



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

**ESPECTROMETRO RAMAN PORTATIL PARA
ESTIMACION DE MADUREZ FENOLICA EN UVA
VINIFERA**

JUAN ANDRES CHACON MATTEO

Tesis presentada a la Dirección de Investigación y Postgrado
como parte de los requisitos para optar al grado de
Magister en Ciencias de la Ingeniería

Profesor Supervisor:
ANDRES GUESALAGA M.

Santiago de Chile, Mayo 2010

© MMX, JUAN ANDRÉS CHACÓN MATTEO



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

**ESPECTROMETRO RAMAN PORTATIL PARA
ESTIMACION DE MADUREZ FENOLICA EN UVA
VINIFERA**

JUAN ANDRES CHACON MATTEO

Miembros del Comité:

ANDRES GUESALAGA M.

MIGUEL TORRES T.

EDMUNDO BORDEU S.

JOSE MANUEL DEL VALLE L.

Tesis presentada a la Dirección de Investigación y Postgrado
como parte de los requisitos para optar al grado de
Magister en Ciencias de la Ingeniería

Santiago de Chile, Mayo 2010

© MMX, JUAN ANDRÉS CHACÓN MATTEO

A mi familia

AGRADECIMIENTOS

Agradezco a todos los que me brindaron su apoyo durante el desarrollo de la investigación y redacción de tesis, especialmente a mi familia y amigos, por su constante aliento. En general agradezco a todos los que de alguna forma ayudaron a sacar adelante este proyecto, profesores, personal del laboratorio y compañeros. En particular quisiera agradecer al profesor Andrés Guesalaga por su motivación y consejos.

INDICE GENERAL

AGRADECIMIENTOS	iv
INDICE DE FIGURAS	vii
INDICE DE TABLAS	ix
RESUMEN	x
ABSTRACT	xi
1. INTRODUCCION	1
1.1. Descripción del problema	2
1.2. Objetivos	2
1.3. Trabajo realizado	3
2. MARCO TEORICO	4
2.1. Compuestos Fenólicos	4
2.2. Espectroscopía Raman	6
2.2.1. Efecto Raman	6
2.2.2. Teoría básica	6
2.2.3. Aspectos prácticos de la aplicación	9
2.3. Calibración	10
2.3.1. Métodos de regresión multivariable	12
3. MATERIALES Y METODOS	15
3.1. Hardware	15
3.1.1. Espectrómetro	15
3.1.2. Diseño óptico	17
3.2. Software	17
3.3. Muestras	19
3.4. Experimento	19

3.4.1.	Muestreo Raman	20
3.4.2.	Valores de referencia	22
3.5.	Procesamiento de espectros	26
3.5.1.	Resta de espectro de oscuridad	26
3.5.2.	Filtros de ruido	26
3.5.3.	Corrección espectral	28
3.5.4.	Eliminación de fluorescencia (<i>background</i>)	29
3.5.5.	Acondicionamiento de los datos	31
3.6.	Calibración	33
3.6.1.	Eliminación de muestras	33
3.6.2.	Selección conjuntos de calibración y validación	34
3.6.3.	Interpretación	35
4.	RESULTADOS Y DISCUSION	37
4.1.	Modelos de calibración	37
4.2.	Efecto conjuntos de calibración y validación	38
4.3.	Interpretación del modelo	40
4.3.1.	Calibración con segmentación	43
4.3.2.	Calibración usando fluorescencia	46
4.4.	Efecto del tiempo de integración	47
4.5.	Efecto de la estabilidad de los espectros	47
4.6.	Efecto de las posiciones de muestreo	48
4.7.	Modelos finales por analito	49
4.8.	Fuentes de error	50
5.	CONCLUSIONES Y TRABAJO FUTURO	53
	BIBLIOGRAFIA	56

INDICE DE FIGURAS

2.1 Esquema compuestos fenólicos	4
2.2 Estructura base de compuestos fenólicos	5
2.3 Representación energética interacción luz - materia	8
2.4 Espectro típico en espectroscopía Raman	10
3.1 Espectrómetro Raman	16
3.2 Retenedor de muestras	18
3.3 Posiciones de muestreo	20
3.4 Cambio intensidad media espectro de baya	21
3.5 Espectro de una muestra	22
3.6 Resta del espectro de oscuridad (<i>dark</i> o <i>bias</i>)	26
3.7 Efecto del filtro sobre el espectro de una muestra	27
3.8 Efecto de correcciones sobre el conjunto de espectros	30
3.9 Ejemplo segmentaciones de fluorescencia	32
3.10 Porcentaje recuperación por muestra	34
3.11 Conjuntos calibración y validación	35
4.1 Vector <i>loadings</i> (p) de LV 1.	41
4.2 Vectores <i>weights</i> (w) de LV 2 y LV 4.	42
4.3 Vector b y VIP.	43
4.4 Espectro Segmentado	44
4.5 Espectro <i>background</i> (polinomio) y espectro completo.	46
4.6 Gráfico de dispersión para antocianinas	50
4.7 Gráfico de dispersión para polifenoles	50

4.8 Gráfico de dispersión para taninos	51
--	----

INDICE DE TABLAS

3.1 Especificaciones de espectrómetro Raman.	15
3.2 Valores de referencia obtenidos en laboratorio.	25
4.1 Resultados por analito.	38
4.2 Comparación de resultados para modelo de antocianas con y sin eliminación de muestras.	39
4.3 Varianza de X e y descrita por cada variable latente.	40
4.4 Resultados promedio y máximo según tipo de espectro utilizado en modelos de antocianas.	45
4.5 Resultados promedio y máximo según tipo de espectro utilizado en modelos de polifenoles.	45
4.6 Resultados promedio y máximo según tipo de espectro utilizado en modelos de taninos.	45
4.7 Resultados de calibración según tiempo de integración.	47
4.8 Resultados de calibración según estabilidad de espectros.	48
4.9 Resultados de calibración según posiciones de muestreo.	49
4.10 Resultados finales por analito.	49

RESUMEN

En este trabajo se estudia el uso de un espectrómetro Raman portátil para estimar no destructivamente los parámetros de madurez fenólica en uva vinífera. Previo a la calibración, un estudio de los espectros de las muestras permitió identificar sus características, destacando una fuerte componente de fluorescencia y un fenómeno dinámico en la intensidad de los espectros. Se diseñó un experimento para obtener los espectros minimizando fuentes de error: interferencia de luz externa y posición de la muestra.

Se calibró el instrumento para la medición de antocianinas, polifenoles y taninos. En la generación de modelos se usó la técnica PLS y se evaluó en términos de los índices $RMSE_v$ y R_v^2 , obteniéndose valores que validan la técnica para antocianinas totales, indicador de color. Para polifenoles y taninos el resultado fue inferior.

Se analizó la procedencia de la información contenida en el espectro que correlaciona con el analito, obteniéndose evidencia que sugiere que la información proviene del efecto Raman, pero no se descarta que también haya un aporte o perturbación de la fluorescencia debido a eliminación incompleta de este fenómeno.

El resultado más importante es que se demostró que el instrumento puede ser utilizado para estimar concentración de antocianinas en la piel de la uva.

Sin embargo, también se identificaron las limitaciones del instrumento, como son el largo tiempo de muestreo, la falta de estabilidad de los espectros y poca uniformidad de las muestras. Se presenta un análisis de cada uno de estos factores y cómo estos pueden perjudicar la precisión de las mediciones.

Finalmente se proponen acciones que deberían ser consideradas en futuras investigaciones.

Palabras Claves: espectroscopía Raman, compuestos fenólicos, instrumentos portátiles y no invasivos, uva vinífera.

ABSTRACT

This work analyzes the use of a portable and non-destructive Raman spectrometer for the determination of ripeness phenolic parameters in wine grapes. Before the calibration of estimation models, a study of the spectra of samples was carried out to determine their characteristics, such as a strong fluorescence effect and a dynamical phenomenon associated to the probing process. An experimental setup was designed in order to minimize the error sources, i.e. external light disturbances and sample position.

The instrument was calibrated for simultaneous estimation of anthocyanins, polyphenols and tannins. Prediction models were calibrated using the PLS technique and a minimization approach based on the $RMSE_v$ and R_v^2 indices, showing results that validate its use for total anthocyanin content, i.e. color prediction. In the case of polyphenols and tannins, the performance achieved was worse.

The spectra were scrutinized to find those components that correlate with each of the compounds. It was found that most of the information comes from the Raman spectrum, however, the contribution of the fluorescence components to the final result was not negligible, even considering that most of it has been eliminated in the spectrum pre-processing.

The main result of this investigation is the demonstration that the method is effective when trying to determine the anthocyanin content in the grape skin.

Another important result was the identification of limitations in the use of the technique, such as integration time, spectral stability and sample uniformity. An analysis of each of these factors is carried out, showing their effect on the prediction accuracy.

Finally, some suggestions for further work are indicated.

Keywords: Raman spectroscopy, phenolics compounds, portable and non-destructive instruments, wine grapes.

1. INTRODUCCION

La calidad del vino tinto es una de las principales bases de la competitividad de la industria vitivinícola chilena. Esta calidad está dada en parte muy importante por la composición fenólica del vino, a través de su contribución a la astringencia, el cuerpo y otras sensaciones en boca. Al respecto, la calidad de la uva es esencial. El proyecto n° 5 del consorcio empresarial Vinnova S.A. “Calidad del Vino: Color, Astringencia y Estabilidad”, financiado por InnovaCHILE (CORFO) y Vinos de Chile A.G., tiene como objetivo optimizar la composición fenólica de la uva. La metodología del proyecto enfatiza el desarrollo de técnicas instrumentales, función en la que se enmarca esta tesis.

Las técnicas espectroscópicas son utilizadas ampliamente en la determinación de parámetros de madurez de la uva vinífera. En laboratorio generalmente se miden absorbancias a distintas longitudes de onda para determinar índices de madurez fenólica (Bordeu & Scarpa, 2008; Sarneckis et al., 2006). En los últimos años la tendencia ha sido a desarrollar instrumentos portátiles, que aplicando técnicas de medición no invasiva permiten medir parámetro de madurez en terreno y sin destruir la muestra, permitiendo estudiar la evolución en el tiempo de las variables de interés (Herrera et al., 2003; Larrain et al., 2008). En estos trabajos, la estimación del contenido de azúcar y pH han resultado ser alentadores, sin embargo, los parámetros de madurez fenólica presentan resultados comparativamente inferiores (Larrain et al., 2008).

La espectroscopía Raman no ha sido evaluada para la estimación de compuestos fenólicos en la industria vitivinícola, sin embargo su gran éxito en otras áreas como farmacéutica (Henson & Zhang, 2006), obras de arte (Clark, 2007), biomédica (Beger et al., 1999), mineralogía (Wang et al., 2003) entre otras, la presenta como una excelente posibilidad para complementar las herramientas analíticas en enología. Al ser una técnica espectroscópica, puede ser utilizada en forma no invasiva y dados los adelantos tecnológicos en el área de láseres y detectores, ahora también puede ser aplicada en forma portátil.

1.1. Descripción del problema

Se necesitan nuevas y mejores herramientas analíticas para el estudio y optimización de la composición fenólica de la uva. Se han desarrollado herramientas para el uso en terreno, como es el caso de la espectroscopía NIR, no obstante, se necesita mejorar los resultados para los indicadores de madurez fenólica. Por ello es necesario evaluar otras técnicas para la predicción de estos analitos y la espectroscopía Raman, al ser otra forma de espectroscopía muchas veces complementaria a la NIR, es un candidato para solucionar este problema, por lo que se requiere evaluarla.

1.2. Objetivos

El objetivo general es la prueba conceptual y validación de un instrumento de terreno para medir concentración de compuestos fenólicos en la uva en periodo de cosecha. Los objetivos específicos son los siguientes:

- (i) Diseñar el muestreo de bayas para la calibración del instrumento, el cual contempla el diseño de un protocolo de muestreo Raman de las bayas y la utilización de un método estándar para obtener las referencias a nivel de laboratorio.
- (ii) Desarrollar modelos de calibración para polifenoles totales, antocianinas y taninos. En esta etapa también se evalúan los procesamientos de señales a aplicar a los espectros.
- (iii) Validar los modelos de estimación.
- (iv) Determinar la influencia de factores como tiempo de integración, estabilidad de los espectros y homogeneidad de la baya..
- (v) Identificar y analizar de las posibles fuentes de error.

1.3. Trabajo realizado

Lo primero fue estudiar el uso de la espectroscopía Raman en forma práctica, es decir, identificar los factores que es necesario tener en cuenta al momento de tomar los espectros en este tipo de aplicación.

Se planificó el muestreo de las bayas, para lo cual se diseñó un retenedor de muestras que controla las condiciones de muestreo, como posición de la muestra e iluminación externa. Además se desarrolló un programa para agilizar el proceso de muestreo.

Se obtuvieron las referencias a partir de métodos utilizados en el laboratorio de enología.

Se calibraron y validaron modelos para polifenoles, antocianinas y taninos, realizándose numerosas pruebas para determinar los procesamientos que permitieran obtener los mejores modelos.

Se analizaron los modelos y se buscó distintas formas de interpretación para descubrir si la información provenía del efecto Raman o de otro efecto presente en los espectros.

Se estudió el efecto de factores particulares del espectro de las bayas que pueden perjudicar el uso del instrumento en terreno, como el tiempo de muestreo y la estabilidad de los espectros.

2. MARCO TEORICO

2.1. Compuestos Fenólicos

Los compuestos fenólicos tienen una gran importancia en enología debido al papel que desempeñan en la calidad del vino, ya que son el origen del color y de la astringencia. (Cheynier et al., 2003). Su comportamiento químico en la producción del vino es muy sensible a muchos factores por lo que están constantemente sufriendo cambios químicos (Margalit, 2004).

Químicamente, los compuestos fenólicos tienen un núcleo benzénico que lleva uno o varios grupos hidroxilos ,y se clasifican en flavonoides y no-flavonoides. En la uva, los compuestos no-flavonoides se encuentran presentes principalmente en la pulpa, y los flavonoides en los hollejos, semillas y raspones. Para entender mejor la clasificación de estos compuestos se presenta el esquema de la Figura 2.1.

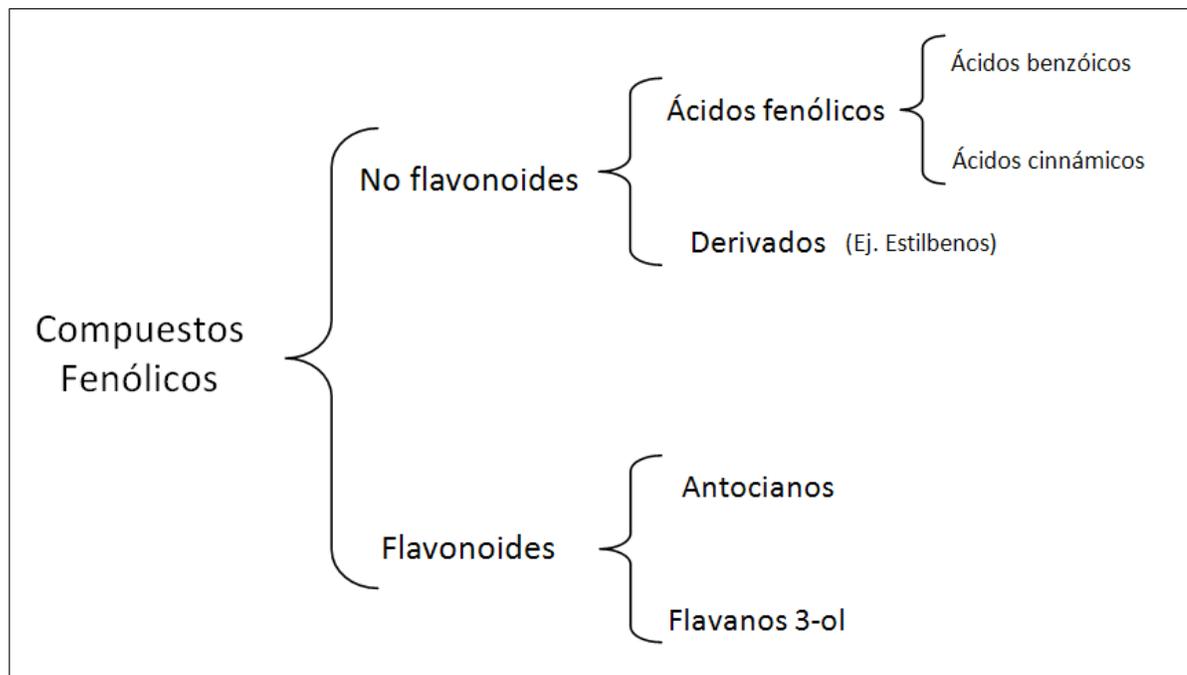


FIGURA 2.1. Esquema compuestos fenólicos.

Los compuestos no flavonoides abarcan a los ácidos fenólicos, divididos en ácidos benzóicos (C6-C1) y ácidos cinnámicos, portadores de una cadena lateral insaturada (C6-C3), pero también a otros derivados como los estilbenos. En la uva los ácidos fenólicos son principalmente ácidos hidroxicinnámicos (Figura 2.2 (a)) que se encuentran en las vacuolas de las células del hollejo y de la pulpa, y participan en la acilación de los sustitutos glucósidos de los antocianos. Los estilbenos se encuentran en el hollejo, pero en menor concentración que los ácidos fenólicos.

Los compuestos flavonoides están caracterizados por un esqueleto de base de 15 átomos de carbono (C6-C3-C6) (Figura 2.2 (b)). Los antocianos (del griego anthos flor y kyanos azul) se ubican en el hollejo y en las 3 ó 4 primeras capas celulares del hipodermo, y contribuyen de manera preponderante al color de las especies tintas. El contenido y composición de antocianos en la uva varían enormemente en función de la especie y variedad.

Los flavanos 3-oles (3-flavonoles) conocidos como taninos en su forma polimerizada, se encuentran localizados principalmente en las semillas. También se encuentran en el hollejo en capas externas del hipodermo. Los principales 3-flavonoles monómeros de la uva son la catequina y epicatequina.

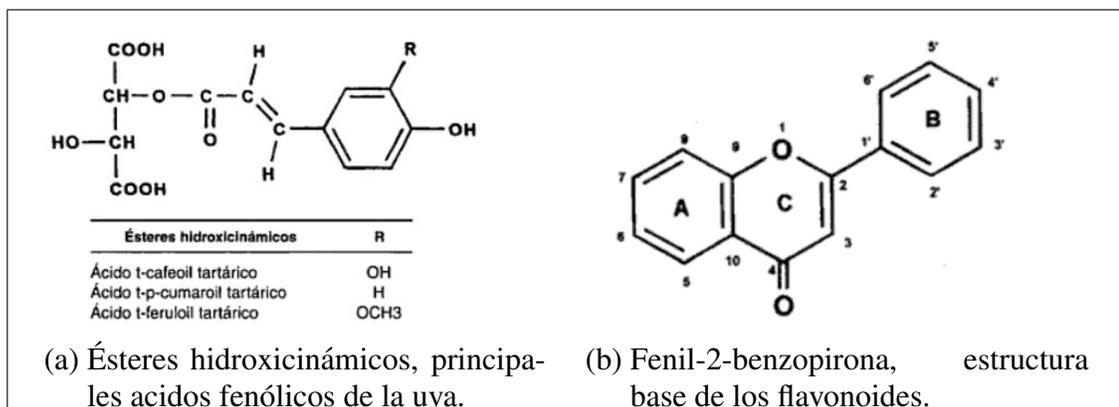


FIGURA 2.2. Estructuras base de compuestos fenólicos.

2.2. Espectroscopía Raman

La espectroscopía Raman es una técnica que permite estudiar la estructura química de una molécula, así como también permite obtener información cuantitativa de ésta en una muestra. Para comprender esta técnica es necesario estudiar el efecto Raman desde el punto de vista teórico, entender la interacción que se produce entre la luz y la materia, y para utilizar esta técnica en forma práctica, es necesario relacionar la teoría con los espectros observados. En esta sección se presentan los fundamentos de la espectroscopía Raman.

2.2.1. Efecto Raman

El efecto Raman es una dispersión inelástica, que ocurre cuando un fotón interactúa con una molécula se genera un fotón de mayor longitud de onda (menos energético) y la molécula absorbe la diferencia de energía. Raman y Krishnan (1928) fueron los primeros en observar experimentalmente este fenómeno. En su experimento observaron que la luz dispersada por una muestra tenía una frecuencia distinta a la luz que incidía sobre ella.

2.2.2. Teoría básica

Si se hace incidir un fotón sobre una molécula pueden ocurrir tres cosas, el fotón: es transmitido a través de la muestra, sin interactuar con ella; es absorbido; o dispersado.

La absorción ocurre cuando la energía del fotón incidente es igual a la diferencia de energía entre 2 estados de la molécula. Al ser absorbido, el fotón pasa su energía a la molécula, haciéndola cambiar a un estado de energía más alto. La espectroscopía de absorción se basa en medir estos cambios de energía.

La dispersión ocurre cuando el fotón interactúa con la molécula, pero no es absorbido, sino que es “desviado”. Para que el fotón sea dispersado no se requiere que tenga una energía que coincida con la diferencia de energía entre 2 niveles de la molécula. El efecto Raman es un tipo de dispersión.

En la espectroscopía de absorción la muestra es iluminada con luz que cubre un rango de longitudes de onda, y se mide el patrón de absorción realizado por la molécula. En

espectroscopía Raman la muestra se excita con una única longitud de onda, y se estudia la diferencia de longitud de onda que presenta la luz dispersada por la muestra con respecto a la de la luz incidente.

Para entender el efecto Raman y otros fenómenos relacionados, es útil usar una representación gráfica como la que se muestra en la Figura 2.3. Las líneas horizontales continuas representan estados energéticos de la molécula, las líneas horizontales segmentadas representan estados virtuales. Las flechas verticales representan saltos en la energía de la molécula. De izquierda a derecha los fenómenos identificados son:

Absorción: Cuando la energía del fotón incidente coincide con la diferencia entre 2 estados energéticos de la molécula, el fotón es absorbido.

Dispersión Rayleigh: El fotón incidente entrega su energía a la molécula haciéndola pasar a un estado virtual (poco estable). La energía es rápidamente devuelta liberando un fotón de la misma energía (longitud de onda) que el incidente. Este proceso también se conoce como dispersión elástica y es el fenómeno predominante.

Raman Stokes: El fotón incidente entrega su energía a la molécula llevándola a un estado virtual, pero a diferencia de la dispersión Rayleigh, la molécula cae a un estado de mayor energía que el inicial, y el fotón liberado tiene una energía menor (mayor longitud de onda) que la energía del fotón incidente. La diferencia de energía entre el fotón de entrada y salida es la energía que ganó la molécula.

Raman Anti-Stokes: El fotón incidente entrega su energía a la molécula que se encuentra en un estado energético mayor al nivel de base. La molécula pasa a un estado virtual del cual cae al estado base, entregando energía al fotón. El fotón liberado tiene mayor energía (menor longitud de onda) que el incidente. La de energía que perdió la molécula, es la diferencia de energía entre el fotón de entrada y de salida.

Fluorescencia: Es similar al efecto Raman Stokes, a diferencia de que la luz incidente es completamente absorbida y la molécula es transferida a un estado excitado del cual baja a varios estados de menor energía sólo después de un cierto tiempo de resonancia. El resultado es que se produce un fotón de distinta energía (longitud de onda) que el incidente.

Otra diferencia con el efecto Raman es que este último puede observarse para cualquier longitud de onda del fotón incidente, no así la fluorescencia.

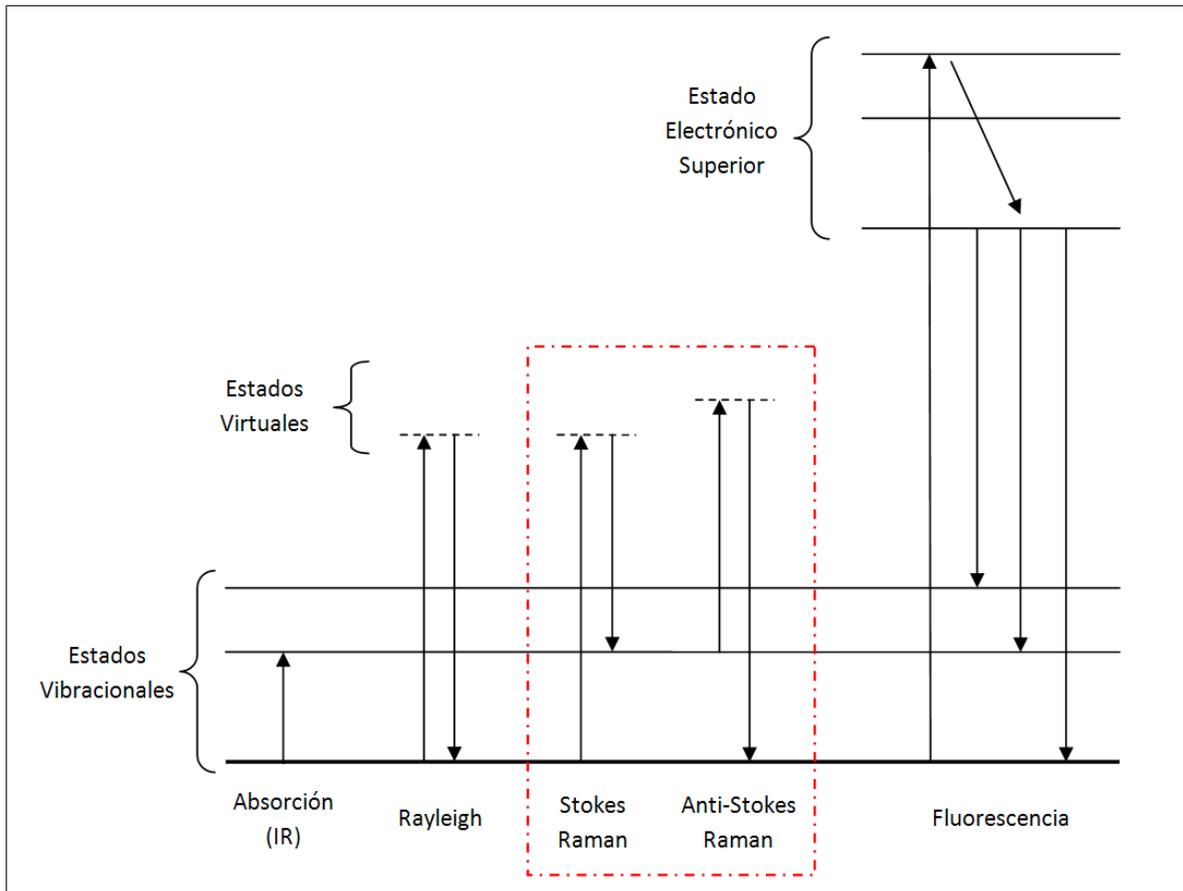


FIGURA 2.3. Representación energética de la interacción luz - materia.

La dispersión Raman es una fracción minúscula de la dispersión Rayleigh, ya que sólo 1 de entre 10^6 a 10^8 fotones es dispersado por el efecto Raman. Para solucionar esto, la muestra se excita con un laser de alta densidad de potencia, pero se producen otros efectos indeseados, como la fluorescencia o la degradación de la muestra. La diferencia de intensidad entre el fenómeno Raman Stokes y Anti-Stokes depende de cuantas moléculas se encuentren inicialmente en un estado base, o en un estado vibracional de mayor energía. A temperatura ambiente hay mayor cantidad de moléculas en estado base, por lo tanto el efecto Raman Stokes es más intenso. La intensidad de la fluorescencia dependerá de la

longitud de onda del láser y podrá ser despreciable, o tan intensa que opaque por completo el efecto Raman.

2.2.3. Aspectos prácticos de la aplicación

Al excitar una muestra con un láser y leer la luz dispersada por ésta mediante un espectrómetro, se obtiene la intensidad de la luz para cada longitud de onda dispersada. En espectroscopía Raman las longitudes de onda se representan con el *Raman Shift*, que es el desplazamiento respecto a la longitud de onda de excitación λ_c , y se calcula según la ecuación 2.1.

$$Raman\ Shift = 1/\lambda_c - 1/\lambda \quad [cm^{-1}] \quad (2.1)$$

En la Figura 2.4 se muestra un gráfico representativo de lo que se debería leer en un espectrómetro Raman. Se pueden distinguir claramente los tipos de dispersión mencionados en la sección anterior:

Dispersión Rayleigh: Se observa como un pico a 0 cm^{-1} , (longitud de onda del láser de excitación). Dado que es más intenso por varios órdenes de magnitud que el efecto Raman, generalmente es filtrado, pero se puede observar un remanente.

Dispersión Raman Stokes: Se observan como picos ubicados en desplazamientos Raman positivos, ya que son de mayor longitud de onda que la del láser de excitación.

Dispersión Raman Anti-Stokes: Se observan como picos ubicados en desplazamientos Raman negativos, ya que son de menor longitud de onda que la del láser de excitación. Generalmente son de menor intensidad y en algunos espectrómetros no alcanzan a ser detectados.

Fluorescencia: Se observa como una componente suave en el sector de desplazamientos Raman positivos, ya que es un fenómeno de pérdida de energía (los fotones de salida son de mayor energía que los de entrada). Sobre ella pueden ir montados los picos Raman.

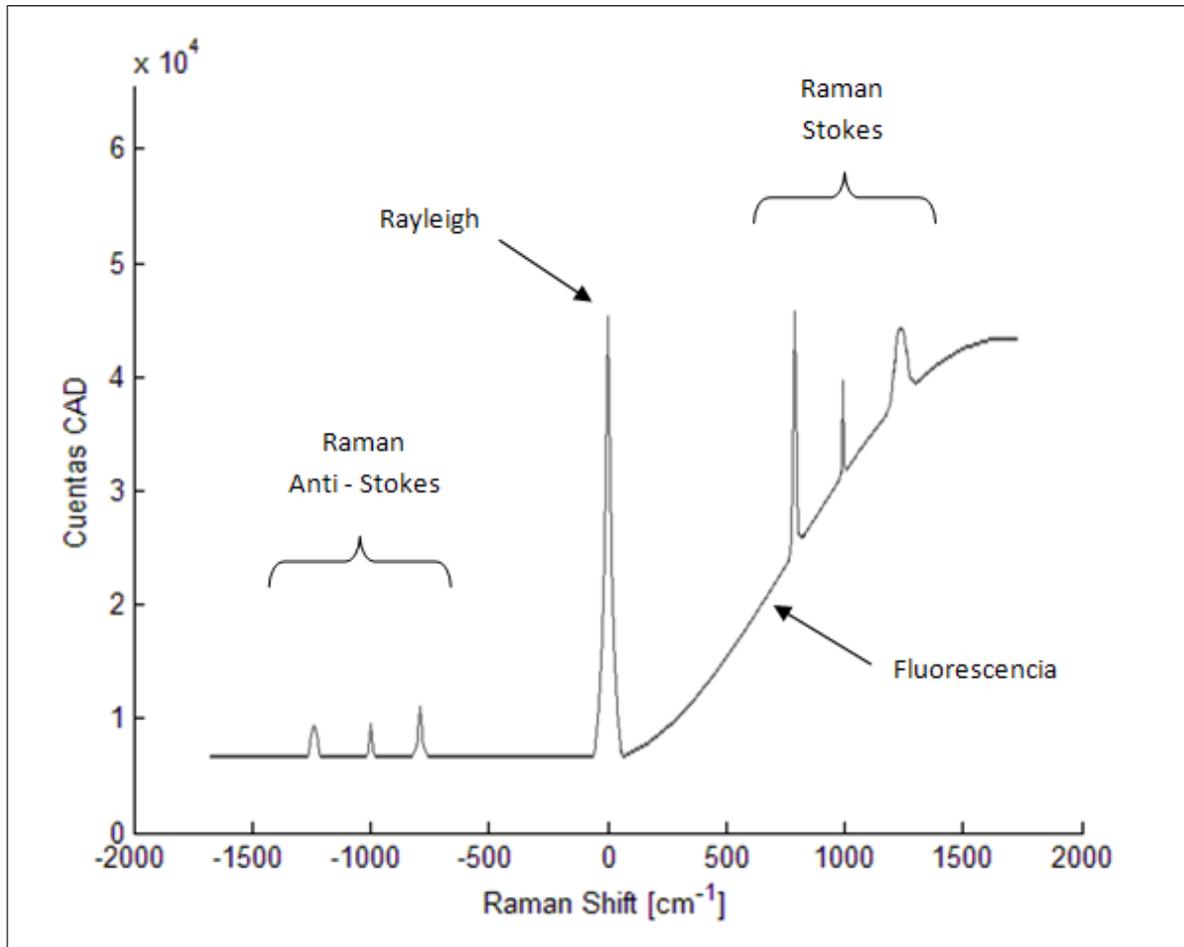


FIGURA 2.4. Espectro típico obtenido en un espectrómetro Raman (distintas dispersiones).

Además de los fenómenos de dispersión, también se detecta ruido proveniente de diferentes fuentes: ruido intrínseco de los fotones, ruido generado por la muestra (poca homogeneidad), ruido generado por el ambiente (cambios de temperatura), ruido generado por la instrumentación (térmico y de lectura), fuentes externas (luz solar, artificial o rayos cósmicos).

2.3. Calibración

Las técnicas de medición basadas en espectroscopía, utilizan la información contenida en el espectro de la muestra para estimar el valor de una variable de interés, por ejemplo, la

concentración de una molécula. En el caso de espectroscopía Raman, el espectro contendrá la cantidad de luz que dispersó la muestra en cada longitud de onda observada.

Desde el punto de vista de la instrumentación, cada longitud de onda es una variable predictora y se requiere un modelo que relacione estas variables para generar una estimación de la variable desconocida. El problema se describe matemáticamente según la ecuación 2.2, donde: $\mathbf{x}^T \in \mathbb{R}^m$ es el espectro de la muestra (vector fila), $\mathbf{b} \in \mathbb{R}^m$ es el modelo o vector regresión (vector columna), y es la concentración de la muestra (escalar), y m es el número de variables de entrada (longitudes de onda observadas).

$$\mathbf{x}^T \cdot \mathbf{b} = y \quad (2.2)$$

En quimiometría, el vector b es encontrado en forma empírica, analizando datos a través del proceso de calibración. El proceso consiste en hacer mediciones de los espectros de un conjunto de muestras y al mismo tiempo medir la concentración de la molécula de interés en cada muestra mediante técnicas analíticas de laboratorio que deben poseer una exactitud alta. Estos valores deben ser obtenidos a partir de un método conocido y se usarán como referencia. El problema se describe matricialmente según la ecuación 2.3 con \mathbf{X} e \mathbf{y} formados según la ecuación 2.4.

$$\mathbf{X} \cdot \mathbf{b} = \mathbf{y} \quad (2.3)$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad (2.4)$$

Donde i indica el número de la muestra y n el número total de muestras. El vector b es tal que al hacer producto punto entre él y el espectro de la muestra i se obtiene su concentración y_i . En la práctica lo que se obtiene es \hat{y}_i , una estimación de la concentración.

Conceptualmente el problema consiste en invertir la matriz \mathbf{X} , para despejar la incógnita b . Sin embargo, en la práctica la matriz \mathbf{X} no es cuadrada, ya que se tendrá menor cantidad de muestras n , que número de variables m (longitudes de onda). Aun cuando el número de muestras pudiera ser igual al número de variables, existe el problema de que cada fila de \mathbf{X} es el espectro de una muestra, y son muy similares entre si, por lo que el rango de la \mathbf{X} generalmente es menor a n . De forma similar, dado que las columnas de \mathbf{X} son longitudes de onda, sus valores adyacentes son parecidos, existiendo colinealidad entre las columnas.

Para resolver el problema existen varias técnicas (Beebe & Kowalski, 1987). Sólo se presentará aquellas que han logrado una mayor aceptación en la comunidad científica.

2.3.1. Métodos de regresión multivariable

A continuación se describen los métodos de regresión MLR (Multiple Linear Regression), PCR (Multiple Linear Regression) y PLS (Partial Least Squares). MLR es el método de regresión multivariable más simple y es la base de los otros 2. PCR y PLS son métodos basados en la descomposición de la matriz \mathbf{X} en factores, también conocidos como Componentes Principales en el caso de PCR o Variables Latentes en el caso de PLS.

2.3.1.1. MLR

La solución al problema de calibración (ecuación 2.3) que propone el método MLR, consiste en despejar el vector \mathbf{b} como lo muestra la ecuación 2.5, donde \mathbf{X}^+ es la pseudo-inversa que se calcula según la ecuación 2.6.

$$\mathbf{b} = \mathbf{X}^+ \cdot \mathbf{y} \quad (2.5)$$

$$\mathbf{X}^+ = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \quad (2.6)$$

La solución minimiza el error entre la concentración estimada \hat{y} y el valor real de la concentración y , para las n muestras. En la práctica, el error experimental y el ruido en las mediciones pueden causar que la predicción de nuevas muestras tenga un error mucho mayor. Para solucionar este problema se usa el método PCR.

2.3.1.2. PCR

PCR se basa en un método de descomposición llamado PCA (Principal Component Analysis) (Jackson, 1980). PCA asume que la información contenida en el espectro no está en el valor individual de una longitud de onda, sino que está en las relaciones entre ellas. Esta técnica encuentra combinaciones de variables (longitudes de onda) o factores, que describen la tendencia de los datos. El resultado de esta descomposición se expresa como la suma de productos exteriores entre los vectores \mathbf{t}_i y \mathbf{p}_i más una matriz de residuo \mathbf{E} .

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + \dots + \mathbf{t}_k\mathbf{p}_k^T + \mathbf{E} \quad (2.7)$$

Los vectores \mathbf{t}_i se llaman *scores* y contienen información de cómo las muestras se relacionan entre ellas. Los vectores \mathbf{p}_i se llaman *loadings* y contienen información de cómo las variables se relacionan unas con otras. Los pares $\mathbf{t}_i\mathbf{p}_i^T$ están ordenados en forma decreciente según la cantidad de varianza (información) que describen de la matriz \mathbf{X} . Los vectores *scores* y *loadings* son ortogonales y los últimos están normalizados.

El método PCR soluciona el problema del mal condicionamiento de la matriz \mathbf{X} haciendo una regresión de las concentraciones \mathbf{y} en los *scores* de las componentes principales de las variables originales, en lugar de hacerlo directamente sobre éstas. El número de componentes principales usadas para generar el modelo se puede optimizar para minimizar el error de predicción de nuevas muestras. A medida que se agregan componentes principales el resultado va convergiendo hacia el entregado por MLR.

2.3.1.3. PLS

La regresión PLS (Wold et al., 1984) se puede pensar como un compromiso entre MLR y PCR. PCR busca los factores que capturan la mayor cantidad de varianza en los espectros (\mathbf{X}), en cambio MLR busca la relación (un solo factor) que tenga la mejor correlación con la concentración (\mathbf{y}). El método PLS elige los factores que correlacionan con \mathbf{y} mientras describen gran cantidad de varianza de los espectros.

PLS se basa en el análisis de la matriz $\mathbf{X}^T \mathbf{y}$, y los *scores* y *loadings* generados son distintos a los obtenidos con PCR, los cuales se basan en el estudio de la matriz $\mathbf{X}^T \mathbf{X}$. Al igual que en PCR, la cantidad de factores utilizados para describir el modelo se pueden elegir de manera de maximizar la capacidad predictiva del modelo.

La descomposición de \mathbf{X} se basa en tres tipos de vectores: los *scores* \mathbf{t} , los *weights* \mathbf{w} y los *loadings* \mathbf{p} , los cuales pueden ser utilizados en la interpretación espectroscópica del modelo 3.6.3. Existen distintos algoritmos para calcular la regresión PLS (Wise et al., 2006), a continuación se presentan los principales pasos del algoritmo NIPALS (*Nonlinear Iterative Partial Least Squares*):

- (i) $\mathbf{w}_i = \frac{\mathbf{X}_{i-1}^T \mathbf{y}}{\|\mathbf{X}_{i-1}^T \mathbf{y}\|}$, el primer *weights* \mathbf{w} es \mathbf{y} proyectado en \mathbf{X}
- (ii) $\mathbf{t}_i = \frac{\mathbf{X}_{i-1} \mathbf{w}_i}{\|\mathbf{X}_{i-1}^T \mathbf{w}_i\|}$, el primer *score* \mathbf{t} es \mathbf{X} proyectado en \mathbf{w}
- (iii) $\mathbf{p}_i = \mathbf{X}_{i-1}^T \mathbf{t}_i$, el primer *loading* \mathbf{p} es \mathbf{t} proyectado en \mathbf{X}
- (iv) $\mathbf{X}_i = \mathbf{X}_{i-1} - \mathbf{t}_i \mathbf{p}_i^T$, \mathbf{X} se modela como el *score* \mathbf{t} veces el *loading* \mathbf{p}

Estos pasos se repiten en forma iterativa con i variando entre 1 y k , donde k es el número de variables latentes (análogo a las componentes principales de PCR). En la primera iteración $\mathbf{X}_0 = \mathbf{X}$. El vector de regresión se calcula como $\mathbf{b} = \mathbf{W}(\mathbf{P}^T \mathbf{W})^{-1}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \mathbf{y}$.

Para mayor información sobre el algoritmo PLS y algoritmos de cálculo puede revisar el manual del *PLS Toolbox* (Wise et al., 2006), donde además se puede encontrar una extensa bibliografía.

3. MATERIALES Y METODOS

3.1. Hardware

En esta sección se describe brevemente el espectrómetro Raman utilizado en la investigación. También se describe el retenedor de muestras diseñado para medir las bayas.

3.1.1. Espectrómetro

La toma de espectros se realizó con el espectrómetro Raman portable *MiniRam BTR111E* fabricado por *B&W Tek Inc.*. La Tabla 3.1 presenta sus principales especificaciones técnicas.

Este instrumento tiene una resolución y rango espectral reducido si se compara con un instrumento diseñado para uso en laboratorio, que puede llegar a tener 1 cm^{-1} y $50 - 8000 \text{ cm}^{-1}$ respectivamente. Sin embargo, el *MiniRam BTR111E* es un instrumento muy liviano y compacto, especialmente pensado para uso en terreno (*BTR111E MiniRam Datasheet*, 2008).

TABLA 3.1. Especificaciones de espectrómetro Raman.

Espectrómetro	
Rango espectral	$175 - 3150 \text{ cm}^{-1}$ (Raman Shift)
Resolución	10 cm^{-1} FWHM
CAD	16 bits
<i>Dark noise</i>	$< 60 \text{ RMS}$ at 9 ms, 1 lectura
Laser de excitación	$785 \text{ nm} + / - 0.5 \text{ nm}$
Dimensiones	$22.6 \text{ cm} \times 16.3 \text{ cm} \times 8.3 \text{ cm}$
Peso	$< 3 \text{ kg}$
Sonda	
Distancia focal	8.1 mm
Superficie muestreo	$105 \mu\text{m}$

Como se muestra en la Figura 3.1, el espectrómetro trabaja en conjunto con la sonda. La fibra óptica de excitación lleva la luz del láser desde el espectrómetro a la sonda. La sonda enfoca la luz sobre la muestra, donde se produce la dispersión Rayleigh, Raman y otras interacciones. La misma óptica conduce la luz dispersada por la muestra a un filtro que elimina la mayor parte de la dispersión Rayleigh y la luz restante es enviada al espectrómetro a través de la fibra de colección. En el espectrómetro, un elemento dispersor (por ejemplo un prisma o una grilla) descompone la luz en sus longitudes de onda, las cuales son registradas por el CCD.

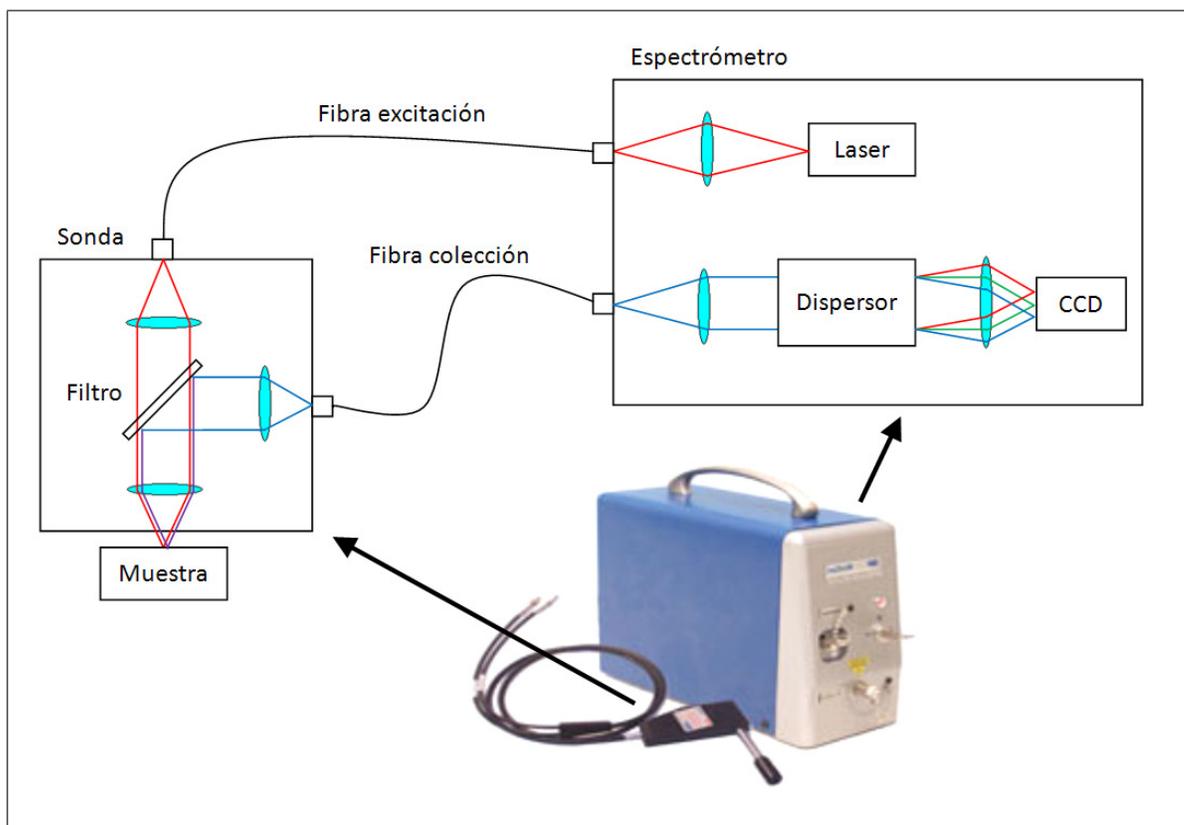


FIGURA 3.1. Espectrómetro Raman.

Es muy importante que la muestra se sitúe a la distancia correcta de la sonda, es decir, en el punto donde el láser es enfocado. Por ejemplo, si se muestrea un líquido, este punto debe ubicarse dentro de la cubeta contenedora. Si la muestra es un sólido, el punto de muestreo debe ubicarse en su superficie (Smith & Dent, 2005).

3.1.2. Diseño óptico

La sonda permite tomar el espectro Raman de una muestra sólida con tan solo apoyarla sobre un accesorio, que asegura que la sonda se ubique a la distancia correcta para que el láser sea enfocado en su superficie. En el caso de líquidos existe otro accesorio que permite muestrearlos dentro de una cubeta de cuarzo o vidrio. Dada la forma esférica, su variado tamaño y la consistencia blanda de las bayas, el accesorio para muestras sólidas no fue efectivo al momento de asegurar las condiciones mecánicas adecuadas para el muestreo: posición adecuada y repetitividad de la posición de la muestra.

Además de las consideraciones mecánicas hay otros factores críticos a tener en cuenta desde el punto de vista óptico. Dada la baja intensidad del efecto Raman, es necesario ocluir la muestra para evitar la interferencia de otras fuentes de luz, como la luz solar o luz artificial. Si se usa un recipiente para contener la muestra, es necesario que este no emita dentro del rango espectral Raman y debe poseer una excelente transmisión en la región observada.

Teniendo en consideración los requerimientos mecánicos y ópticos, se diseñó el retenedor de muestras que se presenta en la Figura 3.2. La baya es presionada suavemente contra una ventana para tener localmente una superficie plana, independizando la medición de la curvatura (tamaño) de la baya. La posición de la ventana es fija, asegurando la repetitividad de la posición para todas las muestras. El material de la ventana es *fused silica* de 1 mm espesor, con un revestimiento antirreflejo para el rango espectral VIS-NIR que reduce la reflexión a menos de 1%. La uva es cubierta por una carcasa metálica para evitar la interferencia de luz externa.

3.2. Software

Para la adquisición de los espectros, se programó una interface en el lenguaje C#, la cual usa como punto de partida las aplicaciones modelo proporcionadas por el fabricante del espectrómetro. El programa permite la conexión y configuración del instrumento, ajuste de tiempo de integración, adquisición de espectros y almacenamiento en formato texto. La

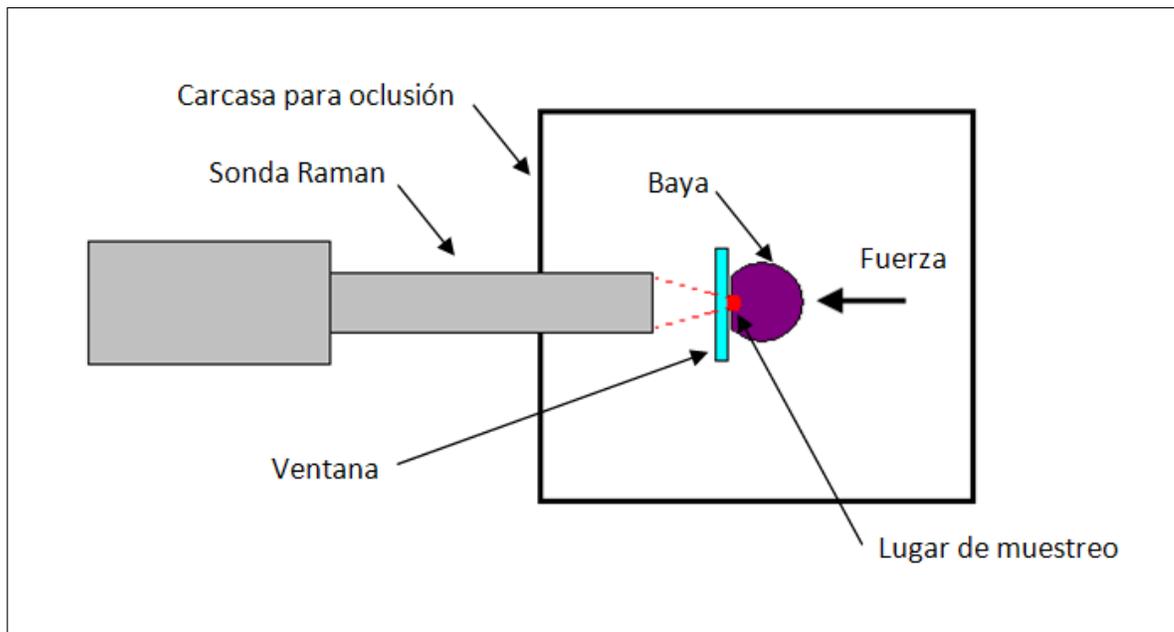


FIGURA 3.2. Retenedor de muestras.

interfaz además permite la comunicación vía puerto serie, protocolo RS232, con un servomotor que abre y cierra el obturador (*shutter*) de la sonda. Al automatizar el obturador es posible tomar secuencias de espectros de oscuridad y espectros completos, además de programar temporizadores para mantener la muestra iluminada durante un tiempo preestablecido, condición necesaria para alcanzar la estabilización de los espectros (ver sección 3.4.1). La optimización del proceso de muestreo fue útil para paralelizar la toma de espectro y el análisis en laboratorio de las muestras, permitiendo que todo el trabajo fuera realizado por el alumno.

Después de adquirir los espectros se utilizó Matlab para extraer la información guardada en los archivos de formato texto. Todos los procesamientos se encuentran programados en rutinas y funciones de Matlab. Para la calibración de modelos se utilizó el set de funciones e interfaz gráfica PLS Toolbox Versión 5.2 desarrollado por la compañía de investigación y desarrollo en quimiometría *Eigenvector Research, Inc.*

3.3. Muestras

Las bayas utilizadas pertenecen a la variedad *Cabernet Sauvignon* y fueron extraídas del fundo la Cornellana de la Viña La Rosa, ubicada en la localidad de Peumo. Todas las bayas utilizadas en el experimento fueron cosechadas el mismo día, 19 de marzo de 2008 y posteriormente fueron congeladas. Las plantas de donde se obtuvieron las muestras fueron utilizadas en ensayos hormonales que buscaban adelantar la madurez fenólica por lo que se espera una diversidad de concentraciones de estos compuestos.

El ensayo consideró 11 plantas, en cada una de las cuales se enumeraron 5 racimos. Debido al volumen mínimo de material (hollejo) necesario para hacer los análisis de laboratorio, se determinó que cada muestra estuviera conformada por 5 bayas. Un número muy bajo de bayas es insuficiente para hacer los análisis y un número muy grande de bayas consumiría mucho tiempo en el muestreo Raman. Las 5 bayas que constituyen una muestra fueron extraídas de la misma parte de un racimo (hombro), con esto se buscó disminuir la variabilidad entre bayas. En los racimos que se conservaron enteros durante la congelación se extrajo además 5 bayas de sus puntas. En total se obtuvo 89 muestras.

3.4. Experimento

En esta sección se describe en detalle los métodos empleados para realizar el experimento. En primer lugar se describe el proceso del muestreo Raman, es decir la obtención de los vectores \mathbf{x}_i^T . Luego se describe el proceso realizado en el laboratorio para la obtención de las referencias y_i .

Como se mencionó en la sección 3.3, existe un compromiso entre número de bayas adecuadas para obtener las referencias y las bayas necesarias para hacer el muestreo Raman. Desde el punto de vista del muestreo Raman, una referencia por baya sería ideal, pero dada las limitaciones del método de laboratorio disponible fue imposible obtener un valor de referencia para una sola baya. Se determinó que 5 es el número mínimo de bayas que permite obtener la cantidad de extracto suficiente para realizar los 3 análisis propuestos. Para cada muestra el proceso consistió en lo siguiente: las 5 bayas que forman una

muestra se descongelaban durante la noche, al día siguiente se secaban con papel y pesaban, luego eran muestreadas con el espectrómetro Raman, finalmente se realizaban los análisis de laboratorio que consistían en la extracción de las moléculas y la medición de concentraciones.

3.4.1. Muestreo Raman

Dado que el muestreo Raman cubre una superficie muy pequeña (≈ 0.1 mm) y asumiendo que la concentración de las moléculas de interés no es completamente uniforme en toda la superficie de la baya, se tomaron espectros en 3 lugares diferentes. Optar por 3 espectros parece arbitrario, sin embargo hay que tener presente que esta técnica tiene como objetivo final ser utilizada en terreno y tomar un gran número de espectros por baya se torna poco práctico. En la Figura 3.3 se muestran las posiciones elegidas para muestrear una baya desde una vista superior (pedicelo apuntando hacia el lector). El espectro de una baya se calcula como el promedio de los espectros de las 3 posiciones. Como una muestra está formada por 5 bayas, el espectro de una muestra se calcula como el promedio de los espectros de las 5 bayas.

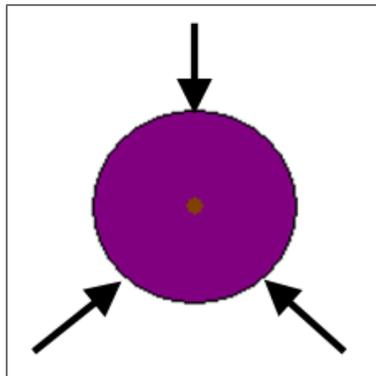


FIGURA 3.3. Posiciones de muestreo.

Al obtener los primeros espectros de bayas se observó que la intensidad del espectro variaba con el tiempo. El problema se puede solucionar iluminando la muestra con el láser hasta que la intensidad del espectro se estabiliza (Smith & Dent, 2005). En el gráfico de la Figura 3.4 se muestra la media del espectro en función del tiempo. El instante $t = 0$

marca la apertura del obturador, observándose que durante el primer minuto el espectro cambia rápidamente y luego comienza a estabilizarse. Para el muestreo final se programó la adquisición de un espectro inmediatamente después de abierto el obturador y otra 1.5 min después. Al obtener el espectro en su estado transiente y estable se pueden calibrar modelos para ambos y determinar cómo influye este factor. Si no influye, el uso del instrumento en terreno será más expedito.

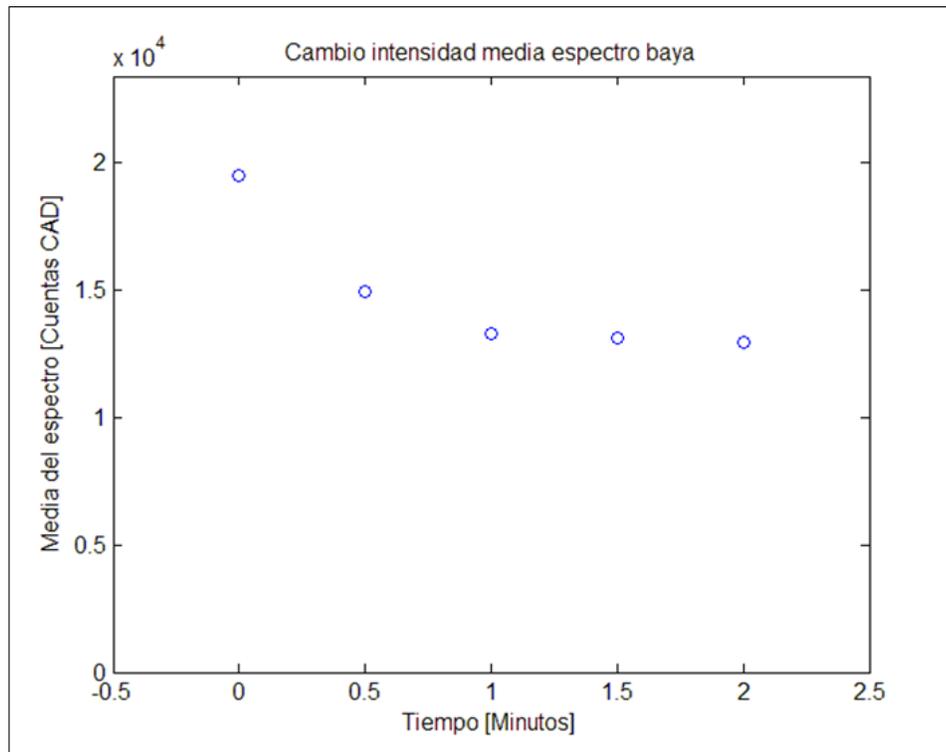


FIGURA 3.4. Cambio de la intensidad media del espectro de una baya.

Por último se determinó el tiempo de integración máximo para que no haya saturación. Este resultó ser de 9 s. Para estudiar el impacto del uso de distintos tiempos de integración se tomaron espectros a 3 y 9 s. Además cada espectro se tomó 3 veces, de manera de poder promediarlos para aumentar la razón señal a ruido. Inicialmente los modelos fueron calibrados para el promedio de 3 espectros de 9 s, y luego se probó como variaba el modelo usando un solo espectro de 9 s y finalmente un espectro de 3 s. Mientras menor es el tiempo de muestreo, más cómodo el uso del instrumento en terreno.

A modo de ejemplo, en la Figura 3.5 se muestra el espectro de una baya. Como se puede observar, destaca una componente importante de lo que parece ser fluorescencia en el rango de 100 a 3000 cm^{-1} . En el rango entre 0 y 100 cm^{-1} hay un cambio abrupto el cual es producido por el filtro que elimina la longitud de onda del láser. En 0 cm^{-1} aparece el remanente de la longitud de onda de excitación. En todo el rango se observa algo de ruido propio de la señal. Finalmente cabe destacar que no se observan picos Raman notorios, sin embargo, en la región comprendida entre 100 y 1500 cm^{-1} se visualiza que sobre la componente suave de fluorescencia (*background*) existen cambios de mayor frecuencia, los que serían provocados por el efecto Raman.

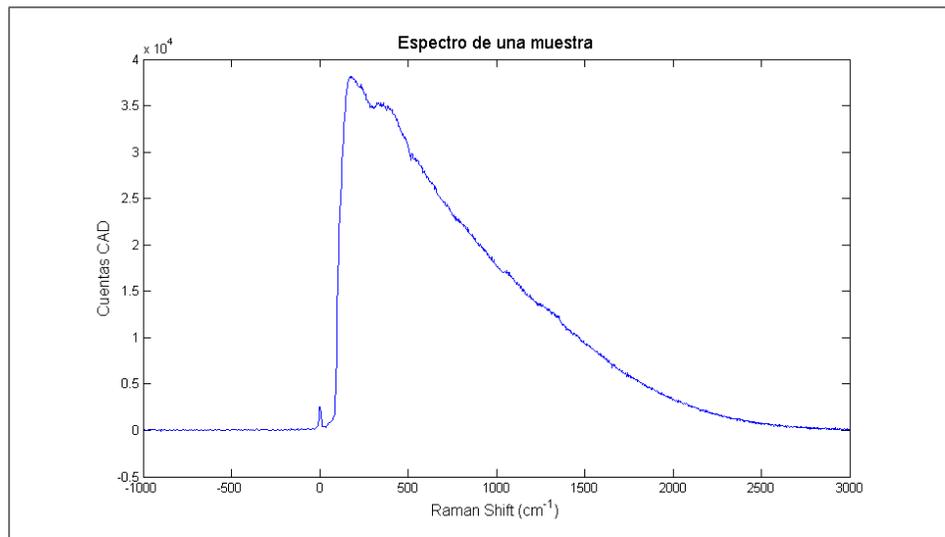


FIGURA 3.5. Espectro de una muestra.

3.4.2. Valores de referencia

Al terminar la toma de espectros de una muestra, ésta fue analizada para obtener las referencias (concentraciones). El proceso se puede descomponer en dos tareas descritas a continuación.

3.4.2.1. Proceso de extracción

Los hollejos son separados de la pulpa y lavados con agua para retirar cualquier resto. Los 5 hollejos pertenecientes a una muestra son secados con papel y pesados. Luego son

sumergidos en nitrógeno líquido durante unos segundos para que se endurezcan. Inmediatamente después son molidos en una moladora de café. Los pequeños pedazos y polvo resultante son pesados nuevamente y guardados en un tubo cónico tipo *Falcon* de 15 ml para ser procesados por una centrífuga. Si se calcula un “Porcentaje de Recuperación” como el peso del material recuperado sobre el peso de los hollejos que ingresaron a la moladora, se obtiene que en promedio se recuperó alrededor del 70% del material. El porcentaje restante corresponde a aquel adherido a las paredes de la moladora.

El hollejo convertido en polvo debe ser macerado para que libere las moléculas que se quieren medir. La maceración se hizo en una dilución 1:10, es decir por cada 0.1 gr de hollejo se agregó 1 ml de solución hidroalcohólica de etanol a 50% en HCl 0.1 M. La preparación se mantuvo por 1 hora en agitación horizontal a una velocidad de 30 RPM. Finalmente la muestra fue centrifugada a 5000 RPM durante 5 min. El sobrenadante o extracto es separado y guardado a 4°C para el análisis de absorbancias al día siguiente.

3.4.2.2. Obtención Absorbancias

El extracto fue analizado para obtener los indicadores de polifenoles totales, antocianinas totales y taninos totales. Cada análisis se hizo con 3 repeticiones para disminuir el error experimental. La repetición con el valor más alejado fue eliminada y las 2 restantes promediadas. Para el análisis de polifenoles totales y antocianinas totales se usó 100 μl de extracto por repetición, para cada análisis. Para el análisis de taninos totales, se usó 200 μl de extracto por repetición (100 para el testigo y 100 para la comparación). Para realizar los análisis se necesitaron 1200 μl de extracto, lo que determinó el número mínimo de bayas por muestra. El peso mínimo de 5 hollejos de bayas fue de 0.28 gr, por lo que si se considera un porcentaje de recuperación promedio de 70% en la molienda, el peso se reduce a 0.2 gr. A este peso se le agrega 2 ml de solución de extracción (1 ml por cada 0.1 gr), de los cuales se usan 1200 μl , dejando un margen de 800 μl para no perder la muestra de ser necesario repetir algún análisis. Por esta razón se consideró usar 5 bayas por muestra.

El análisis de laboratorio utilizado para polifenoles fue el índice DO280, el cuál es ampliamente utilizado en laboratorio por su simpleza (Bordeu & Scarpa, 2008). Para la concentración de antocianinas se utilizó el método de Puissant-Leon por cambio de pH (Puissant & Leon, 1967) y para la concentración de taninos se utilizó el método de precipitación con metilcelulosa (Sarneckis et al., 2006).

Polifenoles totales: 100 μ l de extracto fueron diluidos en 10 ml de agua destilada y mezclados hasta obtener una solución uniforme (factor de dilución = 101). La absorbancia a 280 nm se midió en una cubeta de cuarzo de paso óptico de 10 mm. Como blanco (referencia) se utilizó agua destilada. El índice del extracto DO280 se calculó según la ecuación:

$$DO280 = Abs280 \text{ nm} \times 101 \quad (3.1)$$

Antocianinas totales: 100 μ l de extracto fueron diluidos en 10 ml de HCl 1% p/v y mezclados hasta obtener una solución uniforme (factor de dilución = 101). La absorbancia a 520 nm se midió en una cubeta de poliestireno de 10 mm de paso óptico. Como blanco (referencia) se utilizó HCl 1% p/v. La concentración de antocianinas en mg/l de extracto se calculó según la ecuación:

$$Antocianinas \text{ totales (lm/l)} = Abs520 \text{ nm} \times 22.76 \times 101 \quad (3.2)$$

Taninos totales: se prepararon 2 soluciones, cada una a partir de 100 μ l de extracto, a una se le agregó 1 ml de Metilcelulosa y a la otra no (Testigo). A ambas se le agregó 2 ml de solución saturada de Sulfato de Amonio y se aforó hasta 10 ml con agua destilada (factor de dilución = 100). Se agitó y dejó reposar por 10 min a temperatura ambiente. Finalmente se centrifugó a 5000 RPM durante 5 min Se obtuvo las absorbancias a 280 nm de ambas soluciones, usando cubetas de cuarzo de 10 mm de paso óptico. La concentración de taninos se obtuvo a partir de la diferencia de absorbancias expresadas en mg equivalente de catequina por litro. Para ello se usó la siguiente ecuación (la cual fue calibrada para distintas concentraciones de Catequina):

$$Taninos\ totales\ (mg\ equiv./l) = (83.031 \times (\Delta Abs_{280\ nm}) - 3.8795) \times 100 \quad (3.3)$$

En el caso de antocianinas y taninos se calculó la concentración en gramos de hollejo fresco, dada la concentración del extracto y la cantidad de hollejo por volumen presente en él. En el caso de polifenoles totales, se prefirió expresar la referencia como índice DO280 del extracto. En la Tabla 3.2 se muestra un resumen de los valores obtenidos.

TABLA 3.2. Valores de referencia obtenidos en laboratorio. Antocianinas y taninos expresados en mg de la molécula por gramo de hollejo.

Indicador	Min.	Max.	Prom.	D.Std.
Antocianinas (mg/g)	7.06	24.67	12.70	3.46
Polifenoles (DO280)	41.97	121.20	61.82	12.62
Taninos (mg/g)	6.17	17.38	10.22	2.16

En la literatura es común encontrar datos para concentraciones de antocianinas y taninos expresados en mg de la molécula por kg de baya. Los valores promedio obtenidos en estos términos son 966 mg/kg de baya para antocianinas y 777 mg/kg de baya (en el hollejo) para taninos. Boulton et al. (1996) señala que para una variedad roja oscura como Cabernet Sauvignon, el contenido de antocianinas debe ser alrededor de 1000 mg/kg, lo que se verifica con los valores obtenidos. En el caso de los taninos, al ser registrados por baya es necesario diferenciar si provienen de la piel o semillas. En la literatura se encontró que en la piel las concentraciones pueden ir en el rango de 100 a 500 mg/kg de baya (Blouin & Peynaud, 2005), por lo que los valores obtenidos estarían ligeramente elevados. Por último, cabe señalar que estos valores pueden variar ampliamente dependiendo del método de extracción utilizado (Peyron, 1998), por lo que en trabajos previos es frecuente encontrar valores dispares.

3.5. Procesamiento de espectros

En la sección 3.4.1, se presentó el espectro de una muestra, sin embargo ese espectro ya está procesado con algunos algoritmos básicos. A continuación se describen los algoritmos usados para procesar los espectros, y se indica los que fueron útiles para optimizar los modelos.

3.5.1. Resta de espectro de oscuridad

Esta es la operación más básica de procesamiento de un espectro. Consiste en restar el nivel base de señal que se suma a la señal medida por el instrumento. La resta también elimina los *hot pixels*, que son pixeles que registran una intensidad mayor que los restantes. Esta señal aparece aunque se tome una imagen con el obturador cerrado, es por eso que se llama de oscuridad (*dark*). En la Figura 3.6 se muestra el efecto de la resta del espectro de oscuridad.

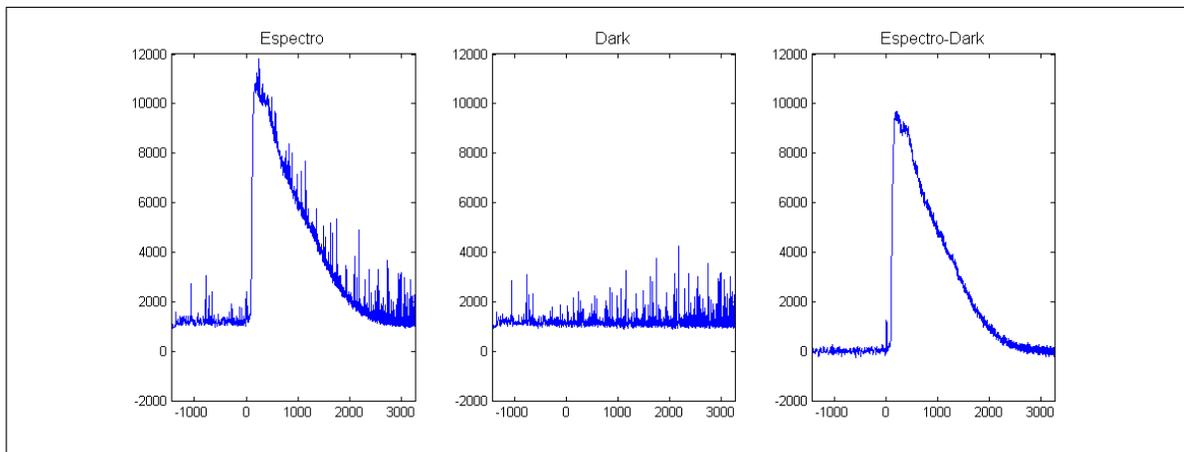


FIGURA 3.6. Resta del dark o bias.

3.5.2. Filtros de ruido

El ruido en la señal es inevitable y proviene de diversas fuentes tales como la señal misma (ruido fotónico), la temperatura en el detector, el ruido de lectura, etc. Existen distintas técnicas para eliminar esta componente indeseable, como por ejemplo la media

móvil (*box car*), y filtros aplicados en el espacio de frecuencia, donde se utiliza la transformada de Fourier. El concepto detrás de estos filtros es reducir la amplitud de oscilaciones rápidas, las cuales son atribuidas a ruido. Sin embargo, en espectroscopía es común que la información también se encuentre en forma de cambios rápidos, como por ejemplo líneas de absorción o picos, como es el caso de la espectroscopía Raman. Para eliminar ruido en este tipo de señales, se usa el filtro de Savitzky-Golay (1964), el cual tiende a conservar las características de estas señales, como anchos, máximos y mínimos relativos, los cuales son normalmente perturbados al usar otras técnicas. El filtro de Savitzky-Golay consiste en hacer ajustes polinomiales locales de grado k en una serie de al menos $k + 1$ valores, permitiendo también calcular derivadas de la señal. En la Figura 3.7 se puede ver cómo cambia el espectro de una muestra al usar un filtro Savitsky-Golay de parámetros: 7 muestras, polinomio de grado 2 y sin derivada. Se observó que el uso de filtros ayudó a homogeneizar el índice R^2 para calibración y validación, sin embargo si se usa un filtro muy fuerte se genera una disminución en la correlación.

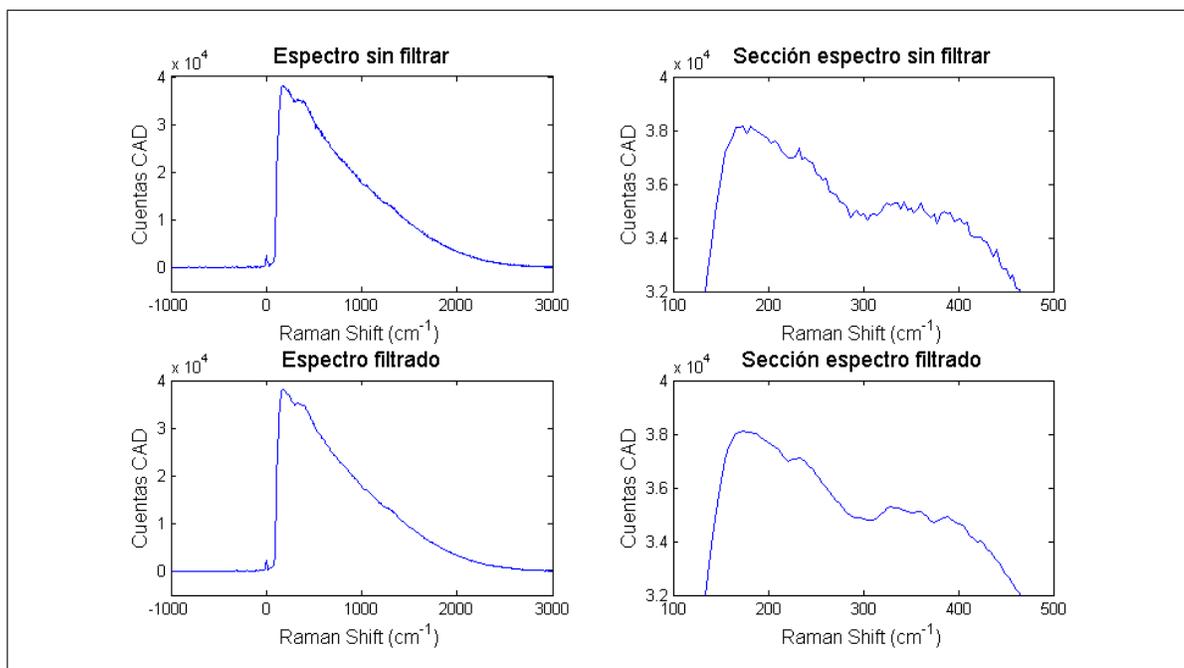


FIGURA 3.7. Efecto del filtro sobre el espectro de una muestra.

3.5.3. Corrección espectral

La corrección espectral es, en general una herramienta muy útil en espectroscopía. La idea central de la espectroscopía es que existe una relación lineal entre la concentración de una molécula y la absorción que ésta genera a una determinada longitud de onda. En espectroscopía Raman esta relación se da en la intensidad de los picos. Sin embargo, existen perturbaciones que afectan esta relación.

Un fenómeno común que puede necesitar corrección es el paso óptico. En espectroscopía de absorción, si dos muestras poseen la misma concentración de una molécula, pero tienen distinto paso óptico, generarán distintas absorciones y por lo tanto si no se corrige este factor, la medición de absorbancias no medirá concentración adecuadamente. Análogamente, en espectroscopía Raman puede suceder que el láser ilumine distintos volúmenes de muestra, generando distintas intensidades en los espectros. En algunas muestras se puede medir su grosor, pero en otras, como es el caso de las bayas no. A continuación se señalan dos técnicas muy utilizadas para intentar corregir este fenómeno.

Sample Thickness Correction: requiere conocer una banda en el espectro donde la muestra de interés no interactúe. Un cambio en la intensidad de esta banda corresponderá a un cambio en el paso óptico. La solución propuesta por este método es normalizar el espectro según su intensidad en aquella banda. La desventaja de este método es que se requiere un conocimiento del espectro de la molécula de interés.

Unit Area Normalization: Se usa cuando no se puede aislar una banda para aplicar el método anterior. Consiste en normalizar el espectro usando el área bajo la curva en todo el rango espectral.

Por otro lado, existen otros fenómenos que pueden afectar el espectro de la muestra, como la distribución no homogénea y tamaño de las partículas o cambios en el índice de refracción de la muestra. El efecto en el espectro se puede manifestar como un nivel base (*baseline*), inclinación (*tilt*) o curvatura y puede ser más pronunciado en longitudes de onda altas. A continuación se señalan dos métodos conocidos.

SNV (Standard Normal Variate): La corrección consiste en normalizar cada espectro por la desviación estándar de la respuesta a lo largo de todo el rango. Primero se calcula la media de cada espectro (ecuación 3.4) y luego se normaliza (ecuación 3.5).

$$\mu_i = \sum_{j=1}^p \mathbf{x}_{i,j} \quad (3.4)$$

$$\mathbf{x}_{i(SNV)} = \frac{(\mathbf{x}_i - \mu_i)}{\sqrt{\frac{\sum_{j=1}^p (\mathbf{x}_{i,j} - \mu_i)^2}{p-1}}} \quad (3.5)$$

MSC (Multiplicative Scatter Correction): El método requiere tener un espectro modelo, el cual puede ser calculado como un promedio de los espectros disponibles ($\bar{\mathbf{x}}$). Luego para cada espectro \mathbf{x}_i se estiman los parámetros a_i y b_i de la ecuación 3.6 por regresión. Finalmente se corrige el espectro según la ecuación 3.7.

$$\mathbf{x}_i = a_i + \bar{\mathbf{x}}b_i \quad (3.6)$$

$$\mathbf{x}_{i(MSC)} = \frac{(\mathbf{x}_i - a_i)}{b_i} \quad (3.7)$$

La Figura 3.8 presenta un resumen con el efecto que tienen las diferentes correcciones sobre los espectros. Se observa que todas tienen un efecto similar que consiste en una agrupación de los espectros, resaltando las diferencias de forma entre ellos, en lugar de sus tamaños absolutos. Así como visualmente el resultado es parecido, también lo es al aplicar estos métodos en la calibración de modelos. Se observó que la corrección espectral aumentó el valor de índice R^2 de validación, con ligeras mejorías para el método de corrección por camino óptico. La longitud a la cual se corrigió fue determinada por prueba y error, hasta obtener el mejor resultado.

3.5.4. Eliminación de fluorescencia (*background*)

El factor que mayormente perturba los espectros Raman es la fluorescencia. En algunas aplicaciones se puede tener buenos resultados sin eliminar este efecto, pero en otras es

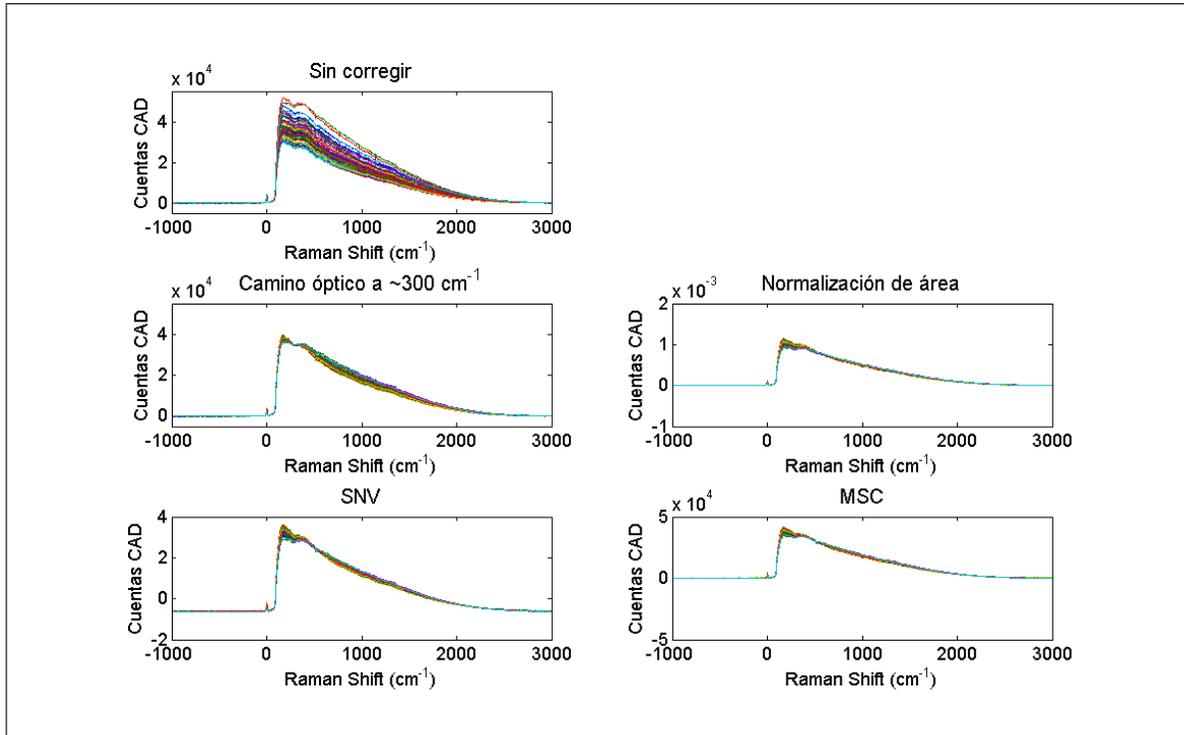


FIGURA 3.8. Efecto de correcciones sobre el conjunto de espectros.

necesario eliminarlo (Smith & Dent, 2005). Es común referirse a este procedimiento como “segmentación”, ya que el efecto Raman es separado de la fluorescencia. En esta sección se revisan distintas técnicas para realizar este proceso.

Todas las técnicas se basan en la diferencia gráfica que tienen los espectros Raman y los espectros de fluorescencia. Como se vio en la sección 2.2.3, la fluorescencia se caracteriza por tener una forma suave, mientras que el efecto Raman se caracteriza por cambios abruptos o picos. A continuación se describen tres métodos.

1ra y 2da derivada: La primera derivada puede eliminar una fluorescencia que se presente en forma de *offset* y la segunda puede eliminar un *offset* y *tilt* (Smith & Dent, 2005). Visto de otra forma la derivada resalta la información contenida en cambios rápidos. Un inconveniente de esta técnica es que puede amplificar el ruido. El filtro de Savitzky-Golay permite derivar la señal sin amplificar excesivamente el ruido.

Ajuste Polinomial: Consiste en ajustar un polinomio, que puede ser una recta o parábola. Este tipo de segmentación permite eliminar fluorescencia en forma de curvatura, además de *offset* y pendiente. Si los picos Raman se pueden identificar y tiene una amplitud similar a la de la fluorescencia se pueden ajustar un conjunto de curvas (Vilaboy et al., 2003).

Rolling Circle: Gráficamente esta técnica consiste en hacer rodar un círculo por debajo de la señal que contiene picos y fluorescencia. El diámetro del círculo debe ser tal que éste no pueda ingresar a los picos, pero pueda rodar por la curva suave de fluorescencia (Mikhailyuk & Razzhivin, 2003).

En la Figura 3.9 se muestra un resumen con algunos resultados obtenidos al segmentar los espectros con las distintas técnicas descritas. En la primera derivada se observa que la componente más importante del espectro segmentado es producto de la subida abrupta del espectro cerca de los 100 cm^{-1} , pero el resto del espectro es prácticamente ruido. En el espectro segmentado por el método del círculo se observa que éste corresponde a los pequeños picos sobre la parte superior del espectro. Sin embargo, en todo el resto del rango no se segmentaron los cambios rápidos que se observaron en la sección 3.4.1, incluso en pruebas que se hicieron con distintos radios del círculo. Por último, cabe destacar que la segmentación con ajuste polinomial, si bien no se logró aplicar en todo el rango espectral, sí logró segmentar relativamente bien la información contenida en el rango entre los 200 y 2000 cm^{-1} .

3.5.5. Acondicionamiento de los datos

Este procedimiento se basa en el supuesto de que la información está contenida en la relación entre intensidades para diferentes longitudes de onda en lugar de la intensidad absoluta. El método de regresión PLS también apunta al mismo supuesto. A continuación se describen los dos métodos más usados (Wise et al., 2006):

Mean Centering: Se calcula la media de todos los espectros del set de calibración. Luego este valor es restado a cada espectro. Lo mismo se hace con las concentraciones. El resultado es que todas las columnas de la matriz \mathbf{X} y vector \mathbf{y} tienen media 0. Si no se

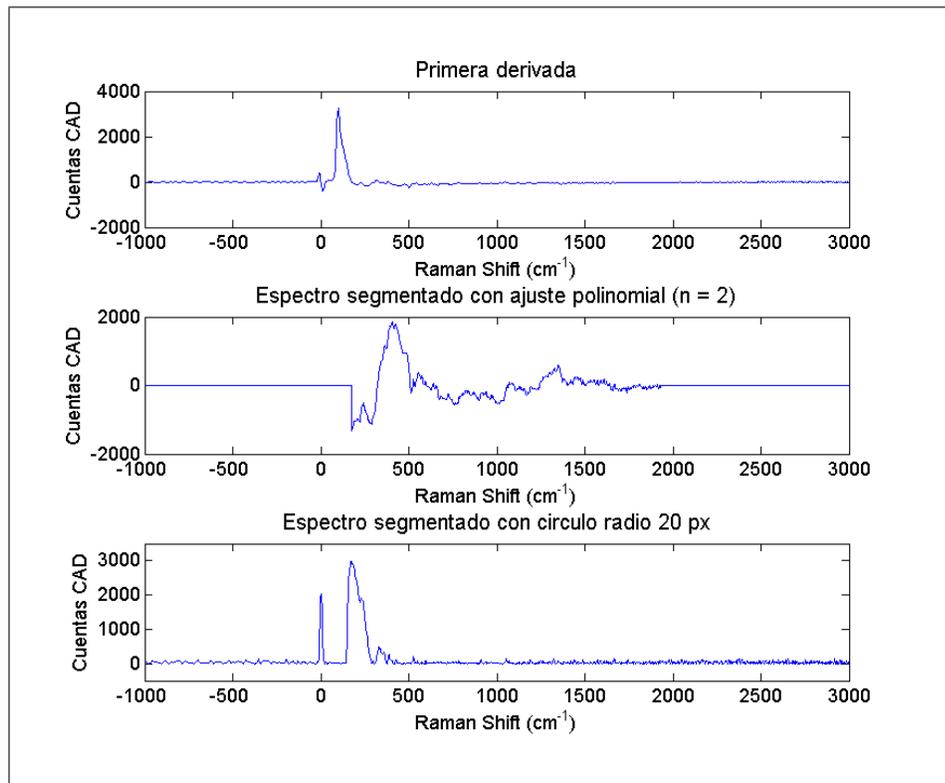


FIGURA 3.9. Ejemplo segmentaciones de fluorescencia con los métodos descritos.

centran los espectros, la primera componente principal describirá la media de los datos, en lugar de la dirección de mayor variación. El uso del centrado dependerá de la aplicación.

Variance Scaling: A los datos previamente centrados por la media, se les escala por la varianza. Este método es más usado cuando los datos contienen variables con distintas unidades, como presión, temperatura, etc.. El resultados es que todas las columnas de la matriz \mathbf{X} y vector \mathbf{y} tienen varianza unitaria.

Para acondicionar los espectros de bayas se usó el método *mean centering*, ya que entregó mejores resultados que con *variance scaling*. Además, en el caso de espectros, todas las variables son longitudes de onda, por lo tanto teóricamente no es necesaria la corrección por varianza unitaria, ya que no hay cambios en la unidad de las variables.

3.6. Calibración

Para la calibración se utilizó el método de regresión PLS. El resultado de la calibración es el vector de regresión o modelo. Para evaluar el desempeño de cada modelo se utilizaron los índices de correlación múltiple R^2 y error cuadrático medio $RMSE$, los cuales se calculan según las ecuaciones 3.8 y 3.9 respectivamente, donde: y es el valor real del analito, \hat{y} es la estimación generada por el modelo, σ_i es la desviación estándar de i ($i = y, \hat{y}$), y $\sigma_{y\hat{y}}$ es la covarianza entre el valor real y el estimado.

$$R^2 = \left(\frac{\sigma_{y\hat{y}}}{\sigma_y \sigma_{\hat{y}}} \right)^2 \quad (3.8)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (3.9)$$

Dependiendo si el índice se usa para evaluar la calibración o validación, se agrega la letra c o v respectivamente, así el índice R_c^2 es el índice de múltiple correlación de calibración. En un buen modelo el índice R^2 tiende a 1 y $RMSE$ a 0.

En la calibración se utilizaron dos técnicas para obtener modelos más robustos. A continuación se describe el procedimiento utilizado.

3.6.1. Eliminación de muestras

En el proceso de calibración se detectó la presencia de muestras que no se ajustaban al modelo. Se probaron exhaustivamente distintos grupos de calibración y validación, obteniéndose resultados muy dispares. Se investigó la procedencia de estas muestras, inspeccionando los pesos de los hoyejos y bayas completas, llegando a la conclusión de que eran generadas en el proceso de molienda (sección 3.4.2.1), en el cual presentaban un bajo porcentaje de recuperación, indicando una muestra poco representativa. Los resultados en la calibración y validación mejoraron al eliminar las muestras cuyo porcentaje de recuperación era inferior al 60%. En la Figura 3.10 se muestra el porcentaje de recuperación de

cada muestra. Finalmente, para la calibración, se conservaron 80 de las 89 muestras originales, eliminando así las 9 muestras con menor porcentaje de recuperación. Este criterio fue el único utilizado en la eliminación de muestras.

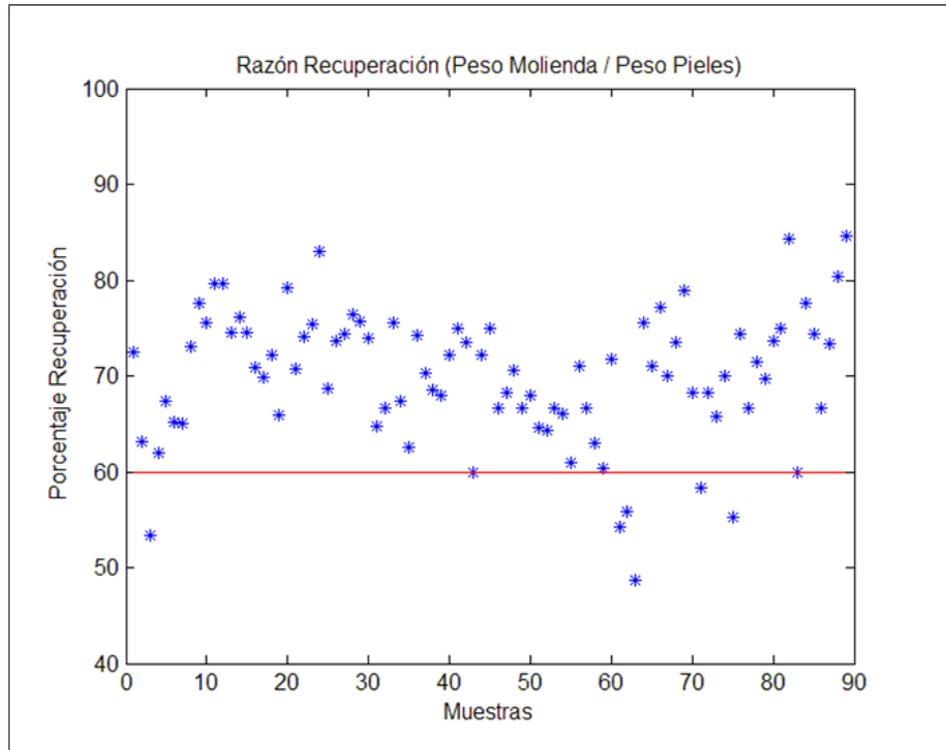


FIGURA 3.10. Porcentaje recuperación por muestra.

3.6.2. Selección conjuntos de calibración y validación

De un total de 80 muestras, el 75% se tomó para el conjunto de calibración y el 25% para el conjunto de validación. Las muestras pertenecientes a cada conjunto deben ser elegidas aleatoriamente. Sin embargo, se requiere que las muestras del conjunto de calibración abarquen todo el rango. También es deseado que ambos conjuntos tengan una distribución similar. Se observó que conjuntos generados aleatoriamente que no cumplen con estos dos requisitos presentan malos resultados en el desempeño del modelo. Por esta razón se ideó un proceso pseudo-aleatorio para crear los conjuntos de calibración y validación. El proceso consiste en ordenar las muestras de menor a mayor según su valor de referencia, luego se recorre el arreglo tomando cuatro muestras y se elige aleatoriamente una de estas para

el conjunto de validación, dejando las tres restantes para el conjunto de calibración. De esta forma, el algoritmo asegura que ambos conjuntos tengan la misma distribución y al mismo tiempo mantiene la aleatoriedad de la selección de muestras. En la Figura 3.11 se muestra un ejemplo con las primeras 16 muestras, donde 4 de ellas fueron escogidas para validación. En el eje de las abscisas se muestra el número de la muestra en el conjunto ordenado, en el eje de las ordenadas se tabula el valor de referencia y sobre cada muestra se identifica el número de la muestra en el conjunto original (para mantener asociado el espectro correspondiente). También se muestran separaciones para los subconjuntos de 4 muestras.

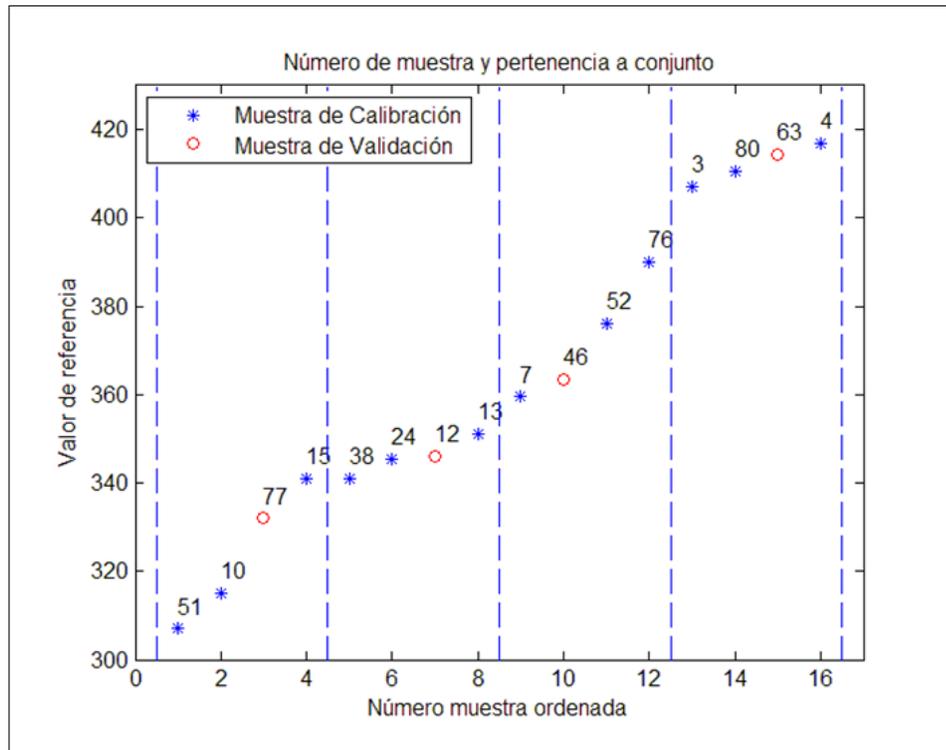


FIGURA 3.11. Conjuntos de calibración y validación.

3.6.3. Interpretación

En la sección 2.3.1.1 se presentaron los vectores \mathbf{t} , \mathbf{w} y \mathbf{p} (*scores*, *weights* y *loadings* respectivamente), a partir de los cuales se calcula el vector de regresión. Sin embargo, estos vectores también pueden ser utilizados para interpretar el modelo. Los vectores *scores* se

pueden graficar uno contra otro como un gráfico de dispersión, y los vectores *weights* se pueden graficar en función de la longitud de onda, como un espectro, de esta forma la contribución de cada dimensión al modelo PLS puede ser interpretada espectroscópicamente (Swerts et al., 1993; Liu et al., 2009). Los vectores *loading*, que se pueden entender como una forma resumida de la matriz \mathbf{X} (Wold et al., 2001), también pueden ser interpretados ya que éstos describen el grado en que cada variable latente contribuye a la representación de los espectros \mathbf{X} (Saariaho et al., 2005; Šašić & Ozaki, 2000).

Además de los vectores mencionados más arriba, también se puede interpretar el vector de regresión \mathbf{b} . Si existen variables para las cuales este vector tiene un valor absoluto relativamente bajo, quiere decir que éstas no se están ponderando mayormente en la predicción de \mathbf{y} . Para determinar el grado de importancia de una variable en el modelo se pueden utilizar los coeficientes VIP (*Variable Importance in Projection*) (Wold et al., 2001). Los VIP se calculan como la suma ponderada de los cuadrados de los vectores \mathbf{w}_i , donde el ponderador es la varianza de \mathbf{y} descrita por la variable latente i . Variables cuyo valor VIP sea inferior a 0.8, y cuyo valor en el vector de regresión sea bajo, son consideradas de poca importancia (Wold et al., 2001; Wold, 2008).

4. RESULTADOS Y DISCUSION

En este capítulo se presentan los resultados de la siguiente manera: en la sección 4.1 se muestran los mejores resultados obtenidos para cada modelo usando los espectros completos (sin segmentar), pero con todos los demás procesamientos aplicados. Con esta información se tiene un punto de partida para los siguientes análisis. En la sección 4.2 se presentan los resultados obtenidos para distintos conjuntos de calibración y validación, lo que se realiza solo para un analito, demostrando lo propuesto en la sección 3.6.1. En la sección 4.3 se hace una interpretación del modelo a partir de análisis de los vectores \mathbf{p} , \mathbf{w} , \mathbf{b} y VIP. Para comprobar las observaciones, se muestran los resultados de la calibración de modelos para los tres analitos, usando espectros completos y espectros segmentados. Para obtener mayor certeza en las conclusiones, se utilizan cinco conjuntos de calibración y validación distintos para cada analito. En las secciones 4.4, 4.5 y 4.6 se analizan factores que tienen incidencia en el uso práctico del instrumento, como el tiempo de integración, la estabilidad de los espectros y las posiciones de muestreo. En la sección 4.7 se presentan los mejores resultados por analito y los respectivos gráficos de dispersión. Finalmente en la sección 4.8 se presenta una breve descripción de las fuentes de error del experimento.

4.1. Modelos de calibración

Como se explicó en la sección 3.4.1 se promediaron tres espectros, cada uno tomado con 9 s de integración después de 1.5 min de exposición a la luz láser (para cada posición de muestreo y para cada baya). Se promediaron los espectros de cada posición para obtener un espectro por baya y se promediaron los espectros de las cinco bayas para obtener un espectro por muestra. El espectro resultante (representativo de la muestra) fue suavizado con el filtro Savitsky-Golay con los siguientes parámetros: siete muestras, polinomio de segundo grado y sin derivada. Se usó la corrección por camino óptico en la región de $280 - 320 \text{ cm}^{-1}$ y se redujo los datos usando centrado de medias. Se usaron los espectros en todo su rango e inicialmente no se usó segmentación para eliminar el *background*

(fluorescencia), de manera de usar toda la información contenida en el espectro. En estas condiciones se obtuvo los resultados que se muestran en la Tabla 4.1.

TABLA 4.1. Resultados por analito.

Analito	# LV	R_c^2	$RMSE_c$	R_v^2	$RMSE_v$
Antocianas totales	6	0.840	1.43	0.822	1.38
Polifenoles totales	6	0.780	6.25	0.608	7.72
Taninos totales	8	0.886	0.74	0.413	1.57

El mejor modelo fue para antocianas totales, donde se obtuvo un R_v^2 de 0.822 y un error de 1.38 mg/g. Este resultado es alentador dado que es superior al alcanzado con el espectrómetro NIR, el cual llegó a un R_v^2 de 0.681 (Larrain et al., 2008). El desempeño del modelo para polifenoles totales fue inferior, llegando a un R_v^2 de 0.608 y un error de 7.72 que en términos comparativos con el rango representa un 9.7%. Por último, el modelo para taninos totales presentó un R_v^2 de 0.413. Es posible que estos resultados estén fuertemente influenciados por la precisión del método de referencia utilizado. Para taninos, el método de referencia tiene un r de 0.79 (Sarneckis et al., 2006) ($R^2 = 0.62$) respecto al método HPLC, con el cuál fue calibrado. En contraste, el método de referencia para antocianas tiene un r de 0.99 (Puissant & Leon, 1967) ($R^2 = 0.98$) en relación a un método de referencia gravimétrico.

Dado que el modelo de antocianas muestra la mejor correlación, se usó para hacer los análisis que se presentan en las siguientes secciones.

4.2. Efecto conjuntos de calibración y validación

Durante el proceso de calibración se observó que los resultados (R_v^2) variaban considerablemente dependiendo del subconjunto de muestras usadas para calibración y validación. Sin embargo, al eliminar las muestras poco confiables (ver sección 3.6.1) este efecto se vio disminuido obteniéndose resultados más uniformes. En la Tabla 4.2 se muestran los resultados para antocianas totales de 10 conjuntos de calibración diferentes (elegidos según el

criterio de la sección 3.6.2) y se comparan los casos en que se eliminaron las muestras y en los que no. Los modelos donde se eliminaron muestras presentan mejores resultados (promedio, mínimo y máximo) y al mismo tiempo menor desviación. El peor resultado obtenido en el caso de eliminación de muestras está cerca del promedio del caso sin eliminación. Esto demuestra que las muestras eliminadas, efectivamente estaban perjudicando la calibración. Se debe recordar que las muestras eliminadas son las mismas para todas las combinaciones y cuya eliminación se basa en que al momento de ser procesadas para el análisis en laboratorio, se obtuvo un bajo porcentaje de recuperación en la molienda. Con esto se intentó utilizar las muestras comparativamente más representativas.

TABLA 4.2. Comparación de resultados para modelo de antocianinas con y sin eliminación de muestras.

# Conjunto	R_v^2 con eliminación	R_v^2 sin eliminación
1	0,696	0,457
2	0,546	0,523
3	0,698	0,628
4	0,764	0,399
5	0,718	0,803
6	0,639	0,627
7	0,796	0,522
8	0,659	0,598
9	0,822	0,413
10	0,510	0,619
Min.	0,510	0,399
Prom.	0,685	0,559
Max.	0,822	0,803
D. Std.	0,098	0,122

4.3. Interpretación del modelo

Para la interpretación del modelo se usaron las herramientas descritas en la sección 3.6.3. Se necesita un modelo con R_c^2 y R_v^2 adecuados (Wold et al., 2001), por esta razón se eligió el mejor modelo de antocianas, cuyos índices R^2 están por sobre 0.8. En la Tabla 4.3 se presenta el porcentaje de varianza de la matriz \mathbf{X} (espectros) y del vector \mathbf{y} (concentraciones) descrita por cada variable latente (LV) utilizada en el modelo. Se observa que la primera LV describe cerca del 95% de la varianza de los espectros. Sin embargo, describe menos del 3% de la varianza de las concentraciones, es decir, la predicción de las concentraciones se hace con sólo el 5% de la información (varianza) de los espectros.

TABLA 4.3. Varianza de \mathbf{X} e \mathbf{y} descrita por cada variable latente.

LV	\mathbf{X}		\mathbf{y}	
	LV	Sum.	LV	Sum.
1	94,91	94,91	2,26	2,26
2	0,51	95,42	37,82	40,08
3	1,92	97,35	9,98	50,05
4	0,48	97,83	25,02	75,07
5	0,46	98,29	2,83	77,90
6	0,27	98,57	6,14	84,04

Para entender cómo la primera variable latente describe los espectros, en la Figura 4.1 se grafica el vector \mathbf{p} (*loadings*) de esta componente. Debido a que las variables (longitud de onda) fueron centradas por su media, la curva toma valores negativos en la región alrededor de 200 cm^{-1} , describiendo más bien la diferencia de cada espectro con respecto al espectro promedio, sin embargo en la región de *Raman Shift* positivos la curva es relativamente suave, al igual que los espectros.

Por otro lado, en la Tabla 4.3 se observa que las variables latentes 2 y 4 son las que describen el mayor porcentaje de la varianza de las concentraciones. Para entender cómo

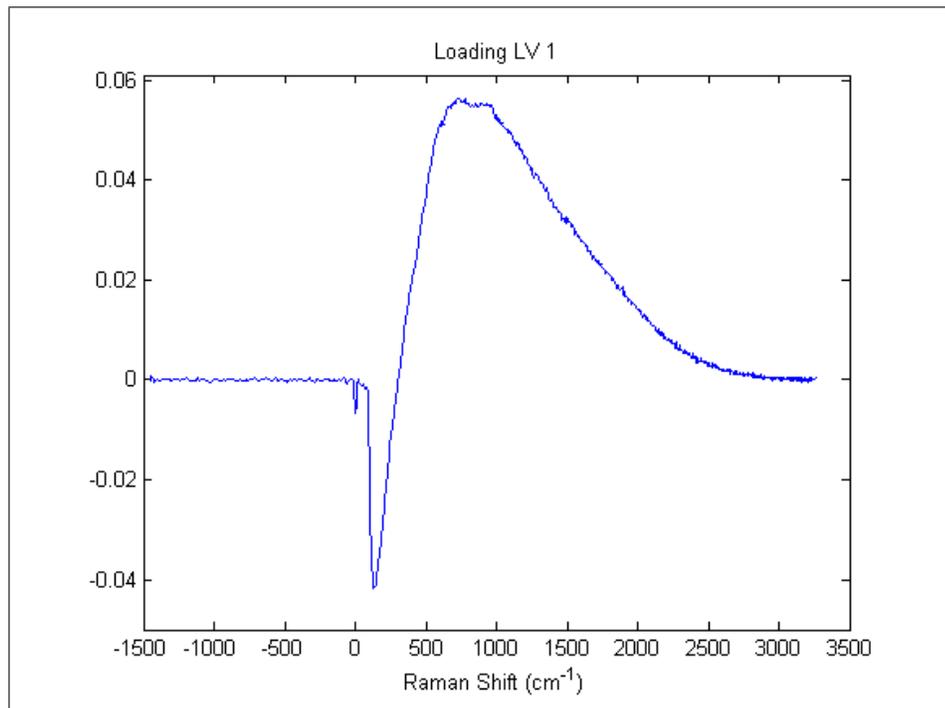


FIGURA 4.1. Vector *loadings* (**p**) de LV 1.

las variables de la matriz **X** (longitudes de onda) se combinan para predecir las concentraciones se deben inspeccionar los vectores *weights* **w**, los cuales se presentan en la Figura 4.2. Se observan características interesantes, ambos vectores tienen forma de pequeñas oscilaciones, o picos en la región comprendida entre 0 y 1500 cm^{-1} , indicando que la información de los espectros, que correlaciona con las concentraciones, se encuentra en las pequeñas fluctuaciones que están superpuestas a la componente suave (Figura 3.9) que los describe mayormente.

Por último, en la Figura 4.3 se muestra el vector de regresión **b** y los coeficientes VIP. El vector de regresión tiene características similares de los vectores **w** analizados más arriba y el vector VIP confirma que la información de interés se encuentra comprendida en la región espectral entre 0 y 2000 cm^{-1} . Sin embargo, la distribución de las variables (longitudes de onda) más importantes no es uniforme en el rango, al contrario, se encuentra distribuida en forma de picos.

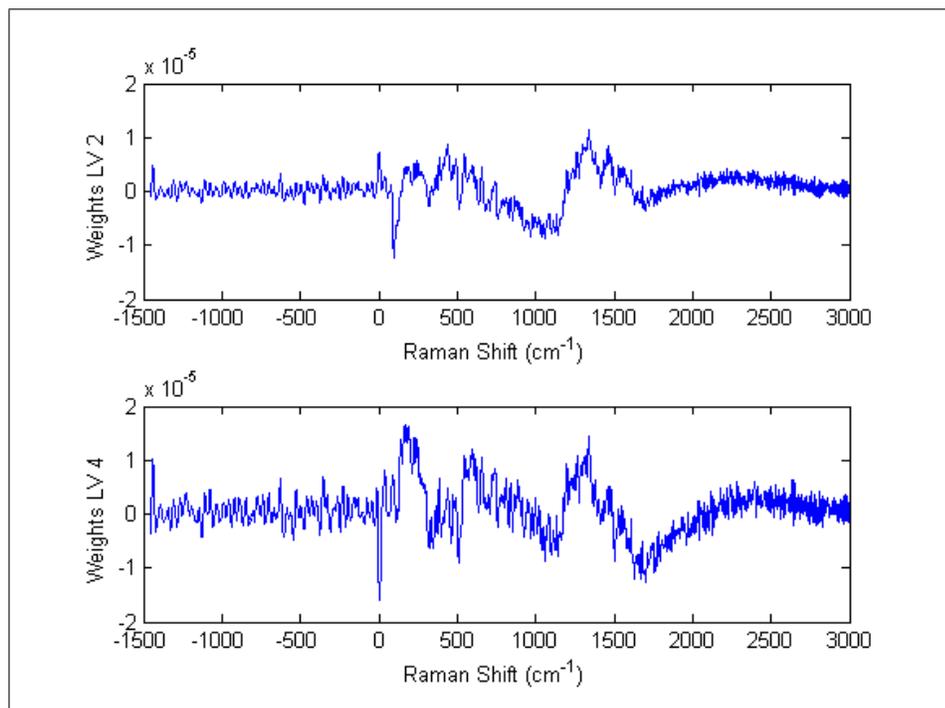


FIGURA 4.2. Vectores *weights* (**w**) de LV 2 y LV 4.

Si se compara el vector VIP, con el ejemplo de segmentaciones en la Figura 3.9 se observa que algunas características de los espectros segmentados coinciden con variables de importancia. En la segmentación realizada con el polinomio se obtiene un espectro con información en el rango donde el vector VIP identifica la mayoría de las variables de importancia. En las otras técnicas de segmentación también se pueden encontrar correspondencias, para la primera derivada el pico cerca de 100 cm^{-1} y para el *Rolling circle* el pico de Rayleigh en 0 cm^{-1} .

En base a estas observaciones, y con el fin de comprobar las conclusiones acerca de la fuente de correlación entre los espectros y las concentraciones, se calibraron modelos para espectros segmentados (sin *background* / fluorescencia) y también para el residuo de la segmentación. Los resultados se muestran en la siguiente sección.

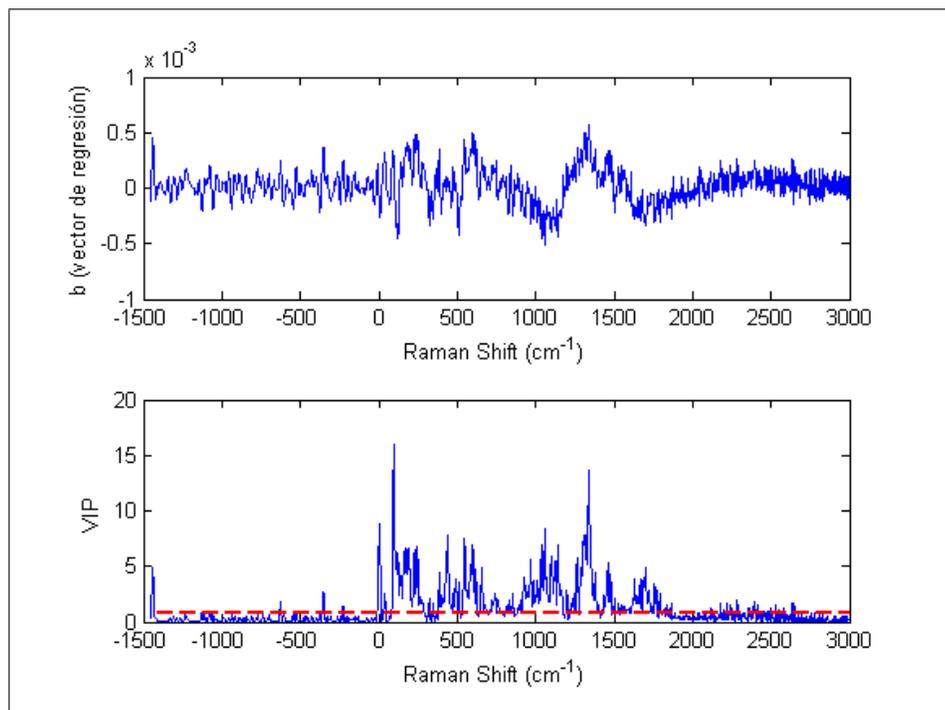


FIGURA 4.3. Vector **b** y VIP.

4.3.1. Calibración con segmentación

Se aplicaron las técnicas para remover la fluorescencia descritas en la sección 3.5.4 y se calibraron modelos para antocianinas totales con los espectros resultantes. La técnica que arrojó los mejores resultados fue el ajuste polinomial. Se ajustó un polinomio a la región del espectro comprendida entre 177 y 1930 cm^{-1} . Intentar ajustar una curva en un rango mayor, hace que se introduzcan errores producto del mal ajuste, dada la gran curvatura en la zona de los 100 cm^{-1} . El grado del polinomio a ajustar no influye mayormente en el modelo en cuanto se mantenga en un rango entre 2 y 4. Un modelo lineal no se ajusta apropiadamente y un grado mayor que 4 demostró empeorar los resultados: probablemente debido a la introducción de oscilaciones superiores a las que representan la forma del espectro. Los otros procesamientos aplicados corresponden a los mismos descritos para espectros completos en la sección 4.1. En la Figura 4.4 se muestra la forma del espectro resultante, pudiéndose apreciar que la segmentación no es perfecta, sin embargo elimina gran parte del *background*.

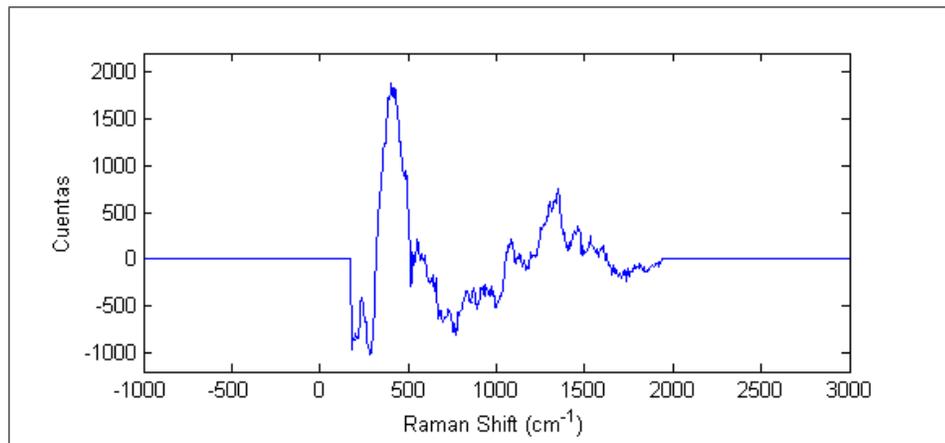


FIGURA 4.4. Espectro Segmentado.

Se calibraron modelos para cinco conjuntos distintos (los mejores resultados de la Tabla 4.2) para cada analito. Los resultados se muestran en forma resumida (promedio y máximo) en las Tablas 4.4, 4.5 y 4.6. Se observa que las diferencias no son significativas si se usa el espectro completo o el espectro segmentado. Este resultado es interesante por dos razones: la primera es que en el caso del espectro segmentado se está usando sólo un sector de éste y la segunda, es que la intensidad máxima de éste es del orden de 2.000 cuentas, mientras que la del espectro completo del orden de 40.000, es decir, con una señal 20 veces menos intensa se obtienen resultados similares.

Esto es una prueba importante de que la información se encuentra en los cambios rápidos de la señal provenientes del efecto Raman, montados sobre la componente suave de fluorescencia. Otro aspecto interesante que se puede observar en el caso de polifenoles, es que al segmentar se redujo el número de componentes necesarias para representar los espectros, otra evidencia de que se eliminó información redundante.

TABLA 4.4. Resultados promedio y máximo según tipo de espectro utilizado en modelos de antocianas.

Promedio resultados	# LV	R_c^2	$RMSE_c$	R_v^2	$RMSE_v$
Espectro completo	7	0.876	0.91	0.760	1.87
Espectro segmentado	7	0.956	0.64	0.751	1.86
Mejores resultados	# LV	R_c^2	$RMSE_c$	R_v^2	$RMSE_v$
Espectro completo	6	0.840	1.43	0.822	1.83
Espectro segmentado	7	0.981	0.49	0.839	1.19

TABLA 4.5. Resultados promedio y máximo según tipo de espectro utilizado en modelos de polifenoles.

Promedio resultados	# LV	R_c^2	$RMSE_c$	R_v^2	$RMSE_v$
Espectro completo	6	0.808	5.17	0.543	9.78
Espectro segmentado	4	0.689	6.64	0.557	9.68
Mejores resultados	# LV	R_c^2	$RMSE_c$	R_v^2	$RMSE_v$
Espectro completo	6	0.780	6.25	0.608	7.72
Espectro segmentado	3	0.628	6.84	0.604	11.6

TABLA 4.6. Resultados promedio y máximo según tipo de espectro utilizado en modelos de taninos.

Promedio resultados	# LV	R_c^2	$RMSE_c$	R_v^2	$RMSE_v$
Espectro completo	6	0.623	1.25	0.348	1.86
Espectro segmentado	5	0.585	1.28	0.351	1.87
Mejores resultados	# LV	R_c^2	$RMSE_c$	R_v^2	$RMSE_v$
Espectro completo	5	0.564	1.36	0.441	1.83
Espectro segmentado	7	0.946	0.51	0.451	1.61

4.3.2. Calibración usando fluorescencia

Para comprobar que no existe correlación entre fluorescencia y analito se calibraron modelos usando el polinomio ajustado para la segmentación (ver Figura 4.5). En la calibración el índice R_c^2 no superó el valor 0.3 para los analitos, incluso al aumentar las componentes principales buscando un mejor ajuste del modelo, lo que demuestra de forma complementaria al análisis anterior, que lo segmentado como fluorescencia (*background*) no tiene correlación con la referencia. Sin embargo, hay que señalar que la segmentación no es perfecta, ya que el resultado no son picos Raman bien definidos sobre un nivel 0, sino más bien oscilaciones en torno a un valor medio, por lo que no se puede asegurar que en la segmentación no haya un remanente del *background*.

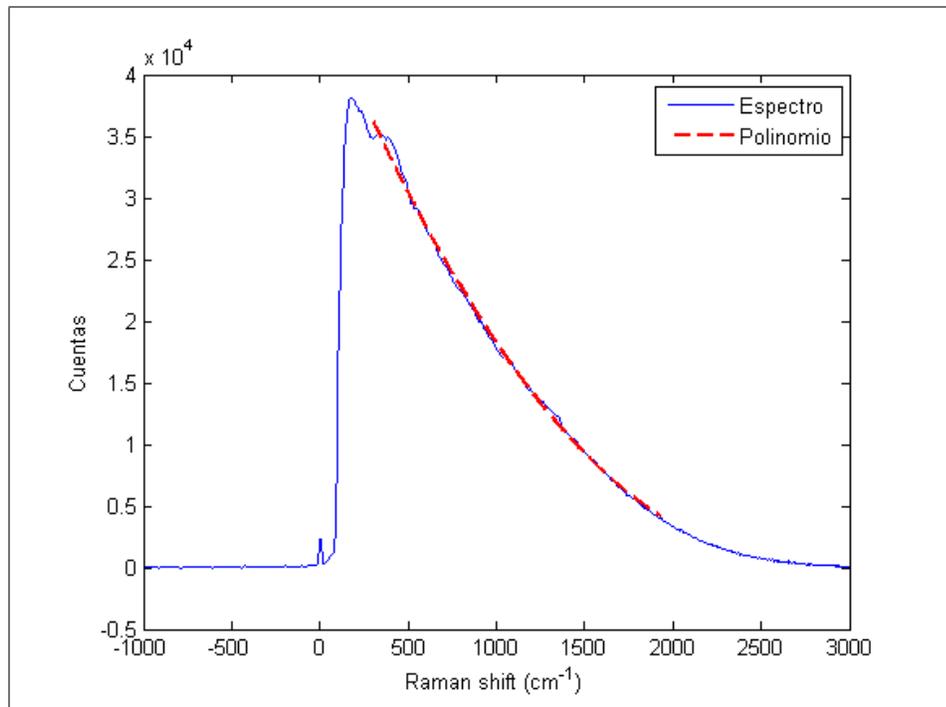


FIGURA 4.5. Espectro *background* (polinomio) y espectro completo.

4.4. Efecto del tiempo de integración

Como se mencionó en la sección 3.4.1, el muestreo consideró la adquisición de espectros con distintos tiempos de integración. Los espectros utilizados hasta ahora son un promedio de tres espectros de 9 s por cada posición de muestreo. En esta sección se hace una comparación de modelos obtenidos usando tres tipos de espectros distintos: el promedio de tres espectros de 9 s (caso de análisis anterior), un único espectro de 9 s y un espectro de 3 s. En todos los casos se promedian los espectros de las 3 posiciones. Los resultados mostrados en la Tabla 4.7, indican que el desempeño del modelo empeora cuando disminuye el tiempo de integración, lo que implica que esta variable es sensible y puede ser optimizada. En la afirmación anterior se deben observar algunas limitaciones ya que algunos espectros saturaban con más de 9 s de integración, quedando como opciones usar un espectrógrafo de mayor rango dinámico o promediar un mayor número de espectros.

TABLA 4.7. Resultados de calibración según tiempo de integración.

Tiempo espectros	# LV	R_c^2	$RMSE_c$	R_v^2	$RMSE_v$
3 esp. 9 s (prom.)	7	0.981	0.49	0.839	1.19
1 esp. 9 s	7	0.987	0.41	0.805	1.37
1 esp. 3 s	7	0.986	0.43	0.737	1.81

4.5. Efecto de la estabilidad de los espectros

Se observó que al iluminar una muestra, su intensidad no era constante (sección 3.4.1), lo cual se solucionó exponiendo la muestra durante 1.5 min min a la luz del laser. Para analizar este efecto, se tomaron espectros junto con exponer la muestra al laser y otros después de 1.5 min. Hasta ahora sólo se han analizado espectros estables (del segundo tipo). En esta sección se compara el resultado del mejor modelo obtenido usando un espectro estable con el obtenido usando un espectro inestable, ambos para 9 s de integración. Los resultados obtenidos sin variar el procesamiento de los espectros se presentan en la Tabla 4.8 observándose una disminución del índice R_v^2 (a 0.736) y un aumento de $RMSE_v$ a 1.39

mg/g. De manera similar a lo que ocurre con el tiempo de integración, ésta es una variable que afecta el desempeño de los modelos y dependiendo de la aplicación puede llegar a ser una limitante. Por ejemplo, si se requiere un muestreo rápido, se puede omitir el tiempo de estabilización (1.5 min), pero las estimaciones obtenidas serán menos precisas.

TABLA 4.8. Resultados de calibración según estabilidad de espectros.

Tipo espectro	# LV	R_c^2	$RMSE_c$	R_v^2	$RMSE_v$
1 Esp. Estable	7	0.987	0.41	0.805	1.37
1 Esp. Inestable	5	0.850	1.39	0.736	1.39

4.6. Efecto de las posiciones de muestreo

En la sección 3.4.1 se determinó muestrear las bayas en tres posiciones, asumiendo que la distribución de los compuestos podría no ser uniforme para cada baya. A continuación se presenta una evaluación de este supuesto.

Si se calibra un modelo de antocianinas con la información proveniente de una posición y se compara con un modelo calibrado con información proveniente de tres posiciones y los resultados son significativamente diferentes, entonces la no homogeneidad de la baya es un parámetro significativo. Se compararán dos modelos: uno calibrado a partir de tres espectros promediados de una única posición de una baya, y el segundo calibrado a partir del promedio de tres espectros tomados en tres posiciones diferentes. De esta manera en los dos modelos se tendrá el promedio de 15 espectros, tres por baya. El resultado se muestra en la Tabla 4.9, observándose una evidente pérdida de correlación que demuestra que la muestra de una posición por baya no es representativa.

TABLA 4.9. Resultados de calibración según posiciones de muestreo.

Tipo modelo	# LV	R_c^2	$RMSE_c$	R_v^2	$RMSE_v$
3 Esp. distintas posiciones	7	0.987	0.41	0.805	1.37
3 Esp. misma posición	7	0.953	0.47	0.470	2.42

4.7. Modelos finales por analito

Finalmente se presenta un resumen con los mejores modelos para cada analito. En el caso de antocianas y taninos se consiguió usando los espectros segmentados y en el caso de los polifenoles se obtuvo usando los espectros completos. Sin embargo, como se explicó anteriormente, la diferencia es muy estrecha, especialmente en el caso de polifenoles y taninos. En la Tabla 4.10 se muestran los resultados y en las Figuras 4.6, 4.7 y 4.8 se presentan gráficos de dispersión correspondientes.

TABLA 4.10. Resultados finales por analito.

Analito	# LV	R_c^2	$RMSE_c$	R_v^2	$RMSE_v$
Antocianas totales	7	0.981	0.49	0.839	1.19
Polifenoles totales	6	0.780	6.25	0.608	7.72
Taninos totales	7	0.946	0.51	0.451	1.61

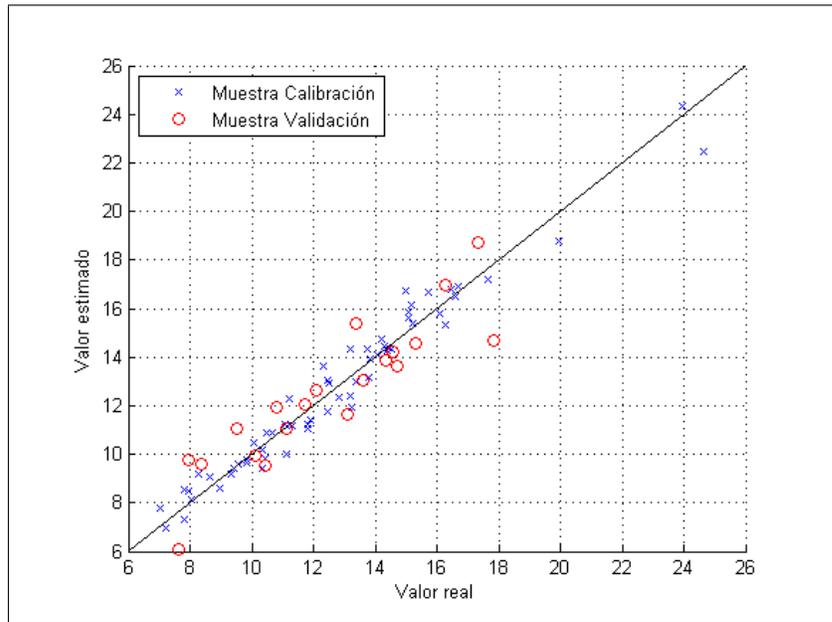


FIGURA 4.6. Gráfico de dispersión para antocianinas. Valores en mg/g de hollejo fresco.

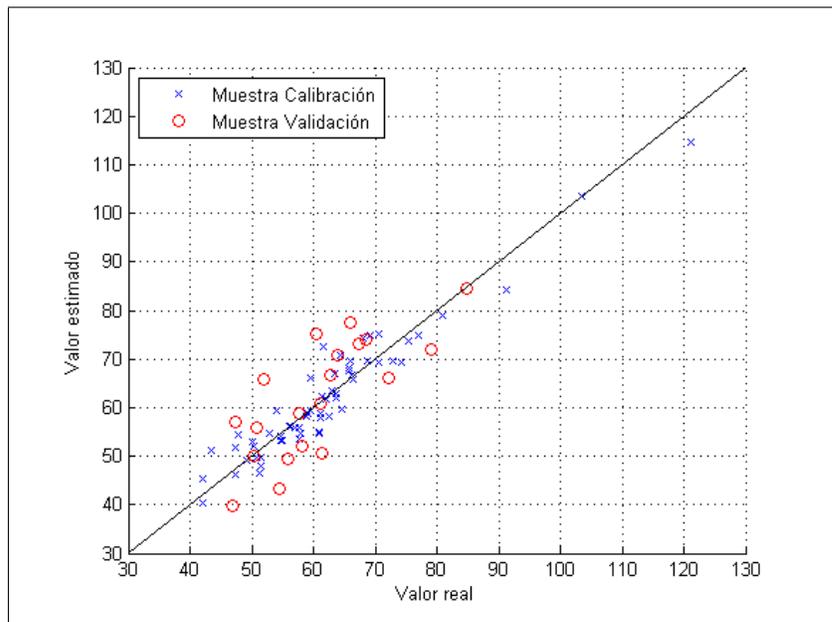


FIGURA 4.7. Gráfico de dispersión para polifenoles. Valores en índice DO280 del extracto.

4.8. Fuentes de error

Sin duda las fuentes de error son numerosas, por lo que a continuación se discuten e identifican aquellas asociadas a las principales etapas del experimento.

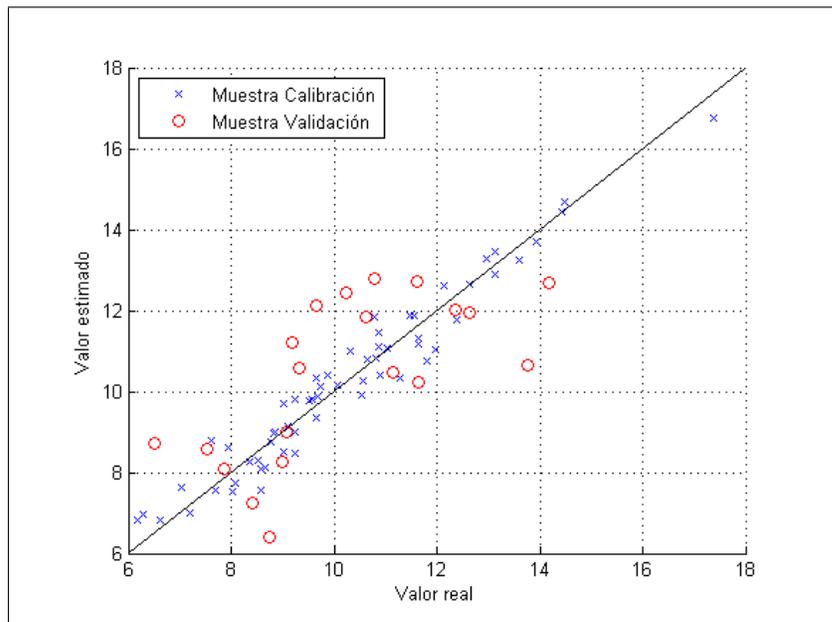


FIGURA 4.8. Gráfico de dispersión para taninos. Valores en mg/g de hollejo fresco.

En el muestreo Raman se tuvo cuidado en eliminar la mayor cantidad de fuentes de error. Por ejemplo, se aseguró la repetitividad en la posición de muestreo de las bayas, se aseguró la oclusión de la muestra y se controlaron los tiempos automáticamente. Sin embargo, a pesar de todos estos cuidados hay factores que no se comprobaron, como por ejemplo la temperatura de la muestra. Se hizo un esfuerzo por mejorar la representatividad de la muestra al tomar espectros en tres lugares de la baya, pero esto no elimina esta fuente de error, solo la disminuye. Por otro lado las posiciones de muestreo pueden no haber sido uniforme entre bayas.

En la obtención de referencias, las fuentes de error son múltiples, comenzando por la separación del hollejo de la baya, la cual se intentó hacer de la manera más prolija posible. Sin embargo, el secado y pesado de los hollejos influye en la concentración final obtenida.

La molienda del hollejo representa una fuente de error, la cual se intentó disminuir eliminando las muestras que presentaban un porcentaje de recuperación inferior a 60%, sin embargo el error persistió. La exactitud de los métodos de referencia, más que un error, presentan una limitante, imponiendo un límite a la calibración. El instrumento calibrado no

puede ser mejor que el método de referencia. En el laboratorio se observó que las mediciones de volumen con las micropipetas introducían error en la dilución de las muestras y por ende en la concentración obtenida. Este error se intentó disminuir haciendo tres repeticiones por medición y eliminando la más alejada, sin embargo, las dos repeticiones restantes rara vez fueron idénticas.

En la calibración y procesos previos a ésta, también se presentan fuentes de error, comenzando por la fluorescencia presente en la muestra, un efecto claramente predominante sobre el espectro Raman. La segmentación de este efecto no es perfecta, dejando residuos que pueden perjudicar la extracción de información por parte del método PLS, o la interpretación de los modelos. El proceso de corrección espectral es ambiguo, dado que se corrigió para un sector espectral el cual fue determinado por prueba y error y no a partir de un conocimiento del espectro Raman de las muestras. Sin embargo, la corrección espectral y filtros fueron escogidos de manera de optimizar la capacidad predictiva del modelo.

5. CONCLUSIONES Y TRABAJO FUTURO

Se calibraron modelos para estimar tres indicadores de madurez fenólica en la uva vinífera (polifenoles, antocianinas y taninos), usando un espectrómetro Raman portátil. Los resultados mostraron un buen desempeño para antocianinas totales, alcanzando un R_v^2 de 0.839. En el caso de polifenoles el resultado fue inferior llegando a un R_v^2 de 0.608. Para taninos el resultado no fue satisfactorio, llegando a un R_v^2 de 0.451.

En términos comparativos, el espectrómetro Raman alcanzó mayor precisión en la medición de antocianinas que el espectrómetro NIR calibrado en trabajos previos para medir parámetros de madurez. Más que una alternativa, el instrumento Raman portátil se presenta como un complemento en la instrumentación en terreno, ya que el instrumento NIR es robusto en la medición de Brix y pH, pero débil en la medición de color, función en la que el instrumento Raman haría un aporte.

Del análisis de los efectos del tiempo de integración y la estabilidad de los espectros, se concluye que existe un compromiso entre exactitud en las mediciones y tiempo de muestreo. Para llevar el instrumento a terreno en forma efectiva, se deben disminuir estos efectos, especialmente el de estabilidad, ya que es el que mayor tiempo involucra y más deteriora la predicción. Si se quiere aplicar los modelos obtenidos, es necesario muestrear la baya en tres posiciones y por cada una se requiere 1.5 min de iluminación con el laser, lo que significa que por cada baya se requieren al menos 4.5 min, además de los tiempos de integración de cada posición. Es de interés investigar métodos que permitan disminuir este tiempo en trabajos futuros. Sin hacer modificaciones al instrumento, un estudio de la evolución del espectro con el tiempo podría llevar a un método de procesamiento de señales que disminuya este efecto. Por otro lado, la eliminación de la fluorescencia podría ser clave en la solución de los problemas de tiempos de integración y estabilidad.

A partir del análisis de uniformidad de las muestras, se concluye que éste es un factor importante al momento de muestrear bayas. Dado que el instrumento mide en un área muy pequeña, es necesario muestrear en varios lugares para obtener un valor representativo de la baya. Este fenómeno se puede abordar de dos formas: la primera es determinar cual

es el número óptimo de mediciones de una baya para tener un valor representativo, y la segunda es modificar el instrumento de manera de medir sobre un área mayor en un solo muestreo, para lo cual es necesario el desarrollo de una sonda especialmente diseñada para la medición de bayas.

Se observó que las bayas presentan un espectro con una alta interferencia producto del fenómeno de fluorescencia. Se investigaron métodos enfocados en el procesamiento de señales para eliminarla, pero ninguno resultó ser completamente efectivo. La calibración de modelos con el espectro segmentado mejoró los resultados para antocianinas y taninos, pero no de forma significativa. Se confirmó que la información no se encuentra presente en el *background* o fluorescencia, sino que en los cambios rápidos que se encuentran superpuestos, los cuales parecen pertenecer al efecto Raman. El método PLS demostró ser robusto al momento de extraer la información espectral, ya que los resultados para espectros segmentados y completos no difieren sustancialmente, a pesar de que los primeros son del orden de 20 veces más débiles. La eliminación de fluorescencia es un proceso donde se pueden obtener mejoras al instrumento. Una forma de eliminarla es usando un láser de mayor longitud de onda, ya que la presencia de la fluorescencia es dependiente de la longitud de onda, no así el efecto Raman. Las ventajas de eliminar la fluorescencia serían las siguientes: no se necesitaría esperar 1.5 min a que el espectro se estabilice y el espectrómetro no saturaría a los 9 s pudiéndose obtener picos Raman más definidos al poder integrar durante más tiempo.

Desde el punto de vista del trabajo en laboratorio, se debería investigar la posibilidad de usar métodos de referencia más precisos. Como se explicó en la sección 4.1, la gran diferencia de desempeño de los modelos para los distintos analitos, se puede deber a la precisión de los métodos de referencia utilizados. Es por eso que se propone investigar la posibilidad de utilizar otras técnicas para la obtención de referencias, como por ejemplo HPLC. Similarmente se debería perfeccionar el método de extracción para obtener idealmente una referencia por baya y no a partir de cinco bayas como fue en esta ocasión.

Finalmente, el instrumento Raman portátil, y en general la espectroscopía Raman, se presentan como una interesante herramienta analítica en enología,. Sin embargo, se requiere complementar esta investigación para mejorar los resultados y llevar el instrumento a terreno de manera efectiva.

BIBLIOGRAFIA

- Beebe, K., & Kowalski, B. (1987). An introduction to multivariate calibration and analysis. *Analytical Chemistry*, 59(17), 1007A-1017A.
- Beger, A., Koo, T., Itzkan, I., Horowitz, G., & Feld, M. (1999). Multicomponent blood analysis by near-infrared raman spectroscopy. *Applied Optics*, 38(13), 2916–2926.
- Blouin, J., & Peynaud, E. (2005). *Connaissance et travail du vin* (4 ed.). Paris, Francia: Dunod.
- Bordeu, E., & Scarpa, J. (2008). *Análisis químico del vino* (1 ed.). Santiago, Chile: Eds. Pontificia Universidad Católica de Chile.
- Boulton, R., Singleton, V., Bisson, L., & Kunkee, R. (1996). *Principles and practices of winemaking* (1 ed.). New York, USA: Springer Verlag.
- BTR111E MiniRam datasheet*. (2008).
- Cheynier, V., Moutounet, M., & Sarni-Manchado, P. (2003). Enología: Fundamentos científicos y tecnológicos. In (2 ed., pp. 114–122). Madrid, España: AMV Eds.
- Clark, R. J. (2007). Raman microscopy as a structural and analytical tool in the field of art and archaeology. *Journal of Molecular Structure*, 834-836, 74–80.
- Henson, M., & Zhang, L. (2006). Drug characterization in low dosage pharmaceutical tablets using raman microscopic mapping. *Applied Spectroscopy*, 60(11), 1247–1255.
- Herrera, J., Guesalaga, A., & Agosin, E. (2003). SW-NIR Spectroscopy for non-destructive determination of maturity of wine grapes. *Journal of Measurement Science and Technology*, 12, 689–697.
- Jackson, E. (1980). Principal components and factor analysis: Part 1 - principal components. *Journal of Quality Technology*, 13(1).

- Larrain, M., Guesalaga, A., & Agosin, 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.
- Liu, F. L., Ye, X., He, Y., & Wang, L. (2009). Application of visible/near infrared spectroscopy and chemometric calibrations for variety discrimination of instant milk teas. *Journal of Food Engineering*, 93(2), 127-133.
- Margalit, Y. (2004). *Concepts in wine chemistry* (2 ed.). Portland, USA: Wine Appreciation Guild.
- Mikhailyuk, I., & Razzhivin, A. (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.
- Peyron, D. (1998). Le potentiel polyphénolique du pinot noir. *Revue Française d'Œnologie*, 170, 42-45.
- Puissant, A., & Leon, H. (1967). La matire colorante des grains de raisins de certains cepages cultivis en anjou en 1965. *Annales de Technologie Agricole*, 16(3), 217-225.
- Saariaho, A. M., Argyropoulos, S. D. S., Jääskeläinen, A., & Vourinen, T. (2005). Development of the partial least squares models for the interpretation of the UV resonance raman spectra of lignin model compounds. *Vibrational Spectroscopy*, 37(1), 111-121.
- Sarneckis, C. J., , Damberg, R. G., Jones, P., Mercurio, M., Herderich, M., et al. (2006). Quantification of condensed tannins by precipitation with methyl cellulose: development and validation of an optimised tool for grape and wine analysis. *Australian Journal of Grape and Wine Research*, 12(1), 39–49.
- Savitzky, A., & Golay, M. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8), 1627-1639.
- Smith, E., & Dent, G. (2005). *Modern Raman Spectroscopy: A Practical Approach* (1 ed.). Chichester, England: John Wiley & Sons, Ltd.

- Swerts, J., Espen, P., & Geladi, P. (1993). Partial least squares techniques in the energy-dispersive X-ray fluorescence determination of sulfur-graphite mixtures. *Analytical Chemistry*, 65(9), 1181-1185.
- Vilaboy, M., Ergin, A., Tchouassi, A., Greene, R., & Thomas, G. (2003, March). Optical measurement of glucose concentrations using raman spectroscopy. *Bioengineering Conference, 2003 IEEE 29th Annual, Proceedings of*, 329-330.
- Šašić, S., & Ozaki, Y. (2000). Band assignment of Near-Infrared spectra of milk by use of partial least-squares regression. *Applied Spectroscopy*, 54(9), 1327-1338.
- Wang, A., Haskin, L. A., Lane, A. L., Wdowiak, T. J., Squyres, S. W., Wilson, R. J., et al. (2003). Development of the mars microbeam raman spectrometer (MMRS). *J. Geophys. Res.*, 108(E1), 5005.
- Wise, B. M., Gallagher, N. B., Bro, R., Shaver, J. M., Windig, W., & Koch, R. S. (2006). *PLS_Toolbox 4.0 for use with MATLAB™*. 3950 West Eaglerock Drive, Wenatchee, WA 98801.
- Wold, S. (2008). PLS for multivariate linear modeling. *QSAR: Chemometric Methods in Molecular Design. Methods and Principles in Medicinal Chemistry*. (ed. H. van de Waterbeemd, Weinheim, Germany: Verlag-Chemie)
- Wold, S., Ruhe, A., Dunn, W., & H.Wold. (1984). The collinearity problem in linear regression. The partial least square approach to generalized inverses. *SIAM Journal on Scientific and Statistical Computing*, 5, 735-743.
- Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2), 109-130.