

**UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA**

FACULTAD DE INDUSTRIAS ALIMENTARIAS



**“CALIBRACIÓN PARA DETERMINAR COMPOSICIÓN
PROXIMAL DE LA QUINUA (*Chenopodium quinoa* W.) USANDO LA
ESPECTROSCOPIA DE TRANSMITANCIA EN EL INFRARROJO
CERCANO”**

Presentada por:

JORGE MIGUEL PEREDA IBAÑEZ

**TESIS PARA OPTAR EL TÍTULO DE INGENIERO EN INDUSTRIAS
ALIMENTARIAS**

Lima – Perú
2016

UNIVERSIDAD NACIONAL AGRARIA LA MOLINA

FACULTAD DE INDUSTRIAS ALIMENTARIAS

**“CALIBRACIÓN PARA DETERMINAR COMPOSICIÓN PROXIMAL DE LA
QUINUA (*Chenopodium quinoa* W.) USANDO LA ESPECTROSCOPIA DE
TRANSMITANCIA EN EL INFRARROJO CERCANO”**

TESIS PARA OPTAR POR EL TÍTULO DE:
INGENIERO EN INDUSTRIAS ALIMENTARIAS

JORGE MIGUEL PEREDA IBAÑEZ

APROBADA POR EL SIGUIENTE JURADO:

Dra. Carmen Velezmoro Sánchez

PRESIDENTE

Mg.Sc. Gloria Pascual Chagman

MIEMBRO

Dra. Patricia Glorio Paulet

MIEMBRO

Dra. Bettit Salvá Ruiz

ASESOR

Dra. Luz Gómez Pando

CO-ASESOR

Lima-Perú

2016

DEDICATORIA

A mis padres Martha y Jorge, que me han apoyado incondicionalmente.

A mis abuelos, con los que he crecido y compartido anécdotas.

Al uso de tecnologías limpias con el fin de utilizar menos reactivos.

AGRADECIMIENTO

Al programa VLIR UNALM, Proyecto de Desarrollo de Cadenas de Valor para la conservación de la biodiversidad y el mejoramiento de las condiciones de vida rurales, Subproyecto de Cereales Andinos; a través del Programa de Investigación y Proyección Social (PIPS) de Cereales y Granos Nativos de la Universidad Nacional Agraria La Molina por el financiamiento de la investigación.

Al Organismo Internacional de Energía Atómica, Proyecto PER 0025 que financió la visita de la experta Lori Oatway – *Research/Scientist, Grain Quality. The Field Crop Development Centre. Alberta – Agriculture and Forestry.*

A la doctora Luz Gómez Pando, quien me brindó la oportunidad y soporte económico para la ejecución de la tesis.

Al profesor Christian Encina, por su insistencia e interés para lograr lo más pronto posible la culminación de la tesis.

A la doctora Bettit Salvá, por su apoyo en la parte final de la investigación.

A mi madre Martha Ibañez, por su apoyo incondicional y facilidades en la parte experimental.

A Maribel Gonzáles y Eduardo, por su apoyo en la parte experimental.

A Ana lucía, por acompañarme durante la parte experimental y darme ánimos durante la redacción.

A la doctora Katherine Flores, por su guía en el desarrollo inicial de la calibración.

A José y Hans de la empresa Camposur Inc, por permitirme trabajar y realizar la tesis de forma simultánea.

A Paolo Temoche y Rolando Tapia de la empresa SCANCO, representantes de la firma FOSS, por su apoyo en el manejo de los *softwares* y soporte técnico de la tecnología NIRS.

A Dios.

Gracias.

ÍNDICE GENERAL

RESUMEN

ABSTRACT

I. INTRODUCCIÓN	1
II. REVISIÓN DE LITERATURA	2
2.1. GENERALIDADES DE LA QUINUA	2
2.1.1. GRANO DE LA QUINUA	2
2.1.2. VALOR NUTRITIVO	2
2.1.3. FACTORES ANTINUTRICIONALES	3
2.2. ESPECTROCOPIA	5
2.2.1. ESPECTRO ELECTROMAGNÉTICO	7
2.2.2. ESTADOS ENERGÉTICOS DE LA MATERIA	8
2.2.3. TRANSICIONES DE NIVELES ENERGÉTICOS EN ESPECTROSCOPIA	10
2.3. ESPECTROSCOPIA INFRARROJA	12
2.3.1. VIBRACIONES MOLECULARES	13
2.3.2. RELACIÓN DE LOS ESPECTROS CON LA ESTRUCTURA MOLECULAR. ...	16
2.4. INSTRUMENTACIÓN	20
2.4.1. LA FUENTE DE RADIACIÓN	20
2.4.2. SISTEMA DE SELECCIÓN DE LONGITUDES DE ONDA	20
2.4.3. COMPARTIMIENTO DE LA MUESTRA	21
2.4.4. DETECTOR	23
2.5. FUNCIONAMIENTO	23

2.5.1. LEY DE LAMBERT Y BEER	23
2.6. BENEFICIOS DE LA ESPECTROMETRÍA EN EL INFRARROJO CERCANO	25
2.7. QUIMIOMETRÍA	26
2.8. ETAPAS DEL PROCESO DE MODELADO	26
2.8.1. SELECCIÓN DE UN CONJUNTO DE CALIBRACIÓN O ENTRENAMIENTO	27
2.8.2. DETERMINACIÓN DE LOS VALORES DE REFERENCIA	28
2.8.3. REGISTRO DE LA INFORMACIÓN ANALÍTICA	28
2.8.4. PRETRATAMIENTO DE LOS DATOS	30
2.8.5. CONSTRUCCIÓN DEL MODELO QUIMIOMÉTRICO	33
2.8.6. ESTRATEGÍAS PARA EL DESARROLLO DEL MODELO QUIMIOMÉTRICO	39
2.8.7. EVALUACIÓN DEL MODELO QUIMIOMÉTRICO	44
2.8.8. SELECCIÓN DEL MODELO QUIMIOMÉTRICO	48
2.9. MANTENIMIENTO DE CALIBRACIONES NIRS	50
III. MATERIALES Y MÉTODOS	53
3.1. LUGAR DE EJECUCIÓN	53
3.2. MATERIA PRIMA	53
3.3. MATERIALES	53
3.4. EQUIPOS	54
3.5. REACTIVOS	54
3.6. MÉTODOS DE ANÁLISIS FÍSICO-QUÍMICOS	55

3.6.1. CONTENIDO DE HUMEDAD.....	55
3.6.2. CONTENIDO DE PROTEÍNA	55
3.6.3. CONTENIDO DE LÍPIDOS	55
3.6.4. CONTENIDO DE CENIZA	55
3.7. MÉTODO DE ANÁLISIS ESPECTROMÉTRICO	55
3.8. PROTOCOLO PARA DESARROLLO EL MODELO.....	56
3.9. METODOLOGÍA PARA VALIDAR EL MODELAMIENTO	59
3.10. HERRAMIENTAS PARA EL DESARROLLO DEL MODELO	60
IV. RESULTADOS Y DISCUSIÓN.....	61
4.1. ANÁLISIS DE HUMEDAD, PROTEÍNA, GRASA Y CENIZA DE LAS MUESTRAS DE QUINUA.....	61
4.2. AFINAMIENTO/PURGA DE ESPECTROS EN LOS COLECTIVOS DE CALIBRACIÓN Y VALIDACIÓN	63
4.3. DESARROLLO DEL MODELO DE CALIBRACIÓN Y EJECUCIÓN DE VALIDACIÓN CRUZADA.....	67
4.4. VALIDACIÓN EXTERNA	72
4.5. ERROR ESTÁNDAR DEL MÉTODO DE REFERENCIA	74
4.6. SELECCIÓN DE LA MEJOR ECUACIÓN DE CALIBRACIÓN.....	77
V. CONCLUSIONES	87
VI. RECOMENDACIONES.....	88
VII. REFERENCIAS BIBLIOGRÁFICAS	89
VIII. ANEXOS	97

ÍNDICE DE TABLAS

Cuadro 1: Composición físico química del grano de quinua (base seca).....	5
Cuadro 2: Comparación del contenido de aminoácidos esenciales en granos de quinua, cereales, frejol, carne, pescado, leche y la proteína patrón.	6
Cuadro 3: Transiciones asociadas a métodos espectroscópicos y longitudes de onda.....	11
Cuadro 4: Grupos funcionales preferentemente detectados según porción del infrarrojo cercano.....	18
Cuadro 5: Diseño Experimental de la investigación	57
Cuadro 6: Estadísticos evaluados para selección de los modelos NIRS.	58
Cuadro 7: Tamaño de colectivos de calibración o validación.	65
Cuadro 8: Medidas de dispersión de los colectivos, evaluados por vía seca.....	67
Cuadro 9: Estadísticos de calibración y validación cruzada para las mejores ecuaciones obtenidas.....	69
Cuadro 10: Evaluación de estadístico SECV basado en el SEC.	71
Cuadro 11: Estadísticos obtenidos de la validación externa.....	72
Cuadro 12: Valores SEL de los métodos de referencia.	74
Cuadro 13: Valores R^2 máximos esperados de los componentes de la calibración.....	74
Cuadro 14: Evaluación del estadístico SEP basado en el SEL.....	75
Cuadro 15: Estadísticos estimadores de calidad de las ecuaciones de calibración.	76
Cuadro 16: Estadísticos de calibración y validación cruzada de la presentación Granos bs (n=99).....	81
Cuadro 17: Características del modelo quimiométrico desarrollado.....	81
Cuadro 18: Sistemas y <i>software</i> de apoyo del Infratec 1241	96

Cuadro 19: Datos técnicos del Infratec 1241.....	97
Cuadro 20: Características de color y variedad de las 120 accesiones de quinua.	99
Cuadro 21: Resultados promedio de los análisis químicos de las muestras de quinua, expresados como porcentaje en base seca.	103

ÍNDICE DE FIGURAS

Figura 1: Sección longitudinal media del grano de quinua.	3
Figura 2: Grano de quinua de la variedad pasankalla.....	4
Figura 3: Diagrama del espectro electromagnético.	7
Figura 4: Diagrama de niveles de energía de tres estados electrónicos.....	9
Figura 5: Vibración de estiramiento de una molécula.	14
Figura 6: Vibración de deformación de una molécula.	14
Figura 7: Vibración de estiramiento de una molécula biatómica.	14
Figura 8: Representación de los estados energéticos, <i>quantum</i> , y relación con los sobretonos.....	15
Figura 9: Espectros infrarrojo del agua (rojo) y del aceite (azul).....	18
Figura 10: Ubicación y sobretonos de grupos funcionales en el espectro electromagnético.....	19
Figura 11: Espectro de infrarrojo de varias muestras de un misma población.....	20
Figura 12: Esquema de funcionamiento del Infratec 1241.....	21
Figura 13: Esquema de los tres modos de registro de espectros NIR.	22
Figura 14: Haz de radiación que atraviesa un cuerpo.....	24
Figura 15: PCA desde un sistema de 3 variables a uno de 2 componentes principales.	32
Figura 16: Protocolo para el desarrollo de la calibración NIT.	56
Figura 17: Gráfico 3D (PCA; 0,0,1,1) de la presentación Granos bs (n=120).....	64
Figura 18: Espectro de la presentación Granos bs (n=120) con tratamiento matemático 1,4,4,1 y <i>scatter none</i>	64

Figura 19: Espectro de la presentación Granos bs (n=99) con tratamiento matemático 1,4,4,1 y <i>scatter none</i>	65
Figura 20: Comparación de los valores calculados vía química (Y) y los vía NIT (X) del componente humedad.	82
Figura 21: Comparación de los valores calculados vía química (Y) y los vía NIT (X) del componente proteína.....	83
Figura 22: Comparación de los valores calculados vía química (Y) y los vía NIT (X) del componente grasa.	84
Figura 23: Comparación de los valores calculados vía química (Y) y los vía NIT (X) del componente ceniza.....	85
Figura 24: Analizador de granos Infratec 1241.	96
Figura 25: Histograma y polígono de frecuencias del porcentaje de humedad en base seca del colectivo de calibración inicial.	107
Figura 26: Histograma y polígono de frecuencias del porcentaje de humedad en base seca del colectivo de validación inicial.	107
Figura 27: Histograma y polígono de frecuencias del porcentaje de proteína en base seca del colectivo de calibración inicial.	108
Figura 28: Histograma y polígono de frecuencias del porcentaje de proteína en base seca del colectivo de validación inicial.	108
Figura 29: Histograma y polígono de frecuencias del porcentaje de grasa en base seca del colectivo de calibración inicial.	109
Figura 30: Histograma y polígono de frecuencias del porcentaje de grasa en base seca del colectivo de validación inicial.	109
Figura 31: Histograma y polígono de frecuencias del porcentaje de ceniza en base seca del colectivo de calibración inicial.	110

Figura 32: Histograma y polígono de frecuencias del porcentaje de ceniza en base seca del colectivo de validación inicial.	110
Figura 33: Espectro anómalo, perteneciente a la muestra POQG.34.1 del colectivo de validación de granos.	111
Figura 34: Espectro anómalo, perteneciente a la muestra POQG.43.1 del colectivo de calibración de granos.	111
Figura 35: Espectro anómalo, perteneciente a la muestra H.POQ.40.1 del colectivo de validación de harinas.	112
Figura 36: Espectro anómalo, perteneciente a la muestra H.POQ.50.1 del colectivo de calibración harinas.	112

ÍNDICE DE ANEXOS

Anexo 1: Información Técnica del Infratec 1241 – FOSS.....	96
Anexo 2: Características de las 120 accesiones.....	98
Anexo 3: Histogramas y polígonos de frecuencia de los datos de referencia	107
Anexo 4: Espectros anómalos	111
Anexo 5: Glosario NIRS	113

RESUMEN

El grano de quinua es consumido en la región andina desde épocas preincaicas. En aquellas épocas se escogieron las plantas que resistían a los diversos ecosistemas de la región, resultando más de 2 000 variedades. Actualmente se desarrollan cultivos que tengan mayor rendimiento en campo, resistencia a plagas y adaptabilidad a nuevas condiciones. En este contexto, es necesario caracterizar grandes cantidades de semillas en corto tiempo y con métodos confiables. El presente trabajo tuvo como objeto desarrollar una calibración multivariable en el equipo Infratec 1241 (850-1048 nm), con el fin de predecir el contenido de humedad, proteína, grasa y ceniza de la quinua; evaluando lecturas de grano y harina. Se utilizó la espectroscopía de transmitancia en el infrarrojo cercano (NIT). Se emplearon 120 accesiones de quinua, provenientes de campos en la costa de Lima y sierra de Ancash, cultivadas por el PIPS de Cereales y Leguminosas de la Universidad Nacional Agraria La Molina. Una vez que se limpió el material, se realizó la lectura NIT de los granos, luego se molieron las muestras y se realizó la lectura NIT de las harinas; finalmente se realizaron por triplicado los análisis químicos empleando los métodos aprobados por la AOAC. Las curvas de calibración fueron obtenidas luego de purgar los espectros anómalos, aplicar regresiones del tipo “PLS”, dispersiones “*SNV and Detrend*” y “*None*”, así como diferentes tratamientos matemáticos. El mejor modelo fue el obtenido de las lecturas en forma de grano, válido solo para colores claros; se obtuvieron valores RSQ de 0,4648, 0,8084, 0,9313 y 0,8517 para humedad, proteína, grasa y ceniza respectivamente. El componente humedad presentó los menores estadísticos de calibración, debido a factores como el control durante la ejecución de los métodos de referencia, almacenamiento y diferencial de tiempo entre análisis por el método de referencia y lectura de espectros.

Palabras claves: espectroscopía, infrarrojo cercano, NIRS, transmitancia, quinua, calibración.

ABSTRACT

The quinoa grain is consumed in the Andean region since pre-Inca times. In those days, they chose the plants that have better resistance to the climatic conditions of the region; this resulted in more than 2000 varieties. Currently, plant breeding programs develop crops that have higher field yields, pest resistance and adaptability to new conditions. In this context, it is necessary to characterize large amounts of seeds in a short time with reliable methods. The present study was aimed at developing a multivariate calibration in the Infratec 1241 (850-1048 nm) equipment, in order to predict the content of moisture, protein, fat and ash from quinoa; evaluating grain and flour. Near Infrared Transmittance Spectroscopy (NIT) was used in the study. A total of 120 quinoa varieties were used, they were cultivated in the Lima's coast and the Ancash's andes by the Plant Breeding Program Cereals and Pulses of the National Agrarian University La Molina. Once the 120 samples were cleaned, the NIT reading of the grains was performed, then the samples were grounded and the NIT reading of the flour was performed; finally the chemical analysis were performed in triplicate using methods approved by the AOAC. The calibration model was obtained after purging of abnormal spectra, applying "PLS" regressions, "SNV and Detrend" and "None" dispersions, as well as different mathematical treatments. The best model was obtained from the readings in the form of grain, but just valid for light colors; RSQ values of 0,4648, 0,8084, 0,9313 and 0,8517 for moisture, protein, fat and ash were obtained respectively. The moisture component presented the lowest statistical of calibration, due to factors such as control over the performance of the reference methods, storage and time differential between the reference method analysis and reading of spectra.

Key Words: Spectroscopy, near infrared, NIRS, transmittance, quinoa, calibration.

I. INTRODUCCIÓN

La quinua es un grano milenario, apreciado por las culturas andinas porque reconocieron su gran valor como alimento. Sin embargo, durante la época colonial su consumo fue suprimido. En los últimos años ha sido revalorizada hasta convertirse en uno de los cultivos con mayor proyección, alternativa al cultivo del arroz y pionera en la exportación de granos andinos. En este contexto son evidentes numerosas investigaciones relativas al manejo agronómico, cruce de variedades, métodos post-cosecha y procesamiento industrial; pero todas finalmente recaen en el rendimiento del cultivo en campo y el valor nutricional que tendrán los granos de quinua.

El análisis de muestras en programas de fitomejoramiento es una labor rutinaria entre cosechas, cada cosecha puede brindar más de un centenar de unidades experimentales. En este ámbito factores como la velocidad, precisión y confiabilidad de los métodos que se utilizan son claves para el desenvolvimiento y objetividad de cualquier investigación.

La tecnología de Espectroscopía de Transmitancia en el Infrarrojo Cercano (NIT por sus siglas en inglés) permite analizar gran cantidad de muestras, luego de una calibración del equipo para las características cuantitativas o cualitativas que se requieren determinar. Las muestras pueden ser líquidas como el petróleo, aceites y vinos; sólidas como las carnes, frutas cortadas y granos enteros; así como también muestras menudas como tierras, aislados proteicos y harinas de diferente origen.

La presente investigación, integra la necesidad de los programas de fitomejoramiento y las características de la tecnología de Espectroscopia en el Infrarrojo Cercano. El objetivo principal fue desarrollar un modelo de calibración multivariable para determinar el contenido de humedad, proteína, grasa y ceniza en quinua; usando la espectroscopía de transmitancia en el infrarrojo cercano. Los objetivos específicos fueron (1) determinar el contenido de humedad, proteína, grasa y ceniza por los métodos de referencia a 120 muestras y (2) evaluar los modelamientos para las lecturas de infrarrojo de la quinua en las presentaciones de grano y de harina.

II. REVISIÓN DE LITERATURA

2.1. GENERALIDADES DE LA QUINUA

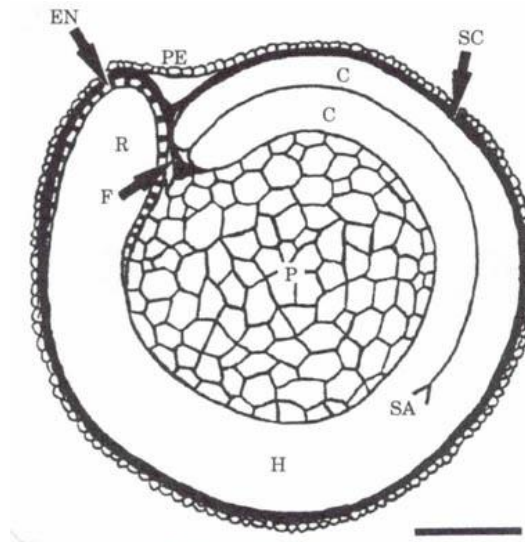
La quinua (*Chenopodium quinoa* Willd) es una planta anual, rústica, herbácea, de crecimiento erecto, que llega a alcanzar diferentes alturas, dependiendo de la especie, cultivar, variedad, factores climáticos y edáficos (Tapia *et al.*, 2014).

2.1.1. GRANO DE LA QUINUA

Quisocala (2000), citado por Tapia *et al.* (2014), indica que el fruto de la quinua es un aquenio cubierto por el perigonio que encierra una sola semilla, el que se desprende con facilidad al frotarlo cuando está seco. El perigonio o envoltura final puede ser verde, púrpura o rojo; este tiene forma estrellada por la quilla que presenta cinco sépalos, estos se separan del grano durante la cosecha. El pericarpio (PE) del fruto está pegado a la semilla, presenta alvéolos y se conforma de colores muy variados, en esta capa se encuentra la saponina que origina el sabor amargo. El episperma (SC), es una membrana delgada de variados colores. El embrión está formado por los cotiledones (C) y la radícula (R), constituyendo la mayor parte de la semilla que envuelve el perisperma como un anillo. El perisperma (P) es almidonoso y de color blanco. Las secciones descritas del grano pueden observarse en la Figura 1. Adicionalmente, en la Figura 2 pueden observarse diferentes fotos del grano de la variedad pasankalla, en la que cada una de sus capas es de diferente color y pueden ser fácilmente diferenciadas.

2.1.2. VALOR NUTRITIVO

El Codex (2008) menciona que la quinua tiene un excepcional balance nutricional. En el Cuadro 1, se presentan resultados de análisis proximales de diversos autores. Además, Gómez y Eguiluz (2012) resaltan el contenido de aminoácidos de la proteína de quinua, información presentada en el Cuadro 2.



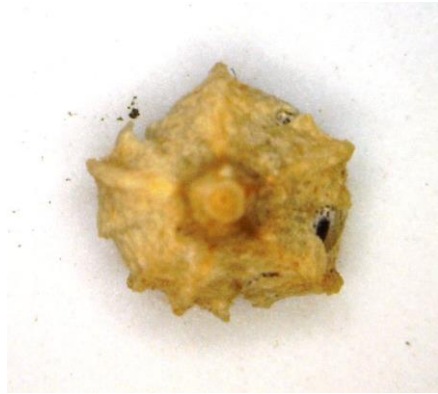
Leyenda: PE: Pericarpio, SC: Cubierta de la semilla o Tegumento o Episperma, EN: Endosperma, C: Cotiledones, H: Hipocolito, SA: Ápice del meristemo, R: Radícula, P: Perisperma, F: Funículo.

Figura 1: Sección longitudinal media del grano de quinua.

FUENTE: Mujica *et al.* (2001).

2.1.3. FACTORES ANTINUTRICIONALES

Según Mosquera (2009) la saponina de la quinua es del tipo glucósido triterpenoide, está ubicada en las primeras membranas y le proporciona el carácter amargo, el cual se ha convertido en un problema a la hora del consumo. Es importante anotar que este glucósido podría afectar la biodisponibilidad de los nutrientes del grano, sin embargo se elimina muy fácilmente en un lavado con agua antes de su consumo. Además indica que es posible clasificar la quinua según su concentración de saponina y esta puede ser: en dulce (sin saponina o con menos del 0,11 por ciento en base húmeda), o en amarga (contiene un nivel mayor al 0,11 por ciento de saponinas).



(a)



(b)



(c)

Leyenda:

(a) Grano con perigonio, es obtenido directamente de la panoja de la planta, el perigonio es la capa que más fácil se desprende durante la trilla en el campo o procesos post-cosecha primarios.

(b) Grano con Pericarpio, es la forma como ingresa a las plantas procesadoras, el pericarpio es una capa delgada dónde se encuentra la saponina; la variedad pasankalla es muy particular y no presenta saponina.

(c) Grano con episperma visible, el episperma es la capa visible luego del procesamiento industrial y la que indica el color que tiene el grano a nivel comercial. Las imágenes fueron tomadas con el uso de un *Dino-lite Digital Microscope*.

Figura 2: Grano de quinua de la variedad pasankalla.

FUENTE: Elaboración propia.

Cuadro 1: Composición físico química del grano de quinua (base seca).

Autor	Componente					
	Humedad	Proteína	Grasa	Cenizas	Fibra	Carbohidratos
1	11,23	14,13	6,90	2,78	7,68	68,52
2	12,61	13,18	7,09	3,15	5,86	76,58
3	12,99	17,51	8,47	3,39	2,82	76,27
4	12,49	12,49	8,66	3,04	6,75	75,82
5	15,31	16,28	7,00	2,74	8,07	73,99
Promedio	12,93	14,72	7,63	3,02	6,24	74,23

FUENTE: (1) Tapia *et al.*, 2014; (2) Collazos *et al.*, 1975; (3) Thoufeek Ahamed *et al.*, 1998; (4) Reyes *et al.*, 2009 y (5) USDA, 2013.

2.2. ESPECTROCOPIA

La espectroscopía se encarga de la producción, medición e interpretación de espectros que surgen de la interacción de la radiación electromagnética con la materia. Hay muchos métodos espectroscópicos diferentes disponibles para la resolución de una amplia gama de problemas analíticos. Los métodos difieren con respecto a la especie a analizar (tales como las especies moleculares o atómicas), el tipo de interacción radiación-materia a monitorizar (por ejemplo, absorción emisión, o de difracción), y la región del espectro electromagnético usado en el análisis. Los métodos espectroscópicos son muy informativos y ampliamente utilizados para los análisis cuantitativos y cualitativos. Los métodos espectroscópicos basados en la absorción o emisión de radiación en los rangos de frecuencia del ultravioleta (UV), visible (VIS), infrarrojo (IR), y radio (resonancia magnética nuclear, NMR); son los encontrados más comúnmente en los laboratorios de análisis de alimentos (Nielsen, 2009).

Se denomina espectroscopio a cualquier instrumento utilizable para producir o estudiar un espectro, en cualquier región espectral. Si el aparato registra el espectro sobre una placa fotográfica, se suele denominar espectrógrafo. En el caso de que proporcione directamente una lectura de la longitud de onda o frecuencia de radiación el aparato se llama espectrómetro; si al mismo tiempo da también una medida de la intensidad de esa radiación se suele llamar espectrofotómetro (Morcillo y Orza, 1979).

Cuadro 2: Comparación del contenido de aminoácidos esenciales en granos de quinua, cereales, frejol, carne, pescado, leche y la proteína patrón.

Alimento / Autor	Aminoácidos esenciales									
	Arginina	Fenilalanina	Histidina	Isoleucina	Leucina	Lisina	Metionina	Treonina	Triptófano	Valina
Quinua (1)	7,3	4,0	3,2	4,9	6,6	6,0	2,3	3,7	0,9	4,5
Arroz (2)	6,9	5,09	2,1	4,1	8,2	3,8	2,2	3,8	1,1	6,1
Cebada (3)	4,8	5,2	2,2	3,8	7,0	3,6	1,7	3,5	1,4	5,5
Maíz (4)	4,2	4,7	2,6	4,0	12,5	2,9	2,0	3,8	0,7	5,0
Trigo (4)	4,5	4,8	2,0	4,2	6,8	2,6	1,4	2,8	1,2	4,4
Frejol (5)	6,2	5,4	3,1	4,5	8,1	7,0	1,2	3,9	1,1	5,0
Carne (6)	6,4	4,1	3,5	5,2	8,2	8,7	2,5	4,4	1,2	5,5
Pescado (6)	5,6	3,7	-	5,1	7,5	8,8	2,9	4,3	1,0	5,0
Leche (6)	3,7	1,4	2,7	10,0	6,5	7,9	2,5	4,7	1,4	7,0
Patrón FAO (7)	-	6,0	-	4,0	7,0	5,5	3,5	4,0	1,0	5,0

FUENTE: Información extraída del Catálogo del Banco de Germoplasma de Quinua, elaborado por Gómez y Eguiluz (2012): (1) Cardozo y Tapia (1979), Mahoney *et al.* (1975), Marroquín (1983), Risi y Galwey (1984) y Romero (1981); (2) Lásztity (1984) y Romero (1981); (3) Lásztity (1984) y Risi Galwey (1984); (4) Lásztity (1984), Risi Galwey (1984) y Romero (1981); (5) Sosulski (1983); (6) Romero (1981); (7) Mahoney *et al.* (1975) y Romero (1981).

2.2.1. ESPECTRO ELECTROMAGNÉTICO

Se llama espectro electromagnético al conjunto de todas las radiaciones electromagnéticas ordenadas por la magnitud de su frecuencia o de su longitud de onda (Barrio *et al.*, 2008). Esta distribución está representada en la Figura 3.

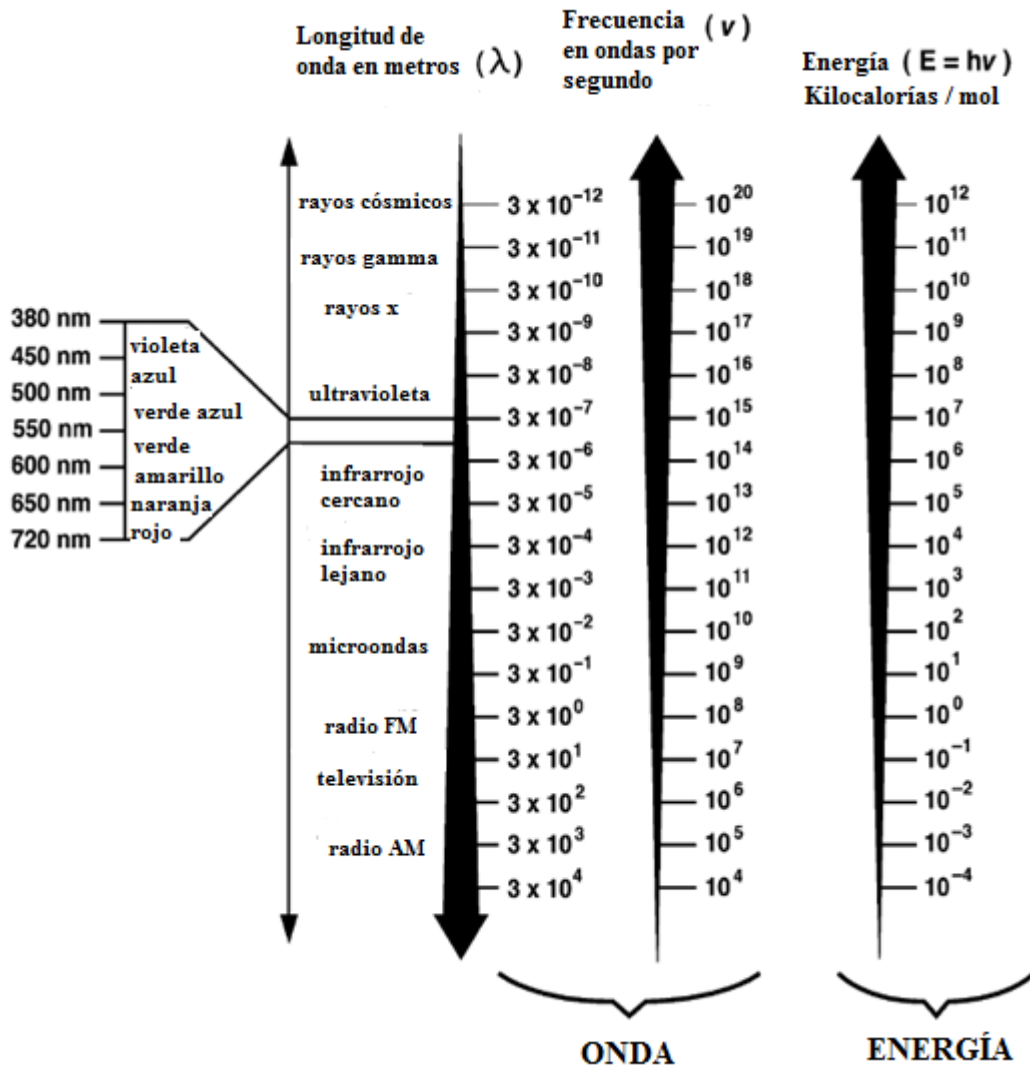


Figura 3: Diagrama del espectro electromagnético.

FUENTE: Nielsen (2009).

2.2.2. ESTADOS ENERGÉTICOS DE LA MATERIA

El contenido de energía de la materia es cuantificada. En consecuencia, el contenido de energía potencial o interna de un átomo o molécula no varía de una manera continua, sino más bien en una serie de pasos discretos (Nielsen, 2009).

a. NATURALEZA CUÁNTICA DE LA MATERIA

Los átomos y moléculas, en condiciones normales, existen predominantemente en el estado basal, que es el estado de energía más bajo. En estado fundamental, los átomos y moléculas pueden ganar energía, en cuyo caso serán elevados a uno de sus estados de energía más altos, referidos como estados excitados. La naturaleza cuántica de los átomos y moléculas pone limitaciones en los niveles de energía que están disponibles para estas especies. En consecuencia, habrá niveles de energía interna específicos "permitidos" para cada especie atómica o molecular. El conjunto de los niveles de energía disponibles para cualquier átomo o molécula será distinto para cualquiera de esas especies. Del mismo modo, las posibles separaciones de energía entre los niveles de energía internos permitidas serán características de una especie. Por lo tanto, el conjunto de espaciados de energía potencial para una especie se puede utilizar cualitativamente como una huella dactilar (Nielsen, 2009).

b. NIVELES ENERGÉTICOS ELECTRÓNICOS, VIBRACIONALES Y ROTACIONALES

La energía potencial relativa de un átomo o molécula corresponde a la diferencia de energía del estado energético en el que las especies existen y su estado basal. La Figura 4 es un diagrama parcial del nivel energético molecular representando los niveles de energía potencial de una molécula orgánica. El menor nivel energético de la Figura 4, la línea inferior de mayor grosor, representa el estado basal. Existen tres estados energéticos electrónicos representados, cada uno con sus correspondientes niveles de energía vibracional y rotacional. Cada uno de los estados electrónicos corresponde al electrón de un orbital. Los electrones en diferentes orbitales son de diferentes potenciales energéticos. Cuando un electrón cambia de orbital, por ejemplo cuando absorbe o emite un fotón de energía apropiada, se denomina transición electrónica pues el electrón está cambiando de niveles energéticos. Sin embargo, cualquier cambio en la energía potencial de un electrón

puede resultar en un cambio correspondiente de la energía potencial de un átomo o molécula en la que el electrón esté asociado (Nielsen, 2009).

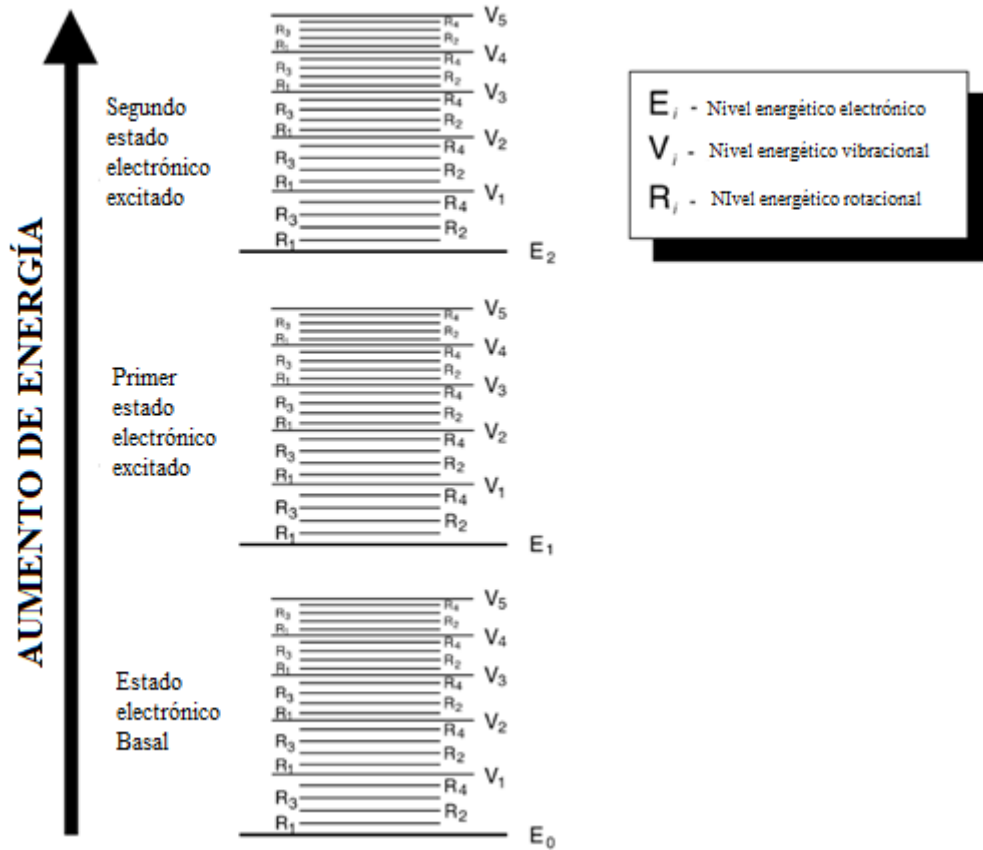


Figura 4: Diagrama de niveles de energía de tres estados electrónicos.

FUENTE: Nielsen (2009).

Nielsen (2009), señala que la energía interna de un átomo es descrita en términos de sus niveles energéticos electrónicos, mientras que la energía interna de una molécula es dependiente de sus energías electrónicas, vibracionales y rotacionales. El investigador hace uso del hecho de que cada una de estas energías asociadas es cuantificable y que las diferentes especies podrán tener algo diferente sus espaciamientos de energía. La forma algebraica de estos argumentos está en las ecuaciones (1) y (2).

$$E_{atom} = E_{electrónica} \quad (1)$$

$$E_{molécula} = E_{electrónica} + E_{vibracional} + E_{rotacional} \quad (2)$$

2.2.3. TRANSICIONES DE NIVELES ENERGÉTICOS EN ESPECTROSCOPIA

Según Albella *et al.* (1993), las técnicas espectroscópicas tratan de la interacción de la radiación con la materia. Esta interacción se traduce, en general, en transiciones entre diferentes niveles de energía de los sistemas y por consiguiente en el cambio de frecuencia de la radiación. Las transiciones, se deben a dos fases a una primera que es la absorción de la radiación y una segunda consecuente de la primera, que es la emisión de la radiación.

a. ABSORCIÓN DE RADIACIÓN

Nielsen (2009), indica que la absorción de la radiación por un átomo o molécula es el proceso en que la energía del fotón de una radiación electromagnética es transferida a las especies absorbentes. Cuando un átomo o molécula absorbe un fotón de luz, su energía interna aumenta el equivalente a la cantidad de energía de ese fotón. Por eso en el proceso de absorción, las especies van de un estado energético bajo a un estado excitado. En la mayoría de los casos, las especies se encuentran en el estado basal antes de la absorción. Como el proceso de absorción puede considerarse cuantitativo (toda la energía del fotón es transferida a las especies absorbentes), el fotón absorbido debe tener un contenido energético que coincida exactamente con la diferencia de energía entre los niveles energéticos en que la transición ocurre. Consecuentemente, si se plotea la energía del fotón versus la absorbancia relativa de una radiación únicamente compuesta de fotones de esa energía, se observará un espectro característico de absorción, la forma estará determinada por la absorptividad relativa de los fotones de diferente energía. La absorptividad de un compuesto es una constante de proporcionalidad dependiente de la longitud de onda que relaciona la concentración de las especies absorbentes con su absorbancia medida experimentalmente en condiciones definidas. Un resumen de las transiciones relativas a las absorciones atómicas y moleculares en espectroscopía, incluyendo sus correspondientes regiones de longitudes onda se presenta en el Cuadro 3.

b. EMISIÓN DE RADIACIÓN

Nielsen (2009), indica que la emisión es esencialmente lo contrario al proceso de absorción, ocurre cuando la energía de un átomo o molécula es liberada en forma de fotón de radiación. Normalmente, una molécula en estado excitado permanecerá así por un periodo muy corto de tiempo, antes de relajarse de nuevo al estado basal. Existen muchos

procesos de relajación a través de los cuales una molécula excitada puede disipar la energía. El proceso de relajación más común para una molécula excitada es disipar la energía a través de una serie de pasos pequeños provocados por colisiones con otras moléculas. La energía es convertida en energía cinética, el resultado neto es la disipación de la energía como calor. Bajo condiciones normales, el calor disipado no es suficiente para afectar al sistema de forma medible.

Cuadro 3: Transiciones asociadas a métodos espectroscópicos y longitudes de onda.

Región de Longitud de onda	Límite de longitud de onda	Tipo de espectroscopía	Rango usual de longitud de onda	Transiciones típicas en sistemas químicos con energías similares
Rayos gamma	0,01 – 1 A	Emisión	<0,1 A	Ordenamientos nucleares de protones y electrones
Rayos X	0,1 – 10 nm	Absorción, emisión, fluorescencia y difracción	0,1 – 100 A	Electrones de la capa interior
Ultravioleta	10 – 380 nm	Absorción, emisión y fluorescencia	180 – 380 nm	Electrones de la capa exterior de átomos, electrones de enlaces moléculas
Visible	380 – 750 nm	Absorción, emisión y fluorescencia	380 – 750 nm	Electrones de la capa exterior de átomos, electrones de enlaces moléculas
Infrarrojo	0,075 – 1000 μ m	Absorción	0,78 – 300 μ m	Posición de vibración de los átomos en enlaces moleculares
Microondas	0,1 – 100 cm	Absorción / Resonancia de spin	0,75 – 3,75 cm / 3 cm	Posición de rotación en moléculas / Orientación de los electrones

«continuación»				
		electrónico		apareados en un campo magnético aplicado
Radio	1 – 1000 m	Resonancia magnética nuclear	0,6 – 10 m	Orientación de núcleos en un campo magnético aplicado

FUENTE: Nielsen (2009).

2.3. ESPECTROSCOPIA INFRARROJA.

La región del infrarrojo comprende el intervalo espectral de 780 a 10^6 nm. Según el fenómeno espectroscópico que provoca la absorción de energía por parte de la materia, podemos dividir esta región en tres zonas: (1) Infrarrojo cercano de 780-2500 nm (NIR), (2) Medio de $2500-4 \cdot 10^4$ nm (MIR) y (3) Lejano de $4 \cdot 10^4-10^6$ nm (FIR). La absorción en el infrarrojo cercano se debe a los armónicos y bandas de combinación de las vibraciones fundamentales de las moléculas, mientras que en el infrarrojo medio proviene de las vibraciones moleculares fundamentales. En el infrarrojo lejano se debe a las rotaciones moleculares (Peguero, 2010).

Nielsen (2009), señala que la radiación infrarroja es una energía electromagnética con longitudes de onda (λ) más largas que las de la luz visible pero menores a las de microondas (ver Figura 3). La radiación IR puede ser medida en términos de su frecuencia, lo cual es útil porque la frecuencia está relacionada directamente a la energía de radiación por la ecuación (3).

$$E = h\nu \tag{3}$$

Dónde:

E= Energía del sistema

h = Constante de Planck

v = Frecuencia en hertz

El mismo autor señala que la frecuencia es expresada comúnmente como número de onda ($\tilde{\nu}$, como su recíproco en centímetros, cm^{-1}). Los números de onda son calculados según la ecuación (4).

$$\tilde{\nu} = \frac{1}{(\lambda \text{ en cm})} = \frac{10^4}{(\lambda \text{ en cm})} \quad (4)$$

Una molécula absorbe radiación infrarroja cuando experimenta un cambio neto en el momento dipolar como consecuencia de su movimiento de vibración o rotación. Si esto ocurre, el campo eléctrico asociado a la radiación puede interactuar con el campo eléctrico originado por la fluctuación del momento dipolar de la molécula. Si la frecuencia de radiación iguala a la frecuencia de vibración natural de la molécula, ocurre una transferencia neta de energía que da lugar a un cambio en la amplitud de la vibración molecular y como consecuencia se absorbe la radiación (Peguero, 2010).

2.3.1. VIBRACIONES MOLECULARES

Nielsen (2009), señala que una molécula puede absorber radiación IR solo si vibra de manera que su distribución de cargas, y por ende su momento dipolo, cambia durante la vibración. A pesar de que existen muchas vibraciones posibles en una molécula poliatómica, las vibraciones más importantes que producen un cambio en el momento dipolo son las de estiramiento (ver Figura 5) y deformación (ver Figura 6).

Según FOSS (2013), cuando una radiación del espectro electromagnético encuentra una sustancia, esta puede ser tanto absorbida como transmitida dependiendo de su frecuencia y la estructura molecular de la sustancia. Las moléculas naturalmente vibran de dos maneras: (1) Estiramiento, siendo cambios continuos en la distancia inter-atómica entre las longitudes de los enlaces de dos átomos. (2) Deformación, cambios en el ángulo entre los enlaces.

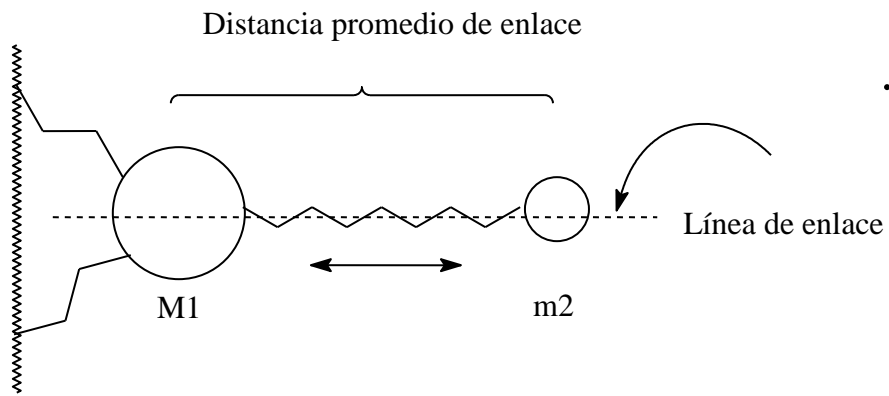


Figura 5: Vibración de estiramiento de una molécula.
FUENTE: Hesse *et al.* (1995).

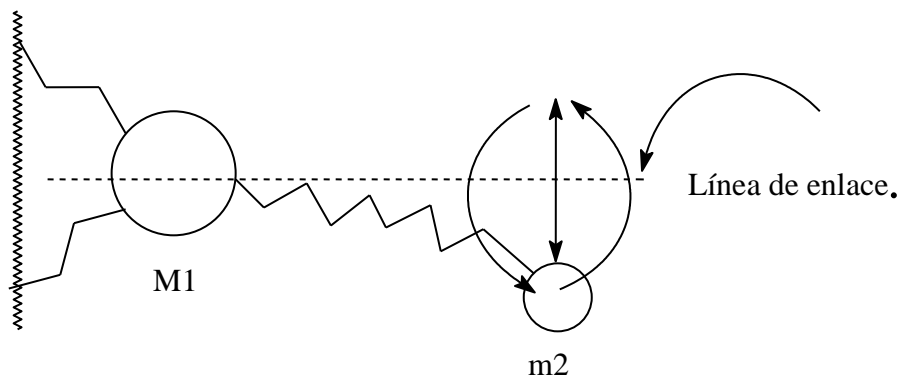


Figura 6: Vibración de deformación de una molécula.
FUENTE: Hesse *et al.* (1995)

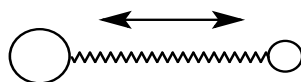


Figura 7: Vibración de estiramiento de una molécula biatómica.
FUENTE: Hesse *et al.* (1995).

Una molécula biatómica como la de la Figura 7, solo poseerá un modo de vibración que será el de estiramiento a lo largo del enlace. La deformación no existe ya que correspondería a una rotación antes que a una vibración (Hesse *et al.*, 1995).

Si la molécula tiene más de dos átomos, los posibles modos de vibración aumentan de manera considerable, llegando a $(3n-5)$ los modos en moléculas lineales y $(3n-6)$ en las no lineales más complejas (Hesse *et al.*, 1995).

a. FACTORES QUE AFECTAN LA FRECUENCIA DE VIBRACIÓN

Según Nielsen (2009), una vibración molecular se puede representar como un oscilador armónico (Ver Figura 8), dónde cualquier vibración molecular tiene un nivel energético dado por la ecuación (5).

$$E = \left(v + \frac{1}{2} \right) * \left(\frac{h}{2\pi} \right) * \sqrt{\frac{k}{\frac{m_1 * m_2}{m_1 + m_2}}} \quad (5)$$

Dónde:

v = Número quantum vibracional (valores enteros positivos incluyendo cero)

h = Constante Planck

k = Constante de fuerza del enlace

m_1 y m_2 = Masas de los átomos individuales involucrados en la vibración

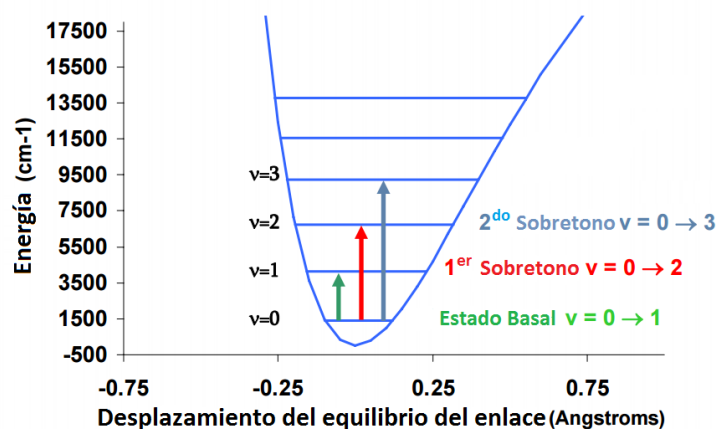


Figura 8: Representación de los estados energéticos, *quantum*, y relación con los sobretonos.

FUENTE: Thermo (2015).

La energía vibracional, y por ende la frecuencia de vibración, es directamente proporcional a la masa del sistema molecular. Por lo tanto, los diferentes grupos funcionales químicos vibrarán a diferentes frecuencias. Un grupo funcional molecular que está vibrando puede absorber energía radiante para moverse del estado vibracional más bajo ($v=0$) al primer estado de excitación ($v=1$), y la frecuencia de la radiación que puede hacer que esto ocurra es idéntica a la frecuencia inicial de vibración del enlace. La frecuencia está referida como la absorción fundamental. Las moléculas también pueden absorber la radiación para moverse a un mayor estado de excitación ($v=2$ ó 3), de manera que la frecuencia de la radiación absorbida es dos a tres veces la de la frecuencia fundamental. Estas absorciones están referidas como sobretonos, y las intensidades de estas absorciones son muy inferiores que a la frecuencia fundamental ya que estas transiciones son menos favorecidas. Las bandas de combinación también pueden ocurrir si dos o más vibraciones diferentes interactúan para dar bandas que son sumas de sus frecuencias fundamentales. El resultado global es que cada grupo funcional dentro de la molécula absorbe la radiación IR en distintas bandas de las longitudes de onda en lugar de ser una absorción continua (Nielsen, 2009).

2.3.2. RELACIÓN DE LOS ESPECTROS CON LA ESTRUCTURA MOLECULAR.

El espectro de infrarrojo cercano contiene información sobre la mayor cantidad de enlaces químicos X-H en productos agrícolas y alimenticios. El espectro es la sumatoria de las bandas de absorción NIR/NIT de los grupos funcionales y/o los componentes sensoriales de una sustancia. El espectro contiene además información de las interacciones ligeras con la muestra y los componentes del instrumento como la lámpara de espectro, estabilidad del instrumento; y recopilación de datos y los errores de cuantificación, tales como la resolución de bits de datos, integración de señales, suavizamiento de datos, etc. (Infrasoft International, 2005).

Las bandas de absorción observadas en las regiones NIR son principalmente sobretonos y combinaciones. Por ello, las absorciones tienden a ser de intensidad débil. Sin embargo, esto es realmente una ventaja desde que las bandas de absorción que tienen suficiente intensidad para ser observadas en la región NIR se originan principalmente de grupos funcionales que tienen un átomo de hidrógeno unido al carbono, nitrógeno u oxígeno. Estos son grupos comunes en los mayores componentes de los alimentos tales como agua,

proteínas, lípidos y carbohidratos. El Cuadro 4 muestra las bandas de absorción relacionadas con un número importante de componentes de alimentos (Nielsen, 2009).

Según Willard *et al.* (1991), el espectro del infrarrojo de un compuesto es esencialmente la superposición de bandas de absorción de sus grupos funcionales específicos, aunque las interacciones débiles con los átomos circundantes de la molécula le confieren un sello de individualidad al espectro de cada compuesto. En la Figura 9 se puede apreciar el espectro infrarrojo de compuestos sencillos como lo son el agua (rojo) y el aceite (azul), de esta manera se observa que cada compuesto tiene un espectro que lo caracteriza. Por otro lado, en la Figura 10 se puede apreciar en qué lugares del espectro del infrarrojo se encuentran los sobretonos de los grupos funcionales, estos se distribuyen en diferentes longitudes de onda del espectro electromagnético.

En efecto, se puede observar la absorción de todos los componentes de una muestra compleja, por ejemplo: un alimento, como un conjunto de curvas. Al analizarse los espectros de una gran cantidad de muestras (ver Figura 11) de un mismo alimento (quinua), se observa que ninguna tiene un espectro igual que otra. Sin embargo, tienen gran parecido en la distribución de las curvas.

En algunas situaciones los métodos de referencia aceptados actuales, usados para calibrar los instrumentos NIR/NIT no están bien definidos químicamente y pueden ser muy difíciles de relacionar con los datos espectroscópicos. Estos métodos encajan en tres categorías generales: (1) Análisis proximal, desarrollado , hace más de 100 años para estimar el valor nutritivo general; (2) Las técnicas detergentes de Van Soest, desarrolladas a fines de 1960 como un camino analítico con las mejores definiciones químicas nutricionales y; (3) Metodologías analíticas específicas diseñadas para medir componentes químicos que pueden ser usados para describir mejor el perfil nutricional de una sustancia alimenticia (Infrasoft International, 2005).

Cuadro 4: Grupos funcionales preferentemente detectados según porción del infrarrojo cercano.

Rango espectral (nm)	Naturaleza de la vibración
2200 – 2450	C-H alargamiento del enlace
2000 – 2200	N-H, O-H, alargamiento del enlace
1650 – 1800	C-H, alargamiento y 1 ^{er} sobretono
1400 – 1500	N-H, O-H, alargamiento y 1er sobretono
1100 – 1225	C-H, alargamiento y 2 ^{do} sobretono
950 – 1100	N-H, O-H, alargamiento y 2 ^{do} sobretono
850 – 950	C-H, alargamiento y 3 ^{er} sobretono
775 – 850	N-H, O-H, alargamiento y 3 ^{er} sobretono

FUENTE: Revista Mexicana de Ciencias Farmacéuticas (2007).

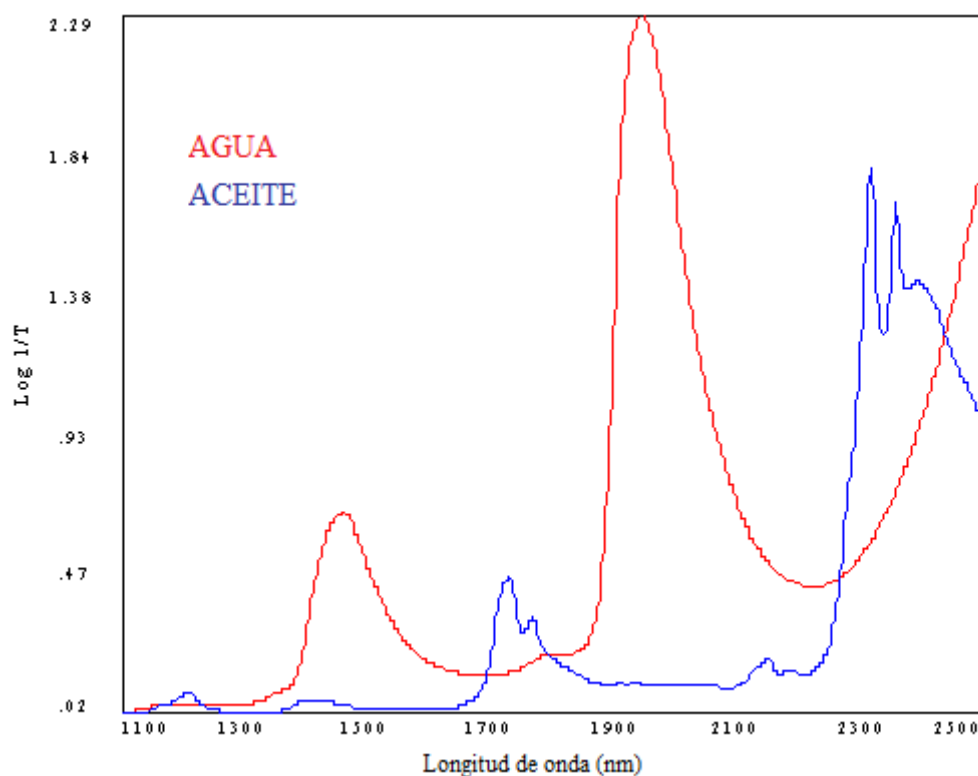


Figura 9: Espectros infrarrojo del agua (rojo) y del aceite (azul).

FUENTE: FOSS (2012).

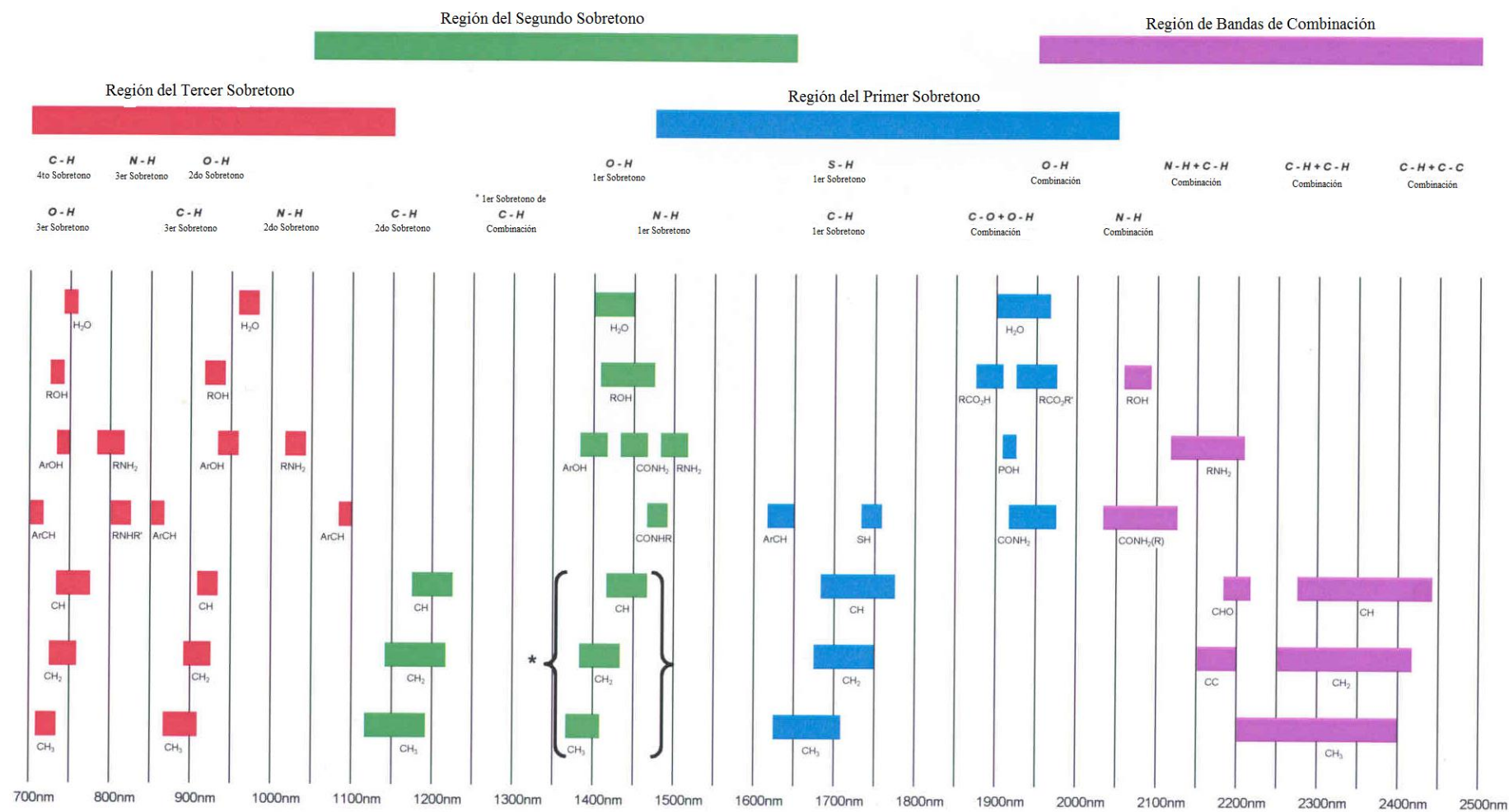


Figura 10: Ubicación y sobretonos de grupos funcionales en el espectro electromagnético.

FUENTE: FOSS (2012).

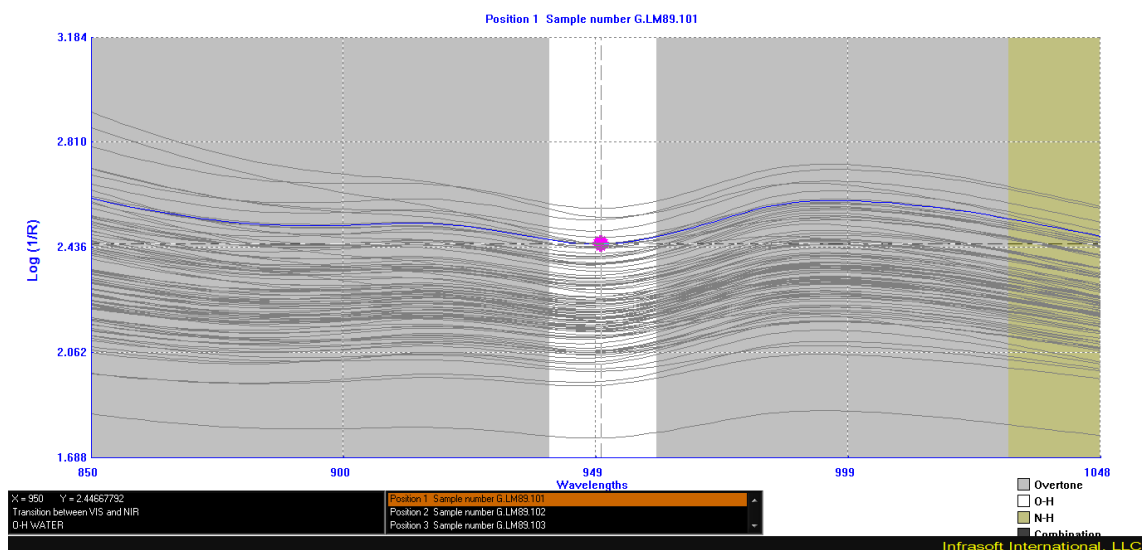


Figura 11: Espectro de infrarrojo de varias muestras de una misma población.

FUENTE: Imagen obtenida de WinISI.

2.4. INSTRUMENTACIÓN

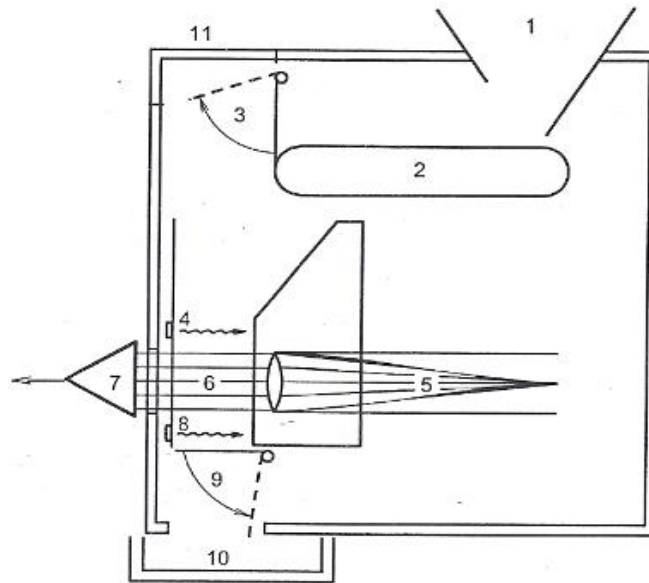
Según Peguero (2010), el esquema básico de un espectrofotómetro NIR consta de: fuente de radiación, selector de longitudes de onda, el compartimiento de la muestra y el detector. En la Figura 12 se expone el esquema de funcionamiento del espectrofotómetro Infratec 1241 de la firma FOSS.

2.4.1. LA FUENTE DE RADIACIÓN

Es un dispositivo capaz de generar el haz de luz necesario para irradiar la muestra. La fuente de radiación más usada es la lámpara halógena de filamento de tungsteno con ventana de cuarzo, capaz de proporcionar un espectro continuo en la región de 320 nm a 2500 nm (Peguero, 2010).

2.4.2. SISTEMA DE SELECCIÓN DE LONGITUDES DE ONDA

Según Peguero (2010), disponer de un sistema adecuado de selección de longitudes de onda es un factor importante para poder obtener un buen nivel de sensibilidad. Un buen sistema de selección de longitudes de onda debe ser capaz de: (1) Proporcionar un ancho de banda preciso y exacto para la longitud analítica, (2) Proporcionar señales altas a fin de conseguir relación señal/ruido satisfactoria.



Leyenda: (1) Tolva de alimentación, (2) Transportador, (3) Obturador superior, (4) Sensor superior, (5) Luz monocromática (monocromador) , (6) Celda de muestra, (7) Detector, (8) Sensor inferior , (9) Obturador inferior, (10) Cajón de recepción de muestras, (11) Cubierta de acceso directo a celda.

Figura 12: Esquema de funcionamiento del Infratec 1241

FUENTE: FOSS (2007)

2.4.3. COMPARTIMIENTO DE LA MUESTRA

Una de las grandes ventajas de la espectroscopía NIRS es la versatilidad y adaptabilidad que presenta para analizar muestras de distinta naturaleza. La ausencia de pretratamiento de la muestra para el registro del espectro, permite disponer de gran cantidad de accesorios adaptables a cada situación (Peguero, 2010).

Según Peguero (2010), existen tres tipos de registros de espectros NIRS: (1) Transmitancia - NIT, (2) Reflectancia – NIR y (3) Transflectancia - NIRT. La Figura 13 muestra los esquemas de los tres modos de registro de espectros.

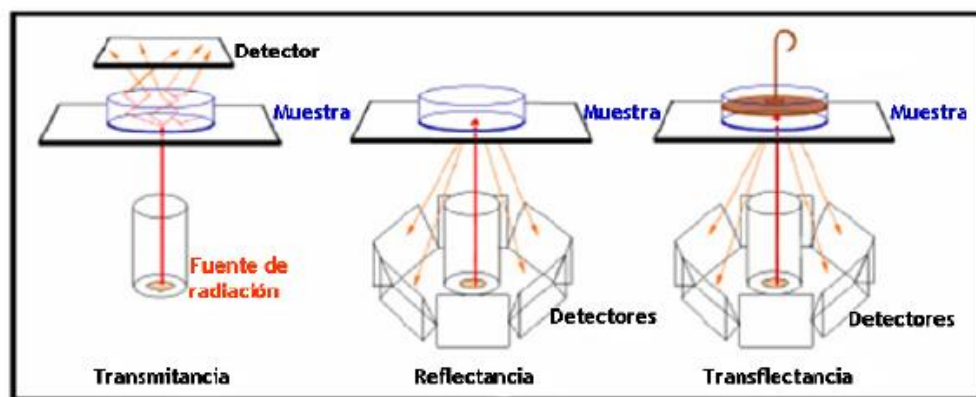


Figura 13: Esquema de los tres modos de registro de espectros NIR.

FUENTE: Peguero (2010).

En modo transmitancia (NIT) se pueden registrar gases, líquidos, semilíquidos y sólidos. En este modo el haz de luz atraviesa la muestra hasta el detector. En las medidas por reflectancia (sólidos y semisólidos), el haz de luz es reflejado por la propia muestra y el haz reflejado es recogido por el detector. Finalmente, un caso intermedio lo ocupa las medidas por transflectancia (líquidos y semilíquidos), donde el haz de luz atraviesa la muestra, se refleja en un reflector que está en contacto con la misma y finalmente llega al detector (Peguero, 2010).

La presente investigación utilizó el equipo Infratec 1241 de la firma FOSS, el cual tiene un registro del espectro en el modo transmitancia – NIT.

Según Peguero (2010), la absorción de la radiación sigue la ley de Lambert-Beer. La transmitancia de la solución es la fracción de radiación incidente transmitida por la solución. Ambas medidas se relacionan en la siguiente ecuación:

$$A = \log \frac{1}{T} = \log \frac{P_o}{P} = abc \quad (6)$$

Dónde A es la absorbancia de la muestra, T la transmitancia, Po la intensidad de la radiación incidente, P la intensidad de la radiación transmitida, a la absorptividad molar, b el camino óptico y c la concentración. Además, Peguero (2010) indica que pueden producirse desviaciones del cumplimiento de esta ley debidas a causas muy diversas (reflexión y/o dispersión de la radiación, concentraciones muy elevadas, interacciones químicas de la muestra, etc.).

2.4.4. DETECTOR

Los detectores más utilizados en espectroscopía NIR están contruidos con semiconductores como In Ga As, Pb S, In As, In Sb y Si. El detector más ampliamente utilizado es el de PbS, que posee sensibilidad adecuada entre 900 y 2600 nm. Para medidas por transmisión en sólidos se utilizan detectores de Arseniuro de Indio y Galio (In Ga As), operativos de 600 a 1900 nm (Peguero, 2010). En la presente investigación se ha utilizado un detector de silicio.

2.5. FUNCIONAMIENTO

Cuando la luz incide sobre una muestra, una parte de los fotones puede transmitirse a través de la muestra y el resto es reflejado o absorbido por algunos enlaces covalentes. Al absorber energía, los enlaces de las moléculas vibran en dos formas fundamentales: se extienden o se doblan. La vibración resultante se disipa, provocando un mero calentamiento de la muestra. La absorción es selectiva y depende de los grupos moleculares involucrados. Así, la absorción de luz se estima por diferencia entre la luz incidente y la reflejada o transmitida. (Murray, 1993).

2.5.1. LEY DE LAMBERT Y BEER

La técnica NIRS obedece a la ley de Lambert – Beer, la cual establece que cuando un haz de radiación atraviesa una especie absorbente de espesor “b” y de concentración “c” se producen interacciones de los átomos y moléculas absorbentes con los fotones, con lo cual la intensidad del haz disminuye. Si consideramos T como la fracción de la radiación incidente transmitida por el medio. Es imprescindible mantener un paso óptico constante, con el fin de minimizar los fenómenos de radiación difusa, la cual modifica el paso óptico efectivo. Esto determina que para un material de naturaleza química heterogénea, el espectro obtenido en la región del infrarrojo cercano es la combinación de bandas de absorciones parciales sobrepuestas o muy cercanas, que suelen confundirse en una línea suavizada, en que se encuentran picos, valles y curvas (Cozzolino *et al.*, 2001).

Según Barnes *et al.* (1989), se requiere una clara definición del término transmitancia, el cual es un método de medida que basado en el siguiente principio: La fuente de la muestra y el detector están colocados sobre una misma recta. Se mide la intensidad del rayo I que atraviesa la materia con respecto a la intensidad de partida I_0 (ver Figura 14). Parte de la

luz transmitida está disminuida por pérdida por reflexión y dispersión.

La transmitancia se expresa a menudo como porcentaje (7), mientras que la absorbancia se expresa como una relación logarítmica de la primera (8).

$$T = \frac{I}{I_0} * 100 \quad (7)$$

$$\text{Absorbancia} = \log \frac{1}{T} \quad (8)$$



Dónde:

I_0 = intensidad de la radiación incidente.

I = intensidad de la radiación transmitida.

Figura 14: Haz de radiación que atraviesa un cuerpo.

FUENTE: Barnes *et al.* (1989).

Siendo “a” una constante de proporcionalidad llamada absortividad. La magnitud de “a” dependerá de las unidades empleadas para “b” y “c”. A menudo “b” es dada en términos de cm y “c” en gramos por litro, entonces la absortividad tiene unidades de $l * g^{-1} * cm^{-1}$. Cuando la concentración se expresa en moles por litro y la longitud de la celda en centímetros, la absortividad se llama absortividad molar, se designa como ϵ y tiene unidades de $l * mol^{-1} * cm^{-1}$ (Day y Underwood, 1989).

Entonces la absorbancia es:

$$A = \epsilon * b * c \quad (9)$$

Además, otros dos términos también son importantes definir: Primero, un espectro es la relación entre la transmisión o absorción de un compuesto con su respectiva longitud de onda; segundo, los valores de referencia son aquellos parámetros fisicoquímicos a ser predecidos usando medidas espectroscópicas. Un espectro puede o no, contener información relacionada a la muestra analizada usando un método de referencia.

La interpretación espectroscópica proporciona una base científica para el establecimiento de una causa conocida y el efecto relacionado entre el instrumento (espectral) y datos de referencia (analito) (Williams y Sobering, 1996).

Según Gómez *et al.* (2010), la ley Lambert y Beer tiene limitaciones, en las que la relación entre la absorción y la concentración pueden ser de ser lineal. Esto debido a que cumple bajo las siguientes condiciones: (1) la concentración de la especie absorbente debe ser baja, (2) la radiación utilizada debe ser monocromática, (3) el único mecanismo de interacción entre la radiación electromagnética y la especie absorbente en estudio es el de absorción, (4) las especies absorbentes no deben interaccionar entre sí ni con el disolvente, (5) las especies absorbentes no deben sufrir transformaciones fotoquímicas, (6) la absorción estudiada corresponde a un volumen de sección transversal uniforme.

2.6. BENEFICIOS DE LA ESPECTROMETRÍA EN EL INFRARROJO CERCANO

Según Peguero (2010), los principales beneficios son los siguientes:

- Es una técnica no destructiva ni invasiva. Permite el análisis *in-line* y *on-line*.
- Se pueden analizar muestras sólidas, líquidas y gaseosas.
- La preparación de la muestra es prácticamente nula y el análisis es rápido.
- El análisis por muestra tiene un costo muy reducido.
- No hay necesidad de usar disolventes y por lo tanto existe generación de residuos.
- Se obtiene tanto información química como física de la muestra.
- A partir de un único conjunto de espectros, se pueden establecer calibraciones paralelas para realizar determinaciones multiparamétricas.

Respecto a los beneficios, Flores (2009), señala que en los últimos años se ha incrementado el interés por realizar evaluaciones NIRS en el lugar de origen de la muestra y no a la inversa. Los equipos usados funcionan con baterías, son de tamaño reducido y carecen de elementos móviles que pudieran dañarse o desalinearse con el traslado. Para la investigación de frutas y hortalizas con NIRS estos equipos pueden ser muy beneficiosos, ya que su portabilidad puede permitir recoger datos espectrales antes o durante la cosecha permitiendo así la posibilidad de realizar ecuaciones de calibración con fines precosecha, y destinados a determinar el momento óptimo de la misma.

Actualmente, la NASA (2016) utiliza la tecnología de infrarrojo cercano para analizar el estado de la vegetación y la composición del suelo desde el espacio aéreo.

2.7. QUIMIOMETRÍA

Peguero (2010) define la quimiometría como la parte de la química que utilizando métodos matemáticos, estadísticos y de lógica formal: (1) diseña o selecciona procedimientos de medida óptimos y (2) proporciona la máxima información relevante de los datos analíticos, esto es logrado mediante la comparación de los datos de referencia y los espectros de las muestras.

Naes *et al.* (2002) junto con William y Norris (2001), indican que el modelo quimiométrico relaciona la información espectral de las muestras que constituyen el colectivo de calibración con sus valores para el parámetro en estudio, proporcionados por un método de referencia el cual suele ser lento, costoso y con un número de muestras a emplear que suele ser limitado. Una vez desarrollado el modelo, este permite predecir el contenido de otras muestras de características similares a las incluidas en el grupo de entrenamiento o calibración.

2.8. ETAPAS DEL PROCESO DE MODELADO

Un buen método de calibración debe satisfacer una serie de requerimientos. En primer lugar los modelos matemáticos aplicados deben realizar una estimación eficiente y por otro lado resolver el problema de colinealidad, que es bastante acusada cuando se trabaja con información espectroscópica NIRS, provocando la inestabilidad de las predicciones. A su vez, los datos de referencia de las muestras obtenidas por laboratorio de vía húmeda deben ser lo más precisos posibles, ya que los resultados obtenidos por un método secundario como NIRS tendrán una precisión similar al del método de análisis convencional usado como referencia y que dio origen a la calibración (Naes *et al.*, 2002).

2.8.1. SELECCIÓN DE UN CONJUNTO DE CALIBRACIÓN O ENTRENAMIENTO.

Es la obtención de un conjunto limitado de muestras de las cuales se conoce la propiedad a determinar. El conjunto debe ser representativo de toda la variabilidad química y física de la muestra con la finalidad de obtener métodos analíticos robustos (Peguero, 2010).

En los modelos cualitativos se deberá recoger variabilidad de diverso tipo: tamaño de partícula y proveedor entre otros; en cambio, para modelos cuantitativos, además de la variabilidad mencionada, se deberá cubrir el rango del parámetro a determinar, por ejemplo, la concentración de analito (Cueva, 2012).

El espectro NIR no solo depende de la composición química de la muestra, sino también de algunas propiedades físicas, tales como el tamaño, la forma, la distribución de las partículas y el grado de compresión de la muestra. Esto es útil en algunos casos, ya que permite la determinación espectroscópica de algunos parámetros físicos de la muestra. Sin embargo, las diferencias físicas pueden dar lugar a efectos multiplicadores en el espectro que, junto con otros efectos aditivos como el cambio de línea base o de la absorción química, si no se toman en cuenta, pueden complicar los modelos de calibración y reducir la calidad de los resultados del análisis cuantitativo (Cueva, 2012).

Según Cueva (2012), una vez preparado el conjunto de muestras, se divide en dos grupos, es decir, el colectivo de calibración y el colectivo de validación. El primero se utiliza para construir el modelo de calibración y el segundo se usa para que evalúe la capacidad predictiva del modelo:

- **Conjunto de entrenamiento o calibración:** son las muestras que formarán parte del desarrollo de la calibración, por lo que interesa que sean muestras que aporten variabilidad y robustez a la calibración.
- **Conjunto de validación:** se trata de muestras redundantes que no van a aportar variabilidad.

Es necesario considerar un tercer grupo de conjunto de anómalas o atípicas, que son muestras que se apartan del promedio global del conjunto total de muestras por diferentes razones. La detección de estas muestras y su posterior clasificación es muy importante, pues se puede tratar de muestras que pertenezcan a algún tipo de variedad distinta a las demás, de un espectro que hemos realizado con el equipo NIR sin estabilizar o averiado, de una muestra con una concentración fuera del rango seleccionado para la calibración, entre otros (Cueva, 2012).

La clasificación de este último conjunto se hace representando cada muestra en un espacio multidimensional que representa los puntos de datos adquiridos. Dicha representación matemáticamente es muy compleja, por lo que aplicamos técnicas de representación en

componentes principales, conocido como PCA, para poder reducir el número de dimensiones y hacerlo más manejable (Cueva, 2012).

2.8.2. DETERMINACIÓN DE LOS VALORES DE REFERENCIA

Se deben usar los métodos de referencia aceptados internacionalmente para la determinación de la humedad, grasa, proteína y cualquier otro componente o parámetro de interés. El método de referencia usado para la calibración debe estar bajo control estadístico, por ejemplo para cualquier muestra, la variabilidad debe consistir en variaciones aleatorias de un sistema reproducible. Es esencial saber la precisión del método de referencia (ISO 12099, 2010). La elección óptima será el método de referencia que proporcione la mayor exactitud y precisión posible de cada analito (Cueva, 2012).

2.8.3. REGISTRO DE LA INFORMACIÓN ANALÍTICA

Para desarrollar un modelo de calibración, debemos “entrenar” al equipo NIRS mediante la adquisición y almacenamiento de una serie de espectros de muestras de datos analíticos conocidos, habiendo sido determinados previamente por métodos tradicionales analíticos de referencia (Cueva, 2012).

Es de gran importancia aportar variabilidad en la concentración del parámetro a analizar para poder obtener un rango adecuado de concentraciones sin dejar huecos que pudiesen afectar el cálculo de interpolación. El tiempo transcurrido entre el análisis de referencia y el análisis por NIRS debe ser tenido en cuenta para, en caso necesario, almacenar la muestra lo mejor posible y evitar variaciones de humedad, color, etc. (Cueva, 2012).

a. ERROR ESTÁNDAR DE LABORATORIO (SEL)

Dardenne (2008), menciona que lo primero que se debe saber al empezar una calibración NIR es el SEL (Error estándar de laboratorio o Error típico del método de referencia). El SEL puede ser el error de repetibilidad o error de reproducibilidad (Si se tiene más de un laboratorio, se puede usar la prueba de SEL de anillos).

Para análisis de muestras por duplicado, Burns y Ciurczak (2007) recomiendan usar la siguiente ecuación para el cálculo del SEL

$$SEL = \sqrt{\frac{\sum_{i=1}^N (L_{i1} - \bar{L}_i)^2}{M}} = \sqrt{\frac{\sum_{i=1}^N (L_{i1} - L_{i2})^2}{2M}} \quad (10)$$

Dónde:

L_{i1} : Valor de laboratorio de la muestra i para la repetición 1.

L_{i2} : Valor de laboratorio de la muestra i para la repetición 2.

\bar{L}_i : Promedio del valor de laboratorio para dos resultados $(L_{i1}+L_{i2}/2)$.

M : Número de muestras

Según García (2002) y Burns y Ciurczak (2007), si se realiza un análisis de múltiples réplicas (más de tres), el SEL se calcula de la siguiente manera.

Siendo:

$$SEL = \sqrt{\frac{\sum_{j=1}^M \sum_{i=1}^N (L_{ij} - \bar{L}_j)^2}{M(N-1)}} \quad (11)$$

Dónde:

L_{ij} : Valor de laboratorio de la muestra j para la repetición i .

\bar{L}_j : Valor medio de los valores de laboratorio.

M : Número de muestras.

N : Número de repeticiones.

Cueva (2012), señala que el SEL nos da una idea de la precisión y exactitud del proceso analítico utilizado para calcular el valor de referencia que se utilizará para el “entrenamiento” del NIR. Al comparar los datos estadísticos calculados por el NIR, con el error de laboratorio se tendrá un dato importante a la hora de intuir cómo se comportará el modelo de calibración desarrollado.

Dardenne (2008) indica que si se conoce la desviación estándar (SD) de los valores de referencia de la población, y se reemplaza el Error Estándar de Calibración (SEC) por el SEL, considerando el error NIR = 0. Se obtendrá el máximo valor R^2 que se puede esperar aún sin empezar el modelamiento de la calibración. Esto queda definido en la siguiente ecuación:

$$R^2 = \frac{(SD)^2 - (SEC)^2}{SD^2} \quad (12)$$

Dónde:

R^2 : Coeficiente de determinación.

SD: Desviación estándar del parámetro evaluado en el colectivo de calibración

SEC: Error estándar de calibración o error típico de calibración

2.8.4. PRETRATAMIENTO DE LOS DATOS

Según Peguero (2010), una vez obtenidos los espectros NIRS de las muestras seleccionadas, puede ser necesario aplicar un pretratamiento espectral que permita aumentar la relación señal/ruido y por ende acrecentar la señal correspondiente al analito de interés. No todos los pretratamientos consiguen ambos objetivos simultáneamente, por lo que la selección del más apropiado es puramente experimental.

Según Peguero (2010), los pretratamientos espectrales más utilizados son:

- a. **Promedios de espectros:** el cálculo del espectro promedio de una muestra reduce el ruido espectral, ya que éste es considerado aleatorio.
- b. **Centrado de los datos:** se calcula el valor medio de cada variable de la matriz y se resta en cada punto de la columna. El valor medio corresponde al centro del modelo, y los valores de todas las variables se encuentran ahora referidos a este centro.
- c. **Autoescalado de los datos:** posteriormente al centrado de cada columna, se divide el resultado por la desviación estándar de la columna.
- d. **Suavizado:** Existen filtros que permiten minimizar el ruido espectral mediante cálculos polinómicos de convolución o por transformada de Fourier.
- e. **Derivadas:** la baja diferenciación que normalmente presentan los espectros NIR puede ser mejorada derivando los espectros, lo cual permite aumentar las diferencias entre las bandas anchas y solapadas del espectro. La primera derivada elimina los desplazamientos de línea base, mientras que la segunda derivada elimina los términos que varían linealmente con la longitud de onda.
- f. **Variable normal estándar (SNV, *Standard Normal Variate*):** se utiliza principalmente para corregir el efecto de dispersión (*scattering*) producido por ejemplo, por las diferencias en el tamaño de partícula en muestras sólidas.
- g. **Corrección de la señal ortogonal (OSC, *Orthogonal Signal Correction*):** permite eliminar toda aquella información contenida en los espectros que no está correlacionada con la propiedad a determinar. Su aplicación permite obtener modelos de calibración más simples.

h. **Análisis en componentes principales (PCA, *Principal Component Analysis*):** Es uno de los métodos de reducción de variables más utilizado, ya que permite condensar, sin pérdidas relevantes, el gran volumen de información que proporcionan los datos espectroscópicos a la vez que elimina la correlación espectral de variables fuera de interés. Se considera la matriz X de datos, donde cada fila corresponde al espectro de cada muestra m y cada columna a la absorbancia a una longitud de onda k . Matemáticamente, las m muestras representan puntos en el espacio de dimensión k . Normalmente, se dispone de un número elevado de muestras m y de un número elevado de longitudes de onda k . Por lo que la matriz X está formada por $m \cdot k$ valores de absorbancia.

El objetivo del PCA es reducir la dimensionalidad de la matriz X calculando un número inferior de nuevos ejes que expliquen la información original. Estos nuevos ejes se llaman Componentes Principales (PC, *Principal Components*). La característica principal de los PC's es que el primero recoge la máxima variabilidad del conjunto de muestras, el segundo recoge una variabilidad inferior al primero y es ortogonal a éste y así sucesivamente hasta que se explica la máxima información relevante. En la Figura 15 se muestra un grupo de muestras en sus ejes originales y en los nuevos PC's.

Numéricamente, la matriz de datos X se descompone en el producto de 2 matrices: una de *scores*, T , y una de *loadings*, P , quedando un residual representado por la matriz E .

$$X = T * P^t + E \quad (13)$$

T son las coordenadas de cada punto en el nuevo espacio definido por los PC's y P son los cosenos de los ángulos formados entre los ejes originales y los nuevos PC's. El conjunto de datos X queda definido por un nuevo conjunto de variables (PC's) no correlacionadas entre sí y en un nuevo sistema de ejes ortogonales. Uno de los gráficos que se obtiene es el llamado gráfico de *scores*, el cual es un conjunto de puntos (muestras) representado en un número reducido de ejes.

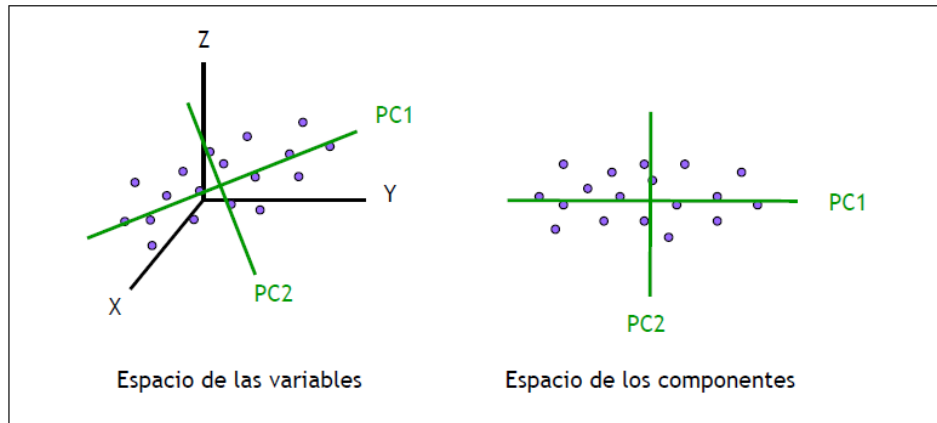


Figura 15: PCA desde un sistema de 3 variables a uno de 2 componentes principales.

FUENTE: Peguero (2010).

a. OUTLIERS Y ESPECTROS ANÓMALOS

William y Norris (2001) indican que la detección y tratamiento de “outliers” es una actividad muy importante durante el proceso de desarrollo de ecuaciones robustas. En general un outlier o muestra anómala es aquella que difiere del resto de las muestras de la población, bien sea por sus características químicas, físico-químicas o por un error de los datos espectroscópicos o de referencia.

La detección, interpretación y posible eliminación de las muestras anómalas es una etapa crítica en el desarrollo de una ecuación de calibración, debido a la gran influencia que provoca su presencia en los resultados del modelo. Los mismos autores hacen un listado de los factores que generan *outliers*:

- Composición química
- Interacción entre los componentes
- Selección de la longitud de onda
- Estado de humedad
- Textura física
- Tamaño y forma de partículas
- Densidad de la masa

En general, las muestras anómalas (en calibración) son indicativas de errores de laboratorio o preparación de las cápsulas. Durante el análisis de rutina de muestras desconocidas, podría ser indicativo de muestras escasamente o no representadas en el colectivo de calibración. Si la precisión del método de referencia y del dato espectroscópico es adecuada no cabría esperar más de un 10 a 15 por ciento de muestras anómalas en un determinado colectivo de calibración (Lucena y Cáceres, 2006). Durante el desarrollo de calibraciones, el especialista en NIRS debe tomar la decisión de mantener dicha muestra en la calibración, retirarla o esperar acumular más muestras similares a ella para su inclusión definitiva (Bendezú, 2010).

En general, las etapas de pretratamiento son útiles para visualizar grupos de muestras, presencia de *outliers* e identificar que tratamientos matemáticos y dispersiones serán las más adecuadas para el desarrollo del modelo.

2.8.5. CONSTRUCCIÓN DEL MODELO QUIMIOMÉTRICO

Según Cueva (2012), el cálculo del modelo quimiométrico seleccionado (cualitativo o cuantitativo) se realiza con un conjunto de muestras llamadas colectivo de calibración. Durante el desarrollo del modelo se evalúan distintos parámetros implicados en el cálculo (pretratamiento, intervalo espectral, umbral de identificación, número de factores, errores de calibración, entre otros).

a. ANÁLISIS CUALITATIVO

Massart *et al.* (1998) mencionan que el análisis cualitativo puede ser realizado de un modo rápido y simple por un conjunto de técnicas quimiométricas llamadas métodos de reconocimiento de pautas (PRM, *Pattern Recognition Methods*).

Los PRM's son métodos basados en la comparación de cuan semejante es la muestra de análisis con otra tomada como referencia. Matemáticamente, esta comparación se realiza según criterios de correlación o distancia. Los PRM's se pueden clasificar según sean supervisados o no supervisados. En los métodos supervisados se tiene en cuenta la pertenencia de los objetos a distintas categorías, que pueden ser conocidas o supuestas, mientras que en los métodos no supervisados no se conoce o se ignora deliberadamente la existencia de categorías (Ramís y García, 2001).

a. Análisis en componentes principales (PCA)

En el acápite 2.8.4. Pretratamiento de los datos (ver numeral h), se ha realizado una explicación amplia del PCA, a continuación algunas aclaraciones.

Peguero (2010) menciona que no existe unanimidad en considerar el PCA como un método de clasificación, por el hecho de que no establece fronteras entre las diferentes clases ni aplica ningún criterio estadísticos que permita discernir las clases. No obstante, es una técnica de reducción de variables que permite visualizar en un espacio de 2 o 3 dimensiones, cuan similares o diferentes son un grupo de muestras entre sí.

La interpretación de los resultados obtenidos con el PCA para una clasificación es puramente subjetiva y se lleva a cabo a partir de la representación de los *scores* de las muestras de un componente principal frente a los *scores* de otro(s) componente(s). Si existe una relación entre las muestras, en el gráfico de los *scores*, los puntos aparecerán agrupados; mientras que si las muestras no se asemejan entre sí, los puntos aparecerán dispersos entre sí (Peguero, 2010).

b. Análisis de agrupamientos (*clusters*)

Bajo el nombre de análisis de agrupamientos se engloban las técnicas de análisis de datos cuya finalidad es descubrir los agrupamientos naturales existentes en una tabla de datos experimentales. Este tipo de análisis constituye el nivel cero de los PRM's y dado su carácter no supervisado es difícil evaluar su rendimiento. Por el contrario, el análisis por agrupamientos es muy eficaz en la investigación científica por la capacidad de generar hipótesis de trabajo. Es conveniente recordar que un algoritmo es inmune a los prejuicios del investigador que examina los datos experimentales y es capaz de manejar las tablas ingentes de datos que genera la actual instrumentación química (Ortiz y Saravia, 2007).

Es necesario introducir el concepto de disimilaridad, el cual es clave en el análisis de agrupamientos porque formaliza operativamente la idea de objetos (variables) parecidas. La disimilaridad es una cantidad, $d(i,j)$, que asocia cada par de objetos (el i -ésimo y el j -ésimo) en base a la matriz Z y es una medida de lo distintas que son sus respectivas filas. La matriz de disimilaridades entre objetos será por lo tanto cuadrada y simétrica (Peguero, 2010).

Las medidas de disimilaridad se realizan mediante el cálculo de distancias entre los objetos o variables. Peguero (2010) menciona diferentes expresiones matemáticas para calcular estas distancias:

- Distancia Euclidea: Es la más difundida de las distancias y está definida mediante la siguiente ecuación:

$$d(Z_i, Z_j) = \sqrt{(Z_{il} - Z_{jl})^2 + \dots + (Z_{in} - Z_{jn})^2} \quad (14)$$

- Distancia de Manhattan: La distancia de Manhattan entre 2 vectores p y q en un espacio vectorial real n -dimensional con un sistema de coordenadas cartesianas fijo, es la suma de las longitudes de las proyecciones del segmento de línea entre los puntos sobre el sistema de ejes. Más formalmente,

$$d_i(p, q) = |p - q|_1 = \sum_{i=1}^n |p_i - q_i| \quad (15)$$

dónde $p = (p_1, p_2, \dots, p_n)$ y $q = (q_1, q_2, \dots, q_n)$ son vectores.

- Distancia de Mahalanobis: la distancia de Mahalanobis entre 2 variables aleatorias con la misma distribución de probabilidad se define mediante la siguiente ecuación:

$$d(p, q) = \sqrt{(p - q)^T S^{-1} (p - q)} \quad (16)$$

dónde $p = (p_1, p_2, \dots, p_n)$ y $q = (q_1, q_2, \dots, q_n)$ son vectores y S es la matriz de covarianza.

La distancia la Mahalanobis se diferencia de la distancia euclidea en que tiene en cuenta la correlación entre las variables aleatorias mediante la matriz de covarianza S .

b. ANÁLISIS CUANTITATIVO

El proceso de calibración permite establecer la relación entre la respuesta instrumental y la propiedad a determinar, utilizando a tal efecto un conjunto de muestras representativas. La espectroscopía NIR proporciona un gran número de variables respuesta para cada muestra, variables que en general no pueden ser asignadas a un solo analito. Esto ha propiciado el desarrollo de métodos de calibración capaces de relacionar múltiples variables con la propiedad a determinar (Peguero, 2010).

a. Regresión lineal múltiple (MLR)

Es un método de análisis cuantitativo que requiere de una selección previa de las variables significativas a partir de las cuales se construye la ecuación de regresión. La regresión lineal múltiple (MLR, *Multiple Linear Regression*) establece una relación lineal entre la señal (longitudes de onda discretas) y la propiedad a determinar mediante un ajuste de mínimos cuadrados (Peguero, 2010).

En dicha regresión se asume una relación entre la absorbancia y la concentración que es inversa a la ley de Lambert-Beer. La concentración a determinar es función de la absorbancia por unos coeficientes de proporcionalidad obtenidos durante la etapa de calibrado (Peguero, 2010).

El desarrollo de modelos MLR requiere una adecuada selección de las longitudes de onda que son utilizadas para su cálculo. Dicha tarea es compleja ya que muchas veces la variación de la concentración a determinar, no puede ser asignada a longitudes de onda discretas y es por ello que su uso en espectroscopía NIR es reducido. Además, la selección de un excesivo número de longitudes de onda puede acarrear problemas de colinealidad, pues pueden existir diferentes variables independientes altamente correlacionadas que aportan la misma información (Peguero, 2010).

b. Regresión en componentes principales (PCR)

Es un método cuantitativo en el que previamente se ha realizado una reducción de variables basándose en un PCA. De este modo, se consigue condensar la información inicial en un número de variables inferior al original. Este método es un caso particular de la regresión lineal múltiple (MLR) en el cual se utilizan los scores obtenidos de un PCA como variables independientes (Peguero, 2010).

Según Peguero (2010), el primer paso del PCR consiste en realizar una descomposición de la matriz X en sus componentes principales:

$$X = TP^T + E \quad (17)$$

Dónde T es la matriz de *scores*, P^t la matriz transpuesta de *loadings* y E el residual.

Una vez elegido el número de componentes principales que se considere como óptimo para describir la matriz X, ésta se puede representar por su matriz de *scores* T:

$$T = XP \quad (18)$$

Hasta aquí lo realizado corresponde a un PCA, al obtener a partir de la matriz de datos X la matriz de *scores* T y la de *loadings* P. A continuación, la matriz de datos Y se puede calcular según la expresión:

$$Y = TB + E \quad (19)$$

Siendo B la matriz de regresores que se halla por mínimos cuadrados, previo conocimiento de los valores de Y del conjunto de calibración:

$$\hat{B} = (T^T T)^{-1} T^T Y \quad (20)$$

El símbolo ^ indica valores calculados. Una vez que se han calculado los regresores y por lo tanto establecido el modelo de calibración, se podrán realizar los cálculos para predecir un conjunto de nuevas muestras. Se calcularán los *scores* de las muestras de predicción T* a partir de sus espectros X* y de los *loadings* P calculados en la calibración:

$$T^* = X^* P \quad (21)$$

Se debe utilizar la matriz de regresores calculada también en la calibración, junto con los *scores* de estas muestras, para el cálculo de la propiedad a determinar en las muestras desconocidas:

$$Y = T^* \tilde{B} \quad (22)$$

Massart *et al.* (1998) señalan que uno de los principales problemas con el PCR es que los componentes principales que mejor representan la matriz de los datos espectroscópicos X, pueden no ser los más apropiados para la predicción de las propiedades de los analitos que se quieren determinar en la calibración.

c. Regresión por mínimos cuadrados parciales (PLS)

Según Peguero (2010), el método de regresión PLS (*Partial Least Squares*) también está basado en una reducción de variables, sin embargo, a diferencia del PCR, la

descomposición de la matriz espectral se realiza simultáneamente con la matriz de la propiedad a determinar. El mismo autor señala que la descomposición se realiza según:

$$X = TP^T + E \quad (23)$$

$$X = UQ^T + F \quad (24)$$

Dónde X es la matriz de datos espectrales e Y la matriz de la propiedad a determinar, T y U son las matrices de *scores*, P y Q las matrices de *loadings* y E y F las matrices de residuales.

La descomposición de ambas matrices no es independiente, sino que se realiza de forma simultánea, estableciéndose una relación interna entre los *scores* de los bloques X e Y:

$$\hat{U} = bT \quad (25)$$

Dónde b es el coeficiente de regresión para cada uno de los factores del modelo.

El cálculo del valor Y de una muestra desconocida se realiza utilizando la relación interna:

$$Y = T * \hat{B}Q^T + F \quad (26)$$

Donde T es la matriz de *scores* de la muestra analizada obtenida del modelo calculado, B es el coeficiente de regresión de cada factor, Q la matriz de *loadings* del modelo y F el residual de la predicción.

En el caso de calcular una sola propiedad a determinar de la matriz Y, el algoritmo recibe el nombre de PLS1 y se puede considerar una simplificación del algoritmo global conocido como PLS2. Este último realiza la determinación simultánea de n variables.

d. Redes neuronales artificiales (ANN)

Las redes neuronales artificiales (ANN, *Artificial Neural Networks*) constituyen una de las alternativas más generales y sistemáticas a los procedimientos de calibración multivariable, han demostrado ser especialmente útiles cuando la relación entre el parámetro a determinar y las variables medidas es de naturaleza compleja y, sobretodo, no lineal (Coello y Maspoch, 2007).

Esencialmente, se pueden definir como un sistema iterativo de cálculo, que toma su nombre al intentar reproducir de forma simple y sencilla el sistema de conexiones que existe entre las neuronas del cerebro humano (Coello y Maspoch, 2007).

2.8.6. ESTRATEGÍAS PARA EL DESARROLLO DEL MODELO QUIMIOMÉTRICO

Uno de los principales problemas que se presentan en el desarrollo de ecuaciones de calibración NIRS, es la dependencia del espectro a diversos factores, a veces incontrolables e incluso desconocidos. Así por ejemplo, el espectro de muestras molidas de un mismo producto, analizadas por reflectancia, puede variar simplemente debido a pequeñas diferencias en el tamaño y la distribución de las partículas, lo que se conoce como efecto “*scatter*” (radiación dispersa). En el caso de productos líquidos o semilíquidos pequeños cambios en el paso óptico, pueden causar diferencias importantes entre espectros de una misma muestra (Lucena y Cáceres, 2006).

Todo ello provoca lo que en análisis NIRS se conocen como efectos aditivos y multiplicativos en el espectro, o desviaciones de la línea base, y provoca la no linealidad de la absorbancia con respecto a la concentración. Esto es el incumplimiento de la ley de Lambert y Beer, en la que la tecnología NIRS, reposa al igual que otras técnicas espectroscópicas (Marten *et al.*, 1991).

Para construir un modelo de calibración “anti-ruido”, muchas estrategias han sido desarrolladas agregando simulaciones de ruido a la data espectral (Sáiz-Abajo *et al.*, 2005). El ruido espectral puede ser atribuido a perturbaciones externas como la fluctuación de la temperatura que puede causar un error sistemático, y diferencia instrumental de datos colectados en diferentes temporadas que lleva errores aleatorios.

Durante la construcción del modelo quimiométrico existe la opción de aplicar un *Scatter* (dispersión), la cual es una función no lineal que puede distorsionar la relación entre el espectro NIR / NIT y el valor de referencia. Hay cinco opciones disponibles: La corrección de dispersión “SNV y *Detrend*” puede evaluarse por separado. SNV escala cada espectro para tener una desviación estándar de 1,0 para ayudar a reducir los efectos del tamaño de partículas. *Detrend* elimina la curvatura lineal y cuadrática de cada espectro. Las otras dos correcciones de dispersión son *Normal Multiplicative Scatter Correction* (MSC) y una versión modificada del MSC usando una corrección para la media y una estandarización a

cada longitud de onda. Estos cinco métodos están disponibles para probar y mejorar la precisión de la calibración. *MSC Inverse* (T) es la única corrección de dispersión apoyada por el instrumento Infratec y es para utilizar exclusivamente con los datos del Infratec. Incluso si la corrección no es necesaria, por lo general no disminuye la exactitud de la calibración (FOSS, 2007).

Los espectros de los productos alimenticios suelen ser mostrados en forma de derivadas de la transmitancia o de la absorbancia. Una forma de expresar los tratamientos matemáticos derivativos es D (*Derivative*), G (*Gap*), S1 (*Smooth 1*) y S2 (*Smooth 2*) dónde D es el orden de la derivada, G es el espacio entre las longitudes de ondas usados para la calibración y usados para calcular la derivada, y S1 y S2 corresponden a las longitudes de los segmentos a ser suavizados. Una derivada de primer orden de $\log(1/T)$ resulta en una curva que contiene picos y valles que corresponden al punto de inflexión en ambos lados de la gráfica del $\log(1/T)$, es difícil interpretar la primera derivada de un espectro porque los picos y los valles no coinciden con los de los espectros $\log(1/T)$. Una derivada de segundo orden muestra un patrón picos de absorción bajando en vez de subiendo. La mayor ventaja de la segunda derivada es que genera pocos o nada picos falsos, con respecto a los espectros $\log(1/T)$. Sin embargo, genera dos valles falsos y en escala positiva para cada enlace en una dirección negativa (Burns y Ciurczak, 2007).

La derivada de tercer orden es raramente usada para interpretar el espectro. Ella exhibe las mismas cualidades que una derivada de primer orden y es mucho más difícil de interpretar. La derivada de cuarto orden, por el otro lado, es un tratamiento matemático muy útil para ayudar en la interpretación del espectro. La derivada de cuarto orden puede proveer de nueva información interesante, pero puede causar problemas con la interpretación del espectro para artificios matemáticos como valles falsos (Burns y Ciurczak, 2007).

Infrasoft International (2005) recomienda un tratamiento matemático derivativo para todos los análisis. Respecto al S1 (suavizamiento), indica que jamás debe ser mayor al *gap* y que el S2 casi nunca se usa y es recomendable usar 1, lo cual significa que no habrá segundo suavizamiento. El autor así mismo señala que muchos tratamientos matemáticos han sido probados y una buena lista para evaluar debería incluir 1,4,4,1, o 2,8,6,1 o 3,10,10,1 o 4,10,10,1. No existe un “mejor” tratamiento matemático. El más adecuado solo puede ser hallado por prueba y error. En la presente investigación, se utilizaron los tratamientos: 1,4,4,1; 2,4,4,1; 2,4,6,1; 2,6,4,1; 2,8,6,1; 3,5,5,1; 3,10,10,1 y 4,10,10,1

Según Infrasoftware International (2005), los métodos de regresión más comunes son: Mínimos cuadrados parciales o *Partial Least Squares* (PLS), Mínimos cuadrados parciales modificados o *Modified Partial Least Squares* (MPLS) y regresión de componentes principales o *Principal Components Regression* (PCR). Según Marten *et al.* (1991), el MPLS y el PLS son los métodos de regresión más empleados en aplicaciones agroalimentarias.

El PLS es similar al PCR, en que los datos espectrales son asumidos como una función de unas pocas variables subyacentes. La regresión de componentes principales reduce la data espectral a unas pocas combinaciones de valores de absorbancia que valen para la mayoría de la información espectral. El PLS reduce la data a unas pocas combinaciones de absorciones que no solo hacen valer mucho la información espectral sino que también relacionan con los datos de referencia. El algoritmo PLS relaciona una variable a la vez. La validación cruzada es recomendada para estimar el número óptimo de términos para la calibración (Infrasoftware International, 2005).

Williams (2001) y Burns y Ciureczak (2007) señalan que un estimador de la capacidad predictiva de la ecuación es el propio error de calibración del modelo desarrollado; éste es el error típico de calibración (ETC), que se define como la desviación típica de los residuales (diferencia entre el valor aportado por el método de referencia y el valor estimado por la ecuación) para el colectivo de calibración. Sin embargo, al ser únicamente un estimador del error del modelo quimiométrico empleado, dicho estadístico suele sobreestimar la capacidad real de predicción de una ecuación (Marten *et al.*, 1991).

El término ETC es traducción al español de *Standard Error of Calibration* (SEC), para fines prácticos se reemplazarán las siglas en español por las siglas en inglés en lo que resta del documento. Según el *software* WinISI, el SEC está representado según la ecuación (27).

$$SEC = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{N-1}} \quad (27)$$

Dónde: y_i = valor por el método de referencia
 \hat{y}_i = valor predicho por la ecuación
 N = número total de muestras usadas en el cálculo

La validación cruzada es un algoritmo que selecciona diferentes colectivos de calibración y validación dentro de una población específica. El procedimiento consiste en dividir el colectivo de calibración en grupos (dependiendo del número de muestras). Cada grupo de validación es predicho una vez con la ecuación desarrollada a partir del resto de grupos. El procedimiento se repite hasta que todas las muestras hayan sido predichas una vez. Este procedimiento además previene el sobreajuste del modelo. Los errores de la misma son evaluados mediante el estadístico Error Típico de Validación Cruzada (ETVC) que es similar al Error Típico de Predicción (ETP). El ETP es el estadístico más empleado para estimar la capacidad de predicción (Marten *et al.*, 1991). Adicionalmente, Burns y Ciurczak (2007) indican que el cálculo del SECV es un método para determinar el mejor número de variables independientes para usar en la construcción de una ecuación de calibración.

Los términos ETVC y ETP son traducciones al español de *Standard Error of Cross Validation* (SECV) y *Standard Error of Prediction* (SEP), para fines prácticos se reemplazarán las siglas en español por las siglas en inglés en lo que resta del documento. Según el *software* WinISI, el SECV está representado según la ecuación (28).

$$SECV = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{N}} \quad (28)$$

Dónde: y_i = valor por el método de referencia

\hat{y}_i = valor predicho por la ecuación

N = número total de muestras usadas en el cálculo

La validación cruzada, también es llamada validación interna o pre validación porque lo realiza el computador mediante un *software*, este proceso se realiza con el set de calibración, cuyo fin principal es arrojar las mejores ecuaciones para cada parámetro. Esta validación puede ser realizada mediante la herramienta “Monitor de resultados” del *software* WinISI, en donde la población es dividida en grupos (cada segunda, cada tercera, cada cuarta muestra, etc.). Si la población es dividida en cuartos, se toma una base de un cuarto de la población y la población será predicha sobre cada cuarta muestra con una calibración desarrollada con las otras tres cuartas muestras restantes (Considerando los datos del método de referencia y los espectros almacenados). Las muestras con grandes residuales (valores “T” mayores a 2,5) suelen ser omitidas, y la validación cruzada vuelve

a realizarse. El ciclo de omitir muestras con grandes residuales, seguido por otras validaciones cruzadas puede ser desarrollada dos o tres veces (FOSS, 2007).

La validación cruzada solo puede ser usada con los métodos PCR, PLS y MPLS. Un número por defecto de grupos de validación cruzada son mostrados por el software. Sin embargo, un mínimo de tres validaciones cruzadas es recomendado y más pueden ser usados si el tiempo para los cálculos no es limitado. Este error (SECV) es un verdadero estimador de la exactitud de predicción de la calibración. El error de validación cruzada siempre será mayor al error de calibración (Infrasoft International, 2005).

Según Cuesta (2013), uno de los estadísticos más importantes al desarrollar un modelo es el R^2 (RSQ) o coeficiente de determinación. La ecuación de cómo obtener el estadístico en mención fue explicada en el acápite 2.8.3.a. Error Estándar de Laboratorio (ver ecuación 12). Por otro lado se tiene el estadístico 1-VR (coeficiente de determinación de la validación cruzada) que describe la variación durante el proceso de validación cruzada. Según el *software* WinISI, el 1-VR está representado según la ecuación (29).

$$1 - VR = \frac{(SD)^2 - (SECV)^2}{SD^2} \quad (29)$$

Dónde: SD = Desviación estándar del colectivo de calibración

$SECV$ = Error estándar de validación cruzada

Ye (2003), indica que las mejores ecuaciones de calibración son escogidas en base al menor valor de Error Estándar de Validación Cruzada (SECV) y el más alto valor del Coeficiente de Determinación de la validación cruzada (1-VR).

Cozzolino *et al.* (2001) proponen otra forma de evaluar las ecuaciones NIRS, basándose en su precisión y exactitud. Toman en cuenta la relación entre el error estándar de la validación cruzada y la desviación estándar del análisis químico (SECV/SD), y considera como una ecuación con alto poder de predicción si la relación es menor a 0,33. Además Rodríguez (2013), recomienda que el valor del SECV se encuentre entre 1 a 1,5 veces el SEC.

El modelo estadístico (SEC-SECV) no podrá ser mejor al SEL, pues siempre será mucho más exacto obtener valores de resultados por el método de referencia que por los métodos calibrados, pero con el costo elevado de los métodos de referencia implicados. Es

preferible analizar por separado muestras $2n$ (Muestra dividida en dos cápsulas) que analizar en duplicado muestras n (Muestra que es sometida a lectura dos veces dentro de una misma cápsula) en el NIRS. Asimismo, los métodos de referencia deben ser muy precisos y merecen réplicas de química húmeda (Dardenne, 2008).

Dardenne (2008) indica que si es posible escanear la muestra más de una vez, se debe realizar la calibración con el promedio de los espectros. De lo contrario se tendrá problemas durante la validación cruzada.

Es posible considerar como ejemplo a Tamburini *et al.* (2015) quién estableció la relación entre los valores NIR y los datos obtenidos por química húmeda usando el método de regresión de *Partial Least Squares* (PLS). Además, detectó *outliers* mediante un *software* basado en el criterio de la distancia de Mahalanobis y realizó la validación cruzada utilizando los parámetros por defecto del *software*. Para cuantificar la probabilidad de autocorrelación entre las series de espectro dadas, aplicó el test Durbin-Watson (DW) a los residuales de la regresión PLS.

2.8.7. EVALUACIÓN DEL MODELO QUIMIOMÉTRICO

Según Cueva (2012), la evaluación del modelo se lleva a cabo mediante la aplicación del modelo desarrollado a un conjunto de muestras conocidas llamadas colectivo de validación. Esto se hace con un set de muestras diferentes a las utilizadas en la calibración (validación externa o post validación). El número de muestras para la validación es un valor generalmente arbitrario no existe un numero definido. Al respecto, Cao (2013), utiliza colectivos de validación que representan alrededor del 10 por ciento del colectivo de calibración. Cueva (2012), indica que si los resultados de predicción no son satisfactorios, el modelo debe ser recalculado. Adicionalmente, Cueva (2012), señala que para los modelos cualitativos se conoce su identidad y para cuantitativos se conoce el valor de la propiedad a determinar

Según Cao (2013), un set ideal de validación debe contener muestras que contengan un rango de variación que cubra todos los componentes químicos. Investigaciones citadas por el autor indican el impacto de la variabilidad de las muestras de una cosecha diferente en el proceso de calibración. Para mantener un set de validación independiente y representativo, Cao (2013) tomó muestras de granos de soya de una cosecha diferente a la del colectivo de

calibración. Si un *outlier* fue inadvertidamente incluido en los resultados futuros, la validación también pueda dar una indicación del tipo y respuesta.

Esencialmente, la validación externa consiste en determinar la posible existencia de un error conocido significativo, denominado sesgo, y de un error no explicado, denominado SEP(c) o error típico de predicción corregido por el sesgo (Flores, 2009). Los estadísticos recomendados por Flores (2009) para la evaluación de los resultados del procedimiento de validación externa son: el sesgo (bias), que representa la diferencia sistemática media entre los datos de referencia y los valores predichos NIR; el error típico de predicción (SEP), que es el error de los residuales para las muestras de validación, su valor puede ser bastante variable dependiendo del error del método de referencia (SEL) y de que las muestras de validación estén bien representadas en el colectivo de calibración (Shenk *et al.*, 1992); el SEP(c) o error típico corregido por el sesgo, es un estimador del error aleatorio o no explicado por la regresión y el coeficiente de determinación para la validación (R^2).

a. ESTADÍSTICOS DE VALIDACIÓN EXTERNA (SEP Y BIAS)

El SEP es un indicador confiable de la calidad de la ecuación desarrollada, ya que a diferencia del SEC (Error estándar de calibración), que mejora (disminuye) a medida que se agregan nuevos términos a la ecuación, el SEP mejora solo hasta que comienza a producirse un sobre ajuste de la ecuación, empeorando (aumenta) con cada nuevo término (Williams, 2001). Según el *software* WinISI, el SEP está representado según la ecuación (30).

$$SEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{N}} \quad (30)$$

Dónde: y_i = valor por el método de referencia

\hat{y}_i = valor predicho por la ecuación

N = número total de muestras usadas en el cálculo

Infrasoft International (2005), indica que el SEP tiende a variar con la diversidad del producto. Esto es porque la composición química de las muestras en poblaciones muy diversas difiere en forma cuantitativa y cualitativa. Además manifiesta que las calibraciones para cambios cuantitativos son sencillas; sin embargo, las calibraciones para cambios cualitativos, no lo son.

Según el *Committee for proprietary medicinal products* (2003), los valores recomendados del SEP se deben encontrar entre 1,2 a 1,4 veces el SEL.

Según Infrasoftware International (2005), los valores de SEP para las ecuaciones NIRS caen en tres categorías generales:

- En la primera categoría están aquellos componentes que exhiben una alta correlación entre la información espectral y los valores de referencia. Por ejemplo proteína cruda y humedad en la mayoría de granos; aceite en suplementos proteicos, tales como concentrados de soya; y grasa en varios derivados de origen animal. Estas combinaciones de producto y componente han demostrado los menores SEP para cualquier producto agrícola o alimenticio en los que la espectroscopía NIR/NIT es usada.
- Una segunda categoría de materiales que exhiben un SEP ligeramente mayor, lo que podría incluir proteína cruda y humedad en forrajes y mezclas alimenticias, y en algunos casos, calibraciones de fibra y minerales de materiales similares a forrajes. Para estos componentes, el método NIRS y el de referencia no miden el mismo grupo funcional (por ejemplo NIRS mide N-H, no todas las formas de N). Un mayor error de análisis en esta segunda categoría es debido a la relación imperfecta entre las dos medidas.
- Una tercera categoría puede incluir las medidas de todos los demás componentes que pueden tener alguna exactitud predictiva, suelen ser más exactas que los valores promedios de una tabla de composición de alimentos. Por ejemplo, la predicción de proteína cruda en comida de pescado. El error estándar de analizar la proteína cruda en comida de pescado es mayor que cuando es comparado a la predicción de proteína cruda en trigo (1,5 por ciento para comida de pescado y 0,3 por ciento para trigo). Esto no es un problema de muestreo, estructura poblacional o el método matemático del desarrollo de la calibración, es debido a la falta de concordancia entre la información espectral de la muestra y el método de referencia. Las alternativas para aceptar el error estándar mencionado de 1,5 por ciento son tres: (1) usar el valor promedio de proteína cruda de tabla de composición de alimentos, (2) esperar hasta que la muestra sea analizada por el método de referencia, o (3) calibrar con otro método de referencia que relacione la información en el espectro mejor que con el nitrógeno total determinado por el método de referencia.

Múltiples comparaciones son realizadas entre los diferentes modelos de calibración con respecto a la predicción de muestras espectrales externas (Cao, 2013). Para Burns y Ciurczak (2007), el SEP es el mejor estadístico para evaluar el futuro desempeño del modelo de calibración.

Según Cao (2013) el desempeño de la calibración es evaluada en términos de precisión, exactitud y ajuste del modelo. Mientras que el error estándar de predicción (SEP) mide la precisión, la exactitud puede ser determinada por el bias (promedio de diferencias), indicador de precisión y similitud entre las muestras de validación y de calibración. Según el *software* WinISI, el Bias está representado según la ecuación (31).

$$Bias = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{N} \quad (31)$$

Dónde: y_i = valor por el método de referencia
 \hat{y}_i = valor predicho por la ecuación
 N = número de muestras usadas en el cálculo de la validación externa

Según el *software* WinISI, el SEP(c) o SEP corregido por el Bias está representado según la ecuación (34).

$$SEP_c = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2 - \frac{(\sum_{i=1}^n (y_i - \hat{y}_i))^2}{N}}{N-1}} \quad (32)$$

Dónde: y_i = valor por el método de referencia
 \hat{y}_i = valor predicho por la ecuación
 N = número total de muestras usadas en el cálculo

2.8.8. SELECCIÓN DEL MODELO QUIMIOMÉTRICO

Font *et al.* (2006), utilizan tres estadísticos para evaluar la exactitud de la ecuación: el R^2 (RSQ), el RPD (ratio de la desviación estándar –SD- de las muestras de validación sobre el error típico de predicción –SEP-) y el RER (ratio del rango de las muestras de validación sobre el error típico de predicción–SEP-). Además indican que mientras sea posible se debe calcular el ratio del SEP (Error típico de predicción) sobre el SEL (Error estándar de laboratorio), pues este estadístico permitirá poner en perspectiva el error NIRS con el error del método de referencia.

Cao (2013), indica que el Determinante Predictivo Relativo (RPD) puede ser usado para evaluar la precisión de la ecuación. Según Ye (2003), si el RPD es alto, la predicción NIRS podrá dividir las muestras en subgrupos de bajo, medio y alto contenido del componente escogido. Si el ratio RPD excede el valor de 3, la ecuación de calibración puede predecir el componente. De lo contrario, la habilidad de la ecuación de calibración será limitada. Williams y Sobering (1993) consideran el valor de 2,5 cómo el mínimo recomendable, para realizar lecturas rápidas en programas de mejoramiento. Williams y Sobering (1996) sugieren que los valores de RPD no pueden ser muy elevados si hay poca variación en los datos de referencia. Según Via *et al.* (2003), un valor efectivo de RPD debe estar entre 1,5 a 2,5 o mayor.

Tamburini *et al.* (2015), proponen el uso del RSQ junto con el RPD como la estadística más significativa para la evaluación de la eficiencia analítica del NIRS. Un buen modelo debe ser caracterizado por un valor de SEP mucho menor al valor de la SD, por lo tanto valores altos de RPD son siempre deseados. Los mínimos requerimientos para valores de RPD son 2,0-3,0 para un adecuado *screening* de datos, 3,0-5,0 para una predicción aceptable han sido sugeridos. Valores que excedan 5,0 indican que el modelo de predicción es casi perfecto.

Según Alonso *et al.* (2005), el RER es la relación entre el rango del método de referencia para el grupo de validación ($V_{\text{mayor}} - V_{\text{menor}}$) y el error estándar de predicción (SEP). Williams y Sobering (1996), consideran el valor de 10 cómo el mínimo recomendable.

El valor del RER debería ser tan alto como fuera posible, pero puede ser aumentado de forma ficticia por una presencia de muestras extremas en baja concentración lo cual le resta utilidad y consistencia. Esto no ocurre cuando para estandarizar el SEP o SECV utilizamos la desviación típica, es decir, utilizando el RPD (Williams, 2001).

Según Millmier *et al.* (2000), la exactitud de la predicción puede juzgarse en función de los valores del índice RER. Para valores comprendidos en el intervalo entre 8 y 12 se pueden realizar predicciones cuantitativas de los parámetros para los que se desarrollan las ecuaciones y valores RER mayores a 12 indican una calidad excelente del modelo predictivo.

Williams (2001), indica que para evaluar la validación cruzada también se pueden utilizar los estadísticos RPD y RER. El estadístico RPD, mide la relación entre la desviación típica

de los datos de referencia y el SECV para un determinado componente. El estadístico RER, establece la relación entre el rango de los datos de referencia y el SECV para un determinado componente.

El método de selección de ecuaciones de calibración desarrollado por Marten *et al.* (1989), establece que para cada componente el error estándar de calibración (SEC) y el coeficiente de determinación (RSQ) representan criterios importantes para la selección de cada ecuación individual. Jiménez (2007) señala que como criterio general, se puede afirmar que una ecuación es considerada aceptable cuando la magnitud del SEP es menor a un tercio de la desviación estándar de los datos de referencia

Burns y Ciurczak (2007), indican que solamente el criterio estadístico es usado para seleccionar y validar las ecuaciones de calibración en la espectroscopía NIR/NIT. En el caso ideal, se tiene un buen conocimiento químico de las muestras a analizar. Este conocimiento, permite al usuario determinar la longitud de onda óptima y da mayor seguridad a que el método NIR/NIT está realmente prediciendo los componentes de interés. Sin el conocimiento químico, el usuario debe confiar puramente en el criterio de selección estadístico y está afectado por el uso bueno o malo de las modernas computadoras.

Según Infracsoft International (2005), para cada tratamiento matemático, el SEP de la ecuación alcanzará un valor mínimo como tantos términos sean se añadan a la ecuación. Con el ingreso términos adicionales a la ecuación de calibración, el SEC disminuye con cada valor adicional, mientras que el SEP disminuye hasta que un sobreajuste de la data es evidenciado por un incremento en el valor SEP. Como guías del desempeño del modelo, los valores de SEC y SEP deben estar entre el 20 por ciento el uno del otro. El método de selección de ecuaciones, incluye los siguientes criterios:

1. El tratamiento matemático debe tener el menor valor de SEC y la menor cantidad de términos (longitudes de ondas) para evitar sobreajustes. Una ecuación PLS debe ser escogida en base al menor error estándar de validación cruzada no el SEC.
2. Ninguna longitud de onda debe tener un valor estadístico F parcial menor a 10 para su correspondiente coeficiente de regresión para prevenir el sobreajuste. Esto no aplica las ecuaciones PLS.

3. Los coeficientes de regresión no deben excederse +/- 10,000, con motivo de minimizar los problemas en la transferencia de la ecuación, debidos al ruido del instrumento y otras variaciones inter-instrumentos. Este criterio no ha sido evaluado con las ecuaciones PLS, pero el valor obtenido probablemente sea adecuado también para las ecuaciones PLS.
4. La pendiente de la línea de regresión del valor NIRS, con el valor primario de referencia debe ser cercano a 1.

Infrasoft International (2005), señala que finalmente la selección de la ecuación es completada usando los criterios listados previamente, de manera que se encuentra la “mejor” ecuación del grupo de mejores ecuaciones individuales. Después de seleccionar la mejor ecuación, a menudo es conveniente volver a calcular y desarrollar una nueva ecuación que combina las muestras de calibración y muestras de prueba utilizando el mismo tratamiento matemático y longitudes de onda usadas para desarrollar la mejor ecuación. La ecuación final está lista para entrar a un sistema de seguimiento en el tiempo y de cambios de composición de la muestra.

2.9. MANTENIMIENTO DE CALIBRACIONES NIRS

Infrasoft International (2005), indica que si la ecuación tiene un buen performance al inicio pero luego su exactitud empieza a deteriorarse en el tiempo, una recalibración sería la acción más conveniente. Las muestras deben ser escaneadas con “*Center samples*” y “*Select samples*” opciones usando el “H” de *global* y *neighborhood*. Esas muestras escogidas son analizadas por el método de referencia primario. Se debe agregar estas nuevas muestras al set original de calibración y repetir el proceso de selección de ecuaciones.

Según Infrasoft International (2005), habiendo definido los errores (SEL, SEC, SEP), dos límites de control son usados para comparar los valores de referencia con los predictivos. Las muestras *outlier* son resaltadas por el software WinISI como número grises o rojos. El primer límite es para determinar si es necesario un bias significativo y el segundo para determinar si está sucediendo un aumento significativo del error no explicable. Asumiendo:

- Una diferencia entre los valores predictivos NIRS y los valores de referencia tanto del bias como del error no explicable es detectable con 90 por ciento de confianza.
- Por lo menos 100 muestras son presentadas en el set de calibración.
- Un bias mayor al SEC y un error no explicable mayor a dos veces el SEC es inaceptable.

Se recomienda el siguiente procedimiento:

1. Escoger nueve muestras de forma aleatoria de un grupo de muestras a ser analizadas
2. Obtener resultados de análisis por el método NIR y el de referencia.
3. Calcular el límite de control bias= $BCL=0,6(SEC)$
4. Calcular el límite de control del error no explicable $UCL=1,3(SEC)$
5. Calcular el bias y el error no explicable de los valores analíticos en el set de prueba.

Estas pruebas deben ser realizadas cada vez que existan sospechas de que la ecuación no está prediciendo bien los parámetros calibrados. Como mínimo, una muestra de cada cien debe ser separada para que se le realice el análisis por el método de referencia luego del análisis NIRS. Una de cada 50 sería aún mejor. Cuando se hayan acumulado nueve muestras, se utilizará la metodología descrita a continuación (Infrasoft International, 2005).

Infrasoft International (2005) recalca que ajustar el bias en principio solo arregla de forma temporal el problema; pues otro set de nueve muestras probablemente sugiera una corrección de bias diferente. El problema está en el set de calibración y solo podrá ser corregido ahí.

III. MATERIALES Y MÉTODOS

3.1. LUGAR DE EJECUCIÓN

El trabajo de investigación se realizó en el Laboratorio de Control de Calidad del Programa de Investigación y Proyección Social (PIPS) en Cereales y Granos Nativos de la Universidad Nacional Agraria La Molina.

3.2. MATERIA PRIMA

Ciento veinte muestras de *Chenopodium quinoa* W. extraídas de los bancos de germoplasma del Programa de Investigación y Proyección Social (PIPS) de Cereales y Granos Nativos de la Universidad Nacional Agraria La Molina y del Instituto Regional de Desarrollo (IRD-Sierra) de Sierra San Juan de Yanamucllo en Valle del Mantaro, perteneciente a la Universidad Nacional Agraria La Molina. Estas muestras de quinua son provenientes de diferentes variedades y accesiones (unidades de conservación de diferentes genotipos).

- Proceso de calibración: se utilizaran 100 muestras (colectivo de calibración).
- Proceso de validación: se utilizaron 20 muestras (colectivo de validación).

3.3. MATERIALES

- Bureta de 50 ml
- Beaker de 150 ml
- Crisol
- Campanas desecadoras
- Guantes resistentes al calor
- Guantes de latex

- Lentes de seguridad
- Papel filtro
- Placas petri
- Pinzas
- Mascarillas de seguridad para gases

3.4. EQUIPOS

- Balanza analítica (Metler Toledo)
- Campana de extracción de gases
- Equipo determinador de grasa – Soxtec ST 245 (FOSS)
- Equipo destilador Vapodest 45 (Gerhardt)
- Equipo de digestión TurboTherm (Gerhardt)
- Espectrofotómetro Infratec 1241 (FOSS)
- Estufa con convección de aire (Mettler)
- Molino experimental Cyclotec modelo 1093 (0,5 mm de haz de luz y una capacidad de 4g/s) (FOSS).
- Mufla Lindberg/Blue M (Thermo *Scientific*)

3.5. REACTIVOS

- Ácido bórico (MerckgaA 64271 - Darmstadt)
- Ácido sulfúrico concentrado (Merck)
- Ácido clorhídrico concentrado (Merck)
- Agua destilada.

- Carbonato de Sodio (Merck)
- Etanol Absoluto ACS (Fermont)
- Éter de petróleo (*Petroleum ether* 35 – 60 °C, PR 1428 – 007)(MTEDIA)
- Hidróxido de sodio (Merck)
- Indicador rojo de metilo (Riedel de Haën)
- Indicador verde de bromocresol (Riedel de Haën)
- Tabletas catalizadoras Kjeltabs 3.5g K₂SO₄ y 0.4g CuSO₄ x 5H₂O (FOSS)

3.6. MÉTODOS DE ANÁLISIS FÍSICO-QUÍMICOS

3.6.1. CONTENIDO DE HUMEDAD

Determinado por el método oficial AOAC 925.10 Sólidos (Totales) y Humedad en Harina. Método de horno de aire (estufa).

3.6.2. CONTENIDO DE PROTEÍNA

Determinado por el método oficial AOAC 920.87 Proteína (Total) en Harina – Ajuste de Gerhardt. Se usó el equipo de digestión TurboTherm (Gerhardt) y el equipo destilador Vapodest 45 (Gerhardt).

3.6.3. CONTENIDO DE LÍPIDOS

Determinado por el método oficial AOAC 923.05 Lípidos en Harina – Ajuste del Soxtec (FOSS)

3.6.4. CONTENIDO DE CENIZA

Determinado por el método oficial AOAC 923.03 Ceniza en Harina. Método Directo

3.7. MÉTODO DE ANÁLISIS ESPECTROMÉTRICO

Se utilizó el equipo Infratec 1241 de la firma FOSS, las muestras recolectadas (grano y harina de quinua) fueron escaneadas en modo de transmitancia con el módulo “*cup*” grey – grey de 60 mm de diámetro y 6 5/6 mm de espesor. En promedio, cada módulo “*cup*” tiene

una capacidad de 2,10 gramos para la presentación de grano y 1,25 gramos para la presentación de harina. Los datos de transmitancia (T) fueron almacenados como el logaritmo de (1/T) en el rango de 850 – 1048 nm a un intervalo 2 nm. La luz difusa transmitida por la muestra fue registrada en detectores de silicio y los datos fueron enviados a una computadora interna del Infratec 1241 para su almacenamiento y procesamiento. En forma paralela a la señal digital de la luz transmitida, la computadora interna recibió una señal que representa la longitud de onda utilizada, de modo que a cada valor de transmitancia almacenada le corresponde una longitud de onda. Los datos fueron almacenados en una memoria USB para luego desarrollar el modelo quimiométrico.

La información técnica del equipo Infratec 1241 se encuentra en el Anexo 1.

3.8. PROTOCOLO PARA DESARROLLO EL MODELO

De acuerdo a los objetivos planteados al inicio de la investigación y las condiciones de la línea base, se realizó un plan de trabajo que incluyó 120 unidades muestrales. Durante la investigación se incluyeron otras muestras que no figuran en la presente investigación, esto fue debido a errores instrumentales o cantidad insuficiente de muestra. La totalidad de las 120 muestras hábiles están en el Anexo 2. La recepción de las muestras fue el retiro de semillas del banco de germoplasma. Las muestras de iniciales LM89 y PAS fueron cultivadas dentro del campus universitario de la Universidad Nacional Agraria La Molina (campaña 2012-B), mientras que las muestras de iniciales POQ fueron cultivadas en el Valle del Mantaro en los campos del Instituto Regional de Desarrollo (IRD) Sierra San Juan de Yanamucllo de la Universidad Nacional Agraria La Molina (campaña 2012-B). Luego de la recepción, se trabajó íntegramente dentro de las instalaciones del Laboratorio de Calidad del Programa de Investigación y Proyección Social en Cereales y Granos Andinos. La secuencia de actividades está diagramada en la Figura 16, mientras que el diseño experimental de la investigación está resumido en el Cuadro 5.

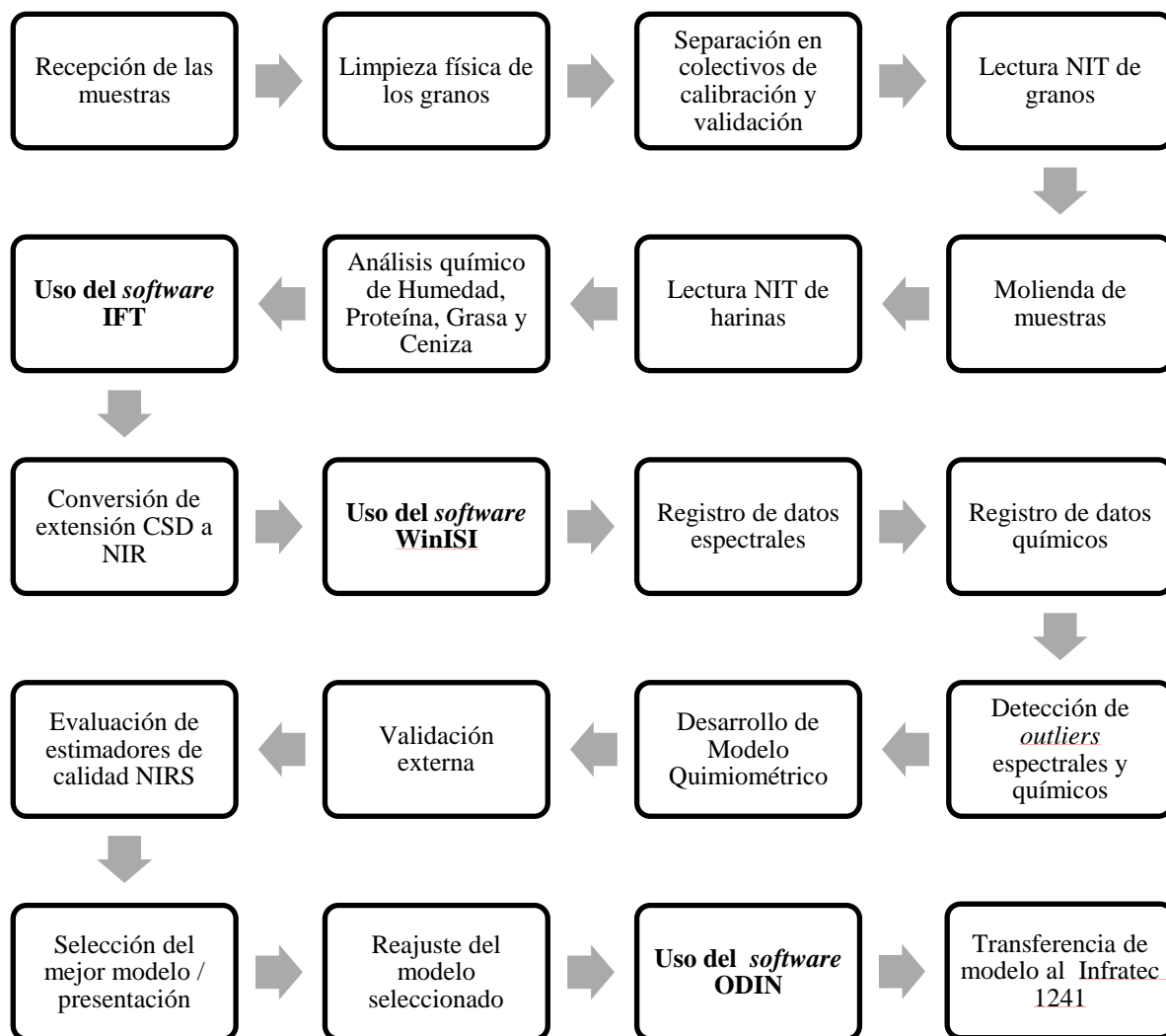


Figura 16: Protocolo para el desarrollo de la calibración NIT.

FUENTE: Elaboración propia.

Cuadro 5: Diseño Experimental de la investigación

Etapas	Recepción y preparación de muestras	Registro de espectros I y II	Análisis químico por métodos de referencia	Selección de colectivos calibración / validación	Desarrollo de los modelos quimiométricos	Validación externa y evaluación de estimadores de calidad	Selección del mejor modelo y reajuste	Transferir modelo al Infratec 1241
Acciones	Recepción y almacenamiento de las muestras de quinua (n=120 mutantes)	Luego de la lectura de granos, se realizó la molienda de cada muestra de quinua	Determinación del SEL de cada método	Selección de muestras de calibración (n1=100) y de validación (n2=20)	Detección de <i>outliers</i> Desarrollo de modelos	Comparación de ecuaciones desarrolladas y colectivos de validación	Escoger el mejor grupo de ecuaciones Realizar nuevo modelo quimiométrico con la población total	Transferir la calibración ajustada al instrumento NIT
Análisis		I. Lectura NIT de cada una de las muestras de grano de quinua. II. Lectura NIT de cada una de las muestras de harina de quinua.	<ul style="list-style-type: none"> • Contenido de Humedad • Contenido de Proteína • Contenido de Grasa • Contenido de Ceniza 		<ul style="list-style-type: none"> • SEC • RSQ • SECV • 1-VR 	<ul style="list-style-type: none"> • RSQ • SEP • RER • RPD 	<ul style="list-style-type: none"> • N • RSQ • SEP 	

FUENTE: Elaboración propia.

3.9. METODOLOGÍA PARA VALIDAR EL MODELAMIENTO

Los modelos individuales para cada componente fueron obtenidos por los diferentes tratamientos matemáticos y dispersiones (*scatters*) durante el desarrollo de la calibración, estos fueron evaluados de acuerdo al valor de sus estadísticos NIRS obtenidos en cada etapa del modelamiento. Los estadísticos utilizados en cada etapa son detallados en el Cuadro 6.

Cuadro 6: Estadísticos evaluados para selección de los modelos NIRS.

Etapa	Estadístico	Parámetros	Decisor	Observación
Metodología de referencia	SEL	El menor posible	No	Se utilizó para saber el valor máximo esperado de RSQ y mínimo esperado de SEC y SEP. Además de los valores recomendados del SEP.
Validación cruzada	N	El mayor posible	Sí	Se prefirió el modelo de mayor cantidad de muestras, significa mayor robustez.
	SEC	Cercano al SEL	No	-
	RSQ	El mayor posible	No	Se buscó el valor más cercano a 1.
	SECV	El menor posible	Sí	Se escogió en base al menor valor SECV y no al SEC.
	1-VR	El mayor posible	Sí	Se buscó el valor más cercano a 1. Fue de mayor validez que el RSQ.
Validación externa	SEP (C)	Cercano al SEL	No	-
	Bias	El menor posible	No	-
	RSQ	El mayor posible	Sí	Se buscó el valor más cercano a 1, y se consideró los valores recomendados por el <i>software</i> WinISI.
	SEP	El menor posible	Sí	-
	<i>Slope</i>	Cercano a 1	No	-
Estimador de Calidad	RER	8-12 o mayor	Sí	Se basó en ratios recomendados del rango del colectivo de validación y el SEP
	RPD	1,5-2,5 o mayor	Sí	Se basó en ratios recomendados de la desviación estándar del colectivo de validación y el SEP.

FUENTE: Elaboración propia.

La selección de ecuaciones fue realizada usando los criterios listados previamente y probando diferentes combinaciones de dispersiones y tratamientos matemáticos, de manera que se encontró la “mejor” ecuación individual por componente del grupo de ecuaciones individuales. Después de seleccionar la mejor ecuación, fue conveniente volver a calcular y desarrollar una nueva ecuación que combine las muestras de calibración y muestras de validación externa; se utilizó el mismo tratamiento matemático y longitudes de onda usadas para desarrollar la “mejor” ecuación. El modelo final, compuesto de cuatro ecuaciones individuales estuvo listo para entrar a un sistema de seguimiento en el tiempo y de cambios en la composición de la muestra.

3.10. HERRAMIENTAS PARA EL DESARROLLO DEL MODELO

- Infratec 1241 *File Tools* -IFT- versión 3.20, usado para visualizar los datos del Infratec 1241 en el computador y convertir los archivos de extensión CDR a NIR.
- WinISI versión 4.7, usado para desarrollo de modelo quimiométrico. La metodología recomendada por WinISI, indicó usar el método de regresión PLS. La descripción del mismo se puede encontrar en el acápite 2.8.5.b.c. Regresión por Mínimos Cuadrados Parciales (PLS).
- Odin versión 4.22, usado para transferir la ecuación al equipo Infratec 1241 y convertir los archivos de extensión EQU a PM a AM y finalmente CAM.
- Infostat *Student Version*, usado para elaborar los histogramas de frecuencias.
- *Microsoft Office* Excel 2010, usado para registrar los resultados de los métodos de referencia.

IV. RESULTADOS Y DISCUSIÓN

Al iniciar la experimentación, la población total de muestras de quinua fue separada aleatoriamente en dos colectivos, un colectivo de calibración y un colectivo de validación (ver Anexo 2).

La lectura de espectros de infrarrojo (equipo Infratec 1241) fue realizada en simultáneo con los análisis químicos de las muestras de quinua. Se inició con la lectura de los granos de quinua, muestra por muestra según se indica en la metodología. Luego, se procedió a la molienda (equipo Cyclotec 1093 - Malla de 0,5 mm) de las mismas para realizar la lectura de la harina de quinua. Finalmente, a partir de la harina de quinua se procedió a realizar los análisis por los métodos de referencia establecidos para determinar la humedad, proteína, grasa y ceniza de las 120 accesiones de quinua.

4.1. ANÁLISIS DE HUMEDAD, PROTEÍNA, GRASA Y CENIZA DE LAS MUESTRAS DE QUINUA

Los análisis químicos fueron realizados según los métodos de referencia propuestos. Estos métodos por defecto brindaron valores en base húmeda (*as is*) para la humedad, proteína y ceniza, mientras que en base seca (*dry matter*) para la grasa, pues se trabajó con muestra anhidra. El modelo quimiométrico se planteó en base seca, por lo que hubo que convertir los valores obtenidos de la humedad, proteína y ceniza de base húmeda a base seca. En los Anexos 2 y 3, se pueden observar las características de color, variedades y resultados de los métodos de referencia de los colectivos de calibración y validación iniciales. En los anexos citados también se visualizan los histogramas y polígonos de frecuencia de los datos de referencia para la humedad, proteína, grasa y ceniza.

Para el caso de la humedad, los valores del colectivo de calibración inicial se encontraron en un rango de 8,84 – 15,01 por ciento en base seca. Mientras que para el colectivo de validación inicial se encontraron en un rango de 10,66 – 15,50 por ciento en base seca. Autores como Tapia *et al.* (2014) y USDA (2013) presentan valores de 11,23 y 15,31 por ciento, rango en que la mayoría de valores se encontró, sin embargo se observaron

valores bajos de humedad; esto pudo deberse a que las muestras fueron de cosechas de años previos y a las condiciones de almacenamiento (refrigeración) del material de germoplasma. En largos periodos de tiempo estas condiciones pudieron causar la pérdida de humedad de las muestras.

Para el caso de la proteína, los valores del colectivo de calibración inicial se encontraron en un rango de 8,35 – 11,72 por ciento en base seca. Mientras que para el colectivo de validación inicial se encontraron en un rango de 8,33 – 11,40 por ciento en base seca. Autores como Reyes *et al.* (2009) y Tapia *et al.* (2014), señalan valores de 12,49 y 14,13 por ciento respectivamente, los cuales son valores superiores a los obtenidos por el método de referencia en la presente investigación. Es posible inferir dos situaciones: la primera es que las 120 accesiones seleccionadas tuvieron un contenido bajo de proteínas y la segunda opción es que existieron deficiencias en el método de referencia.

Para el caso de la grasa, los valores del colectivo de calibración inicial se encontraron en un rango de 4,79 – 9,46 por ciento base seca. Mientras que para el colectivo de validación inicial se encontraron en un rango de 5,62 – 8,72 por ciento en base seca. Autores como USDA (2013) y Reyes *et al.* (2009), señalan valores de 7,00 y 8,66 por ciento respectivamente. Los valores obtenidos en la presente investigación estuvieron muy cerca del rango propuesto por los autores, con mayor frecuencia entre los valores de 6,0 y 7,0 por ciento.

Para el caso de la ceniza, los valores del colectivo de calibración inicial se encontraron en un rango de 2,51 – 4,62 por ciento en base seca. Mientras que para el colectivo de validación inicial se encontraron en un rango de 2,92 – 4,07 por ciento en base seca. Autores como USDA (2013) y Thoufeek Ahamed *et al.* (1998) señalan valores de 2,74 y 3,39 por ciento respectivamente. Estos valores se encuentran dentro del rango de valores obtenidos, observándose mayor frecuencia de los datos entre 2,8 y 3,5 por ciento.

En resumen, los valores químicos de ambos colectivos se encontraron en rangos similares a los presentados por los autores del Cuadro 1. Debe de resaltarse que los valores de humedad se encontraron ligeramente fuera del promedio y los valores de proteínas se encontraron ligeramente por debajo del promedio.

Los resultados de los análisis químicos, se obtuvieron al utilizar el método de referencia por triplicado. Se evaluó el porcentaje de variación medio (%CV) de cada triplicado, en

caso de que este sea mayor a 10, se repitió el análisis hasta obtener un valor menor a 10. Esto se realizó con la finalidad de tener una calibración lo más precisa posible, como lo explican claramente Naes *et al.* (2002), al indicar que los resultados obtenidos por un método secundario como NIRS tendrán una precisión similar al del método de análisis usado como referencia y que dio origen a la calibración.

Las 120 muestras proporcionaron variabilidad al modelo quimiométrico, muchas fueron mutaciones de programas de fitomejoramiento y otras fueron selección del banco de germoplasma. Al respecto de la variación de cada muestra, Gonzales (2002) refuerza el concepto explicando que las variaciones del medio ambiente pueden ser divididas en dos categorías predecibles e impredecibles. La primera incluye las características permanentes del medio ambiente como el clima y el tipo de suelo; así mismo incluye aspectos del ambiente que son determinados por el hombre y pueden ser la época de siembra, densidad de siembra, métodos de cosecha y otras prácticas agronómicas. La segunda categoría incluye fluctuaciones del clima como cantidad y distribución de lluvias, temperaturas y otros factores.

La presente investigación usó como unidad experimental a la quinua con pericarpio, la cual fue: (1) leída en forma de grano en el espectro de infrarrojo, (2) molida, (3) leída en forma de harina en el espectro de infrarrojo y (4) analizada por los métodos de referencia. Se utilizó la quinua con pericarpio (sin pulir), pues es así como se obtiene del campo y se almacena en los programas de fitomejoramiento. Según Mosquera (2009), el contenido de saponina de la quinua con pericarpio tiene una masa poco significativa (un promedio de 0.11 por ciento del peso húmedo), por lo que no interfirió en los resultados obtenidos.

4.2. AFINAMIENTO/PURGA DE ESPECTROS EN LOS COLECTIVOS DE CALIBRACIÓN Y VALIDACIÓN

Una vez finalizada la lectura de espectros de infrarrojo, se analizaron los diferentes espectros obtenidos de la quinua en forma de grano y de harina.

En los colectivos de calibración de granos y de harina de quinua, se observó que las muestras que tenían coloraciones oscuras como las quinuas con pericarpio de color negro y gris no tenían un espectro ordenado, además habían algunos espectros anómalos como POQcVM.34 (ver Anexo 4) POQcVM.39, POQcVM.41, POQcVM.42, POQcVM.43 (ver Anexo 4), POQcVM.45, POQcVM.46, POQcVM.47, POQcVM.48, POQcVM.50,

POQcVM.51, H.LM89.102, H.LM89.118, H.PAS.101, H.PAS.102, H.PAS.107, H.POQ.16, H.POQ.30, H.POQ.40.1 (ver Anexo 4) H.POQ.41, H.POQ.46, H.POQ.47, H.POQ.50 (ver Anexo 4) y H.POQ.51, por lo que estas muestras fueron eliminadas de los colectivos. Por esta razón, se decidió trabajar solo con las quinuas de colores claros como las amarillas, cremas y naranjas. El mismo criterio fue aplicado para los colectivos de validación.

En las Figuras 17 y 18, están claramente diferenciados las muestras y espectros anómalos, en la primera están representados como puntos de un color distinto que están alejados de la población. Mientras que en la segunda están representados como curvas anómalas y desordenadas.

Es necesario indicar que las muestras de quinua en harina H.POQ.39, H.POQ.42, H.POQ.43. y H.POQ.45 no pudieron ser leídas en el espectro de infrarrojo bajo las condiciones del estudio. En cambio, estas mismas muestras en presentación de grano sí pudieron ser leídas. Sin embargo, fueron eliminadas en etapas posteriores. Esto puede deberse al color oscuro de las muestras, por ejemplo, la pasankalla en presentación de grano con pericarpio tiene coloración gris. Pero al estar como harina, el color del pericarpio (gris) está mezclado con el episperma (rojo) y el perisperma (blanco). Los colores oscuros absorben la radiación y altera el resultado del infrarrojo. Estas características de coloración en los granos de quinua, son semejantes con otras accesiones, como lo explica Tapia *et al.* (2014) para lo que él denomina las “razas”: Puka Pachan, Witulla, Misa quinua, Q’ello, Quchiwila, entre otras.

Mediante la presente investigación, se evidenció que muchas de las muestras eliminadas por tener espectros anómalos, tenían igualmente varianzas anómalas en la absorción a lo largo del segmento de espectro analizado.

Respecto a las muestras de la variedad pasankalla, las cuales tienen pericarpio de color gris, pero episperma de color rojo. Se detectó que tenían un patrón ordenado, pero de características muy diferentes al grueso del colectivo de calibración. Esto puede observarse en la Figura 17, en la cual se observa un grupo de muestras juntas (dentro de un círculo), pero alejadas del grueso de la población. Además, en la Figura 18 se observa claramente como las muestras de la variedad pasankalla presentan curvas características al empezar su espectro, todas juntas con mayores niveles de absorbancia y luego se

comportan como las quinuas de colores claros.

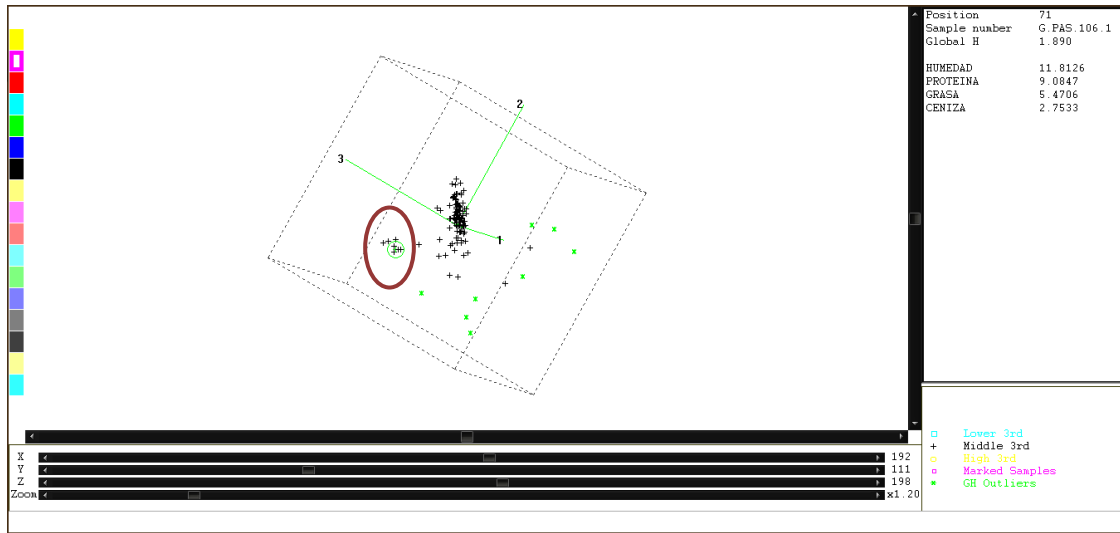


Figura 17: Gráfico 3D (PCA; 0,0,1,1) de la presentación Granos bs (n=120)

FUENTE: Imagen obtenida del WinISI.

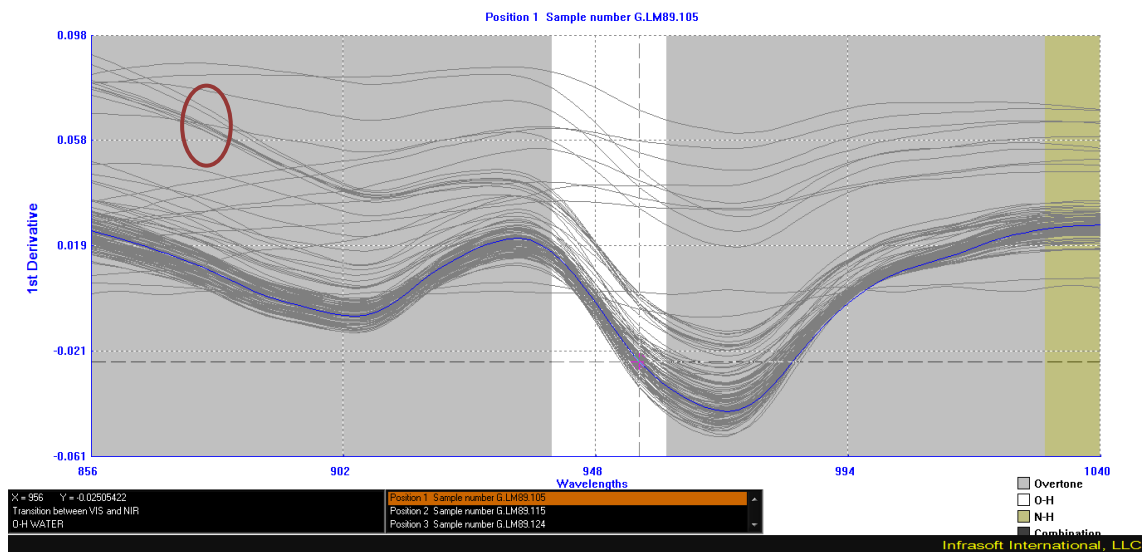


Figura 18: Espectro de la presentación Granos bs (n=120) con tratamiento matemático 1,4,4,1 y scatter none.

FUENTE: Imagen obtenida del WinISI.

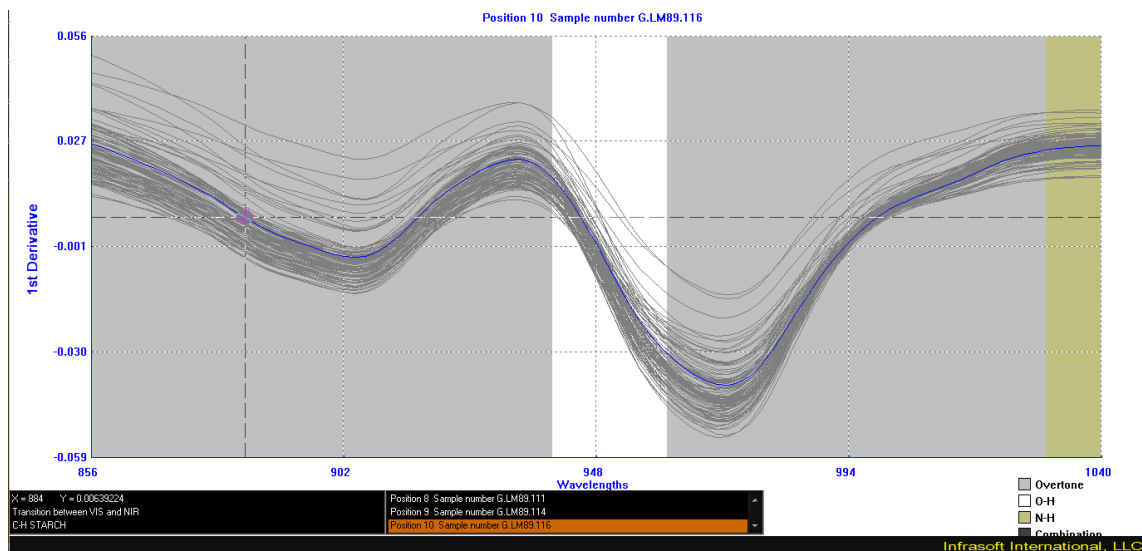


Figura 19: Espectro de la presentación Granos bs (n=99) con tratamiento matemático 1,4,4,1 y *scatter none*.

FUENTE: Imagen obtenida del WinISI.

Luego de la eliminación manual de las muestras, el tamaño de los colectivos fue modificado según el Cuadro 7. Mientras que en la Figura 19, se aprecia la representación de los espectros de las quinuas de colores claros.

Cuadro 7: Tamaño de colectivos de calibración o validación.

Colectivo	Cantidad Inicial	Cantidad Final
Colectivo de Calibración de Harinas	100	79
Colectivo de Validación de Harinas	20	16
Colectivo de Calibración de Granos	100	83
Colectivo de Validación de Granos	20	16

Anderson *et al.* (2006), señalan que la selección del colectivo de calibración es un paso crítico en el procedimiento de optimización. Además, en una calibración multivariable, el tamaño del colectivo tiene un impacto sustancial en lograr una estadística significativa, tanto en poblaciones muestrales grandes o pequeñas. Cao (2013), menciona que para los sets de calibraciones hechos con menor cantidad de muestras, la complejidad de la técnica multivariable puede fácilmente resultar en (1) muy poco poder estadístico de la prueba para identificar realmente resultados significativos o (2) un sobreajuste de la data, de manera que los resultados son artificialmente buenos porque se ajustan a la muestra pero no proporcionan ninguna generalización. Al contrario, para poblaciones muestrales

grandes, la prueba estadística se vuelve sensitiva en términos de la significancia de los resultados.

Ha sido demostrado que el diseño del experimento puede ayudar a escoger el modelo de calibración óptimo (Flaten y Walmsley (2003); Flaten y Walmsley (2004)). El diseño del experimento fue usado para escoger todos los parámetros incluyendo los pretratamientos, el número de componentes, el set de calibración y la selección del subconjunto de variables.

Al momento de realizar la lectura de espectros, se observaron diferencias en las réplicas de una misma muestra, lo cual fue la variabilidad inherente de cada análisis. Foca *et al.* (2011), mencionan que las diferencias pueden ser descritas como un componente de alta frecuencia usualmente conocido como ruido instrumental y un componente de baja frecuencia debido las diferencias específicas en la naturaleza de las muestras. Muchas investigaciones sugieren diferentes estrategias para eliminar el ruido de los espectros NIR en términos de ruido instrumental, pero hay muy poca información sobre como eliminar los componentes de baja frecuencia asociados con la naturaleza de la muestra.

Es necesario tener solo espectros y datos químicos que sean útiles, pues de lo contrario el modelo quimiométrico obtenido no será exacto, esto lo confirma Peguero (2010), el cual indica que mientras más datos se tengan, no querrá decir que más información se tiene del sistema, solo será útiles cuando sean interpretables.

4.3. DESARROLLO DEL MODELO DE CALIBRACIÓN Y EJECUCIÓN DE VALIDACIÓN CRUZADA

Debido al reajuste de los colectivos mostrado en el Cuadro 7, se tuvieron nuevos rangos, promedios, desviaciones estándar y coeficientes de variación. Estos nuevos datos son mostrados en el Cuadro 8.

Se observaron los mayores valores de SD en el componente humedad, y valores medianos en el componente proteína. En el caso de los componentes grasa y ceniza, los valores de la desviación estándar fueron menores a los primeros.

Jiménez (2007), señala que no existe un valor óptimo de desviación estándar para los datos químicos, lo importante es lograr cubrir todo el rango esperado de la calibración;

sin embargo, considera interesante tener en cuenta que la desviación estándar debe ser significativamente mayor que el error de predicción (SEP).

De manera similar, se observó que el CV (Coeficiente de variación medio) fue similar para los cuatro componentes tanto en el caso de los colectivos de calibración como en los de validación. Sin embargo, se observó que es ligeramente mayor para los componentes de grasa y ceniza en las dos presentaciones evaluadas.

Cuadro 8: Medidas de dispersión de los colectivos, evaluados por vía seca.

Presentación	Componente	Medidas de dispersión - Estadísticos				
		Mínimo	Máximo	Promedio	SD	CV
Harina bs Colectivo de Calibración (n=80)	Humedad	10,1020	15,0118	11,90	0,97	0,08
	Proteína	8,3506	11,2966	9,92	0,72	0,07
	Grasa	5,6723	9,4603	6,69	0,66	0,10
	Ceniza	2,5121	4,6242	3,25	0,45	0,14
Harina bs Colectivo de Validación (n=16)	Humedad	10,6590	15,4973	12,30	1,32	0,11
	Proteína	8,3332	11,3981	9,93	0,84	0,08
	Grasa	5,8346	8,7188	6,92	0,79	0,11
	Ceniza	2,9180	4,0701	3,34	0,40	0,12
Granos bs Colectivo de Calibración (n=83)	Humedad	10,1020	15,0118	11,88	0,96	0,08
	Proteína	8,3506	11,2966	9,94	0,72	0,07
	Grasa	5,3461	9,4603	6,71	0,70	0,10
	Ceniza	2,5121	4,6242	3,24	0,45	0,14
Granos bs Colectivo de Validación (n=16)	Humedad	10,6590	15,4973	12,30	1,32	0,11
	Proteína	8,3332	11,3981	9,93	0,84	0,08
	Grasa	5,8346	8,7188	6,92	0,79	0,11
	Ceniza	2,9180	4,0701	3,33	0,40	0,12

Dónde:

Mínimo: Valor mínimo obtenido por el método de referencia.

Máximo: Valor máximo obtenido por el método de referencia.

Promedio: Valor promedio obtenido por el método de referencia.

SD: Desviación Estándar de los datos obtenidos por el método de referencia.

CV: Coeficiente de Variación medio de los datos obtenidos por el método de referencia ($CV=SD/\hat{y}$)

Con los nuevos colectivos de calibración, se realizó el modelamiento quimiométrico que incluyó la detección de *outliers*, la validación cruzada y validación externa. Estas etapas fueron desarrolladas con asistencia del *software* WinISI.

El Cuadro 9 presenta los estadísticos de calibración y de validación cruzada, se observó que las dos presentaciones empleadas presentaron diferentes valores para cada estadístico y además diferentes cantidades de muestras válidas (N) para cada calibración individual, según el *software* WinISI. Esto se debió a los distintos modelamientos quimiométricos empleados en cada una de las presentaciones. Fue importante saber la media de los componentes de las muestras válidas, pues permitió saber dónde estuvieron la mayoría de los datos y hacia dónde tendría preferencia la calibración, la desviación estándar permitió saber que tan alejados están los datos de la media. Mientras que los estimados mínimos y máximos, dieron una idea de los límites de la capacidad de predicción del modelo. Por otro lado, el SEC y RSQ evaluados antes de realizar la validación cruzada dieron una idea previa de cómo sería la calibración. El SECV y 1-VR fueron para evaluar el modelo luego de aplicar la validación cruzada, estos últimos fueron los estadísticos de mayor validez en esta etapa.

A nivel del desarrollo de la calibración, se observaron menores valores de SEC en la presentación Granos bs. Sin embargo, al comparar el valor SEC de humedad con el de otras investigaciones como la de Cozzolino (1998), fueron altos. El autor presenta valores SEC de humedad de 0,38 y 0,27 para los granos de cebada y trigos enteros respectivamente, mientras que presenta valores de 0,38 y 0,43 para las harinas de cebada y trigo respectivamente. Por otro lado, al comparar el trabajo de Bendezú (2011) quién realizó calibraciones de Kiwicha, se puede observar que respecto a la presente investigación obtuvo valores SEC menores en humedad (0,11), pero mucho mayores en proteína (0,50) y grasa (0,37) y valores similares a la presente en ceniza (0,15).

Cuadro 9: Estadísticos de calibración y validación cruzada para las mejores ecuaciones obtenidas

Presentación	Componente	N	Promedio	SD	Mín.	Máx.	SEC	RSQ	SECV	1-VR	#	Trat. Mat.	Regresión y Scatter
Harina bs	Humedad	73	11,85	0,89	9,17	14,53	0,5608	0,6066	0,5934	0,5534	79	2,8,6,1	PLS – SNV and Detrend
	Proteína	75	9,89	0,69	7,81	11,97	0,4838	0,5143	0,5065	0,4604	81	3,5,5,1	PLS – SNV and Detrend
	Grasa	72	6,69	0,64	4,77	8,61	0,1889	0,9128	0,2135	0,8870	93	1,4,4,1	PLS - None
	Ceniza	75	3,22	0,43	1,94	4,50	0,2225	0,7278	0,2312	0,7020	89	2,4,4,1	PLS - None
Granos bs	Humedad	77	11,84	0,89	9,17	14,50	0,5623	0,6002	0,6092	0,5244	81	3,5,5,1	PLS – SNV and Detrend
	Proteína	76	9,88	0,68	7,84	11,92	0,2828	0,8268	0,3232	0,7707	81	3,5,5,1	PLS – SNV and Detrend
	Grasa	73	6,75	0,63	4,84	8,64	0,1646	0,9325	0,1972	0,9014	81	3,5,5,1	PLS – SNV and Detrend
	Ceniza	79	3,22	0,42	1,95	4,50	0,1612	0,8557	0,1973	0,7810	81	3,5,5,1	PLS - None

Dónde: N: El número de las muestras usadas para el desarrollo de la calibración.

Promedio: El promedio de los datos del método de referencia usados para la calibración.

SD: Desviación estándar de los datos del método de referencia

Mín. / Máx.: Valor mínimo / Valor máximo estimado durante el desarrollo de la calibración

SEC (*Standard Error of Calibration*): El error estándar de la calibración.

RSQ (R^2): Es la correlación de los datos de referencia con los datos predichos.

SECV (*Standard Error of Cross Validation*): El error estándar de la validación cruzada.

1-VR: Es 1 menos la variación. Un RSQ tipo de la validación cruzada.

#: Número de longitudes de onda usadas en la calibración

Trat. Mat (Tratamiento Matemático): Conjunto de derivadas y suavizamientos aplicados a los colectivos de calibración

Regresión y Scatter: Tipo de regresión y dispersión usada sobre el colectivo de calibración.

En la presente investigación, los mejores valores de RSQ se encontraron en la presentación de Granos bs, excepto para el componente humedad. Sin embargo, de la misma manera que el SEC, Cozzolino (1998) presentó mejores valores RSQ para la humedad (0,97) de grano de cebada entero. Al respecto, puede indicarse que las investigaciones en cebada y trigo, fueron las primeras en desarrollarse con la tecnología NIRS, globalmente se domina su desarrollo y cada vez se realizan calibraciones mucho más específicas. Incluso calibraciones por tipo de variedad y zonas de origen.

Adicionalmente, Bendezú (2011), obtuvo mejores valores RSQ de humedad (0,95), proteína (0,94), grasa (0,95) y ceniza (0,93). Debe observarse que el valor citado de RSQ para grasa es muy cercano al obtenido en la presente investigación para ambas presentaciones. Por otro lado, Cozzolino (1998) propuso usar otro método para determinar la humedad, que es el de Karl Fisher (titulación) pues determina la humedad total.

A nivel de la validación cruzada, Boisman y Parra (2012) indican que la selección de las ecuaciones se ha de realizar respecto al valor más bajo de SECV y el valor más alto de 1-VR.

Para el 1-VR ocurrió lo mismo que para el RSQ. Los mejores valores se encontraron en los componentes de proteína, grasa y ceniza. En el caso del componente proteína, la presentación de Granos bs (0,77) fue mucho mejor que la presentación de Harinas bs (0,46). Al comparar con investigaciones como la de Bendezú (2011), se han obtenido valores similares de 1-VR para grasa (0,94). En resumen, los mejores valores de 1-VR los tuvo la presentación Granos bs.

La evaluación del estadístico SECV, en el Cuadro 10 evidenció que todos los componentes están dentro del rango recomendado por Rodríguez (2013). Los mejores valores SECV (valores más bajos), están en los componentes de grasa (Granos bs) y ceniza (Granos bs) y), luego está proteína (Granos bs) y muy distante está humedad (Harinas bs / Granos bs). En resumen, los mejores valores de SECV los tuvo la presentación granos bs.

Cuadro 10: Evaluación de estadístico SECV basado en el SEC.

Presentación	Componente	Estadísticos			
		SEC	SECV (VLI)	SECV (VLS)	SECV
Harinas bs	Humedad	0,5608	0,5608	0,8412	0,5934
	Proteína	0,4838	0,4838	0,7257	0,5065
	Grasa	0,1889	0,1889	0,2834	0,2135
	Ceniza	0,2225	0,2225	0,3338	0,2312
Granos bs	Humedad	0,5623	0,5623	0,8435	0,6092
	Proteína	0,2828	0,2828	0,4242	0,3232
	Grasa	0,1646	0,1646	0,2463	0,1972
	Ceniza	0,1612	0,1612	0,2418	0,1973

Dónde:

SEC: Error estándar de calibración

SECV (VLI): Valor límite inferior del Error estándar de Calibración Cruzada, siendo igual a $1 * SEC$

SECV (VLS): Valor límite superior del Error estándar de Calibración Cruzada, siendo igual a $1,5 * SEC$

SECV: Error estándar de validación cruzada obtenido según modelamiento.

4.4. VALIDACIÓN EXTERNA

La validación externa fue desarrollada comparando los valores obtenidos por los métodos de referencia y los valores estimados por las ecuaciones NIRS. Esta etapa llevó las ecuaciones desarrolladas a prueba, pues las muestras utilizadas en la validación externa no fueron consideradas en la elaboración del modelo quimiométrico. Los resultados de la validación se encuentran en el Cuadro 11.

En el caso de la validación externa, fue llamativo que para la presentación de harinas bs, solo el componente grasa obtuvo un RSQ aceptable por el *software* WinISI. Mientras que la presentación de Granos bs obtuvo RSQ aceptables para los componentes de grasa y ceniza. Para el caso del componente humedad, ambos modelamientos obtuvieron valores muy bajos. Por esta razón, el componente de humedad no fue un factor de decisión al momento de comparar las presentaciones de Harinas bs y Granos bs. En cambio, los RSQ de los componentes proteínas, grasa y ceniza sí fueron diferenciadores entre las presentaciones de Harinas bs y Granos bs.

Cuadro 11: Estadísticos obtenidos de la validación externa.

Presentación	Componente	Estadísticos					
		SEP	Bias	Bias (Lím.)	SEP (c)	RSQ	<i>SLOPE</i>
Harinas bs	Humedad	1,197	0,460	0,600	1,142	0,250*	1,071
	Proteína	0,673	-0,092	0,600	0,689	0,399*	1,737
	Grasa	0,363	-0,074	0,600	0,367	0,812	0,849
	Ceniza	0,377	0,074	0,600	0,382	0,296*	0,541
Granos bs	Humedad	1,182	0,429	0,600	1,137	0,262*	1,204
	Proteína	0,647	0,167	0,600	0,646	0,481*	0,721
	Grasa	0,250	0,008	0,600	0,258	0,909	1,148
	Ceniza	0,264	0,020	0,600	0,272	0,654	0,702

Dónde:

SEP: Error Estándar de predicción

Bias ($\hat{\epsilon}$): La diferencia entre la media del valor de referencia \hat{y} y la media del valor predicho por el modelo NIRS \bar{y}

Bias (Límite): Máximo valor que puede obtener el bias

SEP(C): Error estándar de predicción corregido por el Bias

RSQ (R^2): La correlación de los datos de referencia con los datos predichos

SLOPE: La línea de regresión. Entre el valor de referencia (Y) y el valor NIR predicho (X).

Autores como Dhiman (2014), obtuvieron resultados poco satisfactorios en la validación externa. Por ejemplo, el autor obtuvo valores de RSQ menores a 0,28 para todos sus modelos de calibración. Dhiman (2014), indicó que el bajo desempeño de los modelos durante la validación externa podría ser atribuida a la variación entre las muestras, pues no fueron tomadas en cuenta al realizar el modelo. Por el contrario, Cao (2013) recomendó que un set ideal de validación externa debe tener muestras que contengan un rango de variación que cubra todos los componentes químicos, él utilizó muestras de otra cosecha. En el caso de la presente investigación, estos argumentos, no pudieron ser considerados por completo, pues el colectivo de validación fue una fracción escogida al azar de la población muestral y no parte de otras cosechas o lugares de origen como en otras investigaciones.

Especialistas como Nilsson (2011), investigador de la firma FOSS, realizan validaciones de calibraciones NIRS evaluando solo la precisión (expresada como SEP) y el RSQ. Por ejemplo, en el grano de cebada evaluó los parámetros de humedad y proteína (base seca); obteniendo valores SEP de 0,22 y 0,30, y valores RSQ de 0,99 y 0,97 respectivamente para la humedad y proteína. Debe de resaltarse que la investigación de Nilsson (2011) duró más de cinco años y se analizaron más de 4000 muestras. La presente investigación, ha sido desarrollada en un periodo de tiempo mucho más corto y con un número limitado de muestras (n=120), por lo que se espera que a futuro se vaya mejorando el modelo con inclusión de nuevas muestras a la calibración.

Burns y Ciurczak (2007), explicaron que el SEP es el mejor estadístico para evaluar el futuro desempeño de la calibración, además indicaron que está relacionado con la exactitud y la precisión. Este postulado, refuerza lo realizado por Nilsson (2011), quien selecciona sus ecuaciones según los valores obtenidos de SEP y RSQ.

Los valores de SEP, ordenados de menor a mayor según los componentes en cada presentación indicaron que la mejor presentación fue Granos bs. De la misma manera, los valores RSQ de la validación externa indicaron que la presentación de Granos bs fue mejor. En resumen, el modelo desarrollado para la presentación Granos bs fue el que tuvo mejor desempeño en la validación externa.

4.5. ERROR ESTÁNDAR DEL MÉTODO DE REFERENCIA

Los valores SEL fueron determinados según la ecuación (11), ver acápite 2.8.3.a. Error Estándar de Laboratorio, para más de tres repeticiones de Burns y Ciurczak (2007), pues fue la que más se ajustó a la investigación. Esta ecuación fue usada para los resultados obtenidos por los métodos de referencia de los cuatro componentes en estudio, los resultados se encuentran en el Cuadro 12.

Los métodos de referencia empleados según componente, obtuvieron los mejores SEL para la grasa (0.1629) y la ceniza (0.1327), siendo seguidos de manera distante por la proteína (0.3455) y la humedad (0.3554).

Cuadro 12: Valores SEL de los métodos de referencia.

Componente	SEL
Humedad	0,3554
Proteína	0,3455
Grasa	0,1629
Ceniza	0,1327

Dónde:

SEL: Error estándar de laboratorio obtenido de la totalidad de análisis realizados, es decir los colectivos iniciales (n=120).

Según Dardenne (2008), si se sabe el valor de la desviación estándar, el SEL (suponiendo que el SEC = SEL) y se considerando el error NIR=0; es posible obtener el máximo valor R^2 esperado aún sin empezar el modelamiento (ver Cuadro 13).

Cuadro 13: Valores R^2 máximos esperados de los componentes de la calibración.

Presentación	Componente	SEL	SD	RSQ máx.
Harinas bs (n=79)	Humedad	0,3554	0,97	0,8658
	Proteína	0,3455	0,72	0,7697
	Grasa	0,1629	0,66	0,9391
	Ceniza	0,1327	0,45	0,9130
Granos bs (n=83)	Humedad	0,3554	0,96	0,8629
	Proteína	0,3455	0,72	0,7697
	Grasa	0,1629	0,70	0,9458
	Ceniza	0,1327	0,45	0,9130

Dónde:

SEL: Error estándar de laboratorio obtenido de la totalidad de análisis realizados, es decir los colectivos iniciales (n=120).

SD: Desviación estándar del componente evaluado en los colectivos de calibración finales, antes de realizar los tratamientos matemáticos.

RSQ (R^2) máx.: Coeficiente de determinación máximo esperado.

Según la ecuación (11), ver acápite 2.8.3.a. Error Estándar de Laboratorio, casi todos los valores obtenidos durante la validación cruzada y validación externa (a excepción de las

proteínas de la presentación Granos bs), se encontraron por debajo del límite máximo de RSQ propuesto. Se pudo inferir por adelantado dos resultados desde el Cuadro 13, (1) desde un principio se esperó RSQ menores en los componentes de humedad y proteína, (2) el componente grasa podría tener los mejores RSQ. Sin embargo, resultó que el componente proteína de Granos bs obtuvo un valor incluso superior al máximo esperado. Esto pudo deberse a que el cálculo inicial del SEL fue realizado con las 120 muestras y que posteriormente se realizó el cálculo de la desviación estándar con diferentes tamaños de colectivos. Además, que el mismo tamaño del colectivo y valor de la desviación estándar varió durante el modelamiento.

Cuadro 14: Evaluación del estadístico SEP basado en el SEL.

Presentación	Componente	Estadísticos			
		SEL	SEP (VLI)	SEP (VLS)	SEP
Harinas bs	Humedad	0,3554	0,4265	0,4976	1,197*
	Proteína	0,3455	0,4146	0,4837	0,673*
	Grasa	0,1629	0,1955	0,2281	0,363*
	Ceniza	0,1327	0,1592	0,1858	0,377*
Granos bs	Humedad	0,3554	0,4265	0,4976	1,182*
	Proteína	0,3455	0,4146	0,4837	0,647*
	Grasa	0,1629	0,1955	0,2281	0,250*
	Ceniza	0,1327	0,1592	0,1858	0,264*

Dónde:

SEL: Error estándar de laboratorio obtenido de la totalidad de análisis realizados (n=120).

SEP (VLI): VLI= Valor límite inferior, siendo igual a $1,2 * SEL$.

SEP (VLS): VLS=Valor límite superior, siendo igual a $1,4 * SEL$.

SEP: Error estándar de predicción.

Tamburini *et al.* (2015), explicaron que el valor SEL es el mínimo valor que pueden alcanzar los estadísticos de calibración NIRS (SEC y SEP). Esto es debido a las características intrínsecas del NIRS de ser una técnica secundaria. Esta afirmación se cumplió en todos los casos al comparar el SEL con el SEC y SEP, ver Cuadros 9 y 14.

Según el *Committee for proprietary medicinal products* (2003) un valor recomendado del SEP debe estar entre 1,2 a 1,4 veces el SEL. En la presente investigación, todos los modelos desarrollados estuvieron fuera de los valores recomendados. Se revisó individualmente el diferencial de cada componente en las dos presentaciones y se observó que Granos bs tuvo el menor diferencial con el valor máximo recomendado. Respecto a los altos valores obtenidos, Ye (2003), señaló que un SEP por sí solo no refleja claramente la utilidad de la calibración, en especial si los rangos de datos son pequeños. Por lo que este argumentó se consideró para el caso calibraciones para los componentes de grasa y ceniza que su rango de datos era limitado. Sin embargo, al ser solo demostrativa, esta comparación no ha tenido mayor relevancia en la investigación

4.6. SELECCIÓN DE LA MEJOR ECUACIÓN DE CALIBRACIÓN

Luego de la validación externa, se realizó una serie de evaluaciones a las ecuaciones individuales. Estas evaluaciones evaluaron el ratio del rango y la desviación estándar con el error típico de predicción. Los mejores valores de la validación externa, junto con los mayores RER y RSQ fueron los decisores de la mejor presentación para desarrollar la ecuación final, estos decisores se pueden visualizar en los Cuadros 11 y 15.

Cuadro 15: Estadísticos estimadores de calidad de las ecuaciones de calibración.

Presentación	Componente	RER	RPD
Harinas bs	Humedad	4,04	1,10
	Proteína	4,55	1,25
	Grasa	7,94	2,18
	Ceniza	3,06	1,06
Granos bs	Humedad	4,09	1,11
	Proteína	4,74	1,30
	Grasa	11,54	3,17
	Ceniza	4,36	1,51

Dónde:

RER: Es el ratio del rango ($V_{\text{mayor}} - V_{\text{menor}}$) de las muestras de validación sobre el error típico de predicción (SEP).

RPD: Es el ratio de la desviación estándar $-SD-$ de las muestras de validación sobre el error típico de predicción (SEP).

En cuanto al RER, el componente grasa fue el que obtuvo los mejores valores para las presentaciones Harinas bs (7,94) y Granos bs (11,54). De forma global, la presentación Granos bs presentó los mejores valores. Autores como Williams y Sobering (1996), recomiendan un valor mínimo de 10; por otro lado Millmier *et al.* (2000), indican que valores de RER comprendidos entre 8 y 12 permiten realizar predicciones cuantitativas de los parámetros desarrollados y valores mayores a 12 indican una excelente calidad del modelo predictivo. Al considerar estos últimos manifiestos, solo el componente grasa de la presentación de Granos bs estaría cerca del rango aceptable según Williams y Sobering (1996) y Millmier *et al.* (2000).

Según William (2001), el valor RER no debería ser un indicador objetivo de una buena calibración. Esto debido a que si el análisis se basara solamente en el rango del colectivo de validación, la presencia de valores extremos daría valores ficticios del RER. Por esta razón, debe tenerse cuidado al escoger el colectivo de validación.

En cuanto al RPD, el componente grasa fue el obtuvo los mejores valores para las presentaciones Harinas bs (2,18) y Granos bs (3,17). De forma global, la presentación Granos bs presentó los mejores valores. Autores como Williams y Sobering (1993), Cheewapramong (2007) y Ye (2003), recomiendan valores mínimos de 2,5 y 3,0. Por otro lado, Via (2003), indica que un valor efectivo de RPD debe estar entre 1,5 a 2,5 o mayor. Al respecto, Williams y Sobering (1996) sugieren que los valores RPD no pueden ser muy elevados si hay poca variación en los datos de referencia.

Souza *et al.* (2012) citan a Williams (2001), él cual indica que un valor de $RPD > 2,4$ es deseable para satisfacer el modelamiento, mientras que Williams y Sobering (1995) indican que un valor recomendable debería ser de 3 o más. Este resultado, depende del error de predicción (SEP) el cual debe ser menor a la desviación estándar de los valores de referencia. Souza, obtuvo valores $RPD=1,83$ para lípidos y valores $RPD=0,83$ para carbohidratos. En el caso de los carbohidratos obtuvo valores más altos de SEP que de SD ($SEP=3,71$, $SD=3,08$). Souza *et al.* (2012), recalca que evaluar sólo el RPD no determina la capacidad de predicción de los modelos, pues el autor obtuvo varios valores de SEP bajos que al ser comparados con la SD, resultaron en errores de predicción bajos. A pesar de que los resultados Souza *et al.* (2012). son muy altos en ambos estadísticos, presenta similitud con la presente investigación, en el sentido de que los valores de SD están cercanos al SEP (componentes de humedad, proteína y ceniza).

Al considerar los manifiestos de valores recomendados de RPD, la presentación de Harinas bs solo obtuvo un valor aceptable para el componente grasa (2,18). Mientras que la presentación de Granos bs obtuvo valores aceptables para los componentes de proteína (1,30), grasa (3,17) y ceniza (1,51). Los valores obtenidos para los componentes de proteína, grasa y ceniza, permitirían clasificar la quinua en cuanto a su contenido de estos componentes en los programas de fitomejoramiento (Cheewapramong, 2007 y Millmier *et al.*, 2000). En resumen, los mejores valores de RPD se encuentran en la presentación de Granos bs.

Según López (2008), una limitación de la espectroscopía en el infrarrojo cercano es que el tamaño de las muestras analizadas juega un papel decisivo en el análisis. Por esta razón, las muestras de la presentación de Harinas bs fueron sometidas a una molienda fina y uniforme hasta que el tamaño de las partículas sea prácticamente el mismo. Si se hubiera analizado una misma muestra con grupos de distintos tamaños de partículas, el análisis podría haber dado un espectro distinto para cada lote.

Según López (2008), la temperatura tiene un efecto mayor en los enlaces de H-H (agua) de los productos alimenticios. Si la temperatura cambia, la cantidad de agua detectada en la muestra cambiará de una u otra forma, es decir, cambiarán los puentes de hidrógeno en la muestra. Por esta razón, fue recomendable que las muestras sean almacenadas a una temperatura establecida. Las muestras utilizadas en la presente investigación fueron retiradas de una cámara de refrigeración (banco de germoplasma) y analizadas durante finales de invierno e inicios del verano de los años 2014-2015. Durante este periodo, hubo cambios en la temperatura y humedad relativa; estos cambios pudieron afectar los contenidos de humedad y aumentado el error de los resultados en la calibración.

Según Marten *et al.* (1989), la exactitud de la ecuación puede estar limitada por tres motivos. La primera es la exactitud de las muestras de calibración. Los errores aleatorios de la medición pueden tener un efecto mínimo en la calibración, mientras que cualquier error sistemático tendrá un efecto mayor. Segundo, la información espectral disponible puede no ser únicamente de los enlaces químicos de interés. Los enlaces químicos absorben radiación en varias regiones del espectro, esto se observa en la región del infrarrojo cercano en forma sobretonos y bandas combinadas. Incluso, la mayoría de regiones del espectro muestra efectos de absorción para más de un enlace químico. A menudo, las longitudes de onda correspondientes a los enlaces químicos que no son de

interés aparecen en la ecuación para contrarrestar el efecto de otros absorbentes en las longitudes de onda primarias.

Las muestras utilizadas fueron de naturaleza variable: pertenecientes a diferentes accesiones, cosechas, colores y contenido de macronutrientes. Esta característica aportó variabilidad al modelo y le dio la posibilidad de ser usado con una amplia gama de muestras de colores claros a futuro.

De acuerdo a los resultados obtenidos en los Cuadros 9, 11 y 15 se consideró al modelo quimiométrico de la presentación Granos bs como el mejor, tanto en exactitud y precisión según las condiciones del desarrollo de la presente investigación. El principal motivo de esta selección fue la ejecución de la validación externa, al comparar los valores obtenidos por el método de referencia y valores estimados de un set de muestras no considerados para el desarrollo de la calibración.

Siguiendo la recomendación de Infrasoftware International (2005), una vez seleccionada la mejor ecuación fue conveniente volver a calcular y desarrollar una nueva ecuación que combinara las muestras de calibración y muestras de validación. Esto se realizó utilizando los mismos tratamientos matemático, dispersiones y longitudes de onda usadas para desarrollar las mejores ecuaciones para cada componente. Esta nueva ecuación fue transferida al equipo NIRS. En el Cuadro 16, se pueden visualizar los estadísticos de calibración de la nueva ecuación.

El Cuadro 16 presenta los estadísticos de calibración y validación cruzada para el modelo mejorado, este se desarrolló en base a la presentación de Granos bs. Las nuevas ecuaciones desarrolladas presentaron valores ligeramente menores de RSQ que la calibración original de Granos bs. Sin embargo, fue justificable utilizar esta calibración con menores valores de RSQ debido a que presentó mayor cantidad de términos (N) y por ende será más robusta. La calibración mejorada estuvo compuesta por cuatro ecuaciones individuales (una ecuación por componente).

Una vez que se definieron los archivos de las ecuaciones, estos se insertaron manualmente uno por uno al *software* Odin, posteriormente fueron convertidos del archivo de extensión EQU a CAM. Finalmente, el modelo fue transferido al equipo Infratec 1241 para su puesta en marcha. Adicionalmente, en el Cuadro 17 se ha realizado un resumen de las características del modelo quimiométrico desarrollado.

En las Figuras 20, 21, 22 y 23 se pueden visualizar las gráficas de comparación de los valores calculados por los métodos de referencia y los predecidos por las ecuaciones NIT. También se aprecian las ecuaciones empíricas recomendadas por el *software* WinISI.

Cuadro 16: Estadísticos de calibración y validación cruzada de la presentación Granos bs (n=99).

Componente	N	Promedio	SD	Mínimo	Máximo	SEC	RSQ	SECV	1-VR	#	Trat. Mat.	Regresión y Scatter
Humedad	94	11,87	0,87	9,25	14,49	0,6398	0,4648	0,6953	0,3610	81	3,5,5,1	PLS – SNV and Detrend
Proteína	88	9,88	0,69	7,82	11,94	0,3009	0,8084	0,3375	0,7560	81	3,5,5,1	PLS – SNV and Detrend
Grasa	89	6,78	0,66	4,80	8,76	0,1731	0,9313	0,1895	0,9168	81	3,5,5,1	PLS – SNV and Detrend
Ceniza	92	3,22	0,40	2,01	4,42	0,1551	0,8517	0,1915	0,7715	81	3,5,5,1	PLS - None

Cuadro 17: Características del modelo quimiométrico desarrollado

Producto	Componente	Unidades	Método de referencia	Muestras	Mínimo	Máximo	RSQ
Grano de Quinoa (bs)	Humedad	%	Método de Estufa	94	9,25	14,49	0,4648
	Proteína	%	Kjeldahl	88	7,82	11,94	0,8084
	Grasa	%	Soxhlet	89	4,80	8,76	0,9313
	Ceniza	%	Incineración en mufla	92	2,01	4,42	0,8517

Dónde:

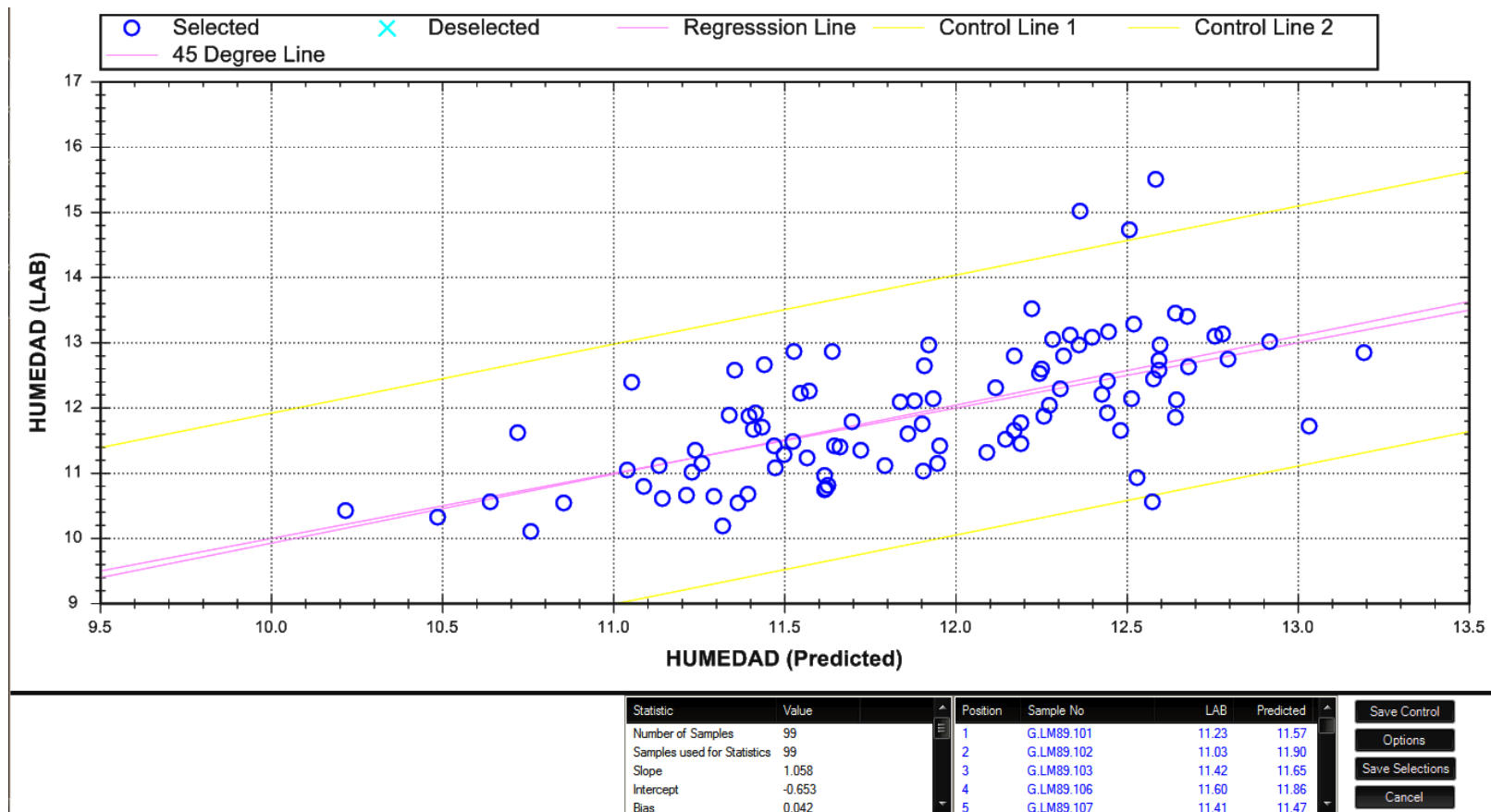
Unidades: La unidad de medida del componente

Método de referencia: El método de referencia con el que fue elaborada la calibración

Muestras: El número de muestras que finalmente fueron considerados para desarrollar cada calibración individual

Mínimo / Máximo: Rango de valores recomendados para el análisis de muestras

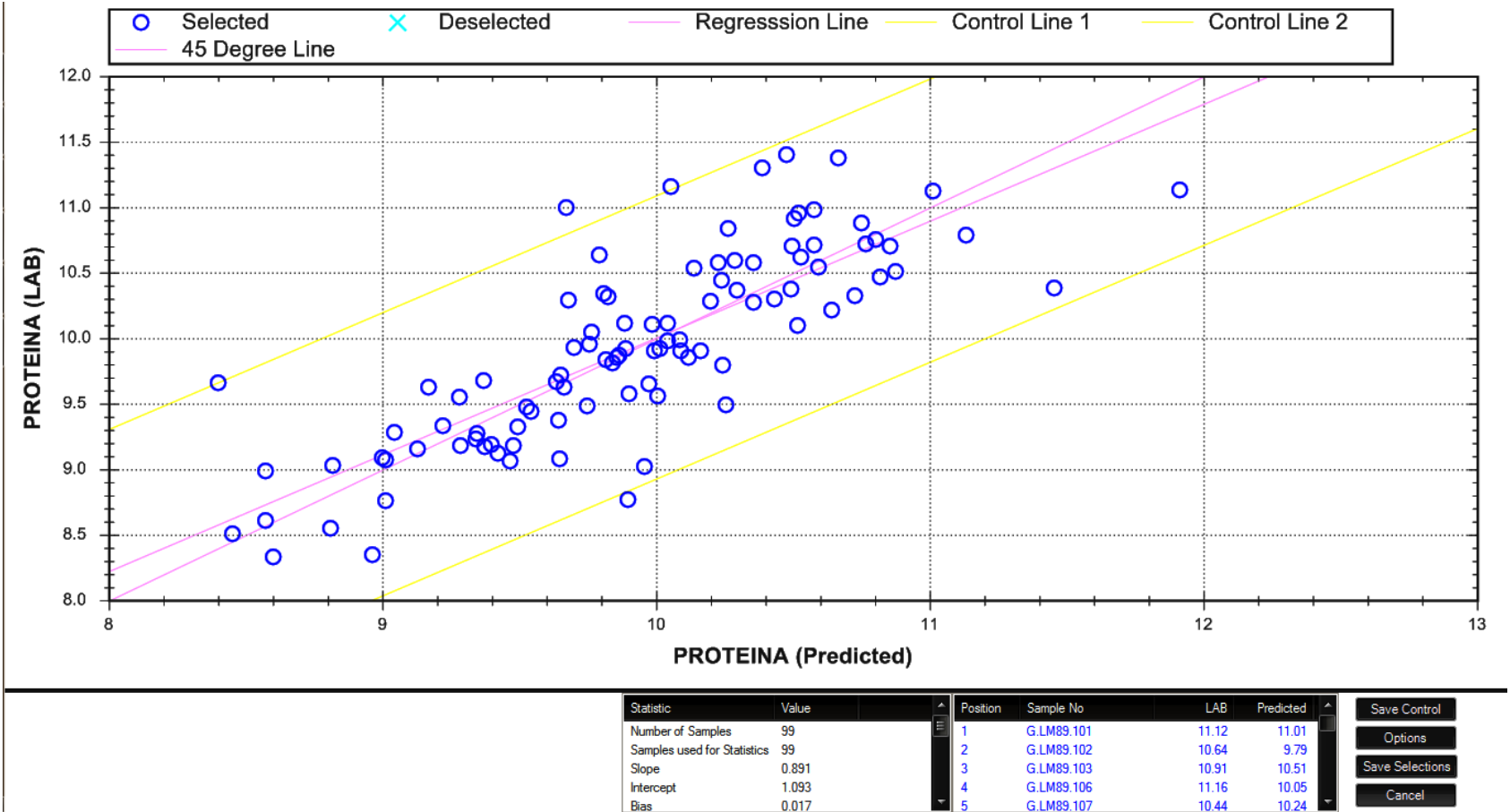
RSQ: La correlación de los datos de referencia con los datos predecidos.



Leyenda: Ecuación Empírica: $Y=1,058X -0,653$, dónde Y es el valor químico y X es el valor predecido vía NIT.

Figura 20: Comparación de los valores calculados vía química (Y) y los vía NIT (X) del componente humedad.

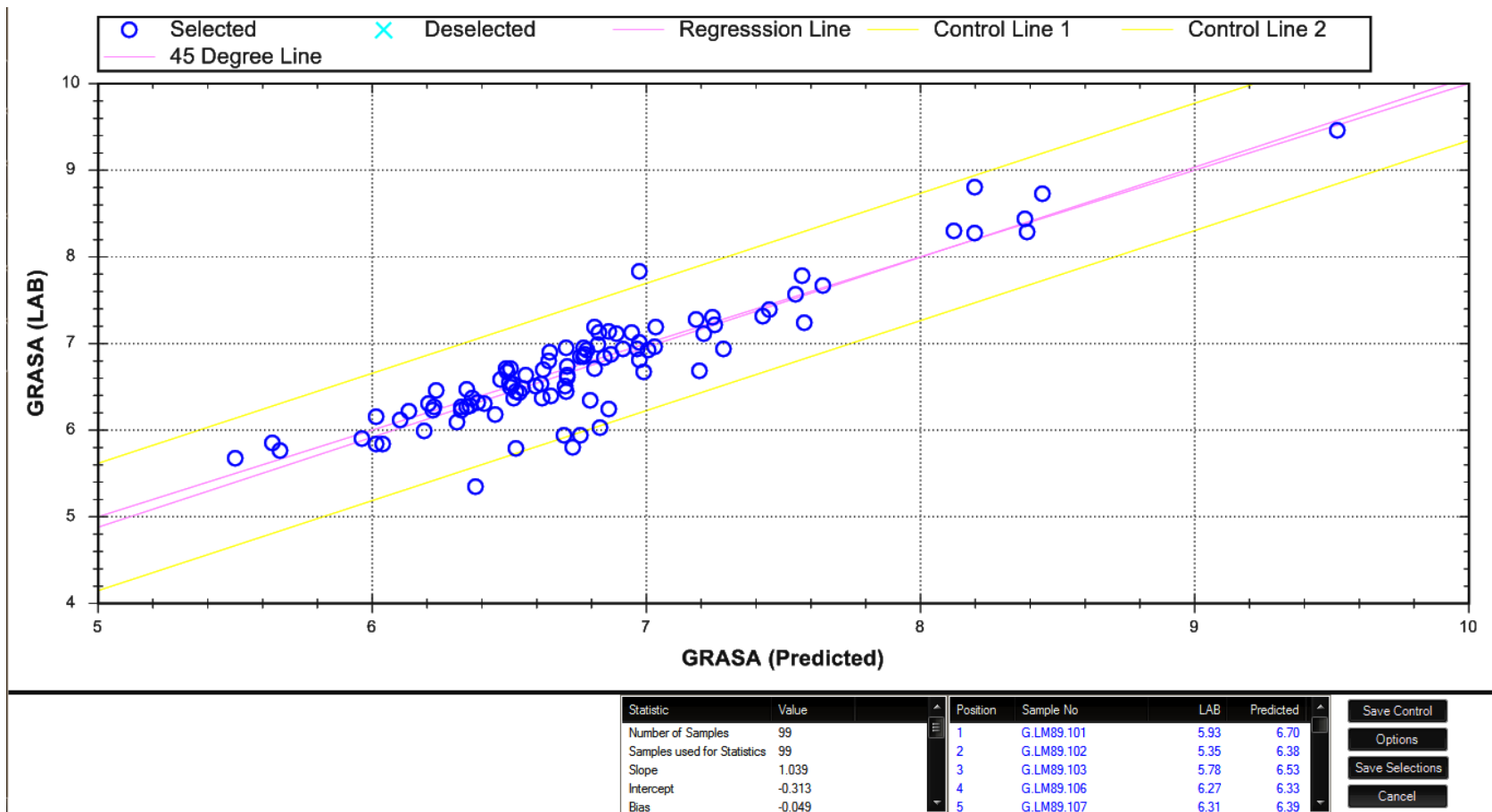
FUENTE: Imagen obtenida del WinISI.



Leyenda: Ecuación Empírica: $Y=0,891X-1,093$, donde Y es el valor químico y X es el valor predecido vía NIT.

Figura 21: Comparación de los valores calculados vía química (Y) y los vía NIT (X) del componente proteína.

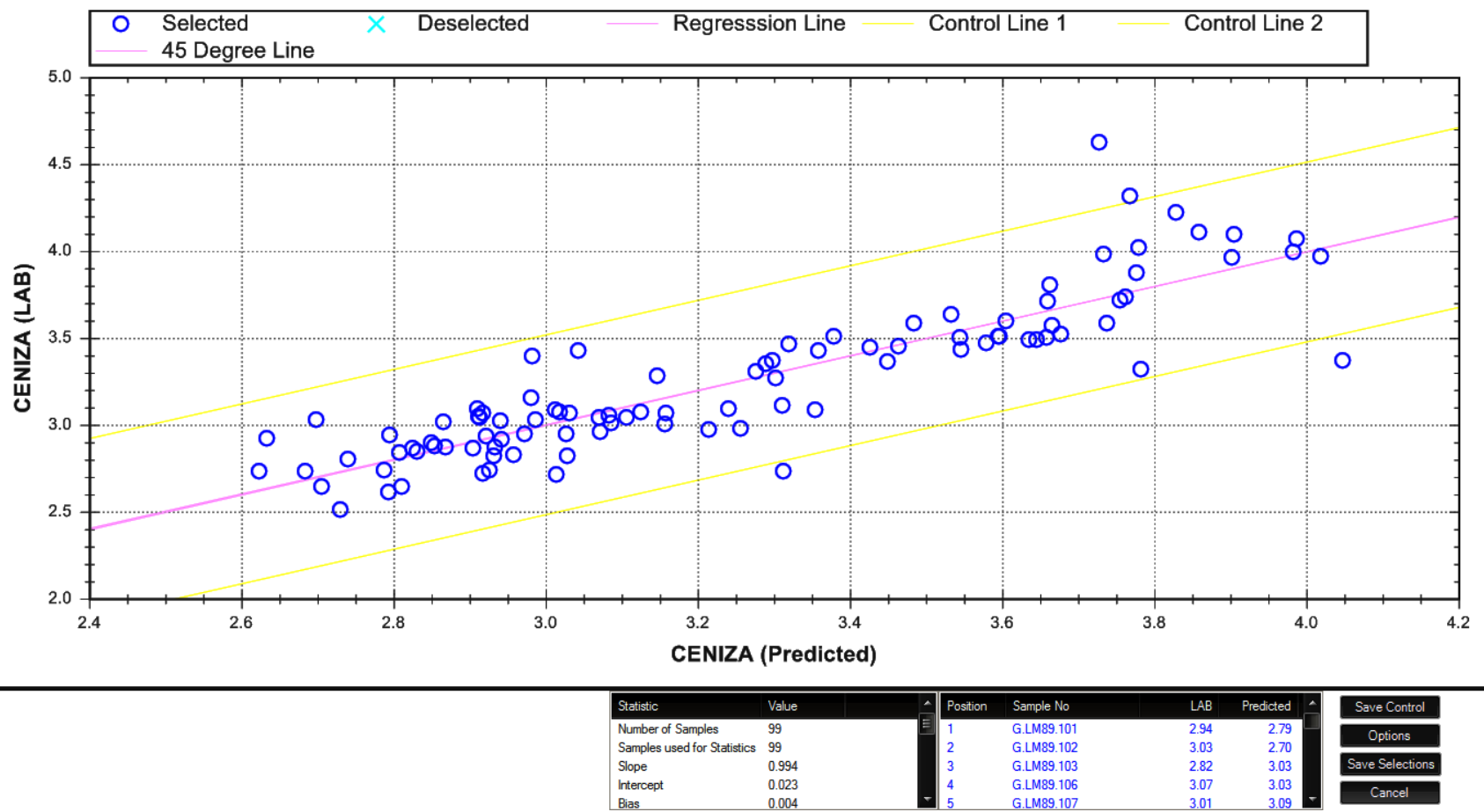
FUENTE: Imagen obtenida del WinISI.



Leyenda: Ecuación Empírica: $Y=1,039X-0,313$, dónde Y es el valor químico y X es el valor predecido vía NIT.

Figura 22: Comparación de los valores calculados vía química (Y) y los vía NIT (X) del componente grasa.

FUENTE: Imagen obtenida del WinISI.



Leyenda: Ecuación Empírica: $Y=0,994X-0,023$, dónde Y es el valor químico y X es el valor predecido vía NIT.

Figura 23: Comparación de los valores calculados vía química (Y) y los vía NIT (X) del componente ceniza.

FUENTE: Imagen obtenida del WinISI.

V. CONCLUSIONES

- El colectivo global de muestras de quinua ($n=120$), analizado mediante los métodos de referencia, presentó en su composición un rango de 8,84 a 15,50 por ciento para la humedad, de 8,33 a 11,72 por ciento para la proteína, de 4,79 a 9,46 por ciento para la grasa y de 2,51 a 4,62 por ciento para la ceniza; todos estos valores son expresados en base seca.
- Se determinó que las muestras de quinua de colores claros y oscuros pertenecían a grupos poblacionales de características diferentes. Dada esta condición, se desarrollaron las diversas calibraciones en base a las quinuas de colores claros (crema, amarillo y naranja), el grupo de mayor tamaño.
- Se determinó que el modelo desarrollado en base a la presentación de “Granos bs” tiene mejor poder de predicción que el de la presentación “Harinas bs”. La decisión fue tomada luego de comparar estadísticos a nivel de validación cruzada y validación externa.
- De acuerdo a los estadísticos estimadores de calidad RER y RPD, el componente grasa del modelo de calibración “Granos bs” tiene valores $RER=11,54$ y $RPD=3,17$. Los cuales permitirán clasificar la quinua en cuanto a su contenido en programas de fitomejoramiento. Además, el estadístico RPD indica que los componentes de proteína ($RPD=1,30$) y ceniza ($RPD=1,51$) tendrían valores de predicción aceptables.
- La calibración mejorada, desarrollada en base a la presentación Granos bs tiene valores RSQ 0,4648, 0,8084, 0,9313 y 0,8517 para la humedad, proteína, grasa y ceniza respectivamente. Esta calibración fue transferida al equipo Infratec 1241.

VI. RECOMENDACIONES

- El desarrollo de una calibración NIRS es un proceso de mejora continua, el cual debe ser ajustado cada cierto tiempo. Es influenciado por las características propias de las muestras como los diferentes años de cosechas, variedades y origen.
- La exactitud y precisión de los resultados obtenidos por el NIRS dependerá especialmente del método de referencia escogido y de los tratamientos utilizados para desarrollar la calibración, pues el NIRS al ser una técnica secundaria siempre tendrá un error intrínseco mayor al del método referencia.
- Para futuras investigaciones, será necesario usar un método para determinación de proteínas que tengan mayor sensibilidad, esta fue una deficiencia encontrada durante el análisis de resultados.
- La ejecución de los análisis por los métodos de referencia y las lecturas de espectros de infrarrojo debe realizarse en el tiempo más próximo posible, para evitar cualquier factor que pueda modificar su composición.
- La cantidad de muestras, el rango y distribución del contenido los componentes de interés son claves para el desarrollo de calibraciones robustas.
- La nitidez del espectro, también debe ser un factor a considerar pues una lectura cada 2 nm, analiza diferentes detalles que una realizada cada 1 nm.

VII. REFERENCIAS BIBLIOGRÁFICAS

- ALBELLA, J.; CINTAS, A.; MIRANDA, T. y SERRATOSA, J. 1993. Introducción a la ciencia de los materiales. Técnicas de preparación y caracterización. Consejo Superior de Investigaciones Científicas. Editorial CSIC. Madrid, España.
- ALONSO, E.; IGARZABAL, A.; OREGUI, L y MANDALUNIZ, N. 2005. Estimación del contenido de nitrógeno en heces de rumiantes mediante espectroscopía en el infrarrojo cercano (NIRS). Revista Producciones agroganaderas: Gestión eficiente y conservación del medio natural (Vol I). Consultado el 04 de octubre del 2015. Disponible en: <http://www.serida.org/seep2005/trabajos/5.pdf>.
- AOAC. 2000. Official Methods of Analysis, Association of Official Analytical Chemists.
- BARNES, R.; DHANOA & M & LISTER, S. 1989. Standard normal variety transformation and de-trending of near -infrared diffuse reflectance spectra. App. Spectrosc. Volume 43, 772.
- BARRIO, J.; ANDRÉS, D. y ANTÓN, J. 2008. Física y Química. Editorial Editex. Consultado el: 01 de Marzo del 2014. Disponible en: http://books.google.com.pe/books?id=bKKlb3z4FgwC&hl=es&source=gbs_navlinks_s.
- BENDEZÚ, E. 2011. Determinación de la correlación entre el análisis químico proximal y el análisis por espectroscopía de transmitancia en el infrarrojo cercano (NIRS) en semillas de kiwicha (*Amaranthus caudatus* L.). Tesis para optar el título de Ingeniero en Industrias Alimentarias. Universidad Nacional Agraria La Molina. Lima, Perú.
- BOISMAN, D. & PARRA M. 2012. Spanish NIR training complete WinISI 4 rev12 - Curso del software para equipos NIR de FOSS. Ejercicios y material de referencia WinISI 4 – ISIscan 3x.
- BURNS, D & CIURCZAK, E. 2007. Handbook of Near-Infrared Analysis. Third Edition. Practical Spectroscopy Series Volume 35. CRC Press. USA.

- CAO, N. 2013. Calibration optimization and efficiency in near infrared spectroscopy. Theses for applying to Doctor of Philosophy in Agricultural Engineering. Iowa State University. Ames, Iowa.
- CHEEWAPRAMONG, P. 2007. Use of Near-Infrared Spectroscopy for Qualitative and Quantitative Analyses of Grains and Cereal Products. Partial Fulfilment of Requirements for the degree of Doctor of Philosophy. Major: Food Science and Technology. Faculty of The Graduate College. University of Nebraska – Lincoln. Lincoln, Nebraska. Consultado el: 01 de Noviembre del 2015. Disponible en: <http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1001&context=foodscidiss>.
- CODEX. 2008. Comisión del *Codex Alimentarius*. Propuesta de elaboración de una norma regional del CODEX para la quinua (*Chenopodium quinoa* Willd). Documento de Proyecto. Acapulco, México. Consultado el: 03 de Marzo del 2014. Disponible en: http://www.cclac.org/documentos/CCLAC/2008/3%20Documentos/Documentos%20Espa%C3%B1ol/la16_13s.pdf.
- COELLO, J. y MASPOCH, S. 2007. Calibración multivariable. En temas avanzados de quimiometría. Universidad de las Islas Baleares. 233-280
- COLLAZOS, C; WHITE, P.; WHITE, H.; VIÑAS, E.; ALVISTUR, E.; URQUIETA, R.; VÁSQUEZ, J.; DÍAZ, C.; QUÍROZ, A.; ROCA, A.; HEGSTED, M. y BRADFIELD, R. 1975. La composición de los alimentos peruanos. Ministerio de Salud. Quinta edición. Lima, Perú.
- COMMITTEE FOR PROPRIETARY MEDICINAL PRODUCTS (CPMP). 2003. Note for guidance on the use of near infrared spectroscopy by the pharmaceutical industry and the data requirements for new submissions and variations. The European Agency for the Evaluation of Medicinal Products (EMA). London. Consultado el 18 de octubre del 2015. Disponible en: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003331.pdf.
- COZZOLINO, D. 1998. Aplicación de la tecnología del NIRS para el análisis de calidad de los productos agrícolas. Editado por la Unidad de Difusión e Información Tecnológica del INIA. Montevideo – Uruguay.
- COZZOLINO, D.; FASSIO, A.; & GIMENEZ, A. 2001. The use of near infrared reflectance spectroscopy (NIRS) to predict the composition of whole maize plants. *Journal of the Science of Food and Agriculture*. Volume. 81: 142 – 146. January 2001.

- CUEVA, R. 2012. Aplicación de la espectroscopía NIR para la determinación de parámetros críticos en la fabricación de comprimidos en la industria farmacéutica. Tesis doctoral. Programa de doctorado de química. Facultad de Ciencias. Universidad Autónoma de Barcelona. Barcelona, España. Consultada el 10 de Febrero del 2014. Disponible en: <http://www.tdx.cat/bitstream/handle/10803/107849/recm1de1.pdf?sequence=1>.
- CUESTA, J. 2013. Nir-Quimiometría – *Blogspot*. 1-VR, squared or coefficient of Determination. Consultado el 01 de Agosto del 2015. Disponible R en: nir-quimometria.blogspot.com/2013_09_01_archive.html.
- DAY, R. y UNDERWOOD, A. 1989. Química Analítica Cuantitativa. Quinta Edición. Editorial Prentice – Hall Hispanoamericana, S. A. México. D.F.
- DARDENNE, P. 2008. Commentary on Nir discussion forum Impublication subject “High RSQ’s with low 1-VR’s???”. Consultado el 15 de Agosto del 2015. Disponible en: <http://www.impublications.com/discus/messages/5/5691.html?1223873743>.
- DHIMAN, J. 2014. Prediction of Heating and Ignition properties of Biomass Dust using Near Infrared Spectroscopy. A thesis submitted to the graduate faculty of Auburn University in partial fulfillment of the requirements for the Degree of Master of Science. Auburn, Alabama.
- FLATEN, G. R. & A. D. WALMSLEY. 2003. Using design of experiments to select optimum calibration model parameters. *Analyst* 128(7): 935-943.
- FLATEN, G. R. & A. D. WALMSLEY. 2004. A design of experiment approach incorporating layered designs for choosing the right calibration model. *Chemometrics and intelligent laboratory systems* 73(1): 55-66.
- FLORES, K.U. 2009. Determinación no destructiva de parámetros de calidad de frutas y hortalizas mediante espectroscopía de reflectancia en el infrarrojo cercano. Tesis Doctoral. Universidad de Cordoba, España.
- FOCA, G.; FERRARI, C.; SINELLI, N.; MARIOTTI, M.; LUCISANO, M.; CARAMANICO, R. & ULRICI, A. 2011. Minimization of instrumental noise in the acquisition of FT-NIR spectra of bread wheat using experimental design and signal processing techniques. *Anal Bioanal Chem* 399(6): 1965-1973.
- FONT, R.; DE HARO, A. & CELESTINO, M. 2006. The use of near-infrared spectroscopy (NIRS) in the study of seed quality components in plant breeding programs. Article in *Industrial Crops and Products* 24 (2006) 307-313. Science Direct.

- FOSS. 2007. User Manual. Version 60022992/Rev 1 Multivariate Data Analysis applying PCA, HB, Cal. IFT. FOSS.AB Sweden.
- FOSS. 2012. Theory – NIR/NIT. Knowledge about the history and theory behind the NIR/NIT technique and how it is used in the instrument. FOSS. Sweden.
- FOSS. 2013. Documentos de Capacitación NIRS.
- GARCÍA, J. 2002. Clasificación y autenticación de canales de Cerdo Ibérico mediante Espectroscopía en el Infrarrojo Cercano (NIRS). Tesis Doctoral presentada para optar el título de Doctor en Ciencias Químicas. Departamento de Bromatología y Tecnología de los Alimentos. Universidad de Córdoba. Córdoba, España.
- GÓMEZ, L. y EGUILUZ, A. 2012. Catálogo del Banco de Germoplasma de Quinoa (*Chenopodium quinoa* Willd) Universidad Nacional Agraria La Molina. Segunda Edición. Edición a cargo del Programa de Cereales y Granos Nativos de La Facultad de Agronomía – Universidad Nacional Agraria La Molina. Lima, Perú.
- GÓMEZ, S. SIERRA, M. y PÉREZ, D. 2009. Análisis Instrumental. Volumen 1. Consultado el 29 de diciembre del 2015. Disponible en: https://books.google.com.pe/books?id=7kqqf47_xB1cC&hl=fr&source=gbs_navlinks_s.
- HESSE, M.; MEIER, H.; y ZEEH, B. 1995. Métodos Espectroscópicos en Química Orgánica. Editorial Síntesis S. A. Valle Hermoso Madrid, España.
- INFRASOFT INTERNATIONAL. 2005. WinISI III Manual.
- ISO - INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. 2010. Animal feeding stuff, cereals and milled cereal products – Guidelines for the application of near infrared spectrometry. ISO 12099:2010. Published in Switzerland.
- JIMÉNEZ, P. 2007. Identificación de Harinas de Yuca (*Manihot esculenta* CRANTZ) con alto contenido proteico mediante espectroscopía de infrarrojo cercano (NIRS). Tesis de Grado presentada como requisito parcial para optar al título de Maestra en Ciencias Agrarias con énfasis en Producción Animal Tropical. Universidad Nacional de Colombia, Sede Palmira. Consultada el 30 de Septiembre del 2016. Disponible en: <http://www.bdigital.unal.edu.co/726/1/07406501.2007.pdf>.
- LÓPEZ, P. 2008. Determinación de las curvas del NIR de alimentos para mascotas. Informe final de cursos en cooperación, como requisito parcial para optar el título de Ingeniero Químico. Decanato de estudios Profesionales, Coordinación de Ingeniería Química. Universidad Simón Bolívar. Sartanejas – Venezuela.

- LUCENA, B y CÁCERES, F. 2006. Análisis prospectivo del uso de la tecnología NIRS en el sector agroalimentario Andaluz. Empresa Pública para el Desarrollo Agrario y Pesquero de la Consejería de Agricultura y Pesca de la Junta de Andalucía. Primera edición. Andalucía, España.
- MARTEN, G.C.; SHENK, J.S. & BARTON, F.E. 1989. Near Infrared Reflectance Spectroscopy (NIRS): Analysis of Forage Quality. Second Edition. USDA Handbook 643, USDA. Washington D.C., United States. 1-96.
- MASSART, D.; VANDEGINSTE, B.; BUYDENS, L.; DE JONG, S.; LEWI, J. SMEYERS-VERBEKE, S. 1998. Handbook of Chemometrics and Qualimetrics: Part B. Editorial Elsevier.
- MILLMIER, A.; LORIMOR, J.; HURBURGH, J.; FULHAGE, C.; HATTEY, J. & ZHANG, H. 2000. Near infrared sensing of manure nutrients. Transactions of the ASAE, 43 (4), 903-908. Consultado el 30 de agosto del 2015. Disponible en: http://works.bepress.com/charles_hurburgh/38/.
- MORCILLO, J. y ORZA, M. 1979. Espectroscopía: estructura y espectros atómicos. Editorial Alhambra, S. A. México D. F.
- MOSQUERA, H. 2009. Efecto de la inclusión de harina de quinua (*Chenopodium quinoa* Willd) en la elaboración de galletas. Trabajo de grado presentado como requisito para optar al Título de especialista en ciencia y tecnología de alimentos. Facultad de Ciencias. Universidad Nacional de Colombia. Bogotá – Colombia.
- MUJICA, A.; JACOBSEN, S.; IZQUIERDO, J. y MARATHEE, J. 2001. Quinua (*Chenopodium quinoa* Willd.). Ancestral cultivo andino, alimento del presente y futuro. Capítulo 8: Valor Nutritivo y Usos de la quinua. FAO. Santiago, Chile. Consultado el: 05 de Octubre del 2015. Disponible en: <http://www.condesan.org/publicacion/Libro03/cap1.htm>.
- MURRAY, I. 1993. Forage analysis by near infrared reflectance spectroscopy. In A. Davies, R.D. Baker, S.A. Grant & A.S. Laidlaw, eds. Sward management handbook.. British Grassland Society, Reading, United Kingdom. 285-312
- NIELSEN, S. 2009. Food Analysis. Fourth Edition. Purdue University. West Lafayette, Indiana, United States.
- NILSSON, T. 2011. FOSS. Comparación de la calibración NIR de FOSS ANN global frente a los métodos de referencia: ¿Un estudio pan-Europeo de cinco años?. P/N 1026582. Primera Edición. Barcelona, España.

- NAES, T., ISAKSSON, T., FEARN, T. & DAVIES, T. 2002. A user-friendly guide to Multivariate Calibration and Classification. NIR Publications Chichester, United Kingdom. 155-175.
- NASA - National Aeronautics and Space Administration, Science Mission Directorate. 2016. Reflected Near-Infrared Waves. Consultado el 05 de Junio del 2016. Disponible en: http://missionscience.nasa.gov/ems/08_nearinfraredwaves.html.
- ORITZ, M.; SARAVIA, L. 2007. Análisis de agrupaciones en temas avanzados de quimiometría. Editorial Blanco y Cerdá. Universidad de las Islas Baleares.
- PEGUERO, A. 2010. La espectroscopía NIR en la determinación de propiedades físicas y composición química de intermedios de producción y productos acabados. Tesis doctoral. Programa de doctorado de química. Facultad de Ciencias. Universidad Autónoma de Barcelona. Barcelona, España. Consultada el 10 de Febrero del 2014. Disponible en: <http://www.tdx.cat/bitstream/handle/10803/3316/apg1de1.pdf?sequence=1>.
- RAMÍIS, G.; GARCÍA, M. 2001. Quimiometría. Editorial Síntesis S.A. Madrid.
- REVISTA MEXICANA DE CIENCIAS FARMACÉUTICAS. 2007. Vol. 38, núm. 1. Pp. 70-74. Consultado el 01 de Junio del 2016. Disponible en: <http://www.redalyc.org/pdf/579/57938109.pdf>.
- REYES, M.; GÓMEZ-SÁNCHEZ, I.; ESPINOZA, C.; BRAVO, F. y GANOZA, L. 2009. Tablas peruanas de composición de Alimentos. Ministerio de Salud. Centro Nacional de Alimentación y Nutrición. Consultado el 30 de Junio del 2016. Disponible en: <http://www.ins.gob.pe/insvirtual/images/otrpubs/pdf/Tabla%20de%20Alimentos.pdf>.
- RODRÍGUEZ, R. 2013. Estadísticas. Consultado el 18 de Octubre del 2015. Disponible en <http://www.slideserve.com/tejana/estadisticas>.
- SÁIZ-ABAJO, M.; MEVIK, H.; SEGTMAN, V. & NAES, T. 2005. "Ensemble methods and data augmentation by noise addition applied to the analysis of spectroscopic data." *Analytical Chemical Acta* 533(2): 147-159.
- SHENK, J.S.; WORKMAN, J. & WESTERHAUS, M.O. 1992. Application of NIR spectroscopy to agricultural products. In: *Handbook of Near-Infrared Analysis*. Editorial. D.A. Burns, E.W. Ciurczak. Marcel Dekker, New York. 383-431.

- SOUZA, D.; AZEVEDO, J. & POPPI, R. 2012. Fourier transform near-infrared spectroscopy (FT-NIRS) application to estimate Brazilian soybean [*Glycine Max* (L.) Merrill] composition. *Food Research International* 51 (2013) 53-58. Science Direct.
- TAPIA, M. 1997. Cultivos Andinos Subexplotados y su aporte a la alimentación. Editorial Organización de las Naciones Unidas para la Agricultura y la Alimentación – Oficina Regional de la FAO para América Latina y el Caribe. Segunda Edición. Santiago, Chile.
- TAPIA, M. E; CANAHUA, A. y IGNACIO, S. 2014. Razas de Quinoa del Perú. ANPE Perú y CONCYTEC. Primera edición. Lima, Perú.
- TAMBURINI, E.; FERRARI, G.; MARCHETTI, M.; PEDRINI, P. & FERRO, S. 2015. Development of FT-NIR Models for the Simultaneous Estimation of Chlorophyll and Nitrogen Content in Fresh Apple (*Malus Domestica*) Leaves. Consultado el 18 de Octubre del 2015. Disponible en: <http://www.mdpi.com/1424-8220/15/2/2662/htm>.
- THERMO ELECTRON CORPORATION. 2015. Theory and Utility of FT-NIR Spectroscopy. An Introduction to Nicolet Antaris FT-NIR Analyzers. Consultado el: 13 de Mayo del 2016. Disponible en: <http://www.cpatc.embrapa.br/eventos/seminario/dequimica/8%BA%20Minicurso%20Infravermelho%20por%20Transformada%20de%20Fourier/nirclaris.pdf>.
- THOUFEEK AHAMED, N.; SINGAL, S.; KULKARNI, P; PAL, M. 1998. A Lesser Known Grain, *Chenopodium quinoa*: Review of the chemical composition of its edible parts. *Food and Nutrition Bulletin*, Vol. 19, No 1, 61-68. The United Nations University.
- USDA. United States Department of Agriculture. 2013. National Nutrient Database for Standard Reference. Consultado el 10 de mayo del 2014. Disponible en: <http://www.ars.usda.gov/ Services/docs.htm?docid=8964>.
- VARGAS, A. 1995. Estadística Descriptiva e inferencial. Editorial Servicio de Publicaciones de la Universidad de Castilla-La Mancha. Segunda Edición. Murcia, España.
- VIA, B.; SHUPE, T.; GROOM, L. STINE, M & SO, C. 2003. Multivariate modeling of density, strength and stiffness from near infrared spectra for mature, juvenile and pith wood of longleaf pine (*Pinus palustris*). *Journal of Near Infrared Spectroscopy*. 11 (5): 365-378.

- WILLARD, H.; MERRITT, L.; DEAN, J. y SETTLE, F. 1991. Métodos Instrumentales de Análisis. Grupo Editorial Iberoamericana S.A. México.
- WILLIAMS, P. y NORRIS K. 2001. Near Infrared Technology in the Agricultural and Food Industry. American Association of Cereal Chemists, St. Paul, Minnesota, United States.
- WILLIAMS, P. & SOBERING, D. 1993. Comparison of commercial Near-infrared Transmittance and Reflectance Instruments for analysis of whole grains and seeds. Journal of Near Infrared Spectroscopy - Volumen 1, Issue 1. Consultado el 10 de abril del 2015. Disponible en: http://www.impuplications.com/content/abstract?code=J01_0025.
- WILLIAMS, P. & SOBERING, D.C. 1996. How we do it: A brief summary of the methods we use in developing near infrared calibrations. In: Near Infrared Reflectance Spectroscopy: The Future Waves, Davies, A.M.C. and Williams, P. (eds). NIR Publications, Chichester, UK, 198-202.
- YE, W. 2003. Application of near-infrared spectroscopy for determination of nutrient contents in manure. Theses for applying to Doctor of Philosophy in Agricultural Engineering. Iowa State University. Ames, Iowa.

VIII. ANEXOS

Anexo 1: Información Técnica del Infratec 1241 – FOSS.



Figura 24: Analizador de granos Infratec 1241.

FUENTE: FOSS (2007).

Cuadro 18: Sistemas y *software* de apoyo del Infratec 1241.

Descripción de sistema	<i>Software</i> de apoyo
Infratec™ 1241 Analizador de granos 220-240V/110-120V, 50-60Hz.	Infratec 1241 <i>File Tools</i> -IFT- versión 3.20, usado para visualizar los datos del Infratec 1241 en el computador
	WINISI versión 4.7, usado para desarrollo de modelo quimiométrico.
	ODIN versión 4.22, usado para transferir la ecuación al equipo Infratec 1241.

FUENTE: FOSS (2007).

Cuadro 19: Datos técnicos del Infratec 1241.

Datos Técnicos	
Voltaje.	220-240V - 50-60Hz - 110-120V
Corriente nominal	1,0A (110-120V) / 0,5A (220-240V)
Dimensiones	W × D × H – 500 × 570 × 375 mm.
Peso	31 kg.
Monocromador	<i>Scanning</i>
Rango de longitud de onda.	850 – 1048 nm.
Amplitud de banda óptica	2 nm
Modo	Transmitancia
Fuente luminosa	Lámpara Halógeno de Tungsteno.
Detector	Silicio
Medios de almacenaje de comunicación	Memoria USB

FUENTE: FOSS (2007).

Anexo 2: Características de las 120 accesiones

Cuadro 20: Características de color y variedad de las 120 accesiones de quinua.

N°	Código	Color Grano con pericarpio	Color Harina	Variedad	Colectivo
1	LM89M4.101	Grupo Crema	Grupo Crema	La Molina 89	Calibración
2	LM89M4.102	Grupo Crema	Grupo Crema	La Molina 89	Calibración
3	LM89M4.103	Grupo Crema	Grupo Crema	La Molina 89	Calibración
4	LM89M4.105	Grupo Crema	Grupo Crema	La Molina 89	Validación
5	LM89M4.106	Grupo Crema	Grupo Crema	La Molina 89	Calibración
6	LM89M4.107	Grupo Crema	Grupo Crema	La Molina 89	Calibración
7	LM89M4.108	Grupo Crema	Grupo Crema	La Molina 89	Calibración
8	LM89M4.109	Grupo Crema	Grupo Crema	La Molina 89	Calibración
9	LM89M4.111	Grupo Crema	Grupo Crema	La Molina 89	Calibración
10	LM89M4.114	Grupo Crema	Grupo Crema	La Molina 89	Calibración
11	LM89M4.115	Grupo Crema	Grupo Crema	La Molina 89	Validación
12	LM89M4.116	Grupo Crema	Grupo Crema	La Molina 89	Calibración
13	LM89M4.118	Grupo Crema	Grupo Crema	La Molina 89	Calibración
14	LM89M4.119	Grupo Crema	Grupo Crema	La Molina 89	Calibración
15	LM89M4.120	Grupo Crema	Grupo Crema	La Molina 89	Calibración
16	LM89M4.121	Grupo Crema	Grupo Crema	La Molina 89	Calibración
17	LM89M4.122	Grupo Crema	Grupo Crema	La Molina 89	Calibración
18	LM89M4.123	Grupo Crema	Grupo Crema	La Molina 89	Calibración
19	LM89M4.124	Grupo Crema	Grupo Crema	La Molina 89	Validación
20	LM89M4.125	Grupo Crema	Grupo Crema	La Molina 89	Validación
21	LM89M4.126	Grupo Crema	Grupo Crema	La Molina 89	Calibración
22	LM89M4.127	Grupo Crema	Grupo Crema	La Molina 89	Calibración
23	LM89M4.128	Grupo Crema	Grupo Crema	La Molina 89	Calibración
24	LM89M4.129	Grupo Crema	Grupo Crema	La Molina 89	Calibración
25	LM89M4.130	Grupo Crema	Grupo Crema	La Molina 89	Calibración
26	LM89M4.131	Grupo Crema	Grupo Crema	La Molina 89	Calibración
27	LM89M4.132	Grupo Crema	Grupo Crema	La Molina 89	Calibración
28	LM89M4.133	Grupo Crema	Grupo Crema	La Molina 89	Calibración
29	LM89M4.135	Grupo Crema	Grupo Crema	La Molina 89	Validación

«continuación»					
30	LM89M4.136	Grupo Crema	Grupo Crema	La Molina 89	Calibración
31	LM89M4.137	Grupo Crema	Grupo Crema	La Molina 89	Calibración
32	LM89M4.138	Grupo Crema	Grupo Crema	La Molina 89	Calibración
33	LM89M4.139	Grupo Crema	Grupo Crema	La Molina 89	Calibración
34	LM89M4.140	Grupo Crema	Grupo Crema	La Molina 89	Calibración
35	LM89M4.141	Grupo Crema	Grupo Crema	La Molina 89	Calibración
36	LM89M4.142	Grupo Crema	Grupo Crema	La Molina 89	Calibración
37	LM89M4.143	Grupo Crema	Grupo Crema	La Molina 89	Calibración
38	LM89M4.144	Grupo Crema	Grupo Crema	La Molina 89	Validación
39	LM89M4.145	Grupo Crema	Grupo Crema	La Molina 89	Validación
40	LM89M4.146	Grupo Crema	Grupo Crema	La Molina 89	Calibración
41	LM89M4.147	Grupo Crema	Grupo Crema	La Molina 89	Calibración
42	LM89M4.148	Grupo Crema	Grupo Crema	La Molina 89	Calibración
43	LM89M4.149	Grupo Crema	Grupo Crema	La Molina 89	Calibración
44	LM89M4.150	Grupo Crema	Grupo Crema	La Molina 89	Calibración
45	LM89M4.151	Grupo Crema	Grupo Crema	La Molina 89	Calibración
46	LM89M4.152	Grupo Crema	Grupo Crema	La Molina 89	Calibración
47	LM89M4.153	Grupo Crema	Grupo Crema	La Molina 89	Calibración
48	LM89M4.154	Grupo Crema	Grupo Crema	La Molina 89	Validación
49	LM89M4.155	Grupo Crema	Grupo Crema	La Molina 89	Validación
50	LM89M4.156	Grupo Crema	Grupo Crema	La Molina 89	Calibración
51	LM89M4.157	Grupo Crema	Grupo Crema	La Molina 89	Calibración
52	LM89M4.158	Grupo Crema	Grupo Crema	La Molina 89	Calibración
53	LM89M4.159	Grupo Crema	Grupo Crema	La Molina 89	Calibración
54	LM89M4.161	Grupo Crema	Grupo Crema	La Molina 89	Calibración
55	LM89M4.162	Grupo Crema	Grupo Crema	La Molina 89	Calibración
56	LM89M4.164	Grupo Crema	Grupo Crema	La Molina 89	Calibración
57	PAS.M5LM12B.101	Grupo Gris	Grupo Marron	Pasankalla	Calibración
58	PAS.M5LM12B.102	Grupo Gris	Grupo Marron	Pasankalla	Calibración
59	PAS.M5LM12B.103	Grupo Gris	Grupo Marron	Pasankalla	Calibración
60	PAS.M5LM12B.104	Grupo Gris	Grupo Marron	Pasankalla	Validación
61	PAS.M5LM12B.105	Grupo Gris	Grupo Marron	Pasankalla	Validación
62	PAS.M5LM12B.106	Grupo Gris	Grupo Marron	Pasankalla	Calibración

«continuación»					
63	PAS.M5LM12B.107	Grupo Gris	Grupo Marron	Pasankalla	Calibración
64	PAS.M5LM12B.108	Grupo Gris	Grupo Marron	Pasankalla	Calibración
65	PAS.M5LM12B.109	Grupo Gris	Grupo Marron	Pasankalla	Calibración
66	POQ.CVM10B.01	Grupo Naranja	Grupo Amarillo	-	Calibración
67	POQ.CVM10B.02	Grupo Amarillo	Grupo Amarillo	-	Calibración
68	POQ.CVM10B.03	Grupo Amarillo	Grupo Amarillo	-	Calibración
69	POQ.CVM10B.04	Grupo Amarillo	Grupo Amarillo	-	Validación
70	POQ.CVM10B.05	Grupo Amarillo	Grupo Amarillo	-	Validación
71	POQ.CVM10B.06	Grupo Amarillo	Grupo Amarillo	-	Calibración
72	POQ.CVM10B.07	Grupo Amarillo	Grupo Amarillo	-	Calibración
73	POQ.CVM10B.08	Grupo Naranja	Grupo Amarillo	-	Calibración
74	POQ.CVM10B.09	Grupo Amarillo	Grupo Amarillo	-	Calibración
75	POQ.CVM10B.10	Grupo Amarillo	Grupo Amarillo	-	Calibración
76	POQ.CVM10B.11	Grupo Amarillo	Grupo Amarillo	-	Calibración
77	POQ.CVM10B.12	Grupo Amarillo	Grupo Amarillo	-	Calibración
78	POQ.CVM10B.13	Grupo Amarillo	Grupo Amarillo	-	Calibración
79	POQ.CVM10B.14	Grupo Amarillo	Grupo Amarillo	-	Validación
80	POQ.CVM10B.15	Grupo Amarillo	Grupo Amarillo	-	Validación
81	POQ.CVM10B.16	Grupo Amarillo	Grupo Amarillo	-	Calibración
82	POQ.CVM10B.18	Grupo Amarillo	Grupo Amarillo	-	Calibración
83	POQ.CVM10B.19	Grupo Amarillo	Grupo Amarillo	-	Calibración
84	POQ.CVM10B.20	Grupo Naranja	Grupo Amarillo	-	Calibración
85	POQ.CVM10B.21	Grupo Amarillo	Grupo Amarillo	-	Calibración
86	POQ.CVM10B.22	Grupo Amarillo	Grupo Amarillo	-	Calibración
87	POQ.CVM10B.23	Grupo Amarillo	Grupo Amarillo	-	Calibración
88	POQ.CVM10B.24	Grupo Naranja	Grupo Amarillo	-	Calibración
89	POQ.CVM10B.25	Grupo Naranja	Grupo Amarillo	-	Validación
90	POQ.CVM10B.26	Grupo Amarillo	Grupo Amarillo	-	Calibración
91	POQ.CVM10B.27	Grupo Amarillo	Grupo Amarillo	-	Calibración
92	POQ.CVM10B.28	Grupo Amarillo	Grupo Amarillo	-	Calibración
93	POQ.CVM10B.29	Grupo Naranja	Grupo Amarillo	-	Calibración
94	POQ.CVM10B.30	Grupo Amarillo	Grupo Amarillo	-	Calibración
95	POQ.CVM10B.31	Grupo Amarillo	Grupo Amarillo	-	Calibración

«continuación»					
96	POQ.CVM10B.32	Grupo Amarillo	Grupo Amarillo	-	Calibración
97	POQ.CVM10B.33	Grupo Crema	Grupo Crema	-	Calibración
98	POQ.CVM10B.34	Grupo Negro	Grupo Gris	-	Validación
99	POQ.CVM10B.35	Grupo Amarillo	Grupo Amarillo	-	Validación
100	POQ.CVM10B.36	Grupo Naranja	Grupo Amarillo	-	Calibración
101	POQ.CVM10B.37	Grupo Crema	Grupo Crema	-	Calibración
102	POQ.CVM10B.39	Grupo Negro	Grupo Gris	-	Calibración
103	POQ.CVM10B.40	Grupo Negro	Grupo Gris	-	Validación
104	POQ.CVM10B.41	Grupo Negro	Grupo Gris	-	Calibración
105	POQ.CVM10B.42	Grupo Negro	Grupo Gris	-	Calibración
106	POQ.CVM10B.43	Grupo Negro	Grupo Gris	-	Calibración
107	POQ.CVM10B.45	Grupo Negro	Grupo Gris	-	Calibración
108	POQ.CVM10B.46	Grupo Negro	Grupo Gris	-	Calibración
109	POQ.CVM10B.47	Grupo Negro	Grupo Gris	-	Calibración
110	POQ.CVM10B.48	Grupo Negro	Grupo Gris	-	Calibración
111	POQ.CVM10B.50	Grupo Negro	Grupo Gris	-	Calibración
112	POQ.CVM10B.51	Grupo Negro	Grupo Gris	-	Calibración
113	POQ.CVM10B.52	Grupo Amarillo	Grupo Amarillo	-	Calibración
114	POQ.CVM10B.53	Grupo Amarillo	Grupo Amarillo	-	Calibración
115	POQ.CVM10B.54	Grupo Amarillo	Grupo Amarillo	-	Calibración
116	POQ.CVM10B.55	Grupo Amarillo	Grupo Amarillo	-	Validación
117	POQ.CVM10B.56	Grupo Amarillo	Grupo Amarillo	-	Calibración
118	POQ.CVM10B.58	Grupo Amarillo	Grupo Amarillo	-	Calibración
119	POQ.CVM10B.59	Grupo Amarillo	Grupo Amarillo	-	Calibración
120	POQ.CVM10B.60	Grupo Amarillo	Grupo Amarillo	-	Calibración

Leyenda:

Los códigos pertenecen a los libros de campo del Programa de Investigación y Proyección Social (PIPS) de Cereales y Granos Nativos de la Universidad Nacional Agraria La Molina PIPS de Cereales y Granos Andinos:

- LM89M4.XXX: Variedad -La Molina 89- (LM89), Número de generación de mutante -Cuarta Generación- (M4), Número de muestra (XXX).
- PAS.M5LM12B.XXX: Variedad -Pasankalla- (PAS), Número de generación de mutante -Quinta generación-(M5), Lugar de siembra -La Molina- (LM), Campaña de siembra -Segundo semestre del 2012- (12B), Número de muestra (XXX).
- POQ.CVM10B.XX: Plantel de Observación de Quinoa de Color (POQ.C), Lugar de siembra -Valle del Mantaro- (VM), Campaña de siembra -Segundo semestre del 2010- (10B), Número de muestra (XX).

Cuadro 21: Resultados promedio de los análisis químicos de las muestras de quinua, expresados como porcentaje en base seca.

N°	Código	% Humedad bs	% Proteína bs	% Grasa bs	% Ceniza bs
1	LM89M4.101	11,2272	11,1226	5,9345	2,9445
2	LM89M4.102	11,0264	10,6399	5,3461	3,0333
3	LM89M4.103	11,4212	10,9100	5,7845	2,8202
4	LM89M4.105	11,4216	11,3743	6,0822	2,9180
5	LM89M4.106	11,6040	11,1601	6,2652	3,0718
6	LM89M4.107	11,4127	10,4449	6,3125	3,0097
7	LM89M4.108	11,7558	10,7099	6,4957	3,0411
8	LM89M4.109	11,3195	11,1318	5,8423	3,0941
9	LM89M4.111	12,9669	11,2966	6,5463	3,0495
10	LM89M4.114	11,6623	10,7009	6,9178	2,7351
11	LM89M4.115	11,9134	9,7923	6,6963	3,0356
12	LM89M4.116	11,3961	9,9792	6,2760	2,8254
13	LM89M4.118	10,7607	9,9223	6,4289	2,9014
14	LM89M4.119	12,6484	9,1526	5,7921	2,9374
15	LM89M4.120	11,8914	10,7229	6,4982	2,9474
16	LM89M4.121	12,2223	10,2870	6,2998	2,6460
17	LM89M4.122	12,5305	9,3774	6,0233	3,0715
18	LM89M4.123	12,8658	10,5811	6,7014	2,8683
19	LM89M4.124	12,6607	10,3268	6,2094	2,9211
20	LM89M4.125	12,1374	10,0456	7,1066	3,0740
21	LM89M4.126	12,8623	10,1104	6,9150	2,8490
22	LM89M4.127	11,4765	9,9860	6,4377	2,7211
23	LM89M4.128	10,7957	10,3871	6,4806	2,6475
24	LM89M4.129	12,3089	9,1251	6,8487	2,9757
25	LM89M4.130	12,0933	9,5745	6,7339	2,8392
26	LM89M4.131	11,8724	9,0896	6,8732	3,1562
27	LM89M4.132	12,2594	10,2903	6,5263	2,7441
28	LM89M4.133	11,3432	8,3506	6,3674	2,8738
29	LM89M4.135	11,1091	8,3332	5,8346	2,9497
30	LM89M4.136	11,0834	9,7198	6,9487	2,8266

«continuación»

31	LM89M4.137	10,7375	9,2281	6,7043	2,7345
32	LM89M4.138	10,1888	9,1824	6,4371	2,5121
33	LM89M4.139	10,6657	8,7610	6,9343	2,9615
34	LM89M4.140	10,9508	8,5523	7,0123	2,8714
35	LM89M4.141	11,0021	9,4812	6,8700	2,7134
36	LM89M4.142	11,7750	9,0816	6,8377	2,8683
37	LM89M4.143	11,1369	9,4927	6,3045	2,8815
38	LM89M4.144	10,7995	8,7715	7,1889	3,0666
39	LM89M4.145	11,2860	9,2738	6,8998	3,2839
40	LM89M4.146	11,0350	9,0216	6,9444	3,4666
41	LM89M4.147	12,5817	9,8508	6,3944	3,3952
42	LM89M4.148	11,6931	8,6095	6,2385	3,3106
43	LM89M4.149	11,3404	9,2783	6,6667	3,4288
44	LM89M4.150	10,5350	9,4415	6,7881	3,1108
45	LM89M4.151	11,1460	9,9566	5,6723	3,0057
46	LM89M4.152	10,6428	9,6810	6,9518	3,0468
47	LM89M4.153	10,5304	9,6713	6,3663	2,6134
48	LM89M4.154	12,3966	9,8346	6,6627	3,0756
49	LM89M4.155	10,6590	8,9909	7,5631	3,0918
50	LM89M4.156	10,5981	9,9076	7,1798	3,0535
51	LM89M4.157	10,1020	9,1751	7,2990	3,0194
52	LM89M4.158	10,3121	9,6522	6,4676	2,7368
53	LM89M4.159	10,5559	9,9040	5,9851	2,8035
54	LM89M4.161	11,6087	9,0682	6,6354	2,7433
55	LM89M4.162	11,1107	9,0277	5,9305	3,0887
56	LM89M4.164	10,4277	8,5119	6,5301	3,0419
57	PAS.M5LM12B.101	9,3316	10,2498	6,5667	2,8311
58	PAS.M5LM12B.102	8,8377	9,5232	6,4020	3,1083
59	PAS.M5LM12B.103	11,8711	10,4878	5,9886	3,4958
60	PAS.M5LM12B.104	12,1434	9,5788	5,9826	2,9508
61	PAS.M5LM12B.105	11,8065	9,7830	5,6248	3,2218
62	PAS.M5LM12B.106	11,8126	9,0847	5,4706	2,7533
63	PAS.M5LM12B.107	12,8025	10,8102	5,9948	3,0844

«continuación»

64	PAS.M5LM12B.108	12,1462	9,5791	5,4460	3,0799
65	PAS.M5LM12B.109	12,1590	9,3466	6,2993	2,9285
66	POQ.CVM10B.01	10,5503	9,9033	6,8493	4,2252
67	POQ.CVM10B.02	11,4473	10,2156	6,1547	3,9947
68	POQ.CVM10B.03	15,0118	10,3030	6,2232	4,6242
69	POQ.CVM10B.04	15,4973	9,6244	7,1158	4,0701
70	POQ.CVM10B.05	14,7313	10,2779	6,4996	3,9857
71	POQ.CVM10B.06	11,9184	9,5594	5,8993	3,9679
72	POQ.CVM10B.07	12,0406	9,3366	7,6698	4,3204
73	POQ.CVM10B.08	13,4524	9,9271	6,3662	4,1083
74	POQ.CVM10B.09	10,9228	9,4740	6,7045	4,0987
75	POQ.CVM10B.10	11,5064	9,0599	7,7807	3,8744
76	POQ.CVM10B.11	11,7748	9,5474	7,2124	3,9678
77	POQ.CVM10B.12	12,5879	10,7896	8,2661	3,3732
78	POQ.CVM10B.13	12,1090	10,9773	6,6842	3,2714
79	POQ.CVM10B.14	11,6563	11,3981	6,9346	4,0222
80	POQ.CVM10B.15	12,1321	10,5122	8,7188	3,5854
81	POQ.CVM10B.16	11,8654	9,3219	7,8268	3,0307
82	POQ.CVM10B.18	11,6439	10,4664	6,2656	3,7187
83	POQ.CVM10B.19	11,8582	10,9527	6,2284	3,5712
84	POQ.CVM10B.20	13,5151	10,8785	7,1321	3,5052
85	POQ.CVM10B.21	13,0747	10,8363	6,3338	3,8040
86	POQ.CVM10B.22	13,1017	9,1894	6,9301	3,3206
87	POQ.CVM10B.23	12,9534	10,1187	8,7971	3,4550
88	POQ.CVM10B.24	12,6289	9,8549	8,2801	3,5090
89	POQ.CVM10B.25	13,1156	9,6618	8,4312	3,4736
90	POQ.CVM10B.26	12,9545	10,1188	7,2380	3,4254
91	POQ.CVM10B.27	12,7985	10,5748	7,3168	3,5990
92	POQ.CVM10B.28	12,7322	9,6292	9,4603	3,5107
93	POQ.CVM10B.29	12,4041	10,5379	7,3877	3,3658
94	POQ.CVM10B.30	12,2846	10,9945	7,1090	3,3510
95	POQ.CVM10B.31	13,1327	10,3705	6,4594	2,9824
96	POQ.CVM10B.32	11,7221	10,7066	8,2956	3,0900

«continuación»

97	POQ.CVM10B.33	12,1122	9,8095	7,2734	3,5079
98	POQ.CVM10B.34	12,3024	10,5283	6,8070	3,6401
99	POQ.CVM10B.35	12,5676	10,3186	6,5802	3,3706
100	POQ.CVM10B.36	12,4392	10,5410	6,2635	3,7354
101	POQ.CVM10B.37	12,2098	10,7534	7,1213	3,5841
102	POQ.CVM10B.39	13,8801	9,2527	6,8697	3,2624
103	POQ.CVM10B.40	13,1513	10,3722	7,0088	3,2487
104	POQ.CVM10B.41	12,9780	9,8855	7,3188	3,3376
105	POQ.CVM10B.42	13,8126	9,9581	6,5678	3,4152
106	POQ.CVM10B.43	13,7457	10,1897	6,1702	3,4908
107	POQ.CVM10B.45	11,3807	11,3701	6,1055	3,2698
108	POQ.CVM10B.46	11,7102	10,2400	7,2053	3,0809
109	POQ.CVM10B.47	12,4668	11,7152	7,0344	3,2556
110	POQ.CVM10B.48	12,1359	10,0455	6,4028	3,2946
111	POQ.CVM10B.50	13,7425	9,9522	4,7928	3,2616
112	POQ.CVM10B.51	12,2909	10,7611	7,5276	2,9994
113	POQ.CVM10B.52	13,0390	10,5973	6,6289	3,5200
114	POQ.CVM10B.53	13,2739	10,6187	6,8071	3,4917
115	POQ.CVM10B.54	13,1700	10,3739	5,7623	3,7111
116	POQ.CVM10B.55	11,3807	11,3701	6,1055	3,2698
117	POQ.CVM10B.56	13,3928	9,9219	5,8369	3,4912
118	POQ.CVM10B.58	13,0075	9,1819	6,9858	3,6395
119	POQ.CVM10B.59	12,8431	9,8738	6,6031	3,4497
120	POQ.CVM10B.60	12,7390	10,0995	6,1146	3,5014

Anexo 3: Histogramas y polígonos de frecuencia de los datos de referencia

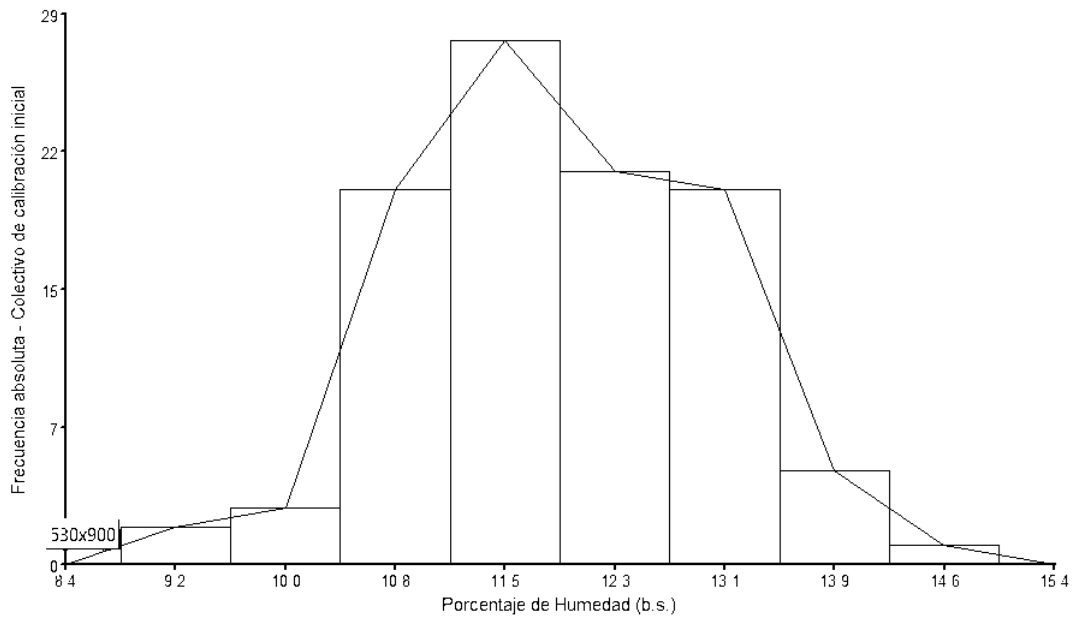


Figura 25: Histograma y polígono de frecuencias del porcentaje de humedad en base seca del colectivo de calibración inicial.

FUENTE: Imagen obtenida del Infostat.

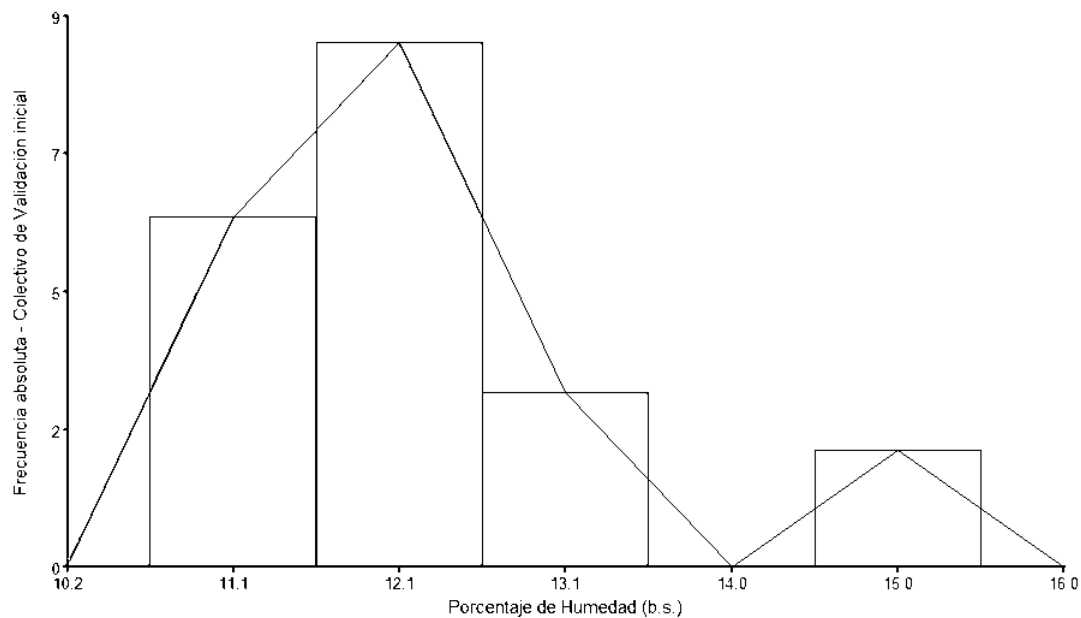


Figura 26: Histograma y polígono de frecuencias del porcentaje de humedad en base seca del colectivo de validación inicial.

FUENTE: Imagen obtenida del Infostat.

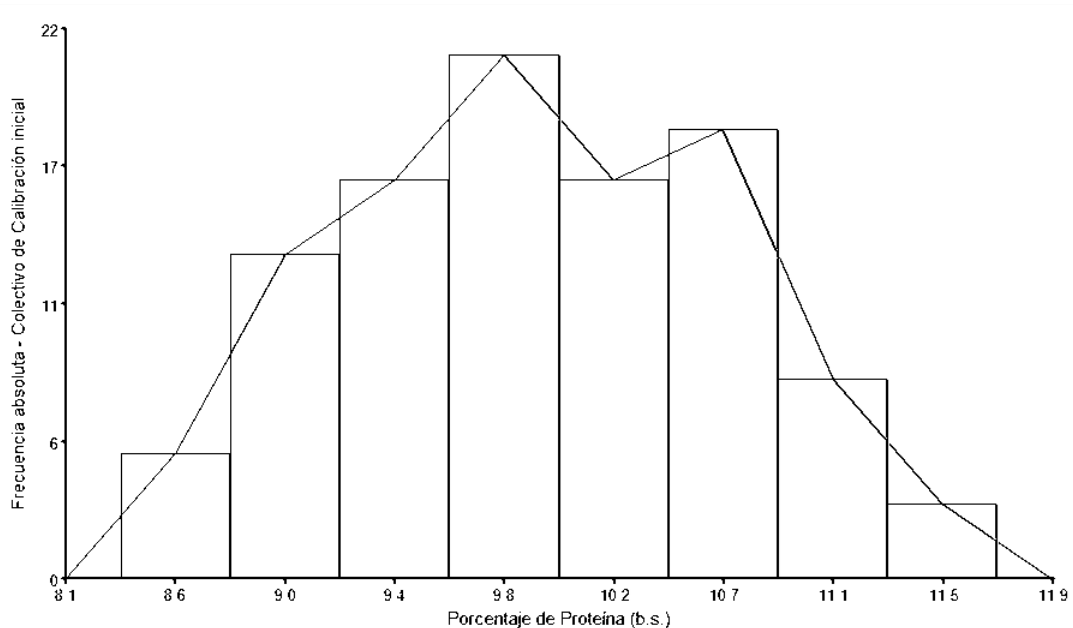


Figura 27: Histograma y polígono de frecuencias del porcentaje de proteína en base seca del colectivo de calibración inicial.

FUENTE: Imagen obtenida del Infostat.

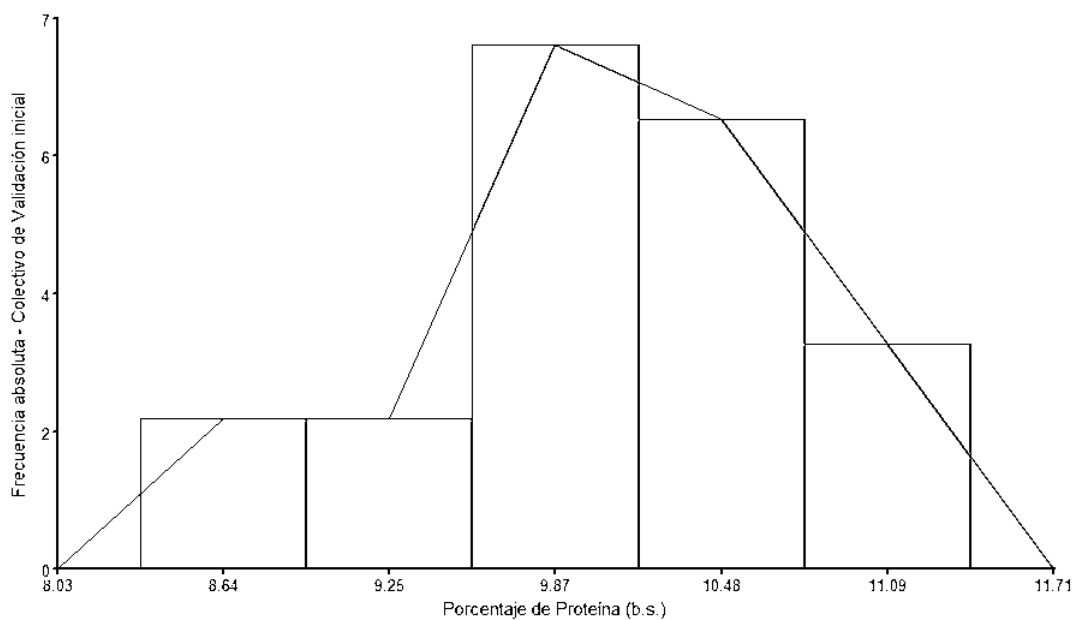


Figura 28: Histograma y polígono de frecuencias del porcentaje de proteína en base seca del colectivo de validación inicial.

FUENTE: Imagen obtenida del Infostat.

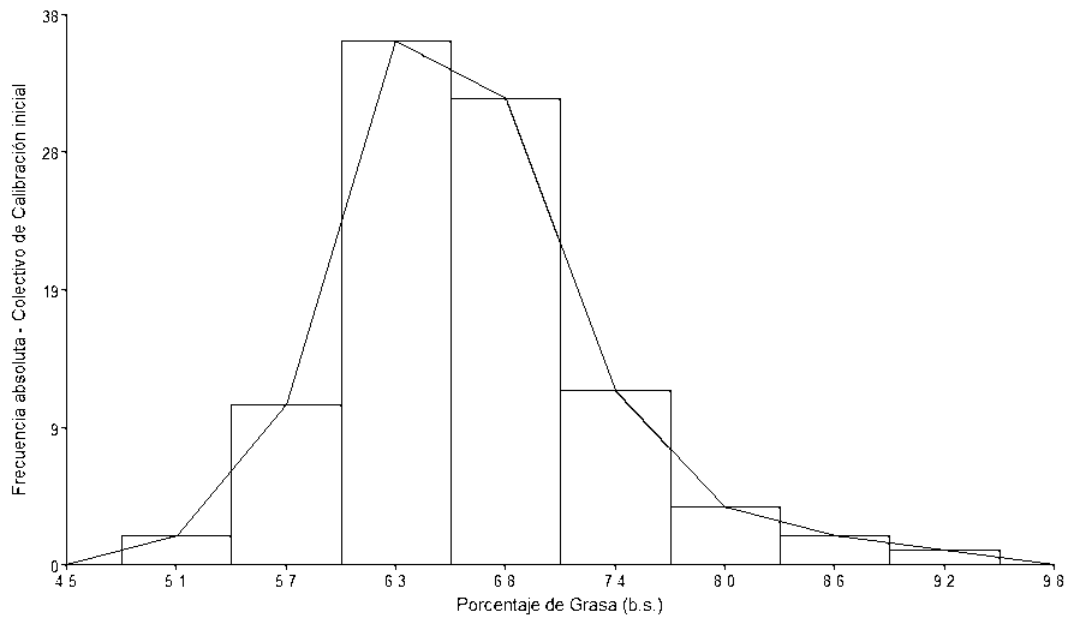


Figura 29: Histograma y polígono de frecuencias del porcentaje de grasa en base seca del colectivo de calibración inicial.

FUENTE: Imagen obtenida del Infostat.

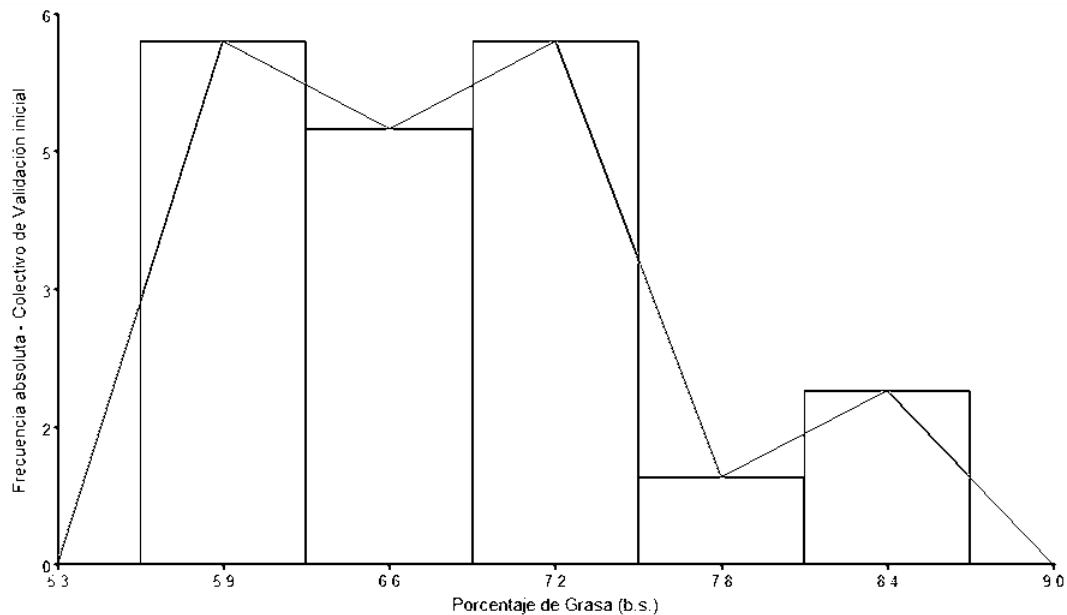


Figura 30: Histograma y polígono de frecuencias del porcentaje de grasa en base seca del colectivo de validación inicial.

FUENTE: Imagen obtenida del Infostat.

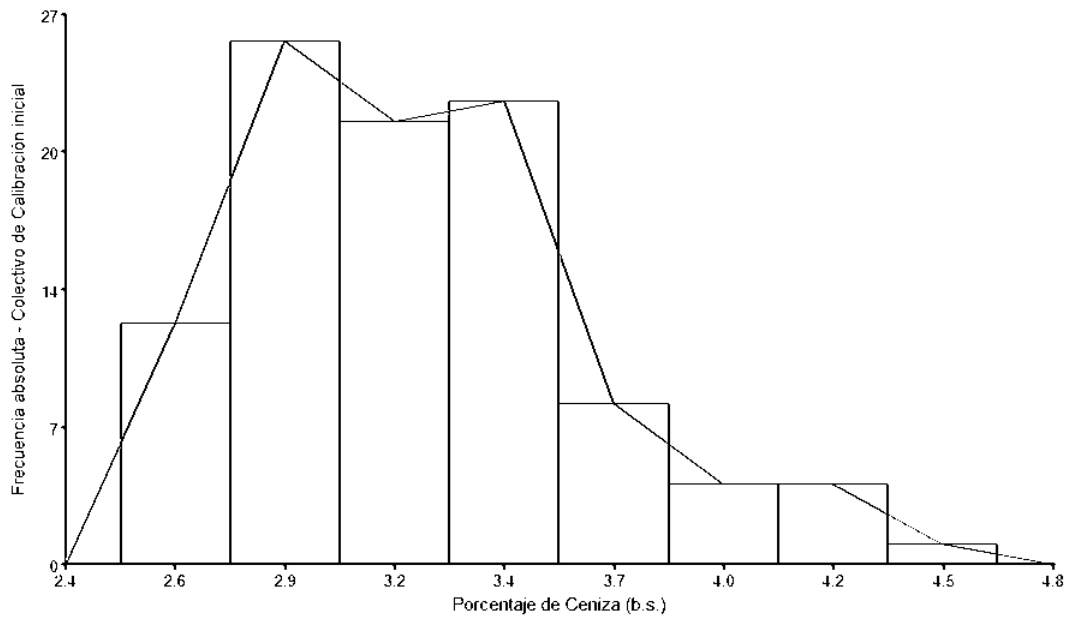


Figura 31: Histograma y polígono de frecuencias del porcentaje de ceniza en base seca del colectivo de calibración inicial.

FUENTE: Imagen obtenida del Infostat.

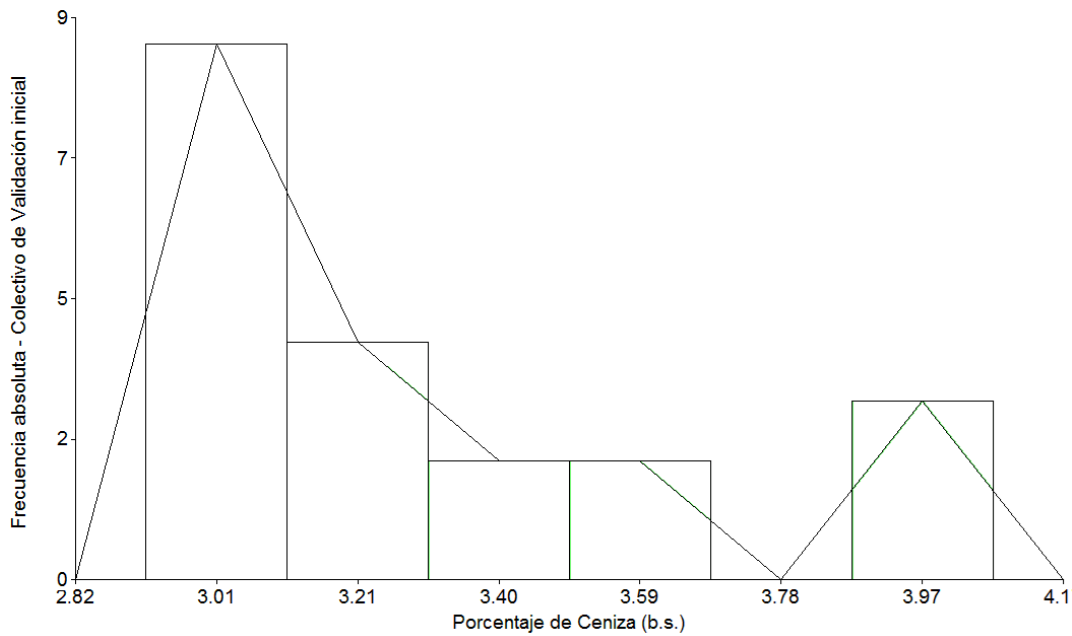


Figura 32: Histograma y polígono de frecuencias del porcentaje de ceniza en base seca del colectivo de validación inicial.

FUENTE: Imagen obtenida del Infostat.

Anexo 4: Espectros anómalos

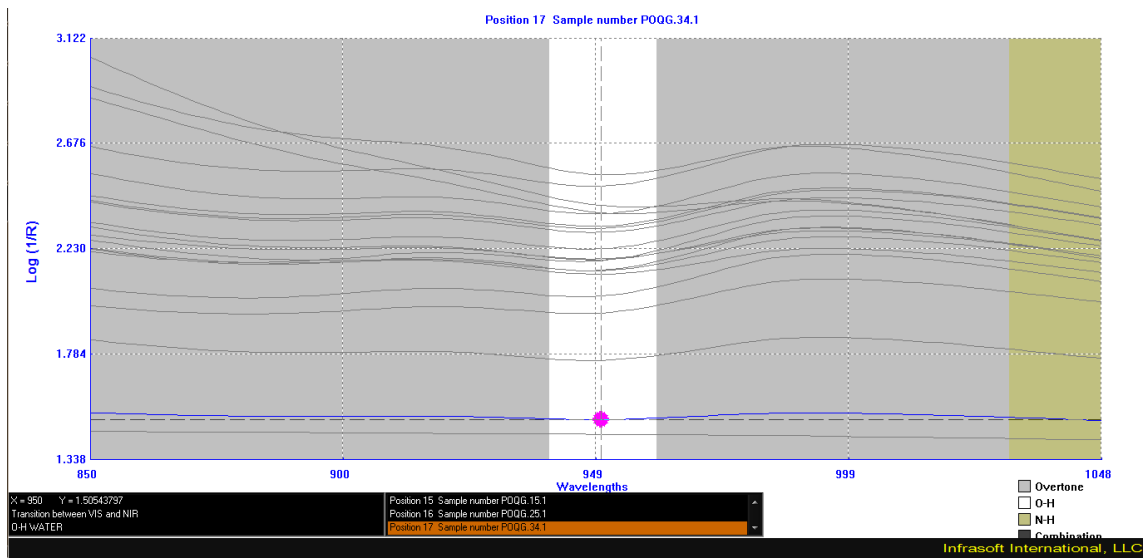


Figura 33: Espectro anómalo, perteneciente a la muestra POQG.34.1 del colectivo de validación de granos.

FUENTE: Imagen obtenida del WinISI.

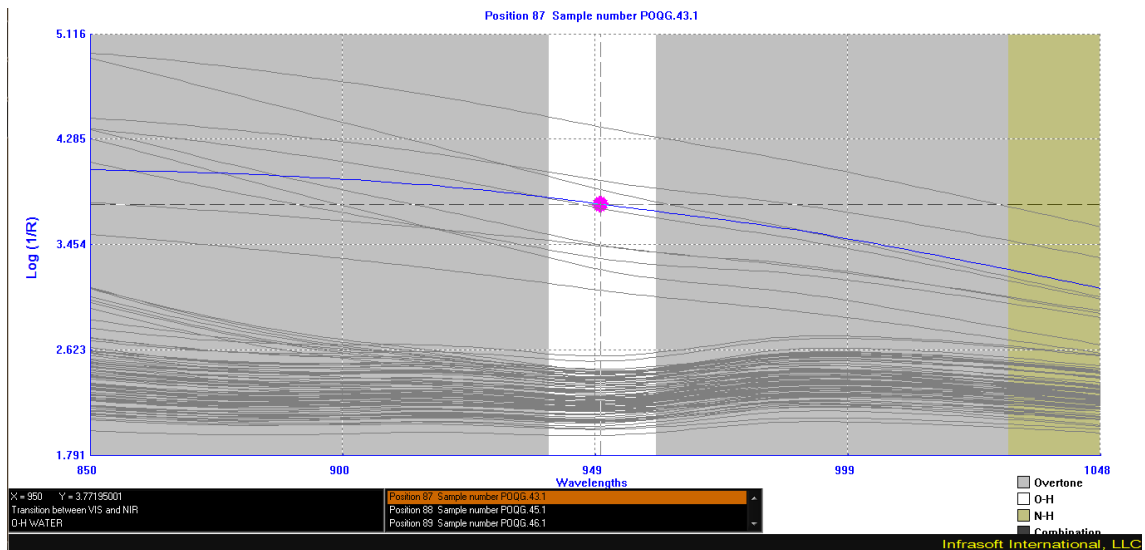


Figura 34: Espectro anómalo, perteneciente a la muestra POQG.43.1 del colectivo de calibración de granos.

FUENTE: Imagen obtenida del WinISI.

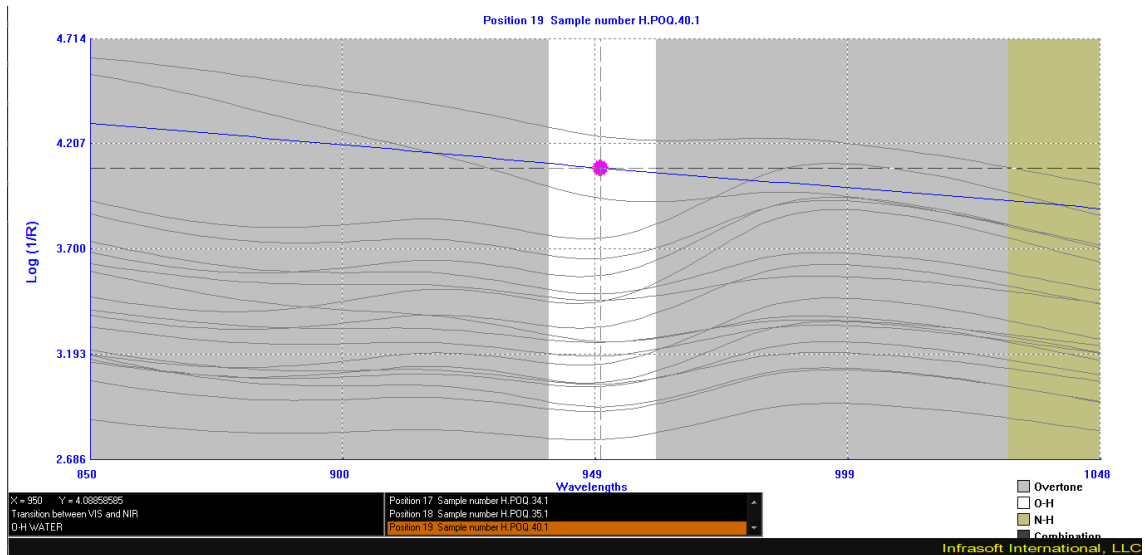


Figura 35: Espectro anómalo, perteneciente a la muestra H.POQ.40.1 del colectivo de validación de harinas.

FUENTE: Imagen obtenida del WinISI.

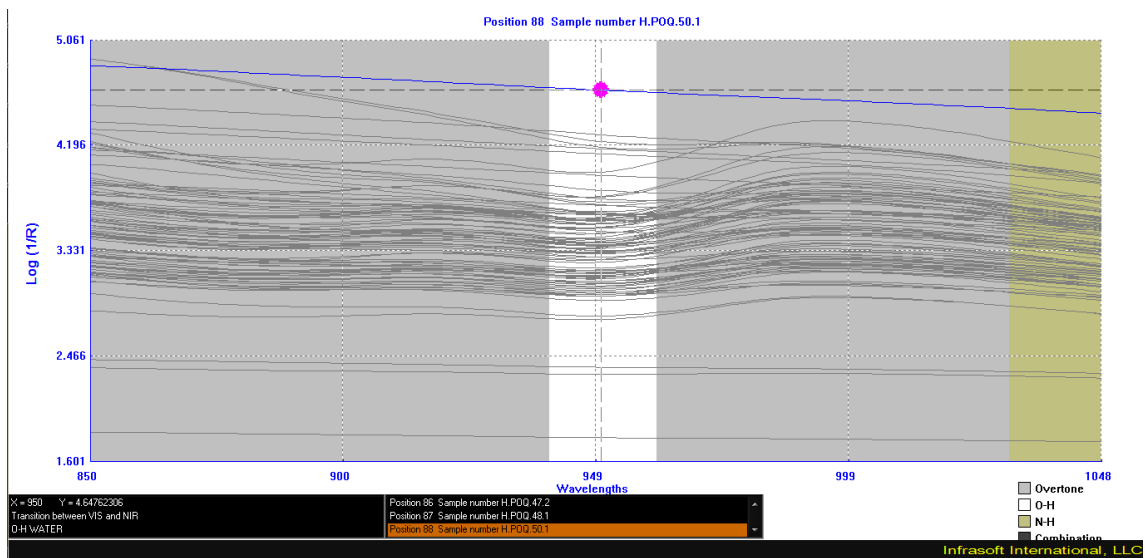


Figura 36: Espectro anómalo, perteneciente a la muestra H.POQ.50.1 del colectivo de calibración harinas.

FUENTE: Imagen obtenida del WinISI.

Anexo 5: Glosario NIRS

- 1-VR: Según Boisman y Parra (2012), es un término de la validación cruzada RQS-tipo que explica cuánto de la variación constitutiva es explicada por la ecuación de la calibración. Si 1-VR es = 1, significa que el 100 por ciento de la variación de los componentes en las muestras de la calibración son explicados por la ecuación durante el proceso de validación cruzada (ver ecuación 29).
- Bias:($\hat{\epsilon}$) Según la ISO 12099 (2010), es la diferencia entre la media del valor de referencia \hat{y} y la media del valor predicho por el modelo NIRS \bar{y} (ver ecuación 31).
- Bias (Límite): Máximo valor que puede obtener el bias.
- CV: Según Vargas (1995), es la representación del coeficiente de variación medio. Se trata de una cantidad sin dimensión, puesto que, al efectuar el cociente, se eliminan las unidades, y nos indica el número de veces que la desviación típica contiene la media. Se toma como absoluto el denominador, ya que se trata de coeficientes de variación positivos. Ya que carece de dimensión, es posible realizar comparaciones entre las dispersiones de dos distribuciones expresadas en distintas unidades.
- Media: El promedio de los datos de referencia del laboratorio.
- RER: Es el ratio del rango ($V_{\text{mayor}} - V_{\text{menor}}$) de las muestras de validación sobre el error típico de predicción–SEP-.
- RPD: Según Burns y Ciurczak (2007), es definido como la Diferencia porcentual relativa. Es el ratio de la desviación estándar –SD- de las muestras de validación sobre el error típico de predicción –SEP- .
- RSQ (R^2): Es la correlación de los datos de referencia con los datos predichos. Según Boisman y Parra (2012), describe la cantidad de variación en el sistema de calibración explicado por la calibración. Un RSQ=1 significa que el 100 por ciento de la variación constitutiva en las muestras de la calibración es explicado por la ecuación. Según la ISO 12099 (2010), es el cuadrado del coeficiente de correlación múltiple entre los valores predichos y valores de referencia (ver ecuación 12).
- SD: Desviación estándar, según Vargas (1995) es la raíz cuadrada positiva de la

varianza y viene en unidades de medida de los datos originales. Es una medida de dispersión, ya que se basa en el promedio de las desviaciones a la mediana o la media. Desviación Estándar de los datos obtenidos por el método de referencia.

- *SEC (Standard Error of Calibration)*: El error estándar de la calibración. Según la ISO 12099 (2010), es una expresión de la diferencia promedio entre los valores predecidos y los valores de referencia para las muestras usadas para desarrollar el modelo. Según Burns y Ciurczak (2007), es la desviación estándar de los residuales de la diferencia de valores entre el valor obtenido por vía química y los valores predecidos por NIRS dentro del set de calibración. Es un indicador del error residual total de la ecuación de regresión que se está utilizando (ver ecuación 27).
- *SECV (Standard Error of Cross Validation)*: Es el error estándar de la validación cruzada. Según la ISO 12099 (2010), para el modelo de calibración, es una expresión del bias corregido de la diferencia promedio entre los valores predecidos y valores de referencia para el subgrupo de las muestras seleccionadas como muestras de predicción durante el proceso de validación cruzada. Estimador verdadero del error de predicción, usualmente mayor al SEC (ver ecuación 28).
- *SEL*: Error estándar de laboratorio. Según Dardenne (2008), el SEL puede ser el error de repetibilidad o error de reproducibilidad (si se tiene más de un laboratorio, uno puede usar la prueba de SEL de anillos) (ver ecuación 11).
- *SEP (Standard Error of Prediction)*: Error estándar de predicción. Según la ISO 12099 (2010), es también llamado desviación estándar de los residuales, expresa la exactitud de los resultados de rutina NIRS corregidos por la diferencia promedio (bias) entre el análisis de rutina NIR y el método de referencia (ver ecuación 30).
- *SEP(C) (Standard Error of Prediction corrected by the bias)*: Error estándar de predicción corregido por el Bias. Según la ISO 12099 (2010), expresión de la diferencia promedio corregida por el bias de los valores predecidos y de referencia, cuando son predecidos por un modelo de regresión del cual no han sido parte de su desarrollo (ver ecuación 32).
- *SLOPE*: Según la ISO 12099 (2010), es la línea de regresión. Representación de la cantidad Y que aumenta por un aumento en X. Dónde el dato de referencia es la

variable dependiente y el valor NIR predicho la variable independiente, para evaluaciones del modelamiento de ecuaciones.