



UNIVERSIDAD DEL AZUAY
FACULTAD DE CIENCIA Y TECNOLOGÍA
ESCUELA DE INGENIERÍA EN ALIMENTOS

“CUANTIFICACIÓN, MEDIANTE INFRARROJO, DE LA
PECTINA RESIDUAL DEL PROCESO DE BIOTECNOLOGÍA DE
CLARIFICACIÓN DE NECTAR Y JUGO DE MANZANA”

TRABAJO DE GRADUACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERO EN ALIMENTOS

AUTOR:

CUMANDÁ DEL ROCÍO MEJÍA PERALTA

DIRECTOR:

DR. PIERCOSIMO TRIPALDI

CUENCA, ECUADOR

2010

DEDICATORIA

Dedico este trabajo a mi madre, quien durante todos estos años me ha apoyado y es el pilar fundamental en mi vida.

AGRADECIMIENTO

Agradezco al Dr. Piercosimo Tripaldi por ser quien me ha brindado sus conocimientos de una manera desinteresada, que gracias a él he podido culminar mi trabajo de graduación.

Resumen

El propósito del presente trabajo ha sido establecer modelos de calibración para la determinación de la pectina residual del proceso de clarificación de néctar y jugo de manzana mediante la espectroscopia FTIR-ATR en jugos de fruta luego de su clarificación enzimática. La técnica infrarrojo que combina la obtención de espectros de alta calidad mediante la Transformada de Fourier con la Reflexión Atenuada es bien establecida. En el campo del análisis de alimentos su uso se restringe al campo de Infrarrojo cercano. Los espectros obtenidos fueron registrados entre 1800 a 700 cm^{-1} . Se seleccionaron las mejores bandas para obtener un modelo mediante algoritmos genético implementado por el software Moby Digs el mismo que ha permitido validar los tres modelos mediante BOOSTRAP. Los modelos obtenidos nos permiten predecir el porcentaje de pectina en un jugo.

Abstract

The aim of the present work was to establish calibration models to determine residual pectin of the process apple juice clarification by means of FTIR-ATR spectroscopy. The substrate was fruit juice after enzymatic clarification. Infrared techniques, combined with the achievement of high quality FTIR spectra with Attenuated Total Reflectance, is a well established method. Applications in food analysis, are related to the near Infrared zone. The spectra studied were registered between 1800 a 700 cm^{-1} . The main bands were selected to obtain a model based on genetic algorithms, implemented with Moby Digs software. This strategy allowed us to validate three models using BOOSTRAP. This theoretical approach predicts the pectin percentage in juices.

ÌNDICE DE CONTENIDOS

Contenidos	Página
Dedicatoria.....	ii
Agradecimiento.....	iii
Resumen.....	iv
Abstract.....	v
Índice de contenidos	vi
Índice de tablas.....	ix
INTRODUCCIÓN.....	1
CAPÍTULO I. PECTINAS	
1.1. Generalidades.....	3
1.2. Estructura y propiedades de la pectina.....	4
1.2.1. Propiedades de las pectinas.....	4
1.3. Usos y Aplicaciones.....	5
1.3.1.Las pectinas como estabilizantes	5
1.4. Grado de Metilación.....	6

1.4.1. Pectina de Bajo Contenido de Metilación.....	6
1.4.2. Pectina de Alto Contenido de Metilación.....	6
1.4.3. Pectina de Gelificación Media y Rápida.....	7
1.4.4. Pectina de Gelificación Lenta.....	7
1.5. Métodos para la medición de la graduación.....	9
1.6. Influencia del ácido en las pectinas.....	10

CAPÍTULO II. MÉTODO ESPECTROMETRÍA INFRARROJO

2.1. Definición.....	12
2.2. Propiedades.....	12
2.2.1 Espectrometría de infrarrojos por transformada de Fourier.....	14
2.2.2. Ventajas.....	16
2.2.3. Espectroscopía de Reflectancia Atenuada.....	16
2.3. Usos y Aplicaciones.....	17
2.4. Preparación y análisis de la muestra.....	18

CAPÍTULO III. CALIBRACION

3.1. Definición y Objetivo.....	20
3.2. La familia de la corrección señalada multiplicativa.....	20
3.2.1. Autoscaling y normalización	21
3.3. Regresión Multivariante.....	22
3.3.1. Regresión Multivariante PLS (Partial Least Squares).....	22

3.4. Mínimos Cuadrados.....	23
3.5. Variables espectroscópicas predictoras utilizadas y sus efectos.....	25
3.6. Correlación interna entre variables predictoras y métodos de eliminación	27
3.7. Correlación entre variables predictoras y variables dependientes.	30

CAPÍTULO IV. VALIDACIÓN

4.1. Generalidades.....	32
4.2. Método Leave One out (LOO).....	33
4.3. Método Leave more out.....	33
4.4. Método Bootstrap.....	34
4.5. Regresión en los Componentes Principales.....	34

CAPÍTULO V. RESULTADOS Y DISCUSIÓN

5.1. Parte experimental.....	36
5.2. Resultados y discusión.....	37
5.3. Método PLS.....	43
5.4. Método PCR.....	45

CONCLUSIONES	48
---------------------------	----

BIBLIOGRAFÍA	49
---------------------------	----

ANEXO	51
--------------------	----

ÍNDICE DE TABLAS

Tabla 1. Asignación de bandas de infrarrojo para pectina	15
Tabla 2. Parámetros del algoritmo genético.....	37
Tabla 3. Bandas seleccionadas por el algoritmo genético.....	38
Tabla 4. Parámetros de calidad del modelo.....	38

Mejía Peralta Cumandá del Rocío

Trabajo de Graduación

Dr. Piercosimo Tripaldi

Noviembre, 2010

CUANTIFICACIÓN, MEDIANTE INFRARROJO, DE LA PECTINA RESIDUAL DEL PROCESO DE BIOTECNOLOGÍA DE CLARIFICACIÓN DE NECTAR Y JUGO DE MANZANA

INTRODUCCIÓN

La pectina es un hidrocoloide por excelencia que tiene la propiedad de absorber una gran cantidad de agua. Pertenece al grupo de los polisacáridos y se encuentra en la mayoría de los vegetales, especialmente en frutas como la manzana y algunos cítricos. La pectina juega un papel fundamental en el procesamiento de los alimentos como aditivo espesante y gelificante. Los geles de pectina son importantes para crear o modificar la textura de compotas, jaleas, confites mermeladas y conservas vegetales (Braverman, 1990).

En los jugos de fruta, la presencia de pectina provoca turbidez por lo que los jugos vienen clarificados mediante diferentes técnicas de hidrólisis químicas o enzimáticas (Gummadi y Kumar, 2006). El conocer el contenido de pectina residual, es un parámetro importante que puede ser controlado en esta etapa de la elaboración de los jugos.

La luz infrarroja nos brinda información especial que no podemos obtener de la luz visible. La espectroscopia infrarroja es ampliamente usada en investigación y en la industria como un método simple y confiable para realizar mediciones y controles de calidad. Al medir a una frecuencia específica a lo largo del tiempo, se pueden medir cambios en el carácter o la cantidad de un enlace particular (Heinz ,2009).

La espectroscopía infrarroja en transformada de Fourier, acoplada a un accesorio de Reflectancia total atenuada, es una técnica que ya ha sido utilizada en análisis de alimentos. Sus principales ventajas son la rapidez de la determinación, la amigabilidad ambiental, en cuanto, por lo menos en este caso, no precisa pre-tratamiento con reactivos químicos (Stuart, 2004). La aplicación de FTIR en

determinación de pectina ha asignado las bandas de absorción entre 1800 y 700 cm^{-1} a los correspondientes grupos funcionales (Schulz, H. and Baranska, M. 2007).

En el presente trabajo se han seleccionado las bandas de absorción utilizando algoritmos genéticos, el mismo que permite seleccionar las bandas con una mínima colinealidad entre si y con el mejor poder modelizante en predicción.

CAPÍTULO I

PECTINAS

1.1. Generalidades

La pectina es un coloide por excelencia, que tiene la propiedad de absorber una gran cantidad de agua. Pertenece al grupo de los polisacáridos y se encuentra en la mayoría de los vegetales, especialmente en frutas como la manzana y algunos cítricos (Braverman, 1990).

Las pectinas son de gran interés para la industria de alimentos ya que se utilizan ampliamente como aditivos por sus propiedades espesantes y gelificantes en productos tales como gelatinas, mermeladas y conservas vegetales.

Son uno de los principales constituyentes de la pared celular de los vegetales y forma parte importante de los componentes característicos de los frutos cítricos. Estas llenan los espacios intercelulares en los tejidos vegetales. En los tejidos jóvenes especialmente en los frutos, las pectinas se encuentran en cantidades abundantes formando canales anchos, apartando entre sí a las células (Kertesz Z.I.1998).

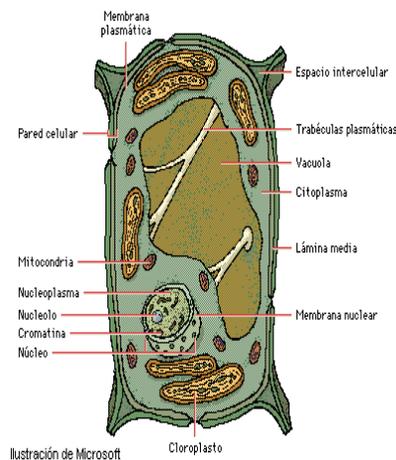


Gráfico 1.1. Célula vegetal.

1.2. Estructura y propiedades de la pectina

Las sustancias pécticas son polímeros lineales del ácido galacturónico, que tienen una parte más o menos amplia de grupos carboxilos.

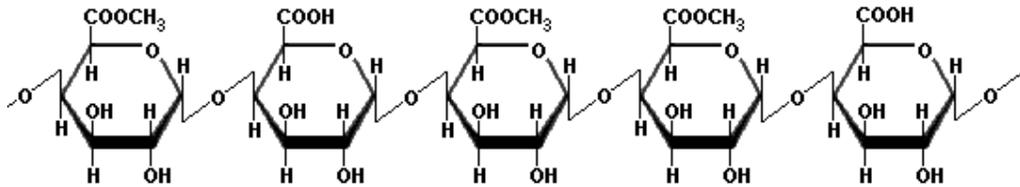


Gráfico 1.2. Pectina.

Fuente: ZAMORA Antonio, 2008, Estructura Química de los Carbohidratos. México, México. <http://www.scientificpsychic.com/fitness/carbohidratos2.html>

La base de su estructura química la constituye el ácido galacturónico, cuyos grupos carboxilos están esterificados por radicales metilos en diferente proporción, lo que le da un grado de metilación del cual dependerá su capacidad de producir geles en condiciones normales con azúcar y ácido.

El grado de metilación mencionado, es lo que hará que a las pectinas se les pueda clasificar entre las de alto contenido de metoxilos o las de bajo contenido de metoxilo (Fizman, 2001.).

1.2.1. Propiedades de las pectinas

La pectina tiene la propiedad de formar geles en medio ácido y en presencia de azúcares. Por este motivo, es utilizada en la industria alimentaria en combinación con los azúcares como un agente espesante.

También son utilizadas en el área alimenticia en concentraciones menores al 1% para dar estabilidad y/o propiedades consistentes.

Los cambios en los valores nutricionales de los productos en que se usan pectinas, son prácticamente imperceptibles. Poseen agentes gelificantes uniformes de gran calidad para alimentos. Son insolubles en alcoholes y parcialmente solubles en jarabes ricos en azúcares. Cuando la pectina es calentada junto con el azúcar se forma una red, que se endurece durante el enfriado (Plocharski y Konopacka 1999).

El grupo de frutas que contienen la suficiente cantidad de pectina para formar un gel es reducido, como son: la pera, higo, la ciruela, cereza, mango, manzana naranja, limón, toronja, albaricoque y el membrillo. Comercialmente, la pectina es fabricada a partir de la pulpa de la manzana y la naranja.

1.3. Usos y Aplicaciones

El uso principal de las pectinas es para la fabricación de jaleas, conservas de frutas y confituras. La pectina es utilizada en la preparación de geles, mermeladas, jaleas, como estabilizantes, etc.

Las industrias de productos de frutas emplean pectina líquida y pectina en polvo, y algunas empresas fabrican su propia pectina. La pectina puede ser utilizada como emulgente para la preparación de ungüentos, polvos, tabletas y otros medicamentos (Amos .2000).

1.3.1. Las pectinas como estabilizantes

Las pectinas se comportan muy bien como estabilizantes de las caseínas frente a los tratamientos térmicos a pH ácido.

Las pectinas, como muchos otros polisacáridos, se hinchan muy rápidamente con el agua, y por eso cuando se añaden de golpe, y especialmente si se añade agua sobre el sólido, forman agregados difíciles de disolver.

La solución es separar las partículas cuando se mezcla el polisacárido con el agua, con sistemas mecánicos o mezclando previamente con otro material no acuoso.

Son relativamente inestables desde el punto de vista químico, especialmente a temperaturas elevadas. Su máxima estabilidad está en torno a pH 4. Pueden perder grupos metoxilo, hidrolizarse, y en medio neutro o alcalino romperse por beta-eliminación. Esto afecta muy negativamente a su viscosidad y capacidad de formación de geles (Pottí D. 1999).

1.4. Grado de Metilación

Como parte de la estructura de la pectina se encuentran los grupos carboxilos, los cuales son esterificados por radicales metilo, a éstos se los conoce como metilación de una pectina.

La importancia de la metilación radica con la propiedad más importante de la pectina que es la capacidad de formar geles (Braverman.1990).

1.4.1. Pectina de Bajo Contenido de Metilación

Tiene un grado de metilación inferior al 50%. No forma geles con adición de azúcar y ácido; pero sí en presencia de iones divalentes como el calcio, magnesio y otros.

Este tipo de gel se denomina "gel iónico" en este caso la pectina tiene gran cantidad de grupos $-\text{COOH}$. Si se agrega una sustancia que pueda formar un enlace entre éstos se formará un gel.

Los geles iónicos se emplean como alimento dietético de bajo valor calórico, útiles para diabéticos.

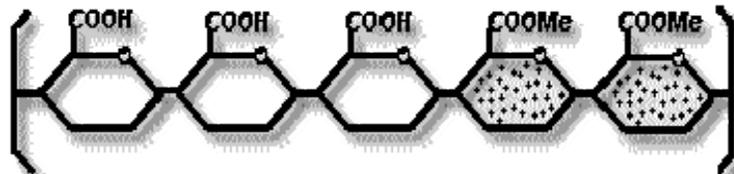


Gráfico 1.3. Pectina con 40% GE

Fuente: ZAMORA Antonio, 2008, Estructura Química de los Carbohidratos. México, México. <http://www.scientificpsychic.com/fitness/carbohidratos2.html>

1.4.2. Pectina de Alto Contenido de Metilación

Son aquellas que forman geles con un grado de metilación superior al 50%. Cuando la pectina tiene un alto contenido de metoxilo el grado de hidratación se reduce mediante la adición de azúcar y la disminución de carga eléctrica que se consigue por un aporte de iones (hidrógeno) H^+ , o dicho de otra forma de ácido, el enlace de unas moléculas pécticas a otras queda básicamente asegurado, en este caso por

uniones de hidrógeno entre grupos hidroxilos, éstos son enlaces débiles y los geles pécticos de este tipo se caracterizan por una gran plasticidad, esto se debe a la movilidad de unas moléculas con relación a otras (Avila J.2001).

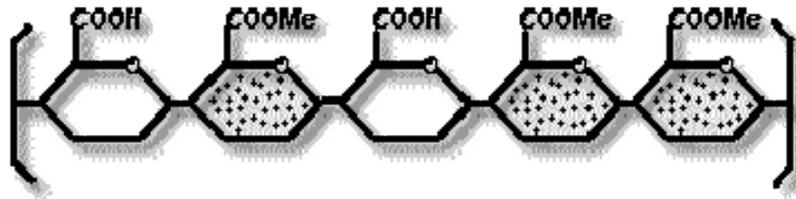


Gráfico 1.4. Pectina con 60% GE.

Fuente: ZAMORA Antonio, 2008, Estructura Química de los Carbohidratos. México, México. <http://www.scientificpsychic.com/fitness/carbohidratos2.html>

1.4.3. Pectina de Gelificación Media y Rápida

Con un grado de metilación de por lo menos el 70%, que forma geles con adición de azúcar y ácidos a pH de 3,0 - 3,5; y a temperaturas superiores a los 75°C.

Esta pectina produce el espesamiento o gelificación al poco tiempo de ser agregada. Esto mantiene las frutas y las partículas de pulpa uniformemente en todo el lote o en los envases, evitando el problema de "flotación".

1.4.4. Pectina de Gelificación Lenta

Con un grado de metilación entre el 50 - 70%, que forma geles con azúcar y ácido a pH óptimo entre 2,8 - 3,2 y su gelificación puede empezar a temperaturas menores a 85°C. El uso de esta pectina evita que la jalea se solidifique antes de ser colocada en los envases (Braverman ,1990).

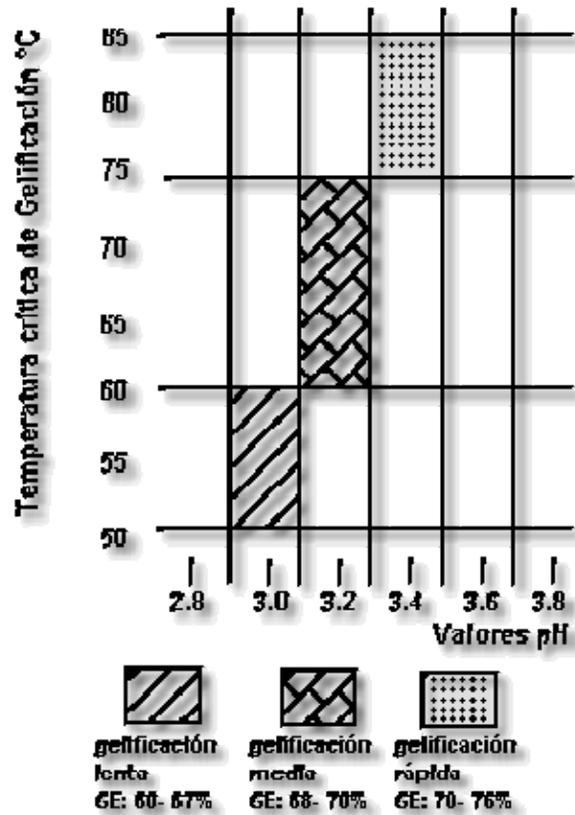


Gráfico 1.5. Intervalos de temperatura y pH a los que gelifican pectinas de alto metoxilo.

Fuente: Universidad Nacional de Colombia, 1999, Espectrometría Infrarroja. Bogotá Colombia. www.virtual.unal.edu.co/cursos/

El gráfico 1.5. presenta los intervalos de temperatura y pH a los cuales gelifican las pectinas de alto metoxilo pero de diferente velocidad de gelificación.

En el gráfico 1.6 se sintetiza la interdependencia de los tres parámetros, pectina, pH y Brix. Se observa que mezclas con altos °Bx gelificarán con mayor facilidad a pH de 3,2 sin necesitar altas cantidades de pectinas y de manera opuesta, mezclas de poco contenido en °Bx necesitan pH mas ácidos (pH cercanos a 2,8) y con altas cantidades de pectina (Pectinas Universidad Nacional de Colombia).

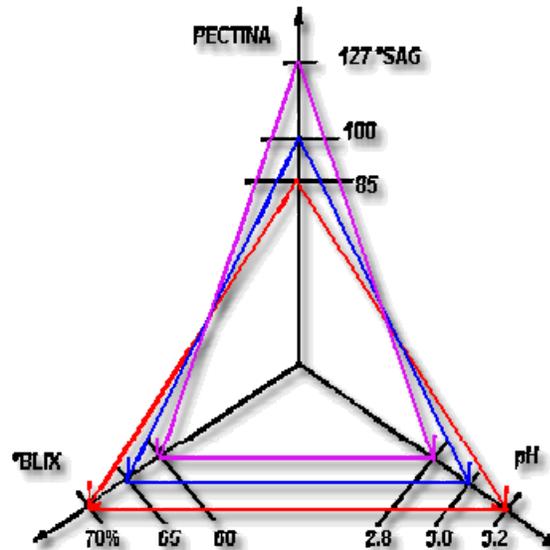


Gráfico 1.6. Equilibrio de los ingredientes para lograr la gelificación.

Fuente: Universidad Nacional de Colombia, 1999, Espectrometría Infrarroja. Bogotá Colombia. www.virtual.unal.edu.co/cursos/

1.5. Métodos para la medición de la graduación

La graduación de una pectina es medida por la consistencia o fuerza del gel obtenido al emplear una formulación determinada. Entre los varios métodos usados para realizar esta medición, se halla la medida de los grados SAG de donde ha tomado el nombre de método SAG (IFT). Este método está recomendado por el Comité de Expertos IFT (International Food Technologist) para la normalización de pectinas.

Este método usa un rigelímetro el cual dispone de un tornillo que permite medir la deflexión de un gel normalizado. Este gel ha sido preparado con la pectina a la cual se le desea medir sus grados SAG. El gel normalizado posee las siguientes características:

- Materia seca refractométrica: 65%
- pH de la masa: 2.3 a 2.4
- Fuerza gelificante normal: 23,5% de penetración medida en un rigelímetro (Pectinas. Universidad Nacional de Colombia).

1.6. Influencia del ácido en las pectinas

El fenómeno de la gelificación está estrechamente ligado a la acidez activa, expresada como pH, que tiene significado y valores diversos de la acidez titulable o total.

Algunas sales contenidas en la fruta, llamadas sales tampones o buffers, tienen poder estabilizante sobre los iones ácidos y básicos de una solución y reducen el efecto de la acidez total. En una solución de alto contenido de ácido, la presencia de sales tampones disminuye la acidez activa e influye negativamente sobre el proceso de gelificación, que requiere el ajuste del pH a valores bien delimitados.

Para cada tipo de pectina y para cada valor de concentración de azúcar existe un valor de pH al cual corresponde el óptimo de gelificación. Este valor óptimo está comprendido entre límites estrechos, que van, para pectinas de alto metoxilo entre $\text{pH}=2,8$ a $3,7$ (gráfico 1.6). Para valores superiores a $3,7$ (o sea para una acidez activa más débil) la gelificación no tiene lugar, mientras que para valores inferiores a $2,8$ (acidez activa más fuerte) se produce la SINÉRESIS.

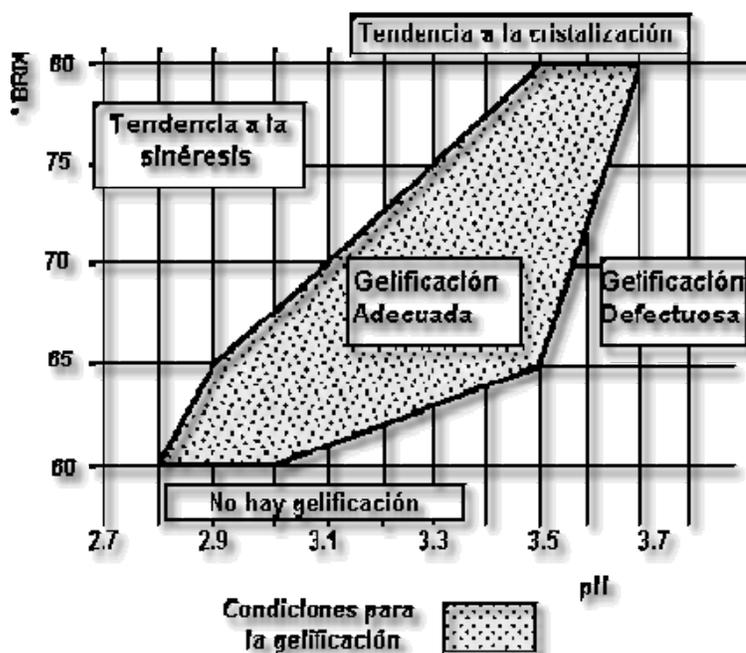


Gráfico 1.7. Condiciones de gelificación de las pectinas de alto metoxilo

El fenómeno de la sinéresis se manifiesta por una exudación de jarabe y es debido al endurecimiento excesivo de las fibras de pectina, que pierden la elasticidad necesaria para retener los líquidos del gel. Entre los factores que disminuyen este fenómeno están el aumento del pH, de la concentración de pectina y los sólidos solubles. De otro lado la sinéresis se ve aumentada por el uso de pectina de rápida gelificación y la adición de jarabe de glucosa.

La exacta valoración del pH es extremadamente importante, ya que una mínima diferencia en la zona del óptimo de gelificación influye definitivamente sobre la rigidez, consistencia y grado de sinéresis de un gel. El modo más práctico para dosificar el ácido es efectuar una pequeña prueba tentativa, De una determinada cantidad de la pulpa o jugo a elaborar, se mide el pH y se lo lleva, con adecuada adición de ácido, a un valor de 0,1 mas bajo del pH considerado para el producto terminado. Por ejemplo si se desea obtener una mermelada de pH 3.2, se calcula la cantidad de ácido que se debe agregar a una muestra de peso conocido para ajustarle el pH a 3.1. El ácido cítrico generalmente es usado en solución al 30% peso-volumen (300g. de ácido seco en un litro de solución), que permite un fácil control de la dosificación.

El ácido debe ser introducido al final de la cocción ya que con esto se crean las condiciones necesarias para la gelificación y se inicia el proceso. Su adición anticipada provocaría fenómenos de pre-gelificación que dañarían el resultado final de la elaboración. Los ácidos más usados son el cítrico, el tartárico y más raramente el láctico y el fosfórico. El ácido cítrico es considerado generalmente más satisfactorio por su agradable sabor; el ácido tartárico es más fuerte, pero tiene un sabor menos ácido (Ahmed, 2002).

CAPÍTULO II

METODO DE ESPECTROMETRIA INFRARROJO

2.1. Definición

El infrarrojo es un tipo de luz que no se puede ver a simple vista. Los ojos pueden solamente ver lo que se llama luz visible. La luz infrarroja brinda información especial que no se puede obtener de la luz visible.

Todas las cosas tienen algo de calor e irradian luz infrarroja. Incluso las cosas que pensamos que son muy frías, como un cubo de hielo, irradian algo de calor. Los objetos fríos irradian menos calor que los objetos calientes.

Entre más caliente sea algo, más es el calor irradiado y entre más frío es algo, menos es el calor irradiado.

Los objetos calientes brillan más luminosamente en el infrarrojo porque irradian más calor y más luz infrarroja. Los objetos fríos irradian menos calor y luz infrarroja, apareciendo menos brillantes en el infrarrojo (Heinz ,2009).



Gráfico 2.1. Espectrofotómetro.

2.2. Propiedades

La lámpara infrarroja es un tipo de lámpara incandescente reforzada con propiedad de radiación infrarroja. Emite menos rayos visibles que una lámpara normal, pero la lámpara infrarroja está diseñada para transformar energía de entrada en rayos infrarrojos con energía térmica y tener una mayor vida.

Los métodos espectroscópicos comúnmente usados para medir las propiedades de los enlaces, tienen como base el hacer incidir radiación electromagnética sobre la molécula a estudiar.

Cuando la radiación infrarroja incide sobre una muestra, es capaz de provocar cambios en los estados vibracionales de las moléculas constituyentes de la misma. La absorción de radiación por parte de una muestra es indicativa del tipo de enlaces y grupos funcionales presentes en la misma. Para que una vibración aparezca en el espectro infrarrojo, la molécula debe someterse a un cambio en su momento dipolar durante la vibración.

Con el fin de hacer medidas en una muestra, se transmite un rayo monocromo de luz infrarroja a través de la muestra, y se registra la cantidad de energía absorbida. Repitiendo esta operación en un rango de longitudes de onda de interés (por lo general, $4000\text{--}400\text{ cm}^{-1}$) se puede construir un gráfico. Al examinar el gráfico de una sustancia, un usuario experimentado puede obtener información sobre la misma.

Esta técnica funciona casi exclusivamente en enlaces covalentes, y se usa mucho en química, en especial en química orgánica. Se pueden generar gráficos bien resueltos con muestras de una sola sustancia de gran pureza. Como un ejemplo de los gráficos que se puede generar a continuación el espectro de absorbancia del ácido láctico (Wiley y Sons, 2000).

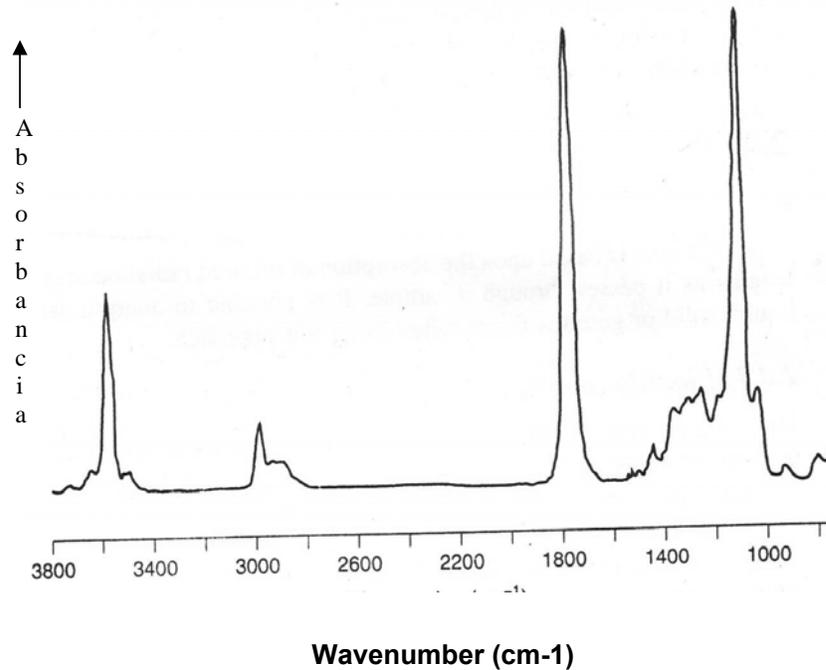


Gráfico 2.2. Espectro de absorbancia del ácido láctico

Fuente: STUART Barbara, 2004, *Infrared Spectroscopy. Fundamentals y Applications*. West Sussex, England.

2.2.1. Espectrometría de infrarrojos por transformada de Fourier

La espectrometría infrarroja por transformada de Fourier (FTIR) es una técnica de análisis para obtener el espectro infrarrojo con mayor rapidez. En lugar de registrar los datos variando la frecuencia de luz infrarroja monocromática, se guía la luz IR (con todas las longitudes de onda de pista utilizada) a través de un interferómetro.

Después de pasar por la muestra, la señal medida da el interferograma. La realización de una transformada de Fourier de la señal produce un espectro idéntico al de la espectrometría infrarroja convencional (dispersiva).

Los espectrofotómetros FTIR son más baratos que los convencionales, porque es más simple construir un interferómetro que un monocromador. Además, la medida de un solo espectro es mucho más rápida en esta técnica, debido a que la información de todas las frecuencias se toman al mismo tiempo.

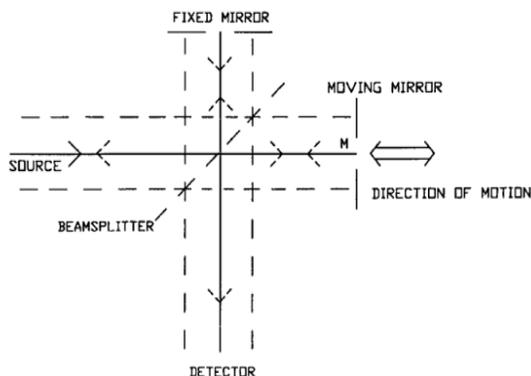


Gráfico 2.3. Interferómetro de Michelson.

Fuente: FERRE Joan, 2006, Interferómetro de Michelson. Tarragona, España.

Esto permite hacer múltiples lecturas de una sola muestra y obtener un promedio, lo que aumenta la sensibilidad del análisis. Debido a sus múltiples ventajas, casi todos los modernos espectrofotómetros de infrarrojos son FTIR (Stuart, 2004).

Se sabe bien que el espectrofotómetro de FTIR es una excelente herramienta para el análisis estructural y cuantitativo de polisacáridos, incluso las pectinas.

La asignación para las bandas de infrarrojo más característico de pectina están expuestas en la siguiente tabla:

Tabla 1. Asignación de bandas de infrarrojo para pectina

Longitud de onda (cm ⁻¹)	Asignación
1745	$\nu(\text{C}=\text{O})$
1605	$\nu_{\text{as}}(\text{COO})$
1444	$\delta(\text{CH})$
1419	$\nu(\text{COO}^-)$
1368	$\delta(\text{CH}_2), \nu(\text{CC})$
1335	$\delta(\text{CH}), \text{ring}$
1150	$\nu(\text{C}-\text{O}-\text{C}), \text{ring}$
1107	$\nu(\text{CO}), \nu(\text{CC}), \text{ring}$
1055	$\nu(\text{CO}), \nu(\text{CC}), \delta(\text{OCH})$
1033	$\nu(\text{CO}), \nu(\text{CC}), \nu(\text{CCO})$
1018	$\nu(\text{CO}), \nu(\text{CC}), \delta(\text{OCH}), \text{ring}$
1008	$\nu(\text{CO}), \nu(\text{CC}), \delta(\text{OCH}), \text{ring}$

972	OCH ₃
963	δ(C=O)

Fuente: Kacurakovà et al.,2001; Schulz and Baranska, 2007. Identification and Quantification of valuable plant substances by IR and Raman spectroscopy.

2.2.2. Ventajas

Los instrumentos FTIR tienen significativas ventajas sobre los viejos instrumentos. Dos de estas son las ventajas de Fellgett (o múltiples) y la Jacquinot (o rendimiento). La ventaja Fellgett es debido a una mejora en el rango señal/ruido en unidad de tiempo proporcional a la raíz cuadrada del número de elementos de la resolución que se monitorea. Este resultado del número grande de elementos de la resolución comienza simultáneamente con el monitoreo. Además, porque la espectrometría FTIR no requiere el uso de una abertura u otro dispositivo restringiendo el rendimiento de la fuente total, puede pasarse continuamente a través de la muestra. Este resulta en una ganancia sustancial en energía para el detector, de aquí se traslada la más alta señal y mejora el rango señal/ruido. Este es conocido como una ventaja Jacquinot.

Otra validez o fuerza que tiene la espectrometría FTIR es la ventaja de la velocidad. El espejo tiene la ventaja de moverse con rapidez a distancias cortas y esto junto con el rango señal/ruido son las ventajas Fellgett y Jacquinot, esto hace posible obtener espectros en un tiempo escalar de milisegundos.

2.2.3. Espectroscopia de Reflectancia Atenuada

La espectroscopia de Reflectancia Atenuada (ATR) utiliza los fenómenos de la reflectancia interna total. Un rayo de radiación entra a un cristal el cual experimenta una reflectancia total interna cuando el ángulo de incidencia en la interfase entre la muestra y el cristal es más grande que el ángulo crítico, donde el último es una función del índice refractivo de las dos superficies. El rayo penetra a una fracción de una longitud de onda más allá de una superficie reflectada y cuando un material que selectivamente absorbe la radiación está en contacto estrecho con la superficie reflectada, el rayo pierde energía donde se absorbió el material.

El resultado de la radiación atenuada es medida y ploteada como una función de la longitud de onda por el espectrofotómetro y da picos de las características espectrales de las muestras. La profundidad de penetración de la espectroscopia ATR es una función de la longitud de onda (λ); el índice refractivo del cristal, n_2 , y el ángulo de radiación incidente (θ). Los cristales usados en las celdas ATR están hechos de materiales que tienen baja solubilidad en agua y son de un alto índice de refractibilidad. Tales materiales incluyen, ZnSe, Ge y KRS-5. Los diferentes diseños de las celdas ATR permiten que las muestras líquidas y sólidas sean determinadas.

Es también posible colocar un flujo continuo de celdas ATR pero incluyendo una entrada y salida del aparato. Esto permitirá un flujo continuo de las soluciones a través de las celdas y permitirá que los cambios espectrales sean monitoreados con tiempo (Stuart, 2004).

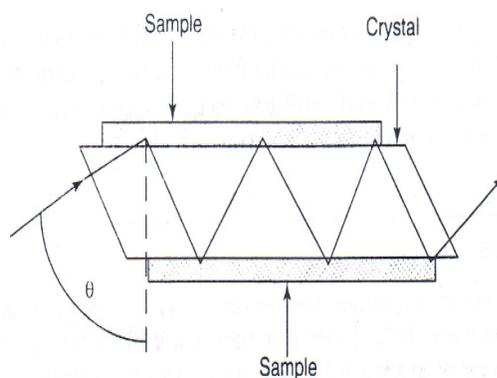


Gráfico 2.4. Esquema típico de una célula de Reflectancia Total Atenuada.

Fuente: STUART Barbara, 2004, Infrared Spectroscopy. Fundamentals y Applications.

2.3. Usos y Aplicaciones

La espectroscopía infrarroja es ampliamente usada en investigación y en la industria como un método simple y confiable para realizar mediciones y controles de calidad. Los instrumentos son en la actualidad pequeños y pueden transportarse fácilmente.

La espectrometría del infrarrojo es sumamente útil para determinaciones cualitativas de compuestos orgánicos y para deducir estructuras moleculares a partir de sus grupos funcionales tanto de compuestos orgánicos como inorgánicos.

En el análisis cualitativo la espectroscopía de infrarrojo puede usarse para la identificación de sustancias puras o para la absorción, localización e identificación de impurezas.

Tanto desde el punto de vista instrumental como de sus aplicaciones es conveniente dividir la región infrarroja en tres regiones denominadas infrarrojo cercano (NIR), infrarrojo medio (MIR) e infrarrojo lejano (FIR). La gran mayoría de las aplicaciones analíticas clásicas de la espectroscopía infrarroja se basan en el empleo del infrarrojo medio (4000-600 cm^{-1}) y el infrarrojo cercano, que proporciona la posibilidad de convertir esta técnica en una técnica cuantitativa.

La técnica de transformada de Fourier supuso una revolución en la espectroscopía en general y particularmente en este tipo de espectroscopía, permitiendo la obtención de espectros de forma rápida, precisa y con relaciones Señal/Ruido (S/N) elevadas (Wiley y Sons, 2000).

Entre las aplicaciones de la espectroscopía de Infrarrojos tenemos:

- Caracterización e identificación de materiales:
 - Polímeros y plásticos
 - Sólidos inorgánicos (minerales, catalizadores, materiales compuestos...)
- Análisis de productos farmacéuticos y de síntesis.
- Análisis de contaminantes
- Ciencia Forense (identificación)
- Biomedicina (análisis de tejidos)
- Conservación artística (análisis de pigmentos, materiales utilizados...)
- Industria del reciclaje (identificación de materiales poliméricos)
- Agricultura y alimentación (IR cercano)
- Seguimiento de procesos químicos
 - Polimerizaciones, curado, ...
 - Reacciones catalíticas (Alhambra, 1999).

2.4. Preparación y análisis de la muestra

Las muestras gaseosas requieren poca preparación más allá de su purificación, pero se usa una celda de muestra con una larga longitud de celda (usualmente 5-10 cm) pues los gases muestran absorbancias relativamente débiles.

Las muestras líquidas se pueden disponer entre dos placas de una sal de alta pureza (comúnmente cloruro de sodio, o sal común), aunque también se utilizan otras sales tales como bromuro de potasio o fluoruro de calcio. Las placas son transparentes a la luz infrarroja y no introducirán líneas en el espectro.

Algunas placas de sal son altamente solubles en agua, y así la muestra, agentes de lavado y similares deben estar completamente anhidros (sin agua).

Las muestras sólidas se pueden preparar principalmente de dos maneras. La primera es moler la muestra con un agente aglomerante para la suspensión (usualmente nujol) en un mortero de mármol o ágate.

Una fina película del agente aglomerante se aplica en las placas de sal y se realiza la medición (Alhambra, 1999).

En un espectrofotómetro IR, la muestra se coloca entre el interferómetro y el detector. La muestra absorbe ciertas frecuencias de la radiación y por tanto el interferograma de la muestra corresponde al de la fuente menos el de la muestra. En la práctica primero se registra el interferograma de la referencia constituida por la celda y el blanco para transformarlo en el espectro. Luego se registra el interferograma de la muestra en la misma celda y se transforma en espectro. El cociente que resulta de dividir el segundo espectro entre el primero corresponde al espectro de la muestra (Wiley y Sons, 2000).

CAPÍTULO III

CALIBRACIÓN

3.1. Definición y Objetivo

Las técnicas de calibración multivariante, tales como regresión parcial de mínimos cuadrados (PLS, por sus siglas en inglés), son capaces de entenderse con procesamiento de espectros y su cuantificación. Una vez que la calibración ha sido realizada, la predicción de futuras muestras con concentraciones es una multiplicación de matrices.

El objetivo de la calibración multivariante es predecir la propiedad de interés utilizando respuestas instrumentales no selectivas, como por ejemplo el espectro de una muestra. El tipo de modelo multivariante más común es el lineal: la concentración del analito se predice multiplicando cada absorbancia del espectro por un coeficiente y sumando (Joan Ferré, Universidad Rovira i Virgili).

3.2. La familia de la corrección señalada multiplicativa

La técnica de la corrección señalada multiplicativa (MSC) se introdujo por Martens en 1983 y mas allá elaborado por Geladi (1985). Originalmente MSC significa “espacio multiplicativo de corrección” pero la abreviación ha cambiado el significado durante los años porque esto se ha encontrado útil para problemas de tipo multiplicativo. El concepto básico del MSC es quitar los no lineales en los datos causado por la dispersión de partículas en las muestras.

El funcionamiento de MSC es dividido en dos pasos: la estimación de los coeficientes de la corrección,

$$X_{\text{org}} = b_0 + b_{\text{ref},1} x_{\text{ref}} + e \quad (2.2)$$

Y corrección de los espectros

$$X_{\text{corr}} = \frac{X_{\text{org}} - b_0}{b_{\text{ref},1}} \quad (2.3)$$

Donde los **b's** son los coeficientes de la corrección, **e** es el error, y x_{org} , x_{ref} , y x_{corr} son el original, la referencia, y el espectro corregido respectivamente.

3.2.1. Autoscaling y normalización

La ecuación básica para el autoscaling de corrección tiene la misma fórmula como la ecuación 2.2 para MSC:

$$X_{\text{corr}} = \frac{X_{\text{org}} - a_0}{a_1} \quad (2.7)$$

Para la normalización, a_0 se igualará siempre el cero, mientras a_1 , depende del tipo de normalización realizado. Hay dos tipos de normalización usados en los datos espectrales: el total de la suma absoluta (Distancia del Taxi) y la raíz cuadrada del total de la suma al cuadrado (Distancia Euclídeana).

Distancia del taxi:

$$a_1 = \sum_{m=1}^M |X_{\text{org},m}| \quad (2.8)$$

Distancia de Euclídean:

$$a_1 = \sqrt{\sum_{m=1}^M X_{\text{org},m}^2} \quad (2.9)$$

Para la técnica de SNV (Variante Normal Estándar), el a_0 es el valor medio del espectro, mientras el a_1 es la desviación normal del espectro:

$$a_0 = \frac{\sum_{m=1}^M X_{\text{org},m}}{M} = \bar{X}_{\text{org}} \quad (2.10)$$

$$a_1 = \sqrt{\frac{\sum_{m=1}^M (X_{org,m} - a_0)^2}{M-1}} = S X_{org} \quad (2.11)$$

En comparación con MSC la técnica SNV no necesita un espectro de referencia, y así ninguna decisión del usuario para el cómputo. Sin embargo, la variación normal estándar es más propensa al ruido que el MSC más robusto.

En general, SNV y los MSC originales llevan los resultados muy similares; es decir los resultados del espectro y los resultados de regresión.

Como en el caso para el MSC y EMSC, el SNV ha transformado los datos de una manera tal, que los datos corregidos generalmente tienen una relación más lineal. La base de Distancia de Euclidean da a menudo los resultados similares a SNV, ya que está basado en la raíz cuadrada de la suma al cuadrado como en SNV (Da-Wen Sun 2009).

3.3. Regresión Multivariante

3.3.1. Regresión Multivariante PLS (Partial Least Squares).

El método PLS es un método para construcción de modelos de predicción cuando los factores son muchos y altamente colineales. El método PLS es una herramienta útil cuando el objetivo es la predicción y no existe una necesidad práctica de limitar el número de factores medidos.

Es un método de calibración multivariante, el cual es capaz de relacionar la concentración de elementos directamente con espectros medidos particularmente con NIR. Las variables espectrales son colectadas como una matriz X con número de filas igual al número de muestras y el número de columnas igual al número de canales en el espectro.

Este método consiste en obtener el máximo de estabilidad del modelo de predicción calculado. El método es utilizado cuando la relación objeto/variable es inferior a uno, en presencia de la variable puesta en correlación entre ellos.

Esta posibilidad es particularmente interesante cuando la respuesta Y es puesta en correlación entre ellos, en cuánto es posible obtener un único modelo a interpretar y no tantos modelos por separado. No obstante si la respuesta mide la cantidad efectivamente independiente entre ellos no hay ventaja para desarrollar un único modelo de regresión. En este caso, PLS tiende a dar un número de componentes, al menos para el número de respuestas analizadas.

Desde el punto de vista matemático, el método PLS utiliza –interactivamente– copias del componente principal X y Y (cuando Y contiene más de una respuesta variable) y entre estos la máxima correlación posible.

(Todeschini, 1998)

3.4. Mínimos Cuadrados

El método de regresión de mínimos cuadrados es un método muy popular en aplicaciones industriales de modelación suave.

En términos de matrices, el problema de la regresión lineal con el método de mínimos cuadrados es representado en el siguiente modelo:

$$y = X\beta + e$$

donde β es el valor del coeficiente verdadero a calcular, X es la matriz del modelo y e es el vector de los errores.

En términos no matriciales, podemos describir explícitamente la relación entre la respuesta y del i -ésimo objeto y su descripción de la variable independiente con la combinación lineal:

$$y_i = b_0 + b_1x_{i1} + b_2x_{i2} + \dots + b_px_{ip}$$

La varianza del vector del coeficiente es una medida de estabilidad del modelo de regresión y es dado en la expresión:

$$V(\mathbf{b}_{OLS}) = \sigma^2 \cdot \text{tr}(\mathbf{X}^T \mathbf{X})^{-1} = \sigma^2 \cdot \sum_{j=1}^p (1/\lambda_j)$$

donde σ^2 es la varianza debido al error experimental, X es la matriz de modelo, λ son los autovalores de la matriz $X^T X$, la matriz de dispersión del modelo. Se observa que la varianza del coeficiente depende del producto de dos contribuciones

separadas, el primero (σ^2) que depende solo de la respuesta, y el segundo que depende solo del espacio de los predictores. Una vez calculado el coeficiente del modelo, es posible determinar el vector de la respuesta calculada \hat{y} .

$$\hat{y} = \mathbf{Xb}$$

Esta última expresión forma el modelo de regresión, constituido entonces de los descriptores definidos en la matriz del modelo y del coeficiente de regresión \mathbf{b} obtenidos. Por cuanto interesa el coeficiente de regresión, de particular interés es el coeficiente de regresión estandarizado b_j' , obtenido de la expresión:

$$b_j' = \frac{b_j \cdot s_j}{s_y}$$

Donde s_y y s_j son la desviación estándar de la respuesta y del j -ésimo predictor. Este coeficiente corresponde al coeficiente de regresión calculado por la variable, escalado y representado los pesos de cada variable del modelo de regresión. Estos coeficientes dan entonces una medida de importancia de cada variable en el modelo y la suma de sus cuadrados pueden ser consideradas una medida de la complejidad del modelo.

Los elementos interesantes de una matriz son los elementos diagonales h_{ii} , llamadas también función de varianza del objeto i -ésimo, para los cuales valen la siguiente propiedad:

$$h_{\min} = 1/n \quad \sum_i h_{ii} = p' \quad \bar{h} = p'/n \quad h^* > 3p'/n$$

Donde \mathbf{h} es el valor de control mas allá del cual el dato puede ser considerado influyente en determinar el parámetro del modelo. El valor de \mathbf{h} para el i -ésimo objeto es dado de:

$$h_{ii} = \mathbf{X}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}_i$$

El valor mínimo de influencia (*leverage*) se encuentra en correspondencia al centro del espacio del modelo, mientras el valor máximo es relativo al punto más distante del centro del modelo. Esta propiedad de influencia es de gran importancia en cuanto permite utilizar el valor de \mathbf{h} para valorar la distancia de cada objeto al centro del espacio del modelo.

Esto se puede obtener con objetos cuyo valor de la variable son obtenidos mediante un diseño experimental, pero no pasa con objetos definidos en espacio experimental no controlado directamente de lo experimentado (Todeschini, 1998).

3.5. Variables espectroscópicas predictoras utilizadas y sus efectos

Para cada modelo de regresión se asume, como situación de referencia o modelo de orden cero, la cantidad referida al valor medio de la respuesta, dicha suma total de los cuadrados (TSS) es:

$$TSS = \sum_i (y_i - \bar{y})^2$$

Un modelo de regresión es mejor cuanto más pequeña es la suma de los cuadrados de los residuos (RSS, suma residual de cuadrados):

$$RSS = \sum_i (y_i - \hat{y}_i)^2$$

Obtenida de la diferencia entre cada valor de la respuesta experimental y la respuesta calculada.

Al mismo tiempo un modelo de regresión es mejor cuanto más grande es la suma del cuadrado del modelo. (MSS, Modelo Suma de Cuadrados):

$$MSS = \sum_i (\hat{y}_i - \bar{y})^2$$

Vale por lo tanto la siguiente ecuación:

$$TSS = MSS + RSS$$

$$\frac{MSS}{TSS} = R^2 = 1 - \frac{RSS}{TSS}$$

Donde R^2 es el coeficiente de determinación y multiplicado por 100 representa la varianza porcentual explicado por el modelo.

La cantidad \mathbf{R} , o comúnmente llamado también \mathbf{r} , es conocido como coeficiente de correlación múltiple y representa la correlación entre la respuesta experimental \mathbf{y} , y el predictorio \mathbf{x}_j . Es definida como:

$$R = \sqrt{1 - \frac{RSS}{TSS}}$$

Las propiedades fundamentales de R son las siguientes:

1. $0 \leq R_{y(1, \dots, p)} \leq 1$
2. Si $R_{y(1, \dots, p)} = 0 \rightarrow R_{y(j)} = 0$ pero $j = 1, \dots, p$, eso es si la correlación es nula, entonces la correlación de la copia entre la respuesta y y cada predictor es nulo.
3. Si $R_{y(1, \dots, p)} = 1$, ahora la respuesta y es una perfecta combinación lineal del p predictor.
4. en caso de un solo predictor x , $R_{y(1)} = |r_{yx}|$, eso es el coeficiente de correlación de y con el solo predictor coincide con el valor absoluto del coeficiente de correlación entre y y x .
5. $R^2_{y(1)} \leq R^2_{y(1,2)} \leq \dots \leq R^2_{y(1,2,3)} \leq R^2_{y(1, \dots, p)}$ eso es para cualquier aumento del número de predicción, el coeficiente de correlación múltiple nunca pueden disminuir.

En la suma al parámetro precedente, es posible definir un parámetro, llamado R^2 ajustado, que tiene también el grado de libertad con que es calculado R^2 :

$$R^2_{\text{adj}} = 1 - \frac{\text{RSS}_p / (n-p')}{\text{TSS} / (n-1)} = (1 - R^2) \cdot \left[\frac{n-1}{n-p'} \right]$$

El parámetro R^2 ajustado es el estado ideal para valorar la conveniencia para agregar una variable al modelo, visto por la propiedad 5, esto no puede ser hecho con el parámetro R . De hecho, el parámetro R^2_{adj} presenta un máximo para la complejidad óptima del modelo y desciende cuando la adición de una variable al modelo no es adecuadamente compensada de un significativo aumento de R . En cambio el parámetro anterior mide exclusivamente la capacidad del modelo para unirlo contemporáneamente con todos los objetos utilizados para la construcción del modelo; estas medidas se llaman de montaje (o fitting). Para obtener del parámetro que mide la capacidad predictiva del modelo obtenido, es necesario utilizar la técnica de validación.

En este caso la suma del cuadrado del residuo viene efectuada utilizando el valor de la respuesta predecida en lugar de la respuesta calculada. Esta grandezza, formalmente es idéntica a RSS, y se llama PRESS (Error predictivo de la suma de cuadrados) y viene calculada por:

$$\text{PRESS} = \sum_i (y_i - \hat{y}_{i/i})^2$$

Donde $\hat{y}_{i/i}$ indica el valor mencionado para el i -ésimo número de un modelo en que el número no es tomado en consideración para la construcción del modelo mismo. Utilizando PRESS en lugar de RSS, obtenemos el porcentual de varianza explicado del modelo en predicción:

$$Q^2 = R^2_{cv} = 1 - \frac{PRESS}{TSS}$$

Diversamente de R^2 y en modo similar a R^2_{adj} , este último parámetro presenta un máximo para la complejidad óptima del modelo y desciende cada vez que agregamos al modelo una variable no predictiva. No obstante, diversamente del parámetro R^2_{adj} , R^2_{cv} es explícitamente valorado respecto al poder predictivo del modelo y no al nivel de montaje del modelo.

Dos parámetros útiles asociados a RSS y PRESS son respectivamente, SDEC (El error de desviación normal en el cálculo, o también SEC) y SDEP (El error de desviación normal en la predicción, o también SEP), definido como:

$$SDEC = \frac{\sqrt{RSS}}{n} \quad \text{y} \quad SDEP = \frac{\sqrt{PRESS}}{n}$$

Este parámetro tiene la ventaja de ser dimensionalmente comparable con la respuesta estudiada (Todeschini, 1998).

3.6. Correlación interna entre variables predictoras y métodos de eliminación

La viabilidad de un modelo de regresión debe ser siempre controlada mediante una prueba entera, métodos y los parámetros en grado de una información completa y detallada en el modelo calculado.

En particular, es usado el residuo y el valor de influencia (*leverage*) para obtener una serie de gráficos y de pruebas muy eficaces para valorar el comportamiento de puntos en el modelo de regresión construido.

El gráfico más común utilizado en el diagnóstico de regresión es el que permite la comparación entre la respuesta experimental y la respuesta calculada (y/o la respuesta en predicción, gráfico.3.2).

El residuo ordinario de un modelo de regresión (los errores) son definidos como la diferencia entre el valor de la respuesta experimental y aquella calculada del modelo:

$$r_i = y_i - \hat{y}_i$$

Un aspecto importante del diagnóstico involucra el control de las hipótesis cuya distribución de los errores que deben tener un valor del promedio igual a cero es una distribución normal entorno al valor medio. La desviación de estas hipótesis son índices de problemas de varias naturalezas del modelo calculado.

Los residuos pueden ser todos aquellos calculados en diversa forma (del modelo calculado o en predicción, estandarizado o no, etc.)

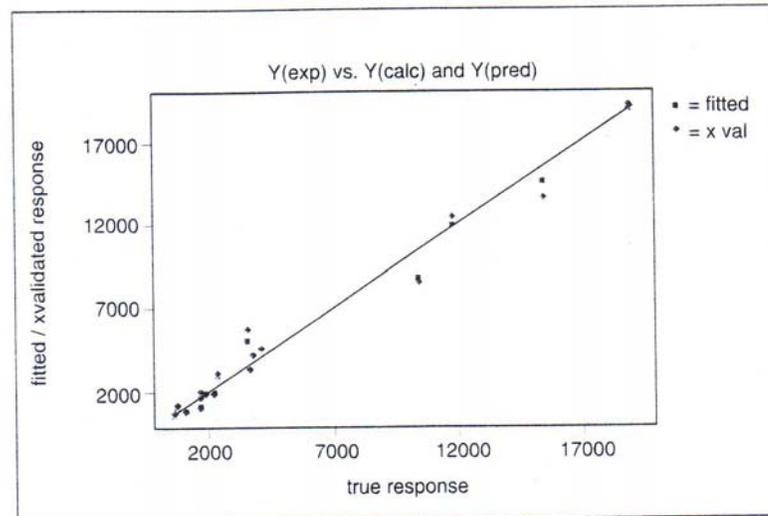


Gráfico 3.1. Comparación entre la respuesta experimental y la calculada

Fuente: TODESCHINI R, 1998, Introduzione alla Chemiometria. Edises, Milano, Italia.

En el gráfico 3.2 muestra un caso del curso ideal. El gráfico **b**, **c** y **d** muestra tres típicas situaciones:

b: el residuo crece cuando aumenta la respuesta, ésa es su varianza no es distribuida homogéneamente

c: la opción del modelo no es conveniente, ya que se observa algunas trayectorias no casuales (como ejemplo, se extraña en el modelo de las condiciones no lineales);

d: la trayectoria del residuo revela un curso constante no casual, debido generalmente a la presencia de un error sistemático. El residuo ordinario no tiene, en general, la misma varianza. Su varianza es de hecho definida en la siguiente expresión:

$$V(r) = \sigma^2 * (I - H)$$

Y, en particular, para cada objeto,

$$V(r_i) = \sigma^2 * (I - h_{ii})$$

Donde H es la matriz de influencia y h_{ii} es su elemento diagonal.

Es entonces evidente que un punto con alta influencia (leverage) tiene una pequeña varianza respecto a un punto de baja influencia (leverage).

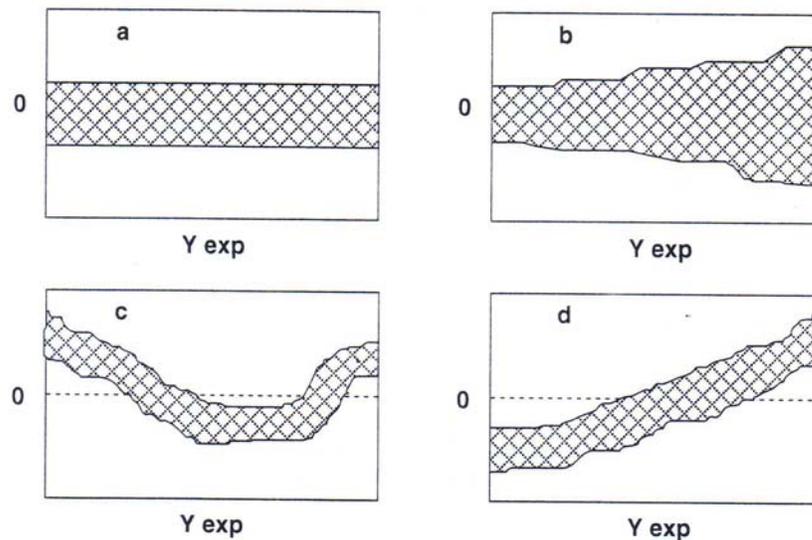


Gráfico 3.2. a) curso ideal, b,c,d) situaciones típicas

Fuente: TODESCHINI R, 1998, Introduzione alla Chemiometria, Edises, Milano, Italia.

Para obtener el residuo que tiene una varianza común es necesario dividir el residuo ordinario para la desviación estándar:

$$r'_i = \frac{r_i}{s * \sqrt{1 - h_{ii}}}$$

donde **s** es la raíz cuadrada de s^2 , la varianza estimada de $RSS/(n - p')$.

Estos residuos son llamados residuos estandarizados y son particularmente indicados para el control de la normativa de la normalidad de los residuos y para individualizar la posible confrontación de los residuos extraños. (Todeschini, 1998)

3.7. Correlación entre variables predictoras y variables dependientes.

Luego de haber efectuado un análisis de regresión es posible verificar si la relación funcional que se encuentra entre la variable dependiente y y la variable independiente x_j es significativa. Ha este propósito, se es capaz de efectuar un análisis de la varianza a fin de obtener el estadístico F de Fisher, donde la hipótesis dada y la hipótesis alternativa se define como:

H_0 : todos los coeficientes de regresión b_j ($j = 1, p$) son igual a cero

H_1 : por lo menos un coeficiente de regresión es diferente de cero, esto es que existe un modelo .

Esto significa que la hipótesis H_0 es verdadera (entonces es falsa la hipótesis alternativa H_1), el modelo de regresión no existe en cuanto todo el coeficiente del modelo no son significativamente diferente de cero. Entonces para un buen modelo de regresión el valor calculado de F debe ser grande.

La variación total debido a la regresión es definida como TSS (suma total de los cuadrados), la variación explicada de la regresión de MSS (Modelo Suma de Cuadrados) y la variación residual no explicada del modelo de regresión de RSS (suma residual de cuadrados).

El grado de libertad para la variación total es debido al modelo y al residuo respectivamente:

$$DF_T = n - 1, \quad DF_M = p' - 1 \quad y \quad DF_r = n - p'$$

Donde n es el número de observaciones y p' es el número total de parámetros del modelo de regresión. La varianza viene por consiguiente calculada por la suma de cuadrados, separando el respectivo grado de libertad.

El valor F para el test de la hipótesis nula es:

$$F = \frac{MSS/(p' - 1)}{RSS/(n - p')} = \frac{MS_M}{MS_r}$$

Y debe ser confrontado con el valor crítico de tabla:

$$F_c (1); (\alpha(p' - 1), (n - p'))$$

El valor de F calculado viene comparado con el valor crítico F de tabla, agregando el nivel de confianza (comúnmente 5% o 1%), relativo al grado de libertad del numerador y denominador. Si el valor calculado de F es al menos igual o superior al valor crítico, la hipótesis es nula, o más bien es estadísticamente probable que en el nivel de selección α , existe una relación de dependencia f entre la variable considerada inconstante en la regresión (Todeschini, 1998).

CAPITULO IV

VALIDACIÓN

4.1. Generalidades

La validación de un modelo consiste, en general, en buscar esa estructura del modelo que aumenta al máximo la capacidad predictiva.

Al mismo tiempo el modelo encontrado tiene que tener alguna característica de estabilidad suficientemente independiente de los datos particulares usados para construirlo.

La varianza de los parámetros estimados que caracterizan un modelo es directamente proporcional a la complejidad del mismo modelo.

Este hecho tiene un curso de importancia extraordinaria en la evaluación de las posibles actuaciones del modelo en la fase predictiva.

En el siguiente gráfico se muestra la varianza:

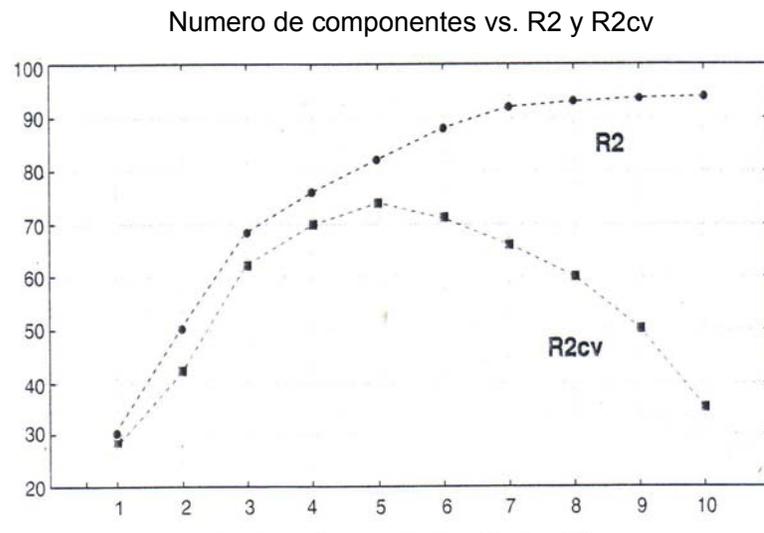


Gráfico 4.1. Número de componentes principales del modelo

4.2. Método Leave One out (LOO)

Según este método, los n objetos, de los datos del modelo, son calculados por separado, en el cuál el objeto se excluye en un momento: cada modelo viene calculado con los objetos restantes $n-1$ y es usado para predecir la respuesta (si es una respuesta cuantitativa de un modelo de regresión o una respuesta cualitativa de un modelo de clasificación): al fin de estimar un parámetro que está efectivamente en relación con el poder predictivo del modelo.

Como una regla, el modelo final real será siempre calculado con todos los n datos disponibles, en modo de poder aprovechar toda la información disponible en todos los datos.

Este método de validación es el que tolera una mínima perturbación, en cuánto cada modelo siempre es calculado utilizando los datos $n-1$ y es lo que proporciona una comparación en predicción entre el modelo obtenido con métodos diversos y los descriptores diferentes. El límite de este método de validación es que extiende al valor de R^2 en fitting al crecimiento del número n de las muestras; por este motivo el método leave-one-out puede todavía acomodar los valores predictivos fijos si el número de muestras es bastante elevado. (Todeschini, 1998).

4.3. Método Leave more out

Este método es una generalización del método leave-one-out. Los datos vienen subdivididos en G grupos de cancelación en modo tal que k objeto para un n/G es asignado a la evaluación entera y por consiguiente excluir del training set: el modelo estimado de un training set constituido de $n-k$ objetos, es utilizado para predecir la respuesta de k excluido el objeto del modelo. Los objetos de k se excluyen un tiempo cada uno (cross-validación); una vez excluido, éstos se insertan de nuevo en el modelo y se excluyen de allí otro k . Cada i -ésimo objeto se asigna por consiguiente a un grupo de cancelación g , según la expresión:

$$G = \text{mod}(i - 1, G) + 1$$

Donde $\text{mod}(a,b)$ es el modulo del operación, eso es a modulo b .

Según esta modalidad la colocación de cada objeto en la evaluación entera va a modo sistemático; el resultado depende obviamente del orden en que los objetos se introducen en la matriz general de los datos.

Una alternativa a la selección sistemática de los objetos es la selección de k objetos al azar entre el conjunto de los datos.

4.4. Método Bootstrap

Según este método, los datos n objetos, todo viene generado por k , todos los n dimensionales, se construye extrayendo al azar con la repetición n objetos junto al original. Esto significa que cada todo de los datos es siempre n -dimensional, pero cada vez algunos objetos aparecerán más veces en el mismo todo, mientras otros no aparecerán en absoluto. El modelo obtenido del primero es utilizado para predecir la respuesta de los objetos excluidos.

Este método es utilizado solo si se dispone de cálculos suficientemente poderosos en cuanto es necesario proceder a la extracción de muchos n -dimensionales (por lo menos mil), para tener una estimación fiable del valor medio del parámetro predictivo seleccionado.

En general se cree que un modelo bueno tiene que conservar un buen poder predictivo aun cuando se sometió a una gran perturbación. Es necesario tener presente que la eliminación del training set de un número excesivo de objetos en la construcción de los modelos pueden extraviar completamente una parte de información fundamental en la construcción del modelo mismo y que por consiguiente esto inevitablemente no puede prever sólo conductas correctamente características de objetos presentes en el juego de la evaluación (Todeschini, 1998).

4.5. Regresión en Componentes Principales (PCR)

El método de regresión en componentes principales es un método de regresión fundado en los principios generales de OLS en que la matriz original del modelo (X) se reemplaza por una representación del espacio del primer componente principal (T_M).

Este método se basa en la idea de representar el espacio del predictor (la variable independiente) en un lugar de reunión de número de componentes principales, eliminando la información ligada a las direcciones de menor varianza se asume como no importante. El empleo en la regresión de componentes principales como la variable independiente posee la ventaja notable para equilibrar la importancia de la variable independiente original, en cuanto la contribución de más variables "similares" es unificado en un único componente.

Una de las ventajas de este acercamiento a la regresión es para tratar problemas cuando el número de objetos es inferior al de la variable, utilizando como nueva variante predictiva un número pequeño de componentes principales, inferior al número de observaciones.

Además PCR se usa para superar los problemas de multi colinealidad que en el método OLS lo reflejan en la dificultad de invertir la matriz $X^T X$. (Todeschini, 1998).

CAPÍTULO V

RESULTADOS Y DISCUSIÓN

5.1. Parte experimental

Para la realización de este trabajo se han construido 132 soluciones patrón de pectina a partir de un producto comercial, la composición de los patrones ha sido variada entre 5 y 20 gr. de pectina en 1000 ml solución y 40gr de sacarosa para permitir la dispersión de la pectina.

Se han obtenido los espectros infrarrojos utilizando un espectrómetro Thermo Nicolet FT-IR 100 con aditamento ATR y promediando sobre 32 corridas. Se ha trabajado con un ATR de ZnSE por lo que se ha limitado el estudio a una longitud de onda de 700 a 1800 cm^{-1} cubriendo así la zona de las huella digitales del espectro. Los espectros han sido digitalizados y pre tratados mediante suavización automática de la línea de base, luego normalizados mediante técnica SNV (Standard Normal Variate). (Da Wen Sun, 2009)

Sucesivamente se han seleccionado las bandas de absorción mas útiles para la construcción del modelo de calibración multivariante inversa (Da Wen Sun, 2009) utilizando algoritmos genéticos (Leardi, R. 2003). En la técnica de selección de las variables mediante algoritmos genéticos se codifican las variables con 1 (presencia) ó cero (ausencia) en el modelo. El conjunto de estos bits se denomina cromosoma y con estos cromosomas se construye la población inicial casual la misma que sufre el algoritmo de evolución genética constituido de etapas iterativas de acoples entre cromosomas y mutaciones de bits en los cromosomas. Los mejores cromosomas siguen perteneciendo a la población mientras que los peores se extinguen procedimiento ha sido realizado mediante programa MOBYDIGS (Todeschini, 2004), implementado con los siguientes parámetros:

Tabla 2. Parámetros del algoritmo genético.

Parámetro	Valor
No. Puntos Trainig Set	66
No. Puntos Test Set	66
Población (No. Cromosomas)	80
Rep/Mutación trade off	0.5
Sesgo selección %	50
Parámetro de calidad del modelo	Q2
No. Máximo de variables a escogerse	5

El mejor modelo obtenido bajo los criterios de la tabla No 1, se ha validado mediante un test set externo (66 patrones externos de composición variable casualmente y no utilizados para la construcción del modelo mismo) y mediante técnica de BOOTSTRAP. (Efron, B. and Tibshirani, R.J. 1993.)

5.2. Resultados y discusión

En el gráfico 5.1. se presentan algunos espectros obtenidos.

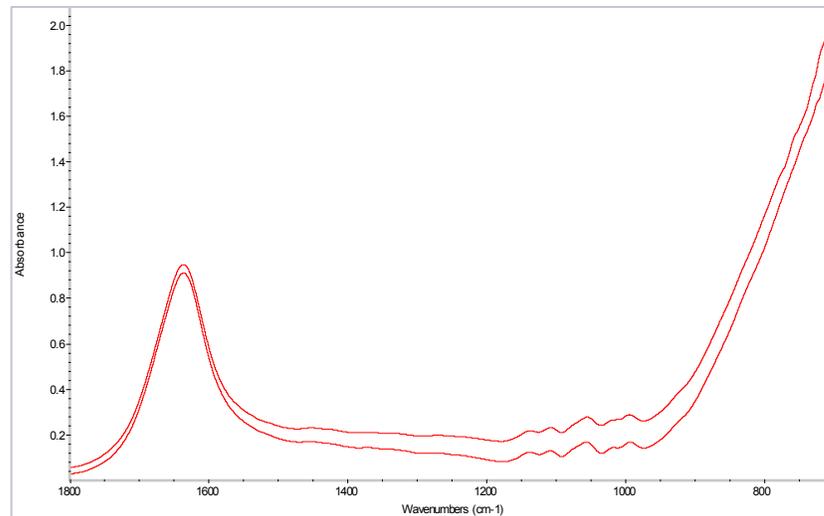


Gráfico 5.1. Espectros Infrarrojos de muestras de solución de pectina

Fuente: Autor

El algoritmo ha seleccionado como mejor modelo de calibración multivariante el presentado en la tabla No 2.

Tabla 3. Bandas seleccionadas por el algoritmo genético.

No	Variable	Coeff. Regresión	Errores Coeff. Reg.	Intervalos Conf. (.95)	Coeff. Reg. Estd.
-	Intercepción	-53.94623	7.95984	15.91967	-
1	1056.8	-174.19141	8.58878	17.17757	-1.66026
2	1106.94	325.78629	11.54592	23.09184	2.59244
3	1170.58	-104.83755	4.88584	9.77168	-0.94069
4	1348	-169.78789	10.9366	21.8732	-0.97423
5	1592.91	39.56481	5.66046	11.32092	0.4599

La tabla No. 3 indica que la banda más influyente positivamente en el modelo de regresión es la correspondiente a 1106.94 cm^{-1} en cuanto presenta el mayor valor del coeficiente estandarizado (2.59244). De acuerdo con Schulz y Baranska, (Schulz, H. and Baranska, M. 2007), ésta banda corresponde a la concentración de las absorciones de los siguientes grupos: $\nu(C-O)$, $\nu(C-C)$, presentes en la pectina. Su valor positivo indica que varía de acuerdo con la cantidad de pectina presente en la muestra.

La segunda banda que influye positivamente es la de 1592.91 cm^{-1} (0.4599) que está muy cerca de la teórica de 1605 a la vibración del grupo carboxílico ($\nu_{as}(COO^-)$) presente.

La primera banda, que influye negativamente 1056.8 cm^{-1} que tiene un coeficiente estandarizado de -1.66026; la segunda banda negativa en importancia se ubica a 1348 cm^{-1} con un coeficiente estandarizado negativo de -0.97423; la tercera banda con influencia negativa se localiza a 1170.58 con un coeficiente estandarizado de -0.94069.

La presencia de aportes negativos en la regresión no es de fácil interpretación, podrían ser interpretados como “correcciones” del valor calculado con la banda principal para obtener un ajuste predictivo óptimo.

La tabla No 4 presenta los parámetros de calidad del modelo en interpolación y en predicción.

Tabla 4. Parámetros de calidad del modelo

No.Obj.Training Set	No.Var.	Q2loo	R2	R2adj	R2-Boo	Kx	Kxy	SDEP
66	5	94.1	95.42	95.03	93.7	49.39	52.07	1.6188

El índice de bondad de la regresión en predicción Q^2_{lo} , (Todeschini, R. Ballabio, D. Consonni, V. Mauri, A. and Pavan, M. 2004). ha sido calculado por el programa MobyDigs, R^2 y R^2_{adj} son respectivamente el porcentaje de varianza explicada por el modelo tal y cual en el primer caso y corregida por la complejidad del modelo en el segundo caso.

El parámetro K_x , indica el porcentaje de correlación multivariante presente en las variables predictoras seleccionadas por el algoritmo genético; su valor debe ser el más bajo posible para que la colinealidad de los predictores no afecte a la inversión de la matriz de dispersión (Draper, N. and Smith, H. 1966), K_{xy} indica la correlación total entre el conjunto de variables predictoras y la respuesta; su valor debe ser mayor de K_x para que el modelo de calibración tenga sentido (Todeschini1998). En el modelo se cumple esta condición por lo que es posible aplicar el método de mínimos cuadrados ordinarios a la regresión.

Adicionalmente el valor de R^2 -Bootstrap ha sido de 93.7 % lo que confirma la estabilidad del modelo en predicción.

En el gráfico 5.1. se presenta la correlación entre el porcentaje de pectina experimental y el predicho por el modelo propuesto.

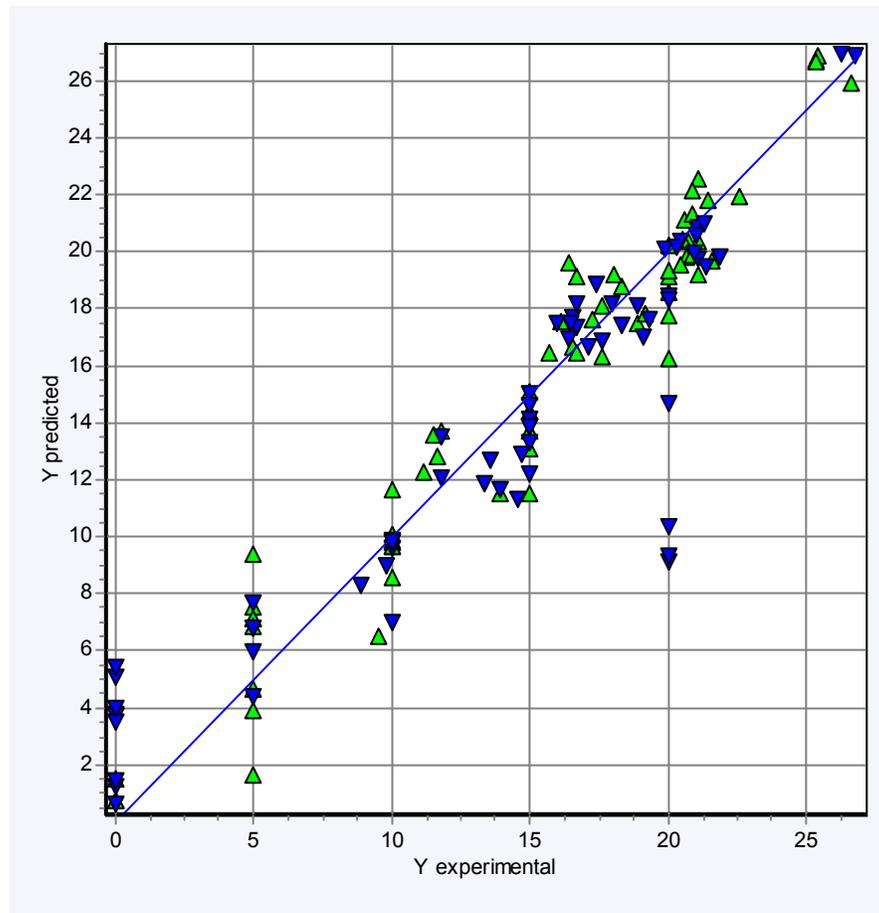


Gráfico 5.2. Correlación entre contenido de pectina experimental y predicho por el modelo (Los triángulos verdes corresponden al training set; los triángulos azules al test set)

Fuente: Autor

En el gráfico anterior la predicción del modelo es buena; también los patrones externos del test set que no han sido utilizados en su construcción.

En el gráfico 5.3 se representa la distribución de los residuos normalizados vs. La respuesta experimental.

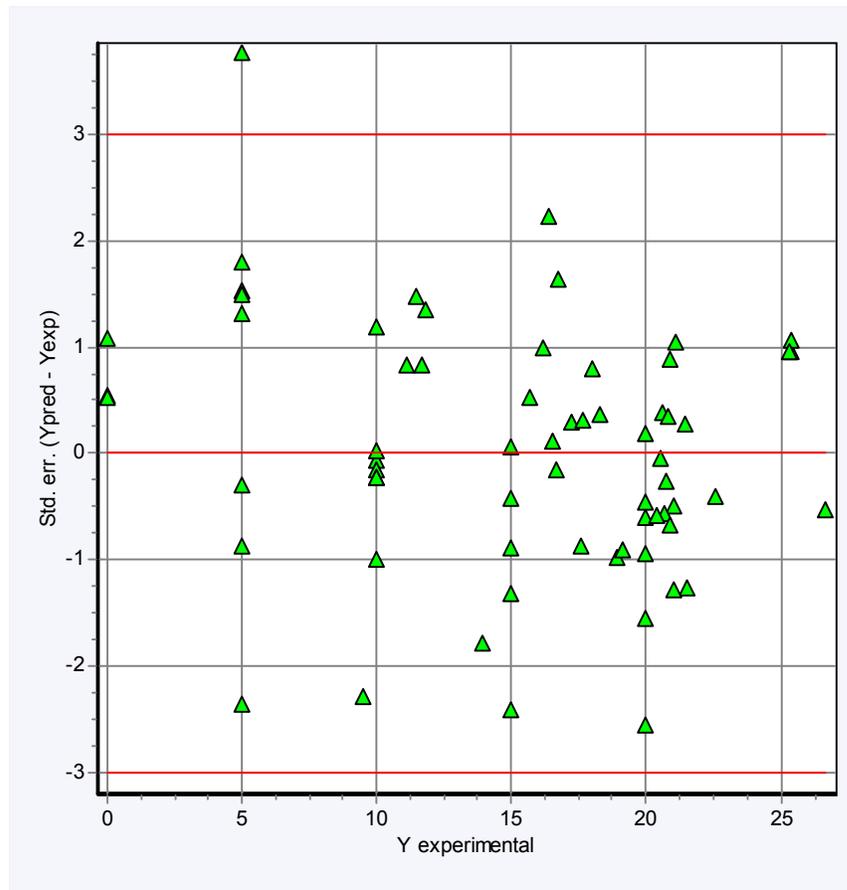


Gráfico 5.3. Distribución de los residuos normalizados

Fuente: Autor

Los residuos se distribuyen en forma casual con tendencia prácticamente uniforme indicando que el modelo es bastante bueno, además solo un valor correspondiente al 5% está sobre la línea de control inferior, demostrando así que el modelo de calibración es bastante estable tomando en cuenta que el modelo fue construido utilizando solo el 50 % de los datos.

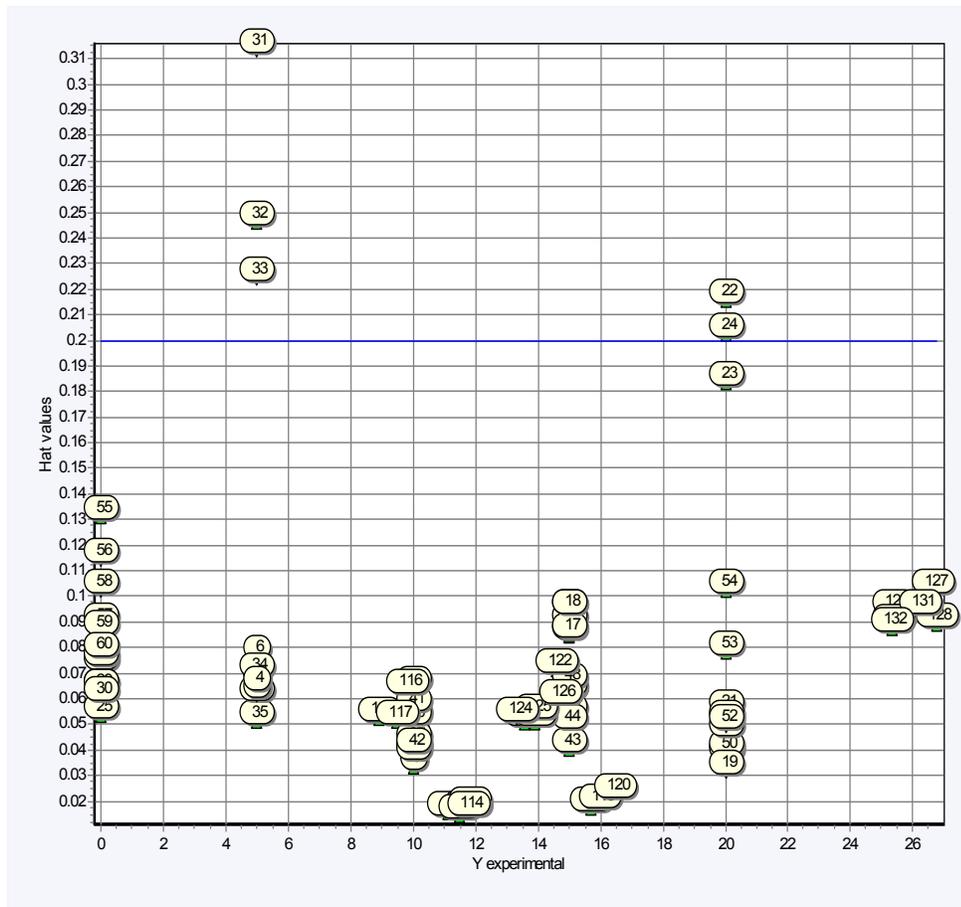


Gráfico 5.4. Distribución de muestras patrón vs Distribución de residuos normalizados.

Fuente: Autor

En este gráfico se observa que solo el 5% está sobre la línea de control indicando la estabilidad del modelo.

5.3. Método PLS

En el gráfico 5.5 representa el modelo de selección que indica cuantos componentes se utilizan para obtener el mejor y más “económico” modelo.

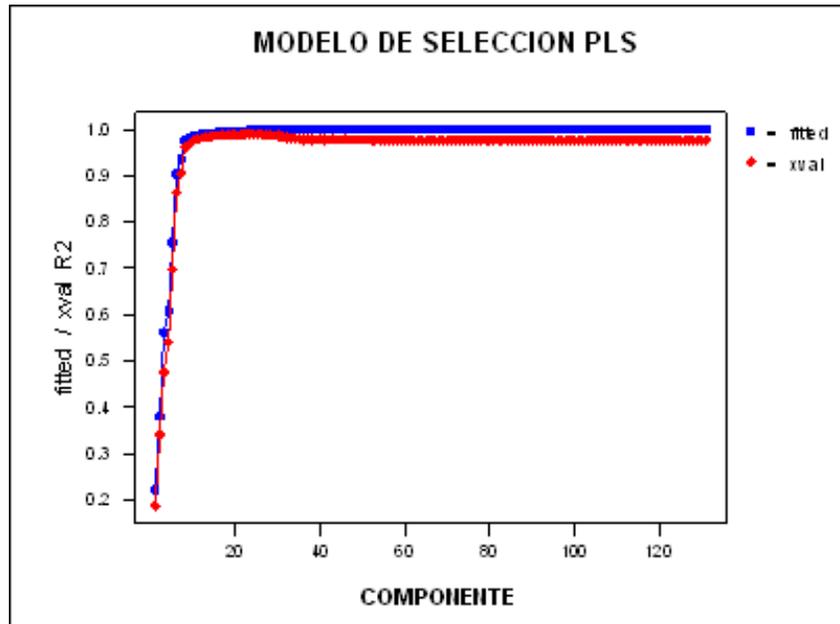


Gráfico 5.5 Modelo de Selección PLS

Fuente: Autor

En el gráfico 5.5 se observa la correlación entre la respuesta experimental y el componente en estudio, observando que, el modelo de calidad, del modelo en validación llega muy lentamente hasta el componente 25, para luego estabilizarse, por lo tanto el mejor modelo PLS será el de 25 componentes.

Un elemento de calidad del modelo es el gráfico 5.6 de los residuos. Si los residuos del valor x son distribuidos según el residuo ordinario el modelo es válido.

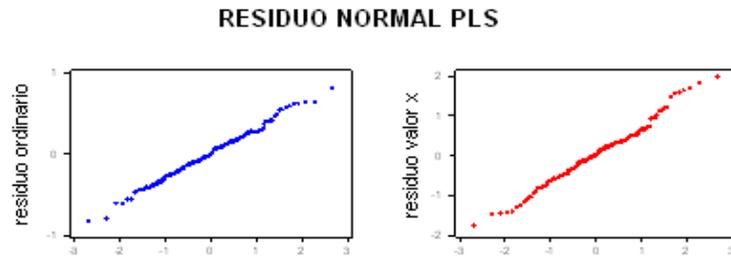


Gráfico 5.6 Residuo Normal en PLS

En el gráfico 5.7 se nota la correlación entre valores experimentales y calculados con el modelo de 25 componentes.

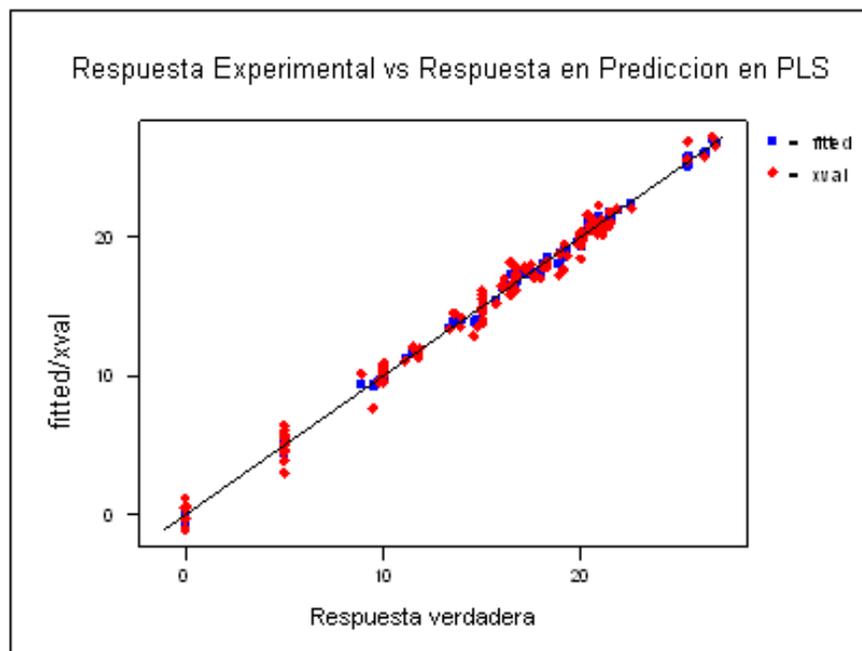


Gráfico 5.7. Respuesta experimental vs respuesta predictiva

Fuente: Autor

La correlación, también en validación es muy buena como indica R^2_{cv} que es de 0.9893.

5.4. Método PCR

En el gráfico 5.8 del método PCR se observa la relación entre el modelo propuesto y el componente (pectina), en predicción; que indica cuantos componentes se utilizan para obtener el mejor modelo.

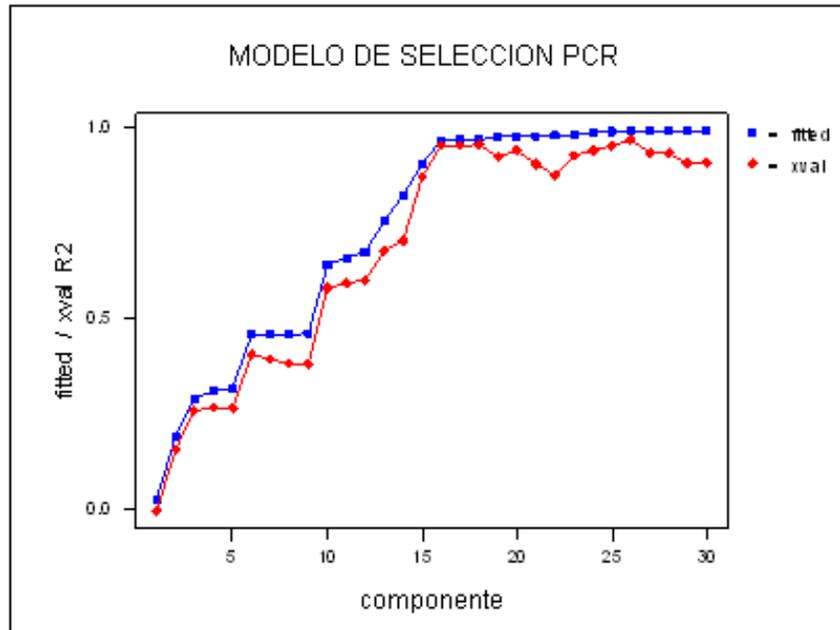


Gráfico 5.8. Modelo de selección

Fuente: Autor

En el gráfico 5.8 se observa la correlación entre la respuesta experimental y el componente en estudio observando que, el modelo de calidad, del modelo en validación llega hasta el componente 26 para luego seguir la trayectoria en forma lineal como se muestra en el fitted, por lo tanto el mejor modelo PCR será del de 26 componentes.

Un elemento de calidad del modelo es el gráfico 5.9 de los residuos. Si los residuos de la interpolación son distribuidos según los residuos de validación el modelo es válido.

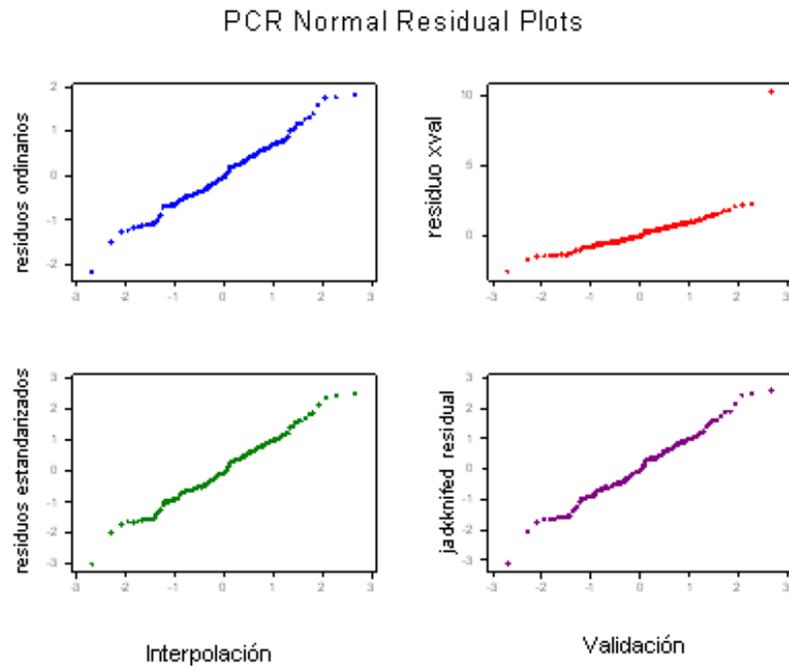


Gráfico 5.9. Residuos Normales en PCR

Fuente: Autor

Siendo los gráficos de interpolación los de la izquierda y los de validación de la derecha

En el gráfico 5.10 se observa la correlación entre valores experimentales y calculados con el modelo de 26 componentes.

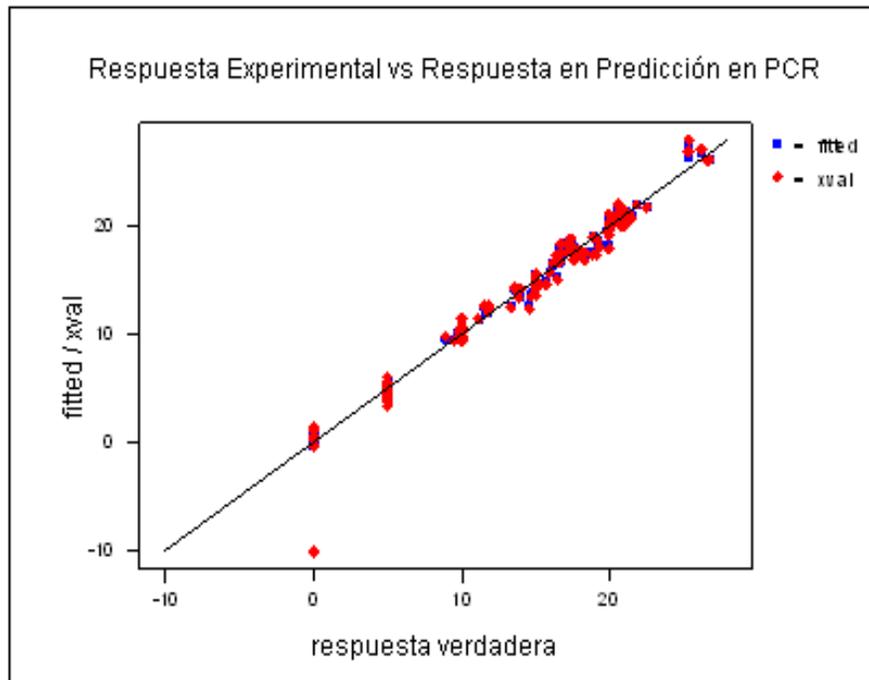


Gráfico 5.10. Respuesta Experimental vs Respuesta en Predicción

Fuente: Autor

La correlación, también en validación es muy buena como indica R^2_{cv} que dio un 0.9678.

Conclusiones

- En el presente trabajo se establecieron las regiones del espectro infrarrojo más útiles para la cuantificación de pectina residual; con la aplicación de FTIR en la determinación de pectina se ha asignado las bandas de absorción entre 1800 y 700 cm^{-1} a los correspondientes grupos funcionales
- Se han seleccionado las bandas de absorción utilizando un algoritmo genético el mismo que permite seleccionar las bandas con la mínima colinealidad entre si y con el mejor poder modelizante en predicción.
- El modelo de calibración multivariante es un método estadístico, resultando ser un método amigable con el medio ambiente y demostrando la validez de esta técnica en el análisis de los alimentos.
- El modelo de regresión brinda una buena calidad, tanto en fitting, cuanto en predicción.
- La aplicación de algoritmos genéticos ha sido de gran utilidad para la codificación de las variables, con lo que se logró la selección del mejor modelo de calibración multivariante.
- Se ha elaborado un Protocolo Analítico para la ejecución del análisis que será de gran ayuda para la cuantificación de pectina residual en la elaboración de jugos de fruta clarificados.
- El presente trabajo demuestra la posibilidad de utilizar la espectroscopia FT-IR-ATR para cuantificar pectina residual en jugos de fruta clarificados.

BIBLIOGRAFIA

Referencias bibliográficas

BURNS Donald and CIURCZAK Emil, 2001, Practical Spectroscopy. Hand book of Near-Infrared Analysis. Editorial. Taylor &Francis Group. Third edition. Los Alamos, New México.

DA-WEN Sun, 2009, Infrared Spectroscopy for food quality Analysis and Control, Editorial Reviews, Dublin, Ireland.

FISZMAN, S. 2001, Propiedades funcionales de los hidrocoloides polisacáridicos. Mecanismo de gelificación. Agroquímica y Tecnología de Alimentos. Editorial Acribia, Madrid, España.

MASSART D.L. and LEWI P.J. 1997, Handbook of Chemometrics and Qualimetric. Editorial Iberoamerica, México, México.

STUART Barbara, 2004, Infrared Spectroscopy. Fundamentals y Applications, Editorial Wiley &Sons Ltd. The Atrium, Southern Gate, Chichester. West. Sussex. England.

TODESCHINI Roberto, 1998, Introduzione alla Chemiometria, Editorial Edises, Milano, Italia.

WILEY John Y SON, 2000, Applied Infrared Spectroscopy Fundamentals. Techniques and analytical, Editorial Wiley &Sons Ltd. Estados Unidos.

Referencias Electrónicas

CAMEJO DE APARICIO Clara. FERRER Alexis. Betzabé de Ferrer². PEÑA Jorge. CEDEÑO Milagros. 2001. Extracción y caracterización de pectina en limones injertados de la región Zuliana. Zulia, Venezuela. (en línea). www.revfacagronluz.org.ve/v13_5/v135z012.html - 17k . Fecha de consulta, Febrero 02, 2010

D'ADDOSIO, R, PAEZ, G, MARIN, M. 2005, Obtención y caracterización de pectina a partir de la cáscara de parchita. Zulia, Venezuela.(en línea). <http://www.scielo.org.ve/scielo.php?script=sci_arttext&pid=S0378-78182005000300004&lng=es&nrm=iso>. ISSN 0378-7818.

Fecha de consulta, Octubre 28, 2009

ESTEVEZ Raúl. 2007. Espectroscopia Infrarroja. San Felipe, Chile. (en línea). www.fq.uh.cu/dpto/especinfrarrojo/estruc_2/ir.html - 18k

Fecha de consulta, Marzo 08, 2010

HURTADO, P. F. 1999.- Ensayo de procesamiento del maracuyá.- Propiedades químicas de la pectina.- Tesis de grado.- Universidad Nacional Agraria la Molina, Lima, Perú.(en línea).

www.producciondepectina/gradosdemetilacion.html

Fecha de consulta, Noviembre 15, 2009

MATSUHIRO Betty _Y RUBIO María José. 2001, Modificación química de pectina de limón, Santiago de Chile, Chile. (en línea).

http://www.scielo.org.ve/scielo.php?script=sci_arttext&pid.

Fecha de consulta, Diciembre 10, 2009

NANZA. REVISTA MEXICANA DE FISICA E. 2005. Modelo de predicción basado en análisis multivariante para la determinación de concentración de azúcar en solución. México, México.(en línea).

http://rmf.fciencias.unam.mx/pdf/rmf-e/51/2/51_2_067.pdf.

Fecha de consulta, Marzo 24, 2010

ANEXO

Anexo1

PROTOCOLO ANALITICO PARA LA EJECUCION PRÁCTICA DEL ANÁLISIS

Para la realización del análisis se debe aplicar los siguientes pasos:

- Preparación de soluciones patrón de pectina a diferentes concentraciones, empezando desde 5gr de pectina hasta llegar a los 20gr de pectina en 1000 mililitros de solución y 40 gr de sacarosa.
- Se realiza las lecturas de las muestras en el infrarrojo en un campo de 700 a 4000 cm^{-1} y promediado sobre 32 corridas. Para la obtención de espectros infrarrojos se empleó un espectrómetro Thermo Nicolet FT-IR 100 con aditamento ATR.
- Los espectros obtenidos son digitalizados y pretratados mediante suavización de la línea de base y sucesivo autoscaling, en un campo de 700 a 1800 cm^{-1} .
- Para conocer la concentración en % de la solución (Pectina) se realiza mediante la siguiente fórmula:

$$\begin{aligned} \text{Concentración \% (Pectina)} = & -53.94623 - 174.19141 \text{ Ab}_{1056.8 \text{ cm}^{-1}} + 325.78629 \\ & \text{Ab}_{1106.94 \text{ cm}^{-1}} - 104.83755 \text{ Ab}_{1170.58 \text{ cm}^{-1}} \\ & - 169.78789 \text{ Ab}_{1348 \text{ cm}^{-1}} + 39.56481 \text{ Ab}_{1592.91 \text{ cm}^{-1}} \end{aligned}$$