventaja de este filtro, en comparación con otros (FT, ventana móvil, LPF) es que tiende a conservar las características de la señal como máximos y mínimos relativos, y anchos de los peaks. El método se basa en ajustar un polinomio de orden *k* en un pequeño intervalo de al menos *k*+1 puntos, con la posibilidad además de aplicarle derivadas de primer o mayor orden, lo cual en algunos casos es útil para eliminar ruidos de fondos constantes y para acentuar las diferencias entre los datos espectrales.
- c) Centrado de medias: el centrado de medias se aplica generalmente cuando se hace un modelo de calibración multivariable. Se calcula un espectro promedio de todo el conjunto de espectros y éste se le resta a cada uno de los espectros individuales. Además, se hace lo mismo con los valores de referencia, se calcula el valor promedio y se resta este valor a cada una de las muestras. Mediante el centrado de la media, las diferencias entre las muestras se mejora considerablemente en términos de la concentración y de la respuesta espectral. Esto lleva a que se realice una calibración del modelo más precisa y con mejores predicciones. Este método se calcula como:

$$X_{i_{corr}} = X_i - \sum_{i} X_i$$
 (3.5)

d) MSC: Multiplicative Signal (o Scatter) Correction es un pre-tratamiento propuesto para datos espectrales en el que se estiman los efectos de la dispersión o los cambios en el paso óptico en relación a un espectro de referencia X_{ref} , el cual generalmente corresponde al espectro promedio de todo el conjunto. En teoría, esta estimación se debería hacer en zonas del espectro que no contengan

información, sin embargo, en la práctica se corrige el espectro completo. El modelo de MSC se puede escribir como:

$$X^T \beta + \alpha = X_{ref}^T \tag{3.6}$$

Donde α y β representan la corrección de la línea base y el paso óptico, en relación al espectro de referencia. Estos parámetros se calculan mediante una regresión simple.

El MSC asume que la totalidad de la información contenida en un espectro se encuentra recogida en X_{ref} y en otro término δ_i que representa un conjunto de variación desconocida. La corrección del espectro se representa como:

$$X_{corr} = (X_{ref} - \alpha)/\beta \tag{3.7}$$

Esta corrección puede llevar a cometer errores, ya que se puede eliminar o modificar la información de interés en los espectros que se corrigen. Además debe ser usado exclusivamente en espectros en donde la información contenida en él varíe linealmente con la concentración.

e) *Rolling Circle* para eliminación de fluorescencia: éste es un método basado en un algoritmo para eliminación de la señal de fondo (*background*) de un espectro (Mikhailyuk y Razzhivin, 2003). En general los métodos de eliminación de *background*, como ajuste de curva, o derivaciones e integraciones sucesivas, requieren de un número considerable de parámetros y consideran muchas restricciones para el polinomio que será ajustado como línea base. Este método está basado en las diferencias geométricas entre el *background* y la señal útil. El algoritmo es el siguiente, por cada punto (*X_i*, *Y_i*) de la señal original, se construye un círculo de radio R con coordenada *X_C* del centro igual a *X_i*. La curva original y cada uno de los círculos, debe tener por lo menos un punto en

común, mientras que el resto de los puntos del círculo debe tener coordenadas Y menores que el respectivo punto en la curva original. Luego se debe calcular la diferencia entre las coordenadas Y_i de la señal con respecto a la parte superior del círculo para cada X_i . La menor de estas diferencias es almacenada en un vector, el cual corresponderá al espectro corregido sin la señal de *background*. Este algoritmo puede ser imaginado intuitivamente como un círculo que se desplaza bajo la curva original y que va eliminando toda la región por la que va pasando, tal como se muestra en la figura 3-8.

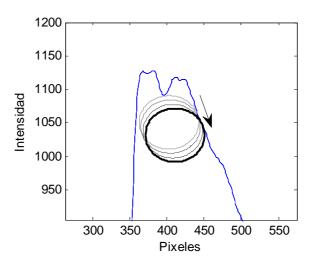


Figura 3-8: Método de eliminación de fluorescencia con círculo

La característica de este método, es que introduce una mínima distorsión si el radio R es igual al mínimo radio de curvatura de la señal de *background*, o menor a ésta, siempre y cuando sea mayor que el ancho de cualquiera de los peaks de la señal útil. Además, tiene la ventaja de que utiliza un solo parámetro R, el cual puede ser elegido fácilmente analizando la condición de distorsión mínima en una curva del área bajo la señal corregida vs el radio R del círculo.

f) Eliminación fluorescencia con ajuste de curva: Los métodos de eliminación de fluorescencia por ajuste de curva, son unos de los más utilizados y los que entregan mejores resultados, ya que aproximan la señal de *background* con un polinomio y excluyen toda la información Raman. Sin embargo no existe un claro consenso en cuál es la curva que más se ajusta al efecto de la fluorescencia. Se han utilizado distintos algoritmos, polinomios de orden 3 y 5, *splines* cúbicos, y funciones por partes, entre otros. Teniendo en cuenta que cada uno de estos es ajustado para un espectro en particular, por lo que lo hace difícil de utilizar de forma general y automatizada (Cao et al., 2007).

3.6 Calibración

Como se mencionó anteriormente, el método de regresión por mínimos cuadrados parciales (PLS) es el método de calibración más apropiado cuando se quiere relacionar una señal multivariable (espectro) con sus respectivos valores de referencia.

Este algoritmo se aplicó a los datos utilizando el software MATLAB a través del *toolbox* PLS_Toolbox 5.0.3 de la empresa Eigenvector Research, Inc., el cual fue obtenido a través de una licencia gratuita por 6 meses para estudiantes.

Se utilizó principalmente el GUI (*Graphic User Interface*) de análisis, el cual se ve en la figura 3-9. Esta interfaz incorpora la mayoría de las funciones necesarias para realizar la calibración de manera simple.

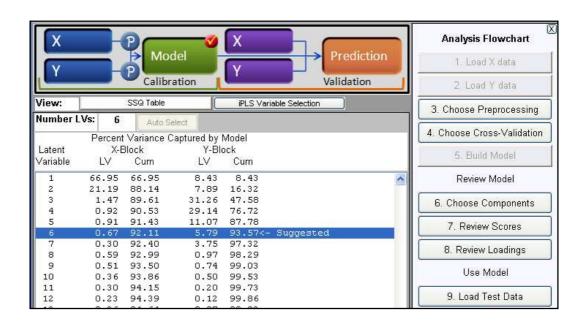


Figura 3-9: Interfaz gráfica de análisis del PLS_Toolbox 5.0.3 para MATLAB

En la calibración, se trabajó por separado con los datos de antocianas, polifenoles y taninos, pero con los mismos espectros, obteniéndose así un modelo para cada uno de estas mediciones.

Los datos de entrada para el modelo son los espectros pre-procesados junto con los correspondientes valores de referencia, mientras que de salida se obtiene el vector de regresión **b**, además de algunos valores estadísticos para la evaluación del modelo.

Los datos de entrada, tanto espectros como referencias deben ser separados en dos grupos para obtener un modelo valido. Uno corresponde a los datos de calibración, mientras que el otro es el de validación. El primero contiene alrededor de dos tercios de las muestras, y el otro el resto. Las muestras de cada uno de los grupos fueron seleccionados para abarcar el rango completo de valores en ambos conjuntos.

Los datos de calibración son los que se usan para construir el modelo en sí, mientras que los de validación sirven para comprobar la efectividad del modelo en muestras que no están en el conjunto de validación.

Luego de ingresar los datos al programa, se ingresan las configuraciones deseadas, se ejecuta el cálculo del modelo, y se debe elegir el número de variables latentes (LV) que éste usará. Para lo cual se realiza una validación cruzada, en donde aleatoriamente se saca un subconjunto del total de las muestras y se utiliza como conjunto de validación, mientras que el resto queda para la calibración. Esto se repite un número predeterminado de veces y se calcula el RMSECV para cada una de ellas. Con ayuda de estos valores a través de un gráfico RMSECV vs LV, como el de la figura 3-10, se elije el número de variables latentes óptimo para el modelo, el cual corresponde al punto en que se quiebra la curva luego de la primera bajada.

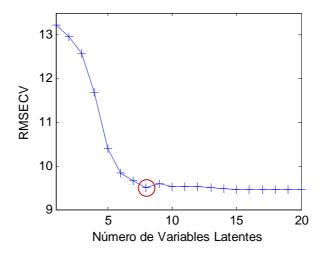


Figura 3-10: Gráfico RMSECV vs LV para selección del número de variables latentes para el modelo PLS

Con el número de variables seleccionado, se pueden analizar los gráficos estadísticos, en los que se ve que tan bueno es el modelo ajustado. Estos gráficos corresponden a la revisión de los *scores* y *loadings* del modelo.

Dentro de los *scores*, se tiene el gráfico de dispersión en donde cada muestra aparece con su valor de referencia predicho vs el real (ver figura 3-11). Las muestras de

calibración se muestran como círculos, mientras que las de validación como triángulos. Además se muestra la recta de tendencia de las muestras junto a la recta de pendiente 1 que representa la correlación perfecta. Junto al gráfico se observan datos estadísticos como R^2 , RMSEC y RMSEV. La importancia de este gráfico es que se puede ver que tan bien ajustados están los datos en el modelo, tanto los de calibración como validación.

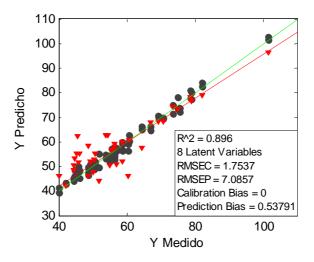


Figura 3-11: Gráfico de dispersión de valores predichos vs medidos

Otro gráfico importante de analizar dentro de los *scores* es el gráfico de residuos de predicción (*y residuals*) para cada muestra, como el que se muestra en la figura 3-12. En este gráfico se muestra básicamente que tan alejado está cada muestra de la recta de tendencia del modelo PLS, por lo que es fácil identificar alguna muestra que no esté representada por éste. Estas muestras son llamadas *outliers* y si se excluyen pueden mejorar considerablemente los resultados de predicción del modelo.

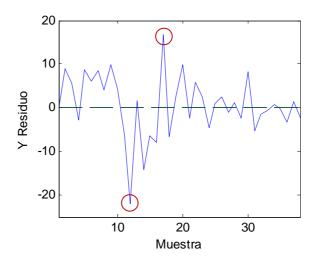


Figura 3-12: Gráfico de residuos de predicción

En el caso del análisis de los *loadings*, se tienen dos gráficos importantes para la interpretación del modelo. El primero es el que muestra el vector de regresión **b**, vector que se multiplica por cada espectro para obtener la respectiva predicción. Este gráfico se puede ver en la figura 3-13.

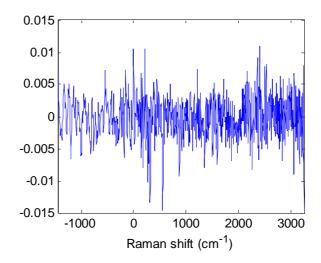


Figura 3-13: Gráfico del vector de regresión **b**

Y el segundo (figura 3-14) es el que muestra las VIP (*Variable Importance in Projection*). Este vector entrega una estimación de la importancia relativa de cada variable independiente, en cada uno de los factores de proyección (LV's) utilizados en el modelo PLS. Estos coeficientes entonces representan la importancia de cada variable **x** en el ajuste de los *scores* de las matrices **X** e **Y**, según el algoritmo del PLS (Chong y Jun, 2005).

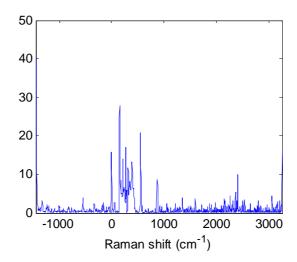


Figura 3-14: Gráfico de las VIP (Variable Importance in Projection)

4. RESULTADOS

4.1 Calibración de espectros

A continuación se muestran los mejores resultados obtenidos al aplicar los diferentes métodos de pre-procesamiento a los espectros obtenidos (ver tabla 4-1). Se realizó la calibración de un modelo PLS para cada uno de ellos, utilizando los valores de referencia de los polifenoles para comparar los efectos en el resultado que tiene cada método. Sus respectivos gráficos de dispersión y espectros procesados se muestran en las figuras 4-1 a 4-7.

Previo a la calibración, para todos los métodos, se aplicó un filtro Savitzky-Golay de orden 2, tamaño 15, sin derivación, a los espectros para eliminar ruido, y se realizó centrado de medias en los espectros y en los valores de referencia para obtener un modelo más preciso.

Tabla 4-1: Resultados calibración PLS utilizando diferentes métodos

Pre-Procesamiento	LV	R_C^2	R_V^2	RMSEC	RMSEV
Ninguno	9	0,975	0,773	2,14	6,40
MSC	6	0,852	0,580	5,16	8,78
Círculo, R=10	7	0,994	0,703	1,08	7,34
Círculo, R=20	8	0,987	0,668	1,52	7,77
Círculo, R=50	6	0,960	0,690	2,69	7,57
Círculo, R=35	8	0,983	0,723	1,75	7,09
Círculo mod., R=35	8	0,982	0,749	1,81	6,73

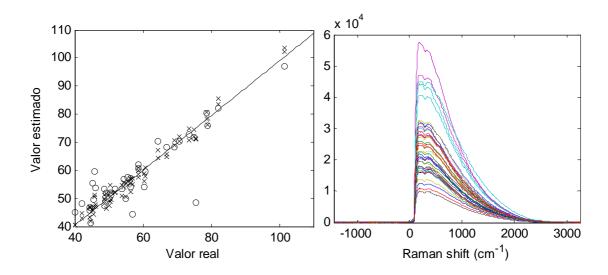


Figura 4-1: Calibración modelo PLS de espectros en bruto, usando 9 LV.

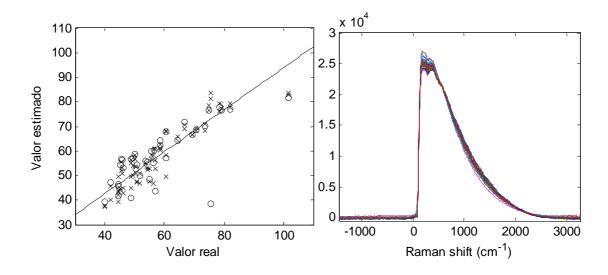


Figura 4-2: Calibración modelo PLS de espectros con MSC, usando 6 LV

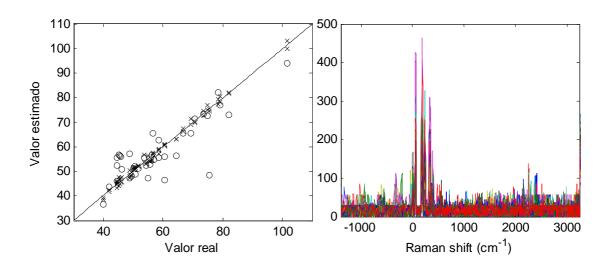


Figura 4-3: Calibración modelo PLS de espectros con eliminación de fluorescencia aplicando método del círculo con R=10, usando 7 LV.

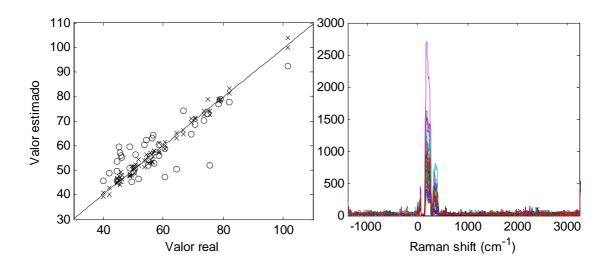


Figura 4-4: Calibración modelo PLS de espectros con eliminación de fluorescencia aplicando método del círculo con R=20, usando 8 LV

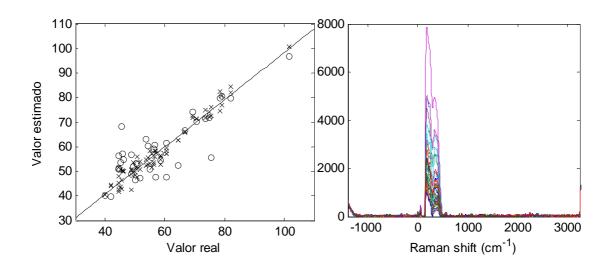


Figura 4-5: Calibración modelo PLS de espectros con eliminación de fluorescencia aplicando método del círculo con R=50, usando 6 LV

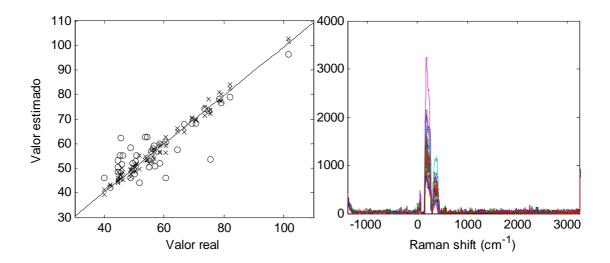


Figura 4-6: Calibración modelo PLS de espectros con eliminación de fluorescencia aplicando método del círculo con R=35, usando 8 LV

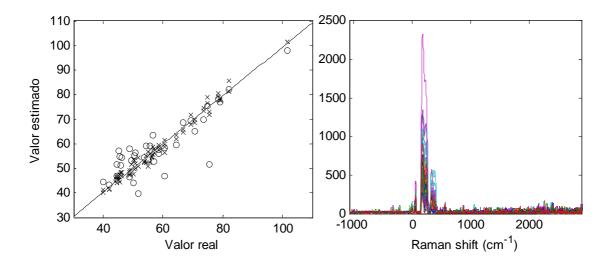


Figura 4-7: Calibración modelo PLS de espectros con eliminación de fluorescencia aplicando método del círculo con R=35 y Raman shift, usando 8 LV

Como se mencionó anteriormente, uno de los principales objetivos del preprocesamiento, es aislar la información relevante del espectro completo. En el caso del
Raman se busca eliminar la fluorescencia que es uno de los fenómenos que más
interfiere en la calibración del modelo. El método con que se obtuvo mejores resultados
fue el método del círculo, el cual tiene como parámetro solamente el radio del círculo.
Como regla general para la elección de este parámetro, se debe conocer a priori el ancho
máximo de los peaks Raman que contienen la información relevante del espectro, y se
debe elegir un círculo de diámetro mínimo igual a este ancho y de radio máximo igual al
radio de curvatura de la señal de fluorescencia. Como estos valores no se conocen, y son
diferentes para cada uno de los espectros, se aplicó el método descrito por Mikhailyuk y
Razzhivin (2003) para encontrar el radio óptimo del círculo que provoca distorsión
mínima en la señal y máxima eliminación de la fluorescencia. Para esto, se grafica el
área bajo la curva de la señal sin fluorescencia en función del radio del círculo, y luego
se elige el punto más a la derecha de la meseta que se forma en la parte central de la
curva, tal como se ve en la figura 4-8.

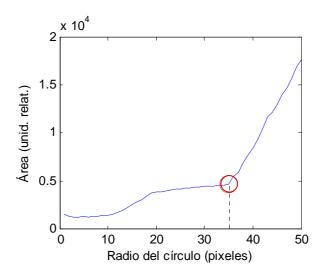


Figura 4-8: Selección del radio óptimo para método del círculo

El efecto que tiene la elección de este parámetro en una zona del espectro $(0-500\ cm^{-1})$ se puede apreciar en la figura 4-9. Considerando que la fluorescencia se manifiesta como una curva suave, claramente se puede ver que ésta no está presente en la figura 4-9: a, b, y c, mientras que en la figura 4-9: d, sólo se aprecia la fluorescencia para los espectros de mayor intensidad. También se ve como en la figura 4-9: a, el método además de la fluorescencia elimina otra información que puede ser relevante para el modelo, ya que el círculo entra en lugares donde ésta puede estar. En el caso de la figura 4-9: b y c, casi no se aprecian diferencias a simple vista y esto se debe a que como se ve en el gráfico de la figura 4-8, ambas se encuentran dentro de la meseta de la curva, por lo que la cantidad de información que recogen es muy similar.

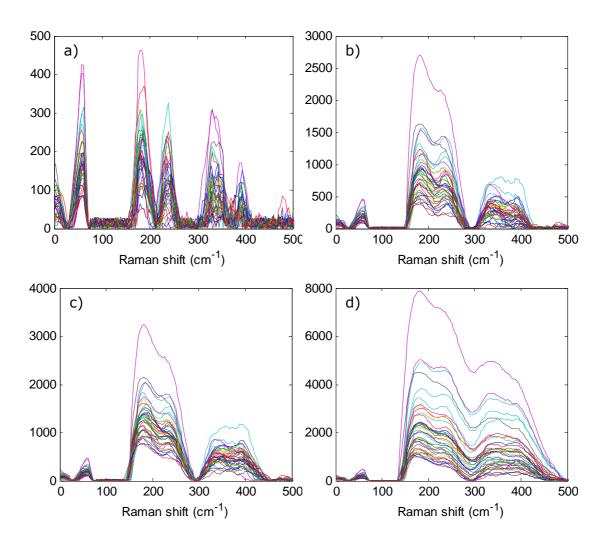


Figura 4-9: Comparación del método de eliminación de fluorescencia con círculo, en una zona del espectro, para distintos radios a) R=10, b) R=20, c) R=35, d) R=50

Al comparar los distintos modelos de calibración, se puede ver que tanto el que incluye el espectro completo como el que no tiene fluorescencia, entregan buenos resultados. Para verificar el origen de la información correlacionada en el modelo se realizó una calibración de la misma manera que los modelos anteriores, pero utilizando solamente la fluorescencia sustraída de los espectro originales. Se usó el modelo con círculo de radio R=35 con 8 LV, obteniéndose los resultados mostrados en la figura 4-

10. Estos resultados son inferiores a los obtenidos con el espectro completo y con el espectro sin fluorescencia, lo cual hace pensar que el fenómeno de la fluorescencia sí aporta a la correlación del modelo, pero el fenómeno Raman por sí sólo, contribuye mayoritariamente con la información relevante para la calibración. Esta comparación se puede apreciar en la tabla 4-2.

Tabla 4-2: Resultados calibración PLS utilizando diferentes segmentaciones del espectro

Espectro	LV	R_C^2	R_V^2	RMSEC	RMSEV
Completo	9	0,975	0,773	2,14	6,40
Sin Fluorescencia	8	0,983	0,723	1,75	7,09
Sólo Fluorescencia	8	0,865	0,702	4,94	7,33

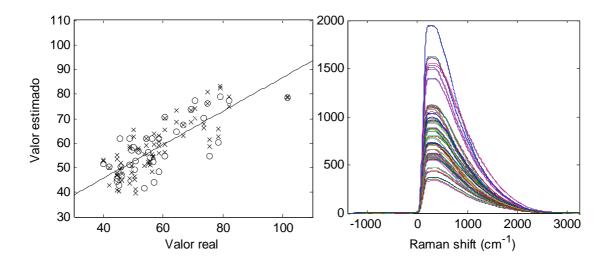


Figura 4-10: Calibración modelo PLS con solo la fluorescencia extraída de los espectros, usando 8 LV

Finalmente se decidió utilizar como modelo definitivo el que utiliza la eliminación de fluorescencia con círculo de R=35, pero con una pequeña modificación, la cual consiste en realizar el mismo algoritmo pero escalando el eje x en función del *Raman shift*. Este modelo fue tomado como base para realizar la calibración final para polifenoles totales, antocianas y taninos. A cada uno de estos se les eliminó los *outliers* identificados en los respectivos gráficos de residuos de predicción, y así se obtuvieron los resultados que se muestran a continuación.

4.2 Polifenoles totales

Los resultados de la calibración de los polifenoles totales fueron los más bajos en términos de R_V^2 , sin embargo son bastante buenos y demuestran una alta correlación del espectro Raman con los resultados de la analítica tradicional. Al calcular el modelo se detectó un solo *outlier* en el conjunto de validación (ver figura 4-11), el cual fue eliminado de éste.

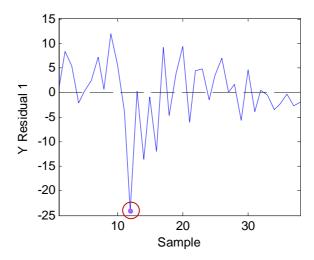


Figura 4-11: Gráfico de residuos de predicción en modelo para polifenoles totales

En la tabla 4-3 se muestran los valores de R^2 y RMSE para el conjunto de calibración y validación del modelo calculado para polifenoles totales. Se debe considerar para los valores de RMSE que el rango de polifenoles en las muestras utilizadas es de 61,5 (mg/lt). El gráfico de dispersión con las muestras del modelo se ve en la figura 4-12.

Tabla 4-3: Resultados modelo de calibración PLS para polifenoles totales

LV	R_C^2	R_V^2	RMSEC	RMSEV
8	0,982	0,829	1,81	5,55

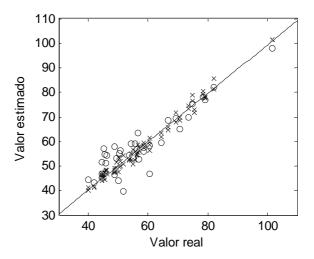


Figura 4-12: Gráfico de dispersión de valores predichos vs medidos para polifenoles totales

Al analizar los *VIP scores* del modelo (ver figura 4-13) se puede observar cuáles son las bandas (expresadas en *Raman shift*) que tienen mayor relevancia en el modelo de calibración, notándose claramente algunos peaks sobretodo en la primera parte del espectro.

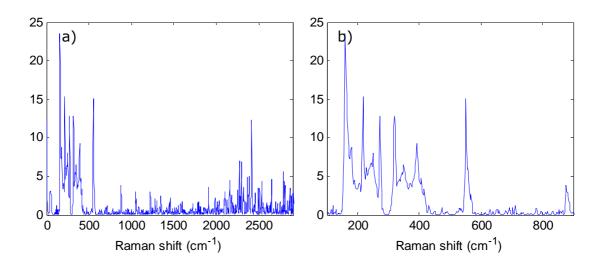


Figura 4-13: Gráfico de las VIP para: a) espectro completo, b) zona 100-900 cm⁻¹

4.3 Antocianas

Los resultados obtenidos en la estimación del nivel de antocianas en las muestras de vino son levemente mayores que los obtenidos con los polifenoles, por lo que igualmente se obtuvo una muy buena correlación entre los valores medidos en laboratorio y los medidos con el espectrómetro Raman. En este modelo se detectaron dos muestras en el conjunto de validación que se alejaban del resto de las muestras (ver figura 4-14), por lo que fueron eliminadas del modelo.

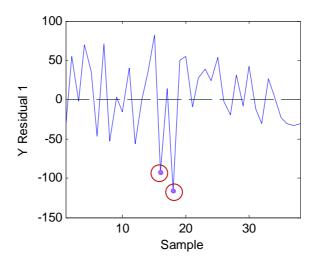


Figura 4-14: Gráfico de residuos de predicción en modelo para antocianas

En la tabla 4-4 se muestran los valores de R^2 y RMSE para el conjunto de calibración y validación del modelo calculado para las antocianas. Se debe considerar para los valores de RMSE que el rango de antocianas en las muestras utilizadas es de 343,66 (mg/lt.).

Tabla 4-4: Resultados modelo de calibración PLS para antocianas

LV	R_C^2	R_V^2	RMSEC	RMSEV
9	0,991	0,844	8,86	38,5

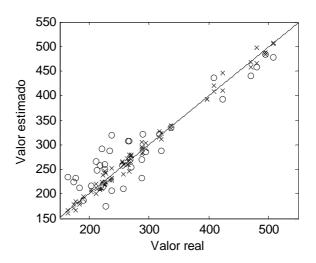


Figura 4-15: Gráfico de dispersión de valores predichos vs medidos para antocianas

El gráfico de dispersión con las muestras del modelo para antocianas se ve en la figura 4-15, mientras que el gráfico de las VIP con la importancia relativa de las bandas del espectro en el modelo se muestra en la figura 4-16.

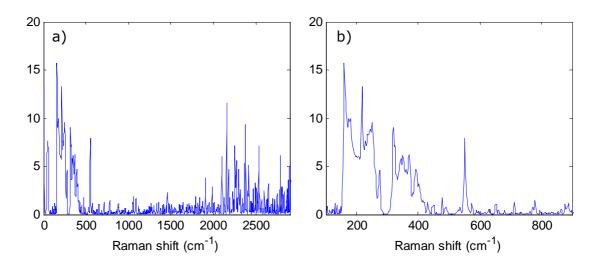


Figura 4-16: Gráfico de las VIP para: a) espectro completo, b) zona 100-900 cm⁻¹

4.4 Taninos

Finalmente para la estimación de taninos en las muestras de vino se obtuvieron los mejores resultados de correlación. Para este modelo no se observaron muestras muy alejadas de la recta de calibración (ver figura 4-17), sin embargo se eliminó del conjunto de validación la muestra más apartada.

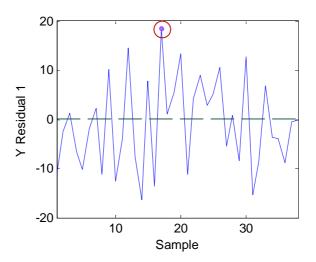


Figura 4-17: Gráfico de residuos de predicción en modelo para taninos

En la tabla 4-5 se muestran los valores de R^2 y RMSE para el conjunto de calibración y validación del modelo calculado para los taninos. Se debe considerar para los valores de RMSE que el rango de la concentración de taninos en las muestras utilizadas es de 138,86 (mg/lt).

Tabla 4-5: Resultados modelo de calibración PLS para taninos

LV	R_C^2	R_V^2	RMSEC	RMSEV
10	0,995	0,895	1,76	8,65

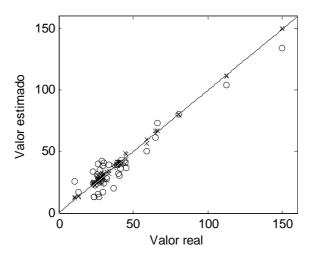


Figura 4-18: Gráfico de dispersión de valores predichos vs medidos para taninos

El gráfico de dispersión con las muestras del modelo para taninos se ve en la figura 4-18, mientras que el gráfico de las VIP con la importancia relativa de las bandas del espectro en el modelo se muestra en la figura 4-19.

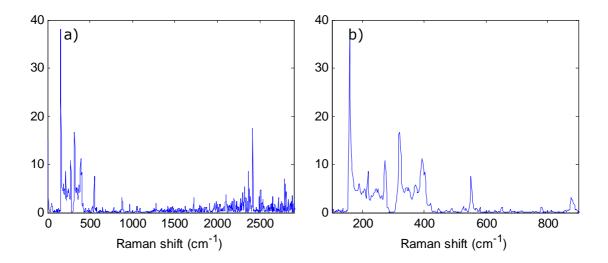


Figura 4-19: Gráfico de las VIP para: a) espectro completo, b) zona 100-900 cm⁻¹

5. CONCLUSIONES

Dados los resultados obtenidos en este trabajo se puede afirmar que la técnica de espectroscopía Raman efectivamente funciona para medir compuestos fenólicos en el vino, específicamente utilizando un espectrómetro portátil de media precisión. Esto abre las posibilidades de utilizar este instrumento como una herramienta de medición en la industria vitivinícola. Su utilidad y sus ventajas son considerablemente mayores que las técnicas analíticas tradicionales, ya que es un instrumento muy rápido, fácil de usar, que entrega resultados instantáneos, y permite una movilidad que para los métodos actuales parece imposible.

Por otro lado se pudo demostrar a través de técnicas de segmentación, que a pesar de que el instrumento captura conjuntamente los fenómenos de fluorescencia y dispersión Raman, ambos fenómenos presentan correlación con los compuestos analizados, siendo mayor ésta, cuando se utiliza la dispersión Raman.

Se obtuvieron resultados muy satisfactorios para polifenoles totales, antocianas y taninos, tanto en la validación como en la calibración de los modelos. Esto, a pesar de los posibles errores que traen implícitos los procedimientos analíticos de laboratorio.

Se obtuvieron valores de R_V^2 entre 0,829 y 0,895 para los distintos componentes medidos y valores de R_C^2 por sobre los 0,982. El compuesto que logró mejor correlación fueron los taninos, mientras que los polifenoles totales fueron los más bajos.

Los modelos obtenidos tienen la ventaja de no ser sesgados ya que fueron calibrados utilizando una variedad de muestras de vino Cabernet Sauvignon, de diferentes años, cosechas, y viñas.

A pesar de que el espectrómetro Raman es un instrumento con un costo inicial elevado, el ahorro que entrega en términos de tiempo y de insumos es bastante alto, ya que como se mencionó anteriormente las muestras no necesitan una preparación previa y se pueden obtener resultados en pocos segundos.

Estos resultados si bien son exclusivos para la cepa Cabernet Sauvignon, es posible extenderlos fácilmente a otras variedades e incluso a un modelo general de vino tinto.

Además aprovechando las ventajas de la espectroscopía Raman es posible ampliar los modelos y utilizar este instrumento para el análisis de otros parámetros y compuestos del vino utilizando las mismas mediciones que se utilizaron en los compuestos fenólicos.

6. BIBLIOGRAFIA

Argyri, K., Proestos, C., Komaitis, M. y Kapsokefalou, M. (2005). Phenolic compounds in red wine digested in vitro in the presence of iron and other dietary factors. *International Journal of Food Sciences & Nutrition*, 56 (3), 213-222.

Ayala, F., Echavarri, J. F. y Negueruela, A. I. (1997). A new simplified method for measuring the color of wines. I. Red and Rose Wines. *American Journal of Enology and Viticulture*, 48 (3), 357-363.

Bordeu E. y Scarpa J. (1998). *Análisis químico del vino*. Santiago, Chile: Ediciones Pontificia Universidad Católica de Chile.

B&W TEK Inc. (2007) *User manual, BTR111 MiniRam miniature NIR Raman spectrometer*. Manual de referencia, Newark, DE.

Cao, A., Pandya, A., Serhatkulu, G., Weber, R., Dai, H., Thakur, J. et al. (2007). A robust method for automated background subtraction of tissue fluorescence. *Journal of Raman Spectroscopy*, *38*, 1199-1205.

Chong, I., Jun, C. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*, 78, 103-112.

Člupek, M., Matějka, P., Volka, K. (2007). Noise reduction in Raman spectra: Finite impulse response filtration versus Savitzky–Golay smoothing. *Journal of Raman Spectroscopy*, 38 (9), 1174–1179.

Edelmann, A., y Lendl, B. (2002). Towards the optical tongue: Flow-through sensing of tannin-protein interactions based on FTIR spectroscopy. *Journal of the American Chemical Society, 124* (49), 14741–14747.

Flanzy, Claude. (2003). *Enología: fundamentos científicos y tecnológicos*. (2da. ed.). Madrid: AMV Ediciones.

Kosir, I. J. y Kidric, J. (2002). Use of modern nuclear magnetic resonance spectroscopy in wine analysis: determination of minor compounds. *Analytica Chimica Acta*, 458 (1), 77-84.

Lampreave, M., Valls, J., Nadal, M. y Arola, Ll. (2000). Importancia de los compuestos fenólicos en la calidad de los vinos tintos de crianza. *Alimentación, equipos y tecnología*, 19 (2), 119-124.

Larraín M., Guesalaga A. y Agosín E. (2008). A multipurpose portable instrument for determining ripeness in wine grapes using NIR Spectroscopy. *IEEE Transactions on Instrumentation and Measurement*, 57 (2), 294-302.

Li-Chan, E. C. Y. (1996). The applications of Raman spectroscopy in food science. *Trends in Food Science & Technology*, 7 (11), 361-370.

Mazza, G. y Miniati, E. (1993). *Anthocyanins in fruits, vegetables, and grains*. Boca Raton, Florida: CRC Press.

Meneghini, C., Caron, S., Poulin, A. C. J., Proulx, A., Émond, F., Paradis, P., Paré, P. y Fougères, A. (2008). Determination of ethanol concentration by Raman spectroscopy in liquid-core microstructured optical fiber. *IEEE Sensors Journal*, 8 (7), 1250–1255.

MicroBeam. Espectrómetros y mini-espectrómetros Raman. Manual de referencia, Barcelona.

Mikhailyuk, I. K., Razzhivin, A. P. (2003). Background subtraction in experimental data arrays illustrated by the example of Raman spectra and fluorescent gel electrophoresis patterns. *Instruments and Experimental Techniques*, 46 (6), 765–769.

Nordon, A., Mills, A., Burn, R. T., Cusick, F. M. y Littlejohn, D. (2005). Comparison of non-invasive NIR and Raman spectrometries for determination of alcohol content of spirits. *Analytica Chimica Acta*, 548 (1), 148-158.

Nose, A. y Hojo, M. (2006). Hydrogen bonding of water-ethanol in alcoholic beverages. *Journal of Bioscience and Bioengineering*, 102 (4), 269-280.

Nose, A., Myojin, M., Hojo, M., Ueda, T. y Okuda, T. (2005). Proton nuclear magnetic resonance and Raman spectroscopic studies of japanese sake, an alcoholic beverage. *Journal of Bioscience and Bioengineering*, 99 (5), 493-501.

Pelletier, M. J. (1999). Analytical applications of Raman spectroscopy. Oxford: Blackwell Science.

Peña, A. (2003). Composición fenólica de uvas y vinos. Aspectos generales. Grupo de Investigación Enológica, Departamento de Agroindustria y Enología, Facultad de Ciencias Agronómicas, Universidad de Chile.

Pérez, J. (2008). Development of an optical fiber probe for mercury detection. Tesis doctoral no publicada, Universitat Rovira i Virgili, Tarragona.

- Picque, D., Cattenoz, T., y Corrieu, G. (2001). Classification of red wines analysed by middle infrared spectroscopy of dry extract according to their geographical origin, *Journal international des sciences de la vigne et du vin, 35* (3), 165–170.
- Porcel, M. (2001). Aplicación de técnicas quimiométricas para el desarrollo de nuevos métodos cinético-espectrofotométricos de análisis. Tesis doctoral no publicada, Universitat Autónoma de Barcelona, Bellaterra.
- Puissant, A., Leon, H. (1967), La matière colorante des grains de raisins de certains cépages cultiviés en Anjou 1965. *Annales de Technologie Agricole, 16* (3), 217-225.
- Rivas-Gonzalo, J. C., Gutierrez, Y., Hebrero, E. y Santos-Buelga, C. (1992). Comparisons of methods for the determination of anthocyanins in red wines. *American Journal of Enology and Viticulture*, 43 (2), 210-214.
- Sarneckis, C. J., Dambergs, R. G., Jones P., Mercurio, M., Herderichm M. J. y Smith, P. A. (2006). Quantification of condensed tannins by precipitation with methyl cellulose: development and validation of an optimized tool for grape and wine analysis. *Australian Journal of Grape and Wine Research*, 12 (1), 39-49.
- Smith, E. y Dent, G. (2005). *Modern Raman spectroscopy: a practical approach*. Hoboken, NJ: John Wiley and Sons.
- Urtubia, A., Pérez-Correa, J. R., Meurens, M. y Agosin, E. (2004). Monitoring large scale wine fermentations with infrared spectroscopy. *Talanta*, 64 (3), 778-784.
- Vega, J. C. (2004). Generalizaciones de mínimos cuadrados parciales con aplicación en clasificación supervisada. Tesis doctoral no publicada, Universidad de Puerto Rico, Mayagüez.
- Versari, A., Boulton, R. y Thorngate, J. (2004). The prediction of the color components of red wines using FTIR, wine analyses, and the method of partial least squares. En *Red Wine Color, Revealing the Mysteries, American Chemical Society, ACS Symposium Series* 886 (pp. 53-67). Washington, DC.
- Wise, B. y Gallagher, N. (2000). *PLS_Toolbox 4.0 for use with MATLAB*. Wenatchee, WA: Eigenvector Research, Inc.
- Wold, S., Sjöström, M., Eriksson, L. (2001). PLS_regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*, 109-130.