

Josep Lluís Ramos Nadal

**PREPARACIÓ DELS MODELS DE CALIBRACIÓ MULTIVARIANT PER A
DIVERSOS TIPUS DE FRUITA SECA**

TREBALL DE FI DE GRAU

**dirigit per la Sra. Betlem Ripoll Sánchez
de l'empresa Unió Corporació Alimentària SCCL**

Grau de Química



UNIVERSITAT ROVIRA I VIRGILI

Tarragona

2015

ÍNDIX DE CONTINGUTS

1. ABSTRACT	5
2. OBJECTIU.....	6
3. INTRODUCCIÓ	7
4. FONAMENTS	8
4.1. Espectroscòpia NIR.....	8
4.2. Humitat.....	9
4.3. Matèria grassa	9
4.4. Àcid linoleic	10
4.5. Determinació de la humitat.....	10
4.6. Determinació del contingut de matèria grassa	11
4.7. Determinació del percentatge d'àcid linoleic	12
4.8. Calibració multivariant PLS	12
4.8.1. Selecció del nombre components principals	13
4.8.2. Selecció dels rangs espectrals.....	14
4.8.3. Pretractament espectral.....	14
4.8.4. Mètodes de validació.....	14
5. PART EXPERIMENTAL	16
5.1. Calibració del NIR.....	16

5.1.1. Visualització d'espectres	17
5.1.2. Assignació dels valors als espectres	18
5.1.3. Anàlisi de components principals (PCA)	18
5.1.4. Calibració del mètode	19
5.2. Validació del mètode.....	23
6. RESULTATS I DISCUSSIÓ	24
6.1. Mètodes de referència.....	24
6.2. Espectres complets sense pretractament	25
6.3. Resultats del PCA	25
6.4. Paràmetres òptims	26
6.5. Resultats de validació i exclusió d' <i>outliers</i>	29
6.6. Predicció de noves mostres	30
7. CONCLUSIONS	32
8. BIBLIOGRAFIA	33
ANNEX 1: Procediment per la determinació de la humitat	34
ANNEX 2: Procediment per la determinació de la matèria grassa	35
ANNEX 3: Procediment per la determinació de l'àcid linoleic	36

1. ABSTRACT

Near Infrared (NIR) reflectance spectroscopy has been established itself as an important analytical technique in the field of food and agriculture. In this case, it is used as a rapid and non-destructive method to determine the content of many parameters in almonds, hazelnuts and their elaborated products. These parameters are moisture, fat content and linoleic acid.

Firstly, an introduction about NIR spectroscopy is shown, followed by an explanation of the parameters that will be determined. Also the importance of the determination of these parameters in food industry is mentioned.

Next, the reference methods and Partial Least-Squares (PLS) concepts are explained. To determine the moisture in solid samples a Halogen Moisture Analyzer is used and for paste samples an oven at 105°C is used. Fat content in solid samples is determined by Nuclear Magnetic Resonance (NMR) with a Bruker NMS Minispec and for paste samples an extraction with petroleum ether is used. Linoleic acid composition is measured by Gas Chromatography (GC), for all samples. These procedures are annexed at the end of this report.

Then practical part is shown, followed by the results and their discussion. A total of 41 samples (36 calibration samples and 5 unknown samples) are scanned in the reflectance mode. Calibration equations for each parameter are developed using the regression method of PLS with internal cross validation.

Finally conclusions are shown, with the appropriate reflections.

2. OBJECTIU

L'objectiu principal d'aquest treball és la realització dels models de calibració multivariant i posada a punt del NIR de l'empresa Unió Corporació Alimentària de Reus.

Els models de calibració multivariant són per a un aparell FT-NIR Tango Sphere de la marca comercial Bruker, amb el qual es volen determinar tres paràmetres diferents (la humitat, el contingut de matèria grassa i el percentatge d'àcid linoleic) de diverses mostres d'avellana, ametlla i els seus elaborats.

La determinació a través del NIR permetrà una ràpida obtenció de resultats utilitzant una sola tècnica, sense l'ús de reactius químics i sense la generació de residus.

3. INTRODUCCIÓ

La realització d'aquest treball ha estat a l'empresa Unió Corporació Alimentària de Reus. La determinació a través de les tècniques tradicionals del contingut de matèria grassa i del percentatge d'àcid linoleic s'han fet a la planta de Vilallonga del Camp.

Les mostres que es volen analitzar són d'avellana i d'ametlla en diversos formats, els quals són: natural (amb pell), repelat (només per ametlles), torrat i pasta. La pasta, tant d'ametlla com d'avellana, és el producte resultant de la molturació de la fruita seca torrada fins a l'estat de líquid dens, el qual posteriorment es fa passar per una refinadora per obtenir la textura cremosa desitjada i amb una mida de partícules baixa (granulometria ≤ 35 micres). A la pasta no s'addiciona sucre ni cap altre producte.

Per a fer les posteriors determinacions dels paràmetres utilitzarem l'espectroscòpia NIR. Com que aquesta tècnica ens dóna molta informació per a cada mostra, el mètode de calibració ideal serà la calibració multivariant.

L'espectroscòpia NIR és molt utilitzada dins de la indústria alimentària ja que s'obtenen resultats ràpidament i sense l'ús de reactius. La rapidesa és important perquè, com a qualsevol indústria, és necessària l'obtenció de resultats de manera gairebé immediata. Evitar l'ús de reactius també és molt important, ja que no s'altera la mostra analitzada i es pot tornar a utilitzar per altres anàlisis.

4. FONAMENTS

Primerament cal fer una breu introducció sobre el fonament i la importància de l'espectroscòpia d'infraroig proper (NIR). Després veurem la importància de conèixer perquè es determinen els paràmetres a analitzar dins de la indústria alimentària, i després els mètodes de referència que s'han utilitzat per a conèixer els valors dels paràmetres.

4.1. Espectroscòpia NIR

La tecnologia NIR involucra llum interactuant amb un material, on es dóna una radiació electromagnètica en forma d'ones. L'infraroig proper és la regió de l'espectre situada entre la regió visible i la regió infraroja. El rang de longituds d'ona que el NIR cobreix va de 750 a 2600nm.

A les molècules, les unions específiques entre els àtoms vibren a una certa freqüència, i cada tipus d'aquestes unions químiques dins d'una mostra absorbeixen raigs de la regió del NIR d'una longitud d'ona específica, mentre que tota la resta de longituds d'ona són reflectides [1]. Com a resultat d'aquestes interaccions entre els rajos i les freqüències de vibració s'obté un espectre característic, el qual a través de l'assignació de les bandes a sobretons i combinacions, de vibracions fonamentals involucrant modes de tensió hidrogenoide, permet identificar els diferents grups funcionals i molècules de la mostra [2].

Hi poden haver tres tipus diferents de mesura per a l'obtenció dels espectres mitjançant un aparell NIR: de transmissió, de reflectància difusa i de transflectància. La mesura per transmissió calcula la fracció de radiació transmesa a través de la mostra, la de reflectància difusa calcula la fracció de radiació reflectida sobre el total incident a la mostra, i la de transflectància és una combinació de les dues anteriors, ja que calcula les fraccions de radiació transmesa i reflectida alhora [3].

A partir dels espectres obtinguts de mostres de concentració coneguda dels paràmetres a determinar i mitjançant tècniques quimiomètriques d'anàlisi, es pot obtenir el model de calibració per a identificar posteriors mostres problema.

El principal factor per a la implementació de l'espectroscòpia NIR és la rapidesa de l'anàlisi. A la indústria alimentària pot ser utilitzada per a la gradació de qualitat de la matèria primera, anàlisi d'ingredients per a la formulació de l'aliment, i verificació de la qualitat del producte final, entre d'altres. No es requereixen habilitats especials o preparació de mostres, els instruments són fàcils d'operar i els resultats estan

disponibles immediatament. No obstant tot això, la preparació de la calibració i especialment els programes de manteniment d'una calibració apropiada són essencials per assegurar resultats fiables continus.

4.2. Humitat

La humitat és un dels paràmetres principals que cal determinar en qualsevol fruita seca, ja que a percentatges elevats s'afavorirà el desenvolupament de fongs. Aquesta aparició de fongs afavoreix el procés de podrit de la fruita. La presència d'un tipus determinat de fongs, del gènere *Aspergillus*, pot generar aflatoxines. Aquest tipus de toxines posseeixen activitat mutàgena i carcinògena que contribueix al risc de patir càncer hepàtic [4]. La principal aflatoxina que pot aparèixer en la fruita seca és l'aflatoxina B₁, l'estructura de la qual podem veure a la Figura 1.

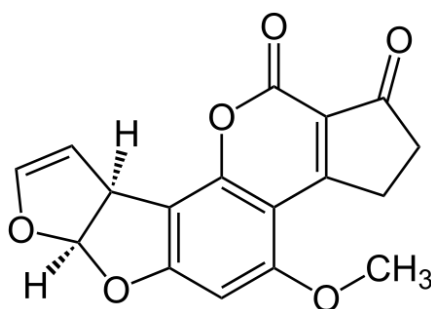


Figura 1. Aflatoxina B₁

4.3. Matèria grassa

Un altre paràmetre a tenir en compte és el percentatge de matèria grassa, tot i que en aquest cas és merament informatiu. Tant l'avellana com l'ametlla tenen un alt percentatge de matèria grassa, més d'un 50% del seu pes, la qual està composta per compostos lipídics. Aquests compostos estan constituïts en la seva majoria per àcids grassos insaturats els qual donen un alt valor nutricional i ajuden a la prevenció de diverses malalties, especialment aquelles relacionades amb el sistema cardiovascular [5].

El contingut de matèria grassa està directament relacionat amb altres paràmetres que es poden determinar, com ara l'acidesa, l'índex de peròxids, l'estabilitat i el percentatge d'àcid linoleic, entre d'altres.

4.4. Àcid linoleic

La determinació del percentatge d'àcid linoleic de la fruita seca té un caràcter informatiu referent a la capacitat d'enranciment d'aquestes. Tot i que l'àcid gras majoritari present en avellanes i ametlles és l'àcid oleic (al voltant del 70%), es determina l'àcid linoleic degut a la seva rapidesa per oxidar-se, ja que és més susceptible a l'oxidació degut a que té més insaturacions. L'àcid oleic és monoinsaturat mentre que el linoleic és poliinsaturat, com es pot veure a la Figura 2.

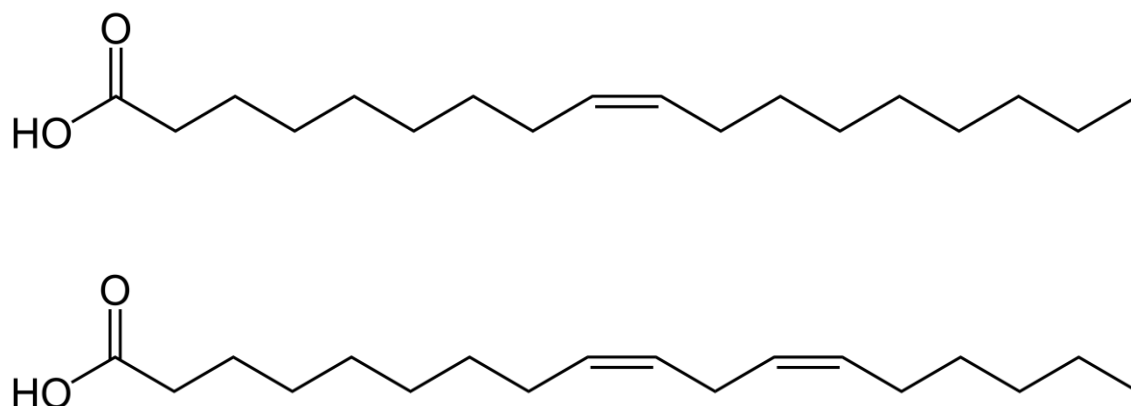


Figura 2. Àcid oleic (a dalt) i àcid linoleic (a baix)

Així doncs sabent la concentració d'àcid linoleic present en una mostra es pot fer una aproximació de la bona qualitat del producte al llarg del temps. Per exemple, si tenim una mostra amb un alt percentatge d'àcid linoleic és probable que es comenci a enrancir més aviat i per tant s'haurà de consumir en un període de temps més curt.

Per a poder crear un bon model de calibració del NIR primer s'han de conèixer els valors dels diferents paràmetres a identificar de les mostres que s'utilitzaran per a fer la calibració. Aquests valors es relacionaran amb els espectres obtinguts amb el NIR per tal de crear el model multivariant. Els valors de referència han de ser el més acurats possible, ja que si no ho són no es podrà construir un bon model. Per això cal utilitzar mètodes de referència, els quals es comenten a continuació:

4.5. Determinació de la humitat

Per determinar la humitat d'una mostra s'utilitzen dues metodologies diferents, depenent si la mostra és sòlida o és en forma de pasta.

Per mostres sòlides s'usa un Analitzador Halogen d'Humitat. Aquest analitzador es basa en el mètode tèrmic, el qual calcula la humitat de la mostra a partir de la pèrdua de pes en aplicar temperatura, en aquest cas amb una làmpada halògena.

L'analitzador que s'utilitza és el HB43 de la marca comercial Mettler Toledo, com el que es mostra a la Figura 3.



Figura 3. Analitzador Halogen d'Humitat HB43 de Mettler Toledo

Per les mostres de pasta es fa amb el mètode tradicional per a la determinació de la humitat mitjançant el mètode tèrmic, és a dir, també es calcula la pèrdua de pes de la mostra en aplicar temperatura però en aquest cas es fa de forma manual, posant la mostra a una estufa a 105 °C i anar pesant fins a tenir pes constant.

4.6. Determinació del contingut de matèria grassa

Com en el cas anterior, s'utilitzen dues metodologies diferents depenent si la mostra és sòlida o en forma de pasta.

Per mostres sòlides s'usa un aparell de ressonància magnètica per a detectar nuclis d'hidrogen. En les nostres mostres el nuclis d'hidrogen que ens pot detectar són els dels àcids grassos i els de l'aigua, per tant prèviament les mostres s'han d'assecar per a eliminar la senyal que donaria l'aigua. L'aparell que s'utilitza és el NMS 110 Minispec de la marca comercial Bruker, com el que es mostra a la Figura 4.



Figura 4. NMS 110 Minispec de Bruker

Per les mostres de pasta es determina la matèria grassa a través de l'extracció de l'oli de la mostra amb d'èter de petroli.

4.7. Determinació del percentatge d'àcid linoleic

Per a poder determinar aquest paràmetre s'utilitza la cromatografia de gasos per a tots els tipus de mostres.

A partir d'una petita quantitat d'oli de la mostra s'extreuen els àcids grassos amb un dissolvent apolar i posteriorment es converteixen en èsters amb KOH. Ens interessa tenir els àcids grassos en forma d'èsters perquè la columna que s'usa és específica per a separar èsters metílics (columna HP-88). Finalment, en el cromatograma s'obtenen els percentatges dels diferents àcids grassos presents a la mostra, tot i que només es tindrà en compte el percentatge de l'àcid linoleic.

Tot seguit veurem una introducció sobre la calibració multivariant, la qual és necessària per a la creació del nostre model, i els factors principals que s'hauran de tenir en compte per a dur-la a terme.

4.8. Calibració multivariant PLS

En el nostre cas utilitzem una tècnica concreta de calibració multivariant, la Regressió per Mínims Quadrats Parcial (PLS). Aquest tipus de tècnica és capaç de fer una relació entre el senyal instrumental i la propietat a determinar, i posteriorment fer una predicció d'aquesta propietat a partir del senyal mesurat en una mostra desconeguda. En el nostre cas les propietats seran la humitat, la matèria grassa i l'àcid linoleic, i el senyal mesurat serà l'espectre obtingut amb el NIR.

El PLS es basa en que la informació continguda en les variables mesurades es pot concentrar en un nombre menor de variables sense perdre informació rellevant. La regressió es realitza sobre les noves variables, eliminant així els problemes de col·linealitat que puguin tenir les dades originals.

Les variables espectrals són col·lectades com una matriu X , amb un nombre de files igual al nombre de mostres i un nombre de columnes igual al nombre de canals de l'espectre. La informació de la propietat a determinar anirà continguda a la matriu Y . Ambdues matrius es descomponen en una suma de factors, com podem observar a les següents fórmules:

$$X = TP' + E = \sum_{a=1}^A t_a p_a^t + E$$

$$Y = UQ' + F = \sum_{a=1}^A u_a q_a^t + F$$

on T és la matriu de *scores*, P la de *loadings* i E la de residuals, corresponents a la matriu X, i U és la matriu de *scores*, Q la de *loadings* i F la de residuals, corresponents a la matriu Y.

La descomposició de les dues matrius es realitza de manera simultània, establint una relació interna entre els *scores* dels blocs X i Y, de manera que per a cada component es compleixi la següent fórmula:

$$u_a = b_a t_a$$

on b_a és el coeficient de regressió entre *scores* dels blocs X i Y per a cada un dels components.

Un cop establert el model de calibració, es pot realitzar la predicció de la propietat modelada per a noves mostres. A partir dels espectres originals, utilitzant els valors calculats per les mostres de calibració, es calculen els *scores* de les mostres a predir (t^*) i a partir de la següent fórmula s'obté el valor de la propietat a determinar [6]:

$$Y^* = \sum_{a=1}^A b_a t_a^* q_a^t$$

Durant la construcció de models de calibració basats en la reducció de variables és important examinar els següents factors:

4.8.1. Selecció del nombre components principals

La selecció del nombre de components principals o factors que configura el model és un punt clau en la utilització de tècniques de reducció de variables, ja que el model ha d'incloure la informació rellevant però ha d'excloure el soroll. L'ús d'un nombre insuficient de components principals resulta una pobre reproducció de les dades espectrals, i per tant el model no serà capaç de reconèixer canvis en les característiques espectrals. Per altra banda, incloure massa components principals només afegeix soroll spectral a la regressió i no incrementa la quantitat d'informació útil [7].

4.8.2. Selecció dels rangs espectrals

Per a una millor identificació dels paràmetres que es volen determinar s'han de conèixer els rangs espectrals en els quals es donarà senyal. També s'ha de tenir en compte que hi poden haver contribucions de soroll spectral o de bandes d'absorció de components interferents que poden deteriorar la qualitat del model. En aquests casos serà recomanable seleccionar els rangs espectrals adients per a minimitzar les interferències, i així aconseguir un model més òptim [7].

4.8.3. Pretractament spectral

Per a que l'algoritme PLS pugui establir una bona correlació entre l'espectre i la dada de concentració, és possible que calgui fer un pretractament del conjunt d'espectres per a que siguin comparables entre ells. Els pretractaments més habituals entre d'altres són: la correcció de la línia base, la normalització vectorial, la primera i segona derivada, etc.

4.8.4. Mètodes de validació

La validació és la fase més important en la construcció dels model de calibració, ja que t'assegura si el model és bo per a predir els valors dels paràmetres a determinar.

Hi ha dos mètodes per a validar la calibració: la validació del conjunt de test (validació externa) i la validació creuada (validació interna).

La validació del conjunt de test defineix dos conjunts de dades diferents: un conjunt de calibració, el qual serà per desenvolupar el mètode, i l'altre de test, el qual serà per validar el mètode. Generalment s'utilitza quan tenim conjunts de dades amb moltes mostres. En el nostre cas, pel nombre de mostres amb el qual fem la calibració utilitzarem la validació creuada.

La validació creuada divideix el conjunt de mostres en segments, els quals estan formats per una sola mostra. Aquest procediment va deixant cada vegada una mostra fora del conjunt i utilitza la resta de mostres per a calcular el model, repetint aquest procés fins que totes les mostres han estat excloses i s'ha predit el seu paràmetre a determinar, tal i com es pot observar a la Figura 5.

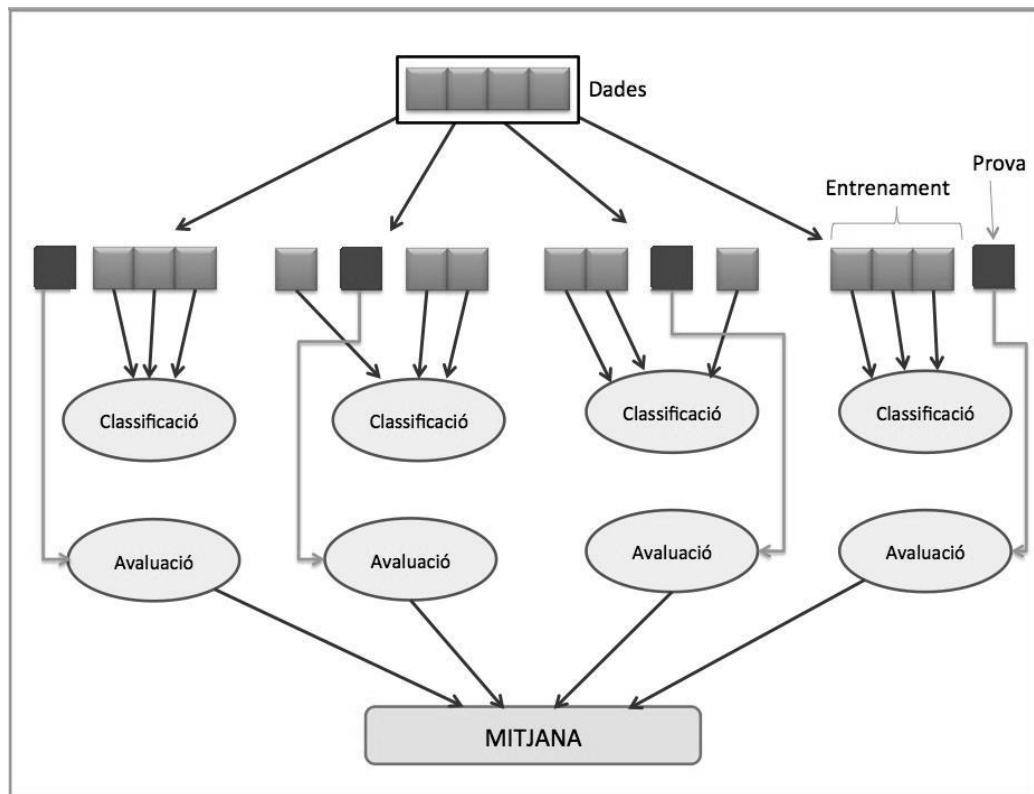


Figura 5. Validació creuada

Un cop finalitzada la validació creuada obtenim l'error de l'anàlisi resultant, anomenat *Root Mean Square Error of Cross Validation* (RMSECV), el qual es calcula a partir de la següent fórmula:

$$RMSECV = \sqrt{\frac{1}{M} \sum_{i=1}^M (Y_i^{mes} - Y_i^{pred})^2}$$

on M és el nombre de mostres, Y^{mes} és la concentració coneguda i Y^{pred} és la concentració predita. Aquest error és l'equivalent a l'error mitjà de predicció, i per tant interessa que aquest valor sigui el més baix possible per a que les prediccions siguin correctes.

A continuació veurem l'aplicació del PLS en la part experimental.

5. PART EXPERIMENTAL

Per a dur a terme la part experimental, primerament veurem la llista de reactius utilitzats en els mètodes de referència anomenats anteriorment. Els podem observar a la Taula 1. Per a la manipulació de tots els reactius s'han utilitzat els equips de protecció adequats: ulleres de seguretat, bata, guants i campana extractora.

Reactius	Marca comercial	Toxicitat
Èter de petroli ($\geq 90\%$)	Sigma-Aldrich	Fàcilment inflamable, nociu, irritant i perillós pel medi ambient
n-Heptà ($\geq 99\%$)	Sigma-Aldrich	
n-Hexà (95%)	J. T. Baker	
Hidròxid de potassi ($\geq 85\%$)	Sigma-Aldrich	Corrosiu i nociu

Taula 1. Llistat de reactius

Els procediments emprats per a la determinació de la humitat, la matèria grassa i l'àcid linoleic dels mètodes de referència els podem trobar al final del treball en forma d'annexos.

Tot seguit veurem el procediment utilitzat per a la creació dels models de calibració per al nostre NIR.

5.1. Calibració del NIR

L'aparell pel qual els models de calibració han estat creats és un FT-NIR Tango Sphere de la marca comercial Bruker, com el que es mostra a la Figura 6.



Figura 6. FT-NIR Tango Sphere de Bruker

El software pel tractament dels espectres i l'anàlisi quimiomètrica utilitzat és l'Opus Spectroscopy Software, versió 7.5. Aquest programa és exclusiu de la marca comercial Bruker.

Primerament s'ha obtingut un conjunt de mostres prou representatiu per a que el model de calibració sigui efectiu. S'ha creat el model amb diverses mostres de cada tipus de producte, tenint un total de 36 mostres. Les mostres sòlides han estat molturades abans de fer les anàlisis i lectures pertinents.

Tot seguit s'ha procedit a fer la lectura de les mostres amb el NIR per l'obtenció dels seus espectres, mitjançant la mesura per reflectància difusa. Per a cada mostra s'ha fet la mesura per triplicat. Els paràmetres òptims per a la lectura venen descrits a la Taula 2.

Paràmetre	Valor
Scans	32
Resolució	8 cm ⁻¹
Rang espectral	11540 – 3952 cm ⁻¹

Taula 2. Paràmetres òptims NIR

Els paràmetres de lectura triats són els recomanables pel tipus de mostra que es vol determinar, ja que donaran uns bons espectres amb un temps ràpid d'anàlisi (30 segons).

Un cop feta la lectura de les mostres, s'han importat els espectres obtinguts al programa OPUS per realitzar l'anàlisi quimiomètrica. Per la creació del model de calibració s'han seguit una sèrie de passos els quals els veurem a continuació. Les figures que es poden veure en els següents apartats són resultats reals, però es comentaran amb més detall a l'apartat 6.

5.1.1. Visualització d'espectres

Per fer una primera revisió i veure si hi ha errors de lectura de les mostres és recomanable fer una visualització dels espectres obtinguts, tant individuals com en conjunt. D'aquesta manera es pot veure si els espectres tenen la mateixa forma o si hi ha algun espectre que sigui molt diferent a la resta. En la Figura 7 podem veure la visualització dels nostres espectres, sense realitzar cap pretractament.

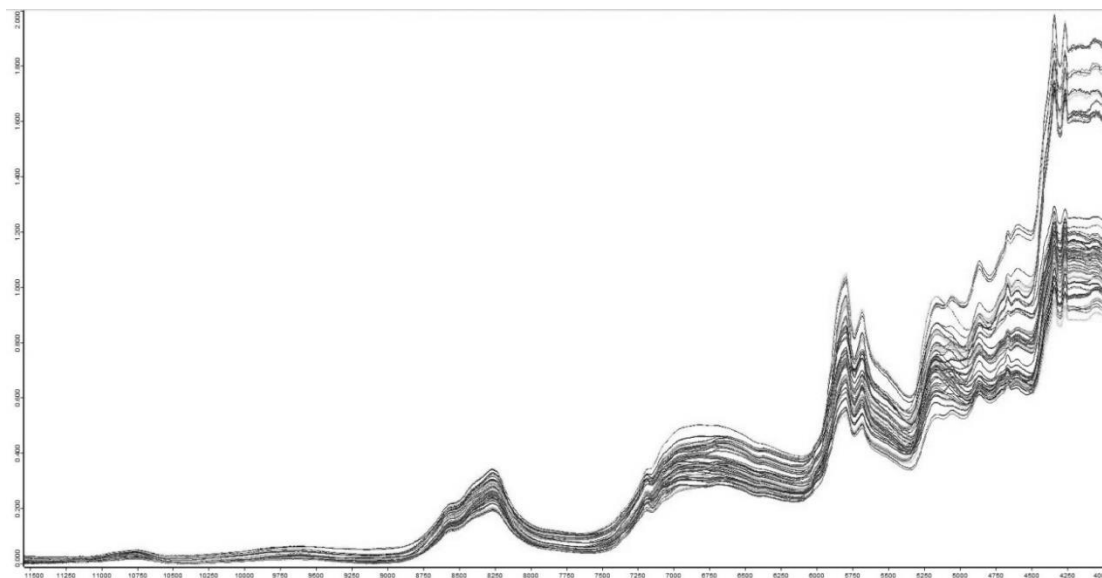


Figura 7. Espectres de les mostres

5.1.2. Assignació dels valors als espectres

Un cop es tenen els espectres classificats i identificats en forma de taula, es procedeix a assignar els valors dels paràmetres determinats amb els mètodes de referència, també anomenats “components”, a cada espectre.

Quan ja es té la taula completa, abans de començar a fer el tractament de dades, es pot fer la mitjana del triplicat d'espectres. En el nostre cas és recomanable ja que hi pot haver una lleugera diferència entre espectres de la mateixa mostra, ja que les nostres mostres no són del tot homogènies.

5.1.3. Anàlisi de components principals (PCA)

Per veure si hi ha una classificació de les mostres o si n'hi pot haver alguna de discrepant, a partir dels espectres obtinguts es pot fer una Anàlisi de Components Principals (PCA). Aquest tipus d'anàlisi no té en compte els valors dels components entrats, és a dir que treballa amb la matriu X que s'ha comentat a l'apartat 4.4. En el nostre cas s'ha fet un PCA amb 5 factors, tenint com a gràfic d'scores resultant dels dos primers PCs el que podem observar a la Figura 8. La informació que conté el PC1 (Score 1) és del 89% i la del PC2 (Score 2) és del 9%.

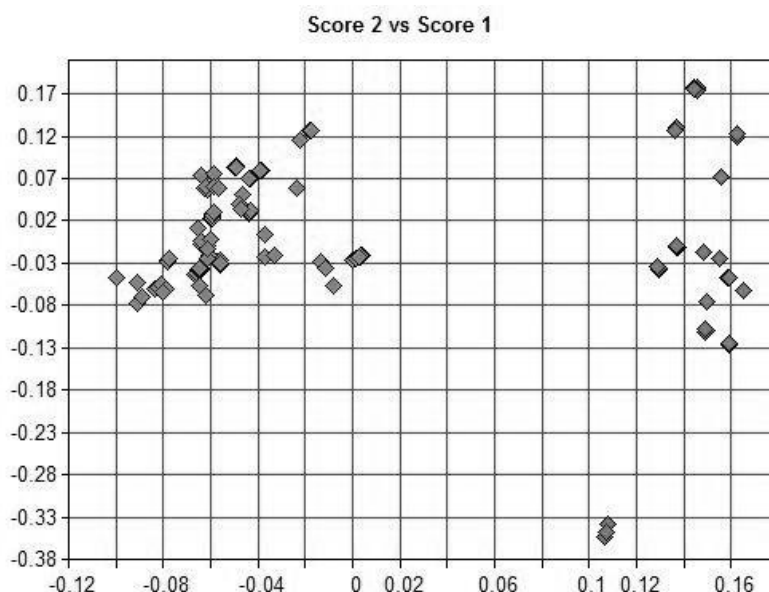


Figura 8. Gràfic d'scores PCA

Observant el gràfic d'scores podem veure que clarament hi ha dos grups diferenciables. En aquest cas els valors de l'esquerra són els de les mostres sòlides i els de la dreta són els de les pastes.

5.1.4. Calibració del mètode

Per a que el nostre mètode sigui el més fiable possible s'han de realitzar els tractaments necessaris als espectres obtinguts. Realitzant els ajustos adients per a cada component que es vol determinar, s'obtidran millors prediccions. Així doncs es realitzaran una sèrie de correccions, les quals s'anomenen a continuació:

- a) Optimització automàtica.** Aquesta opció del nostre software permet fer una ràpida comprovació de com els diferents pretractaments d'espectres afecten a la predicció dels valors de les mostres de calibració. Un cop fetes les comprovacions dels pretractaments en les diferents regions de l'espectre, les quals prèviament s'han seleccionat, el software ens classifica les combinacions de menys a més error (RMSECV) i amb el nombre de factors (PCs) òptims.
- b) Selecció dels paràmetres òptims.** Un cop s'obté la classificació anterior, es seleccionen quins paràmetres poden anar millor pel nostre mètode (pretractament d'espectres i regió d'espectres). Per fer això s'escullen els paràmetres que tenen menys error de predicció, amb un nombre de PCs baix i que el pretractament utilitzat no sigui massa complex. Si el nombre de PCs és massa alt, la informació que ens donarà pot ser que no sigui rellevant ja que tendirà a donar informació sobre el soroll espectral. Si el pretractament és molt complex, pot fer que el model

no sigui del tot precís, ja que tendirà a fer massa correccions en els espectres de calibració. Així doncs cal trobar l'equilibri entre paràmetres per a que les nostres posteriors prediccions siguin el més precises possible.

- c) **Validació.** Un cop seleccionats els paràmetres es procedeix a fer la validació. En el nostre cas s'ha fet una validació creuada, ja que pel nombre de mostres del que es disposen és la més recomanable. Com a resultat de la validació s'obté un gràfic que relaciona el valor del component amb el valor predit, com podem observar a la Figura 9.

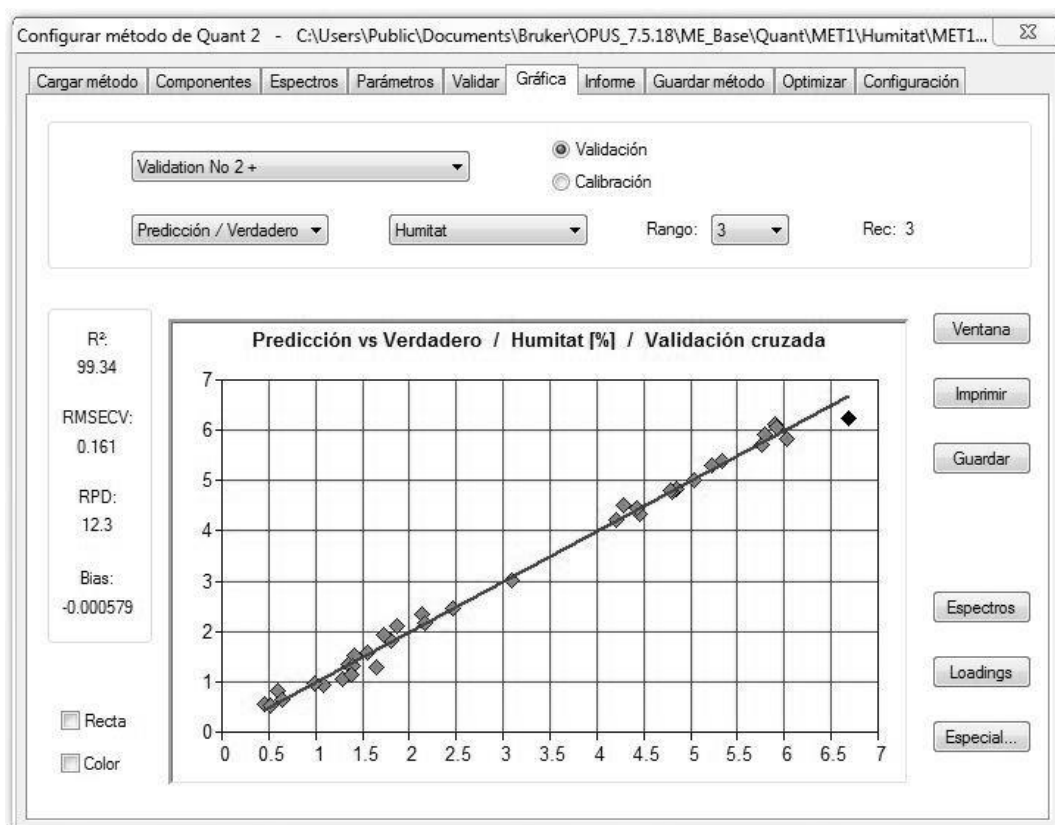


Figura 9. Gràfic predicció/real

En aquest tipus de gràfic es pot identificar amb facilitat si hi ha algun *outlier*, en aquest cas és el punt marcat en negre. També es pot observar el coeficient de determinació (R^2) i la Desviació de Predicció Residual (RPD). El RPD és el valor resultant que ens relaciona la desviació estàndard (SD) amb l'error estàndard de validació (SECV), que en el nostre cas és el RMSECV, a través de la següent fórmula:

$$RPD = \frac{SD}{SECV}$$

A través del valor de RPD es pot classificar el nostre model de la manera que podem veure a la Taula 3.

RPD	Classificació
<1.0	Molt pobre
1.0 – 2.4	Pobre
2.5 – 2.9	Just
3.0 – 3.9	Raonable
4.0 – 5.9	Bo
6.0 – 7.9	Molt bo
8.0 – 10.0	Excel·lent
>10.0	Superior

Taula 3. Classificació del RPD

- d) Exclusió d'outliers.** Finalment, un cop detectat si tenim algun *outlier*, s'exclou la mostra que dóna aquest valor discrepant i es torna a validar. La decisió d'excloure o no un *outlier* dependrà del nostre criteri, ja que pot ser un valor que ens dóna informació rellevant. També dependrà de si el nostre model millora o no, ja que excloure un *outlier* no implica directament una millora del model.
- e) Altres comprovacions.** En el resultat obtingut de la validació, abans d'excloure els possibles *outliers*, hi ha altres tipus de gràfics que també ens poden servir per a determinar si el mètode optimitzat és correcte. Alguns exemples són:
- Diferències respecte valors de referència: la distribució de les desviacions, i especialment el rang entre la desviació mínima i màxima, ajuden a comprovar el model. Podem veure el gràfic a la Figura 10.

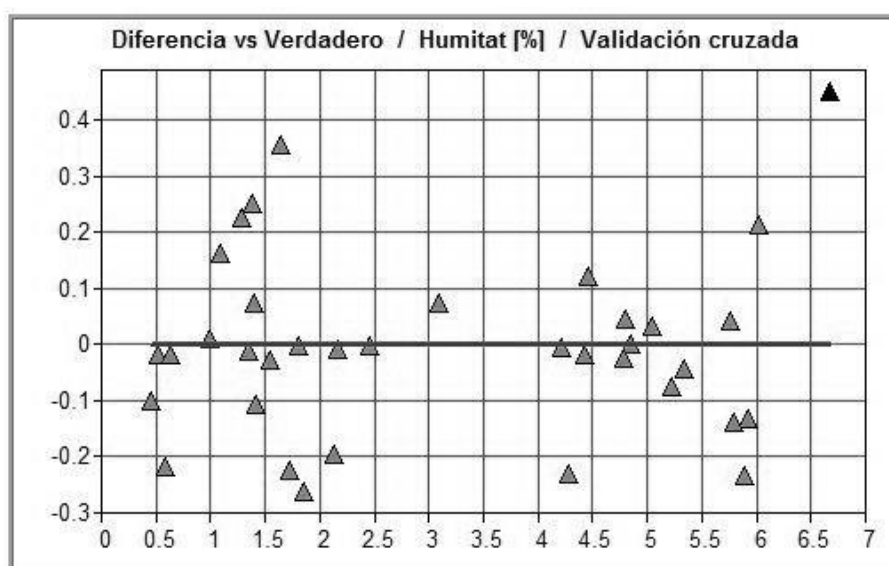


Figura 10. Gràfic diferències/valors referència

- Error respecte nombre de factors: cada factor contribueix amb informació útil per disminuir l'error. Després d'arribar a un mínim d'error, es torna a incrementar (*overfitting*). Podem veure el gràfic a la Figura 11.

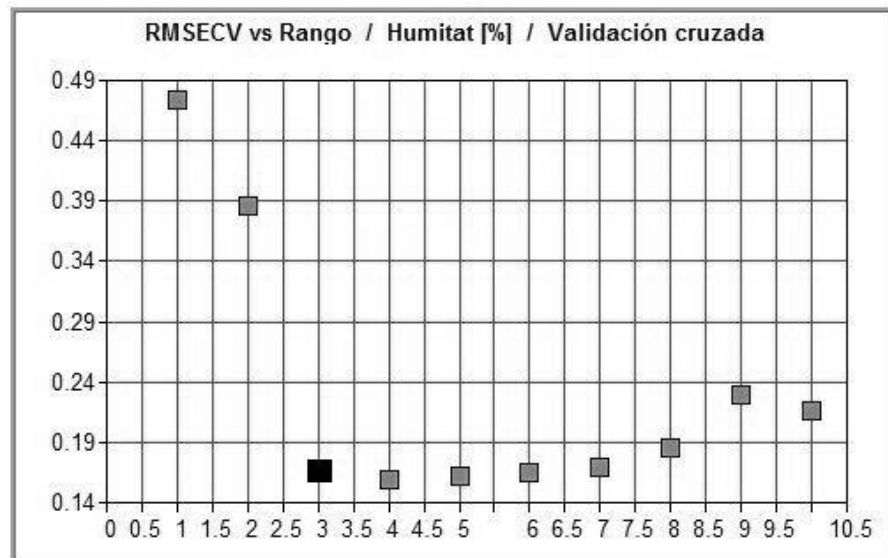


Figura 11. Gràfic error/rang

- Distància de Mahalanobis (MD) respecte els residus espectrals: la distància de Mahalanobis és el valor que ens dóna una idea de la proximitat dels punts entre ells. Així doncs aquest gràfic ens identificarà que els espectres situats a l'extrem superior dret són *outliers* potencials, però no ho seran si els seus valors de referència són molt alts o molt baixos. Podem veure el gràfic a la Figura 12.

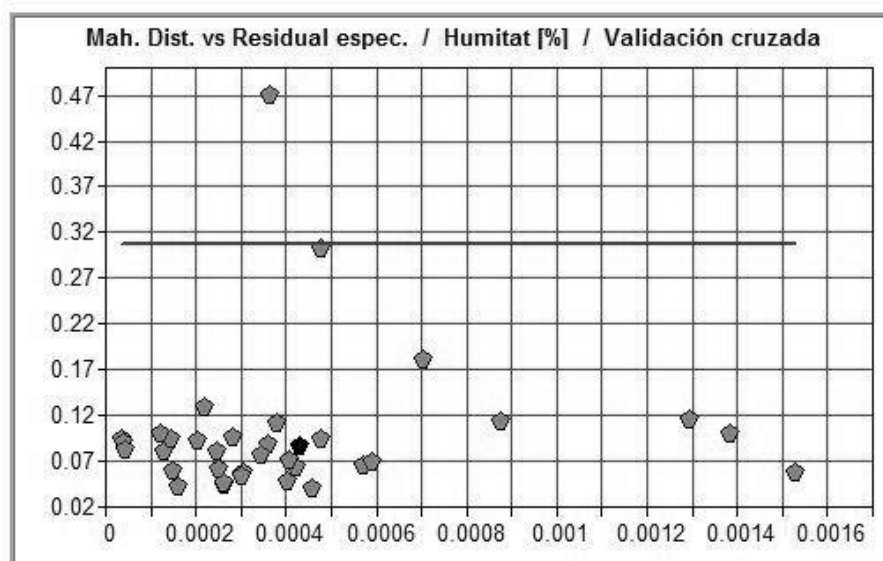


Figura 12. Gràfic MD/residus espectrals

- Gràfic d'scores: també es pot observar el gràfic d'scores, tal i com podem veure a la Figura 13.

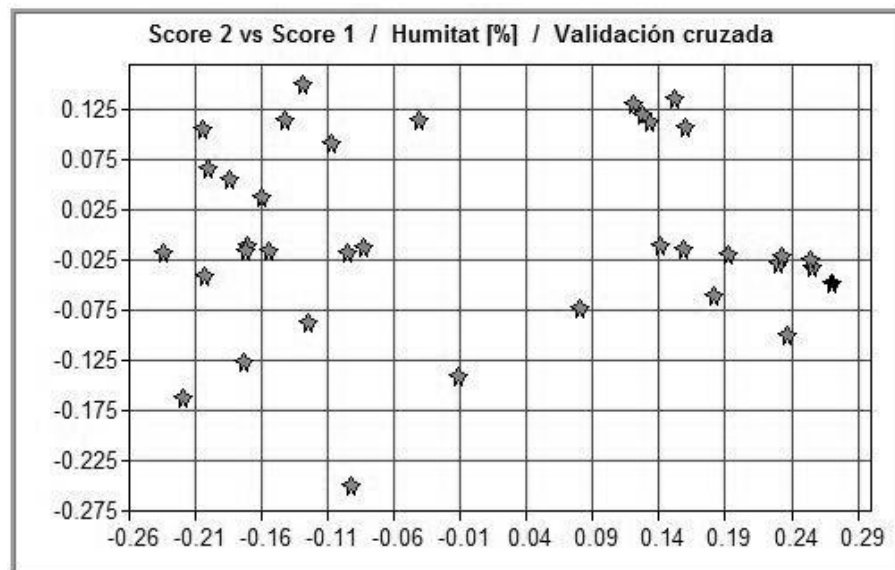


Figura 13. Gràfic d'scores

Un cop realitzades les correccions pertinents per a la creació dels models de calibració, per a cada paràmetre a determinar, s'ha procedit a la validació del mètode.

5.2. Validació del mètode

Per a poder afirmar que el mètode optimitzat és fiable s'han fet lectures de diferents mostres problema dels diversos productes que tenim. Per fer això primer s'ha creat un mètode que agrupa els tres models creats, un per a cada component a analitzar. Aquesta agrupació ens permet obtenir els diferents resultats en la mateixa anàlisi.

Després de la creació del nou mètode s'ha fet la lectura de les noves mostres problema, les quals han estat prèviament analitzades amb els mètodes de referència per a conèixer els valors dels paràmetres que es volen determinar. Un cop feta la lectura de la mostra ens apareixen els valors predits de cada component.

Finalment s'han comparat els valors predits amb el valors reals per veure si estan relacionats.

6. RESULTATS I DISCUSSIÓ

A continuació podem veure els diferents resultats que s'han obtingut al llarg del procediment anteriorment comentat.

6.1. Mètodes de referència

A la Taula 4 podem veure els resultats de la humitat, la matèria grassa i l'àcid linoleic presents en cada mostra, determinats pels mètodes de referència.

Mostra	Humitat (%)	Matèria Grassa (%)	Àcid Linoleic (%)
M1	5,76	55,65	22,02
M2	4,21	49,58	23,67
M3	5,23	55,90	20,96
M4	5,04	56,99	23,52
M5	0,45	59,28	27,89
M6	0,52	58,99	27,40
M7	1,65	61,78	22,39
M8	1,42	62,88	23,43
M9	1,86	51,35	23,27
M10	1,28	66,54	9,87
M11	4,85	56,93	22,90
M12	4,81	64,91	11,21
M13	6,03	56,12	24,88
M14	1,55	66,10	10,22
M15	6,68	54,30	22,30
M16	0,59	68,86	13,29
M17	4,78	65,12	11,63
M18	1,35	66,67	11,60
M19	4,28	65,81	10,32
M20	5,90	54,48	22,10
M21	5,92	54,54	26,51
M22	4,46	66,30	10,32
M23	1,09	57,27	26,02
M24	5,79	56,76	18,80
M25	1,40	60,27	23,18
M26	2,14	56,47	21,14
M27	1,81	63,36	8,21
M28	2,17	57,03	23,07
M29	5,34	60,58	20,60
M30	3,09	55,86	24,31
M31	2,46	66,02	8,23
M32	0,99	68,15	12,95
M33	4,43	66,23	10,91
M34	0,64	72,27	12,44
M35	1,72	51,07	23,26
M36	1,38	56,04	25,84

Taula 4. Resultats pels mètodes de referència

Les mostres marcades de color gris són les corresponents a productes d'avellana i les de blanc són d'ametlla.

Observant els resultats d'àcid linoleic podem veure que les mostres d'avellana tenen un contingut clarament inferior, al voltant del 10%, respecte les mostres d'ametlla, les qual tenen un contingut al voltant del 23%. També podem veure que el contingut de matèria grassa de les avellanes és superior al de les ametlles. Les diferències d'humitat depenen del tractament que a patit la mostra, per exemple les mostres més seques són les que han patit un procés de torrat i les més humides són les naturals.

6.2. Espectres complets sense pretractament

A la Figura 14 podem veure els espectres obtinguts de les mostres, amb la seva posterior mitjana dels triplicats i sense aplicar cap tipus de pretractament.

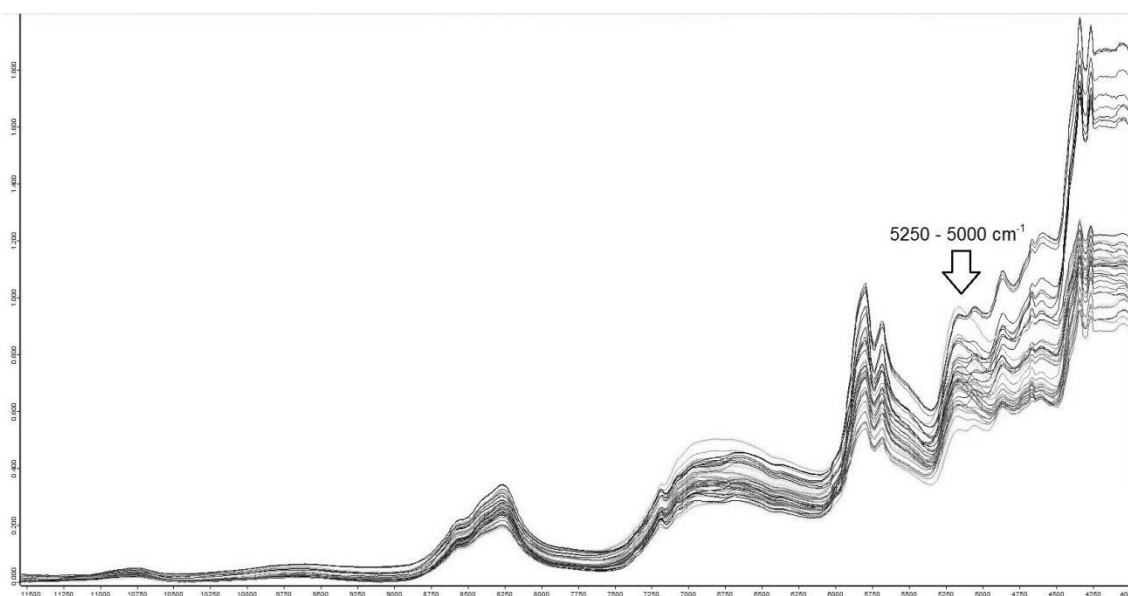


Figura 14. Representació espectral de les mostres

Observant els espectres de les mostres es poden veure les bandes i sobretons característics que ens indiquen el contingut dels paràmetres que es volen analitzar. Per exemple, la banda de combinació de l'aigua, la qual ens apareix entre els 5000 i 5250 cm^{-1} , té diferent absorbància depenent de la humitat de cada mostra.

6.3. Resultats del PCA

Com s'ha comentat anteriorment (Figura 8), observant el gràfic d'scores resultant del PCA s'observa una classificació de dos grups diferents, essent el grup esquerre les mostres sòlides i el dret, les pastes. Veient aquesta classificació podríem pensar que caldria fer dos mètodes diferents, un per a cada grup, ja que inclou els dos grups

dins del mateix mètode pot comportar un error de predicció elevat. Per saber quin cas és el més adient s'ha fet l'optimització i la validació de dues maneres diferents: incloent totes les mostres i excloent les mostres de pasta. Observant els errors dels dos casos s'ha arribat a la conclusió que es pot crear un sol mètode que englobi totes les mostres, incloses les pastes.

6.4. Paràmetres òptims

Per a cada component que es vol determinar s'han optimitzat els paràmetres a través d'una optimització automàtica, la qual ha donat els resultats que podem veure a les Figures 15, 16 i 17.

Configurar método de Quant 2 - C:\Users\Public\Documents\Bruker\OPUS_7.5.18\ME_Base\Quant\MET1\Humitat\MET1...

Cargar método Componentes Espectros Parámetros Validar Gráfica Informe Guardar método Optimizar Configuración

Usar parámetros Humitat NIR Optimizar

Número	RMSECV	Rango	Regiones	Preprocesamiento
254	0.154	7	7500 - 5448	Primera derivada + Normalización vectorial (SNV)
117	0.161	3	5452 - 4248	Normalización vectorial (SNV)
280	0.166	3	7500 - 6472	Primera derivada + Normalización vectorial (SNV)
98	0.169	8	9400 - 7500 6100 - 5448	Normalización vectorial (SNV)
262	0.169	5	7500 - 4600	Primera derivada + Normalización vectorial (SNV)
255	0.169	6	9400 - 5448	Primera derivada + Normalización vectorial (SNV)
250	0.17	4	7500 - 6100	Primera derivada + Normalización vectorial (SNV)
251	0.17	4	9400 - 6100	Primera derivada + Normalización vectorial (SNV)
260	0.171	4	6100 - 4600	Primera derivada + Normalización vectorial (SNV)
263	0.172	5	9400 - 4600	Primera derivada + Normalización vectorial (SNV)
152	0.174	2	6100 - 4248	Normalización de mín. - máx.
261	0.175	5	9400 - 7500 6100 - 4600	Primera derivada + Normalización vectorial (SNV)
249	0.175	5	9400 - 7500	Primera derivada + Normalización vectorial (SNV)
129	0.176	8	9400 - 7500 6100 - 5448	Normalización de mín. - máx.
114	0.177	9	9400 - 7500 6100 - 5448 4600 - ...	Normalización vectorial (SNV)
278	0.177	5	7500 - 4248	Primera derivada + Normalización vectorial (SNV)

Estado de la optimización

Figura 15. Optimització automàtica per la humitat

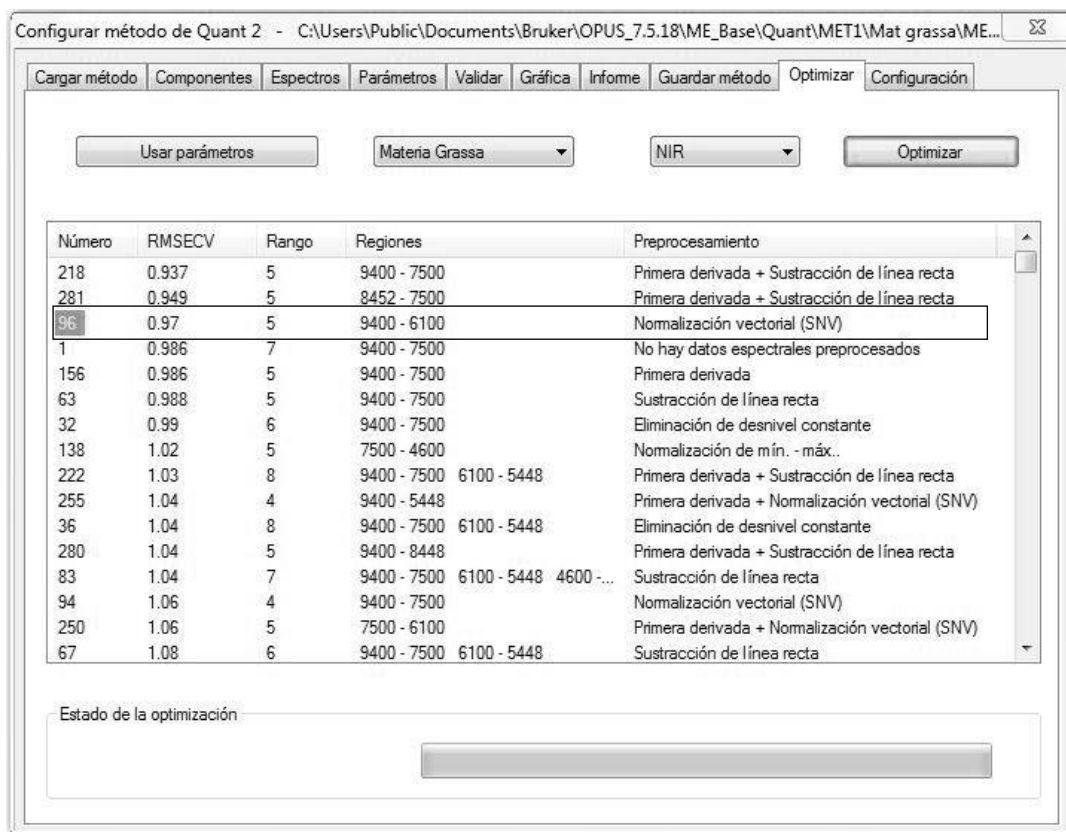


Figura 16. Optimització automàtica per la matèria grassa

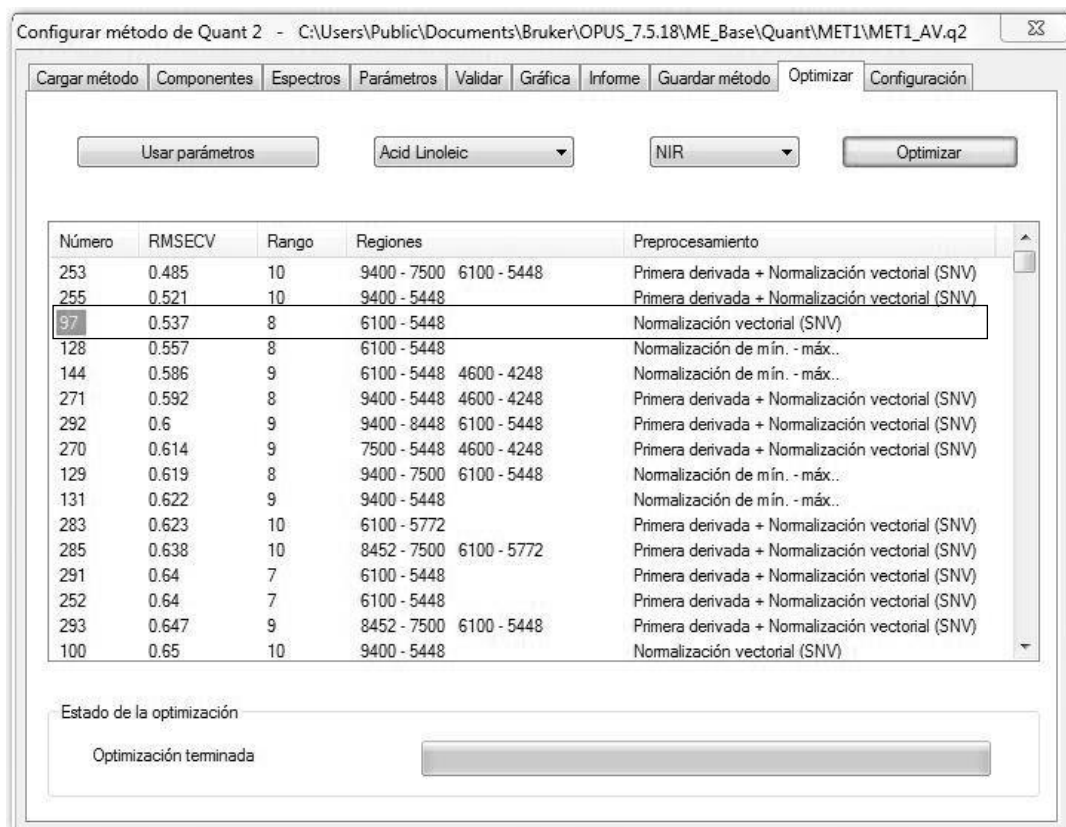


Figura 17. Optimització automàtica per l'àcid linoleic

Per a cada component s'han seleccionat els paràmetres que apareixen emmarcats. Han estat triats aquests paràmetres perquè són els que donen menys error, amb el nombre de PCs més baix i utilitzen els pretractaments menys complexos.

Utilitzant aquests paràmetres s'han obtingut els espectres de les Figures 18, 19 i 20.

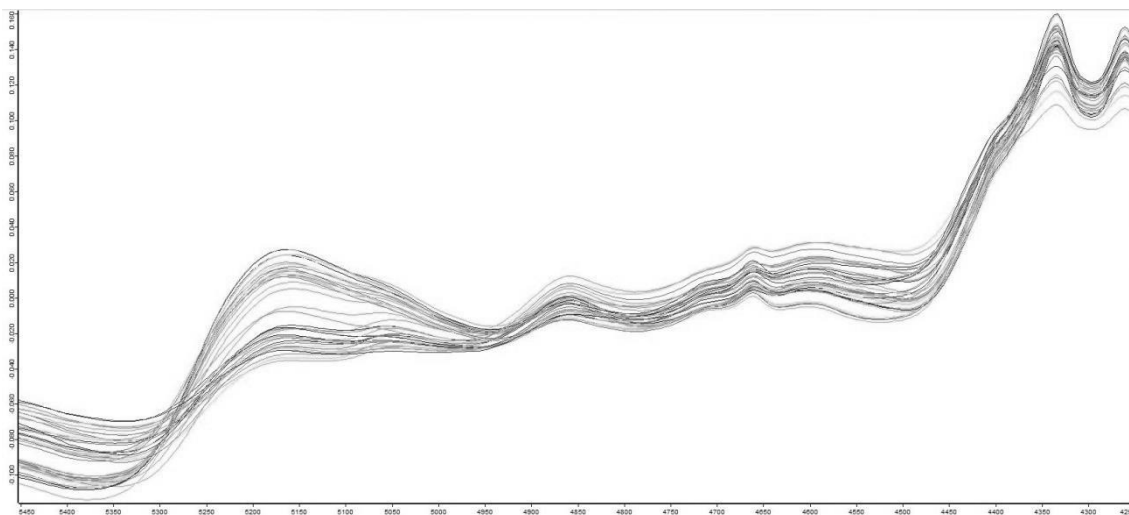


Figura 18. Regió entre 5452 - 4248 cm^{-1} i fent una normalització vectorial com a pretractament (humitat)

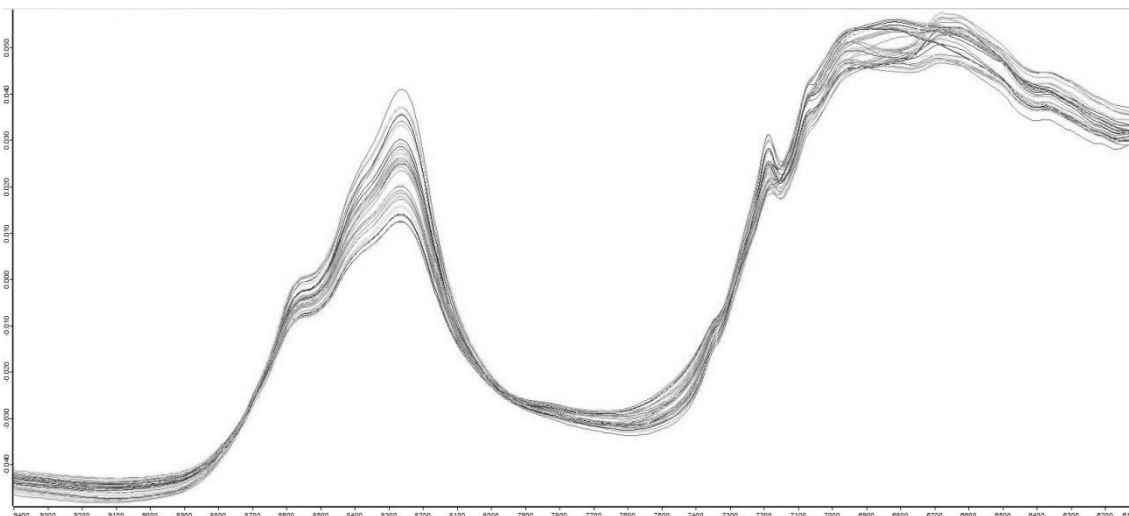


Figura 19. Regió entre 9400 - 6100 cm^{-1} i fent una normalització vectorial com a pretractament (matèria grassa)

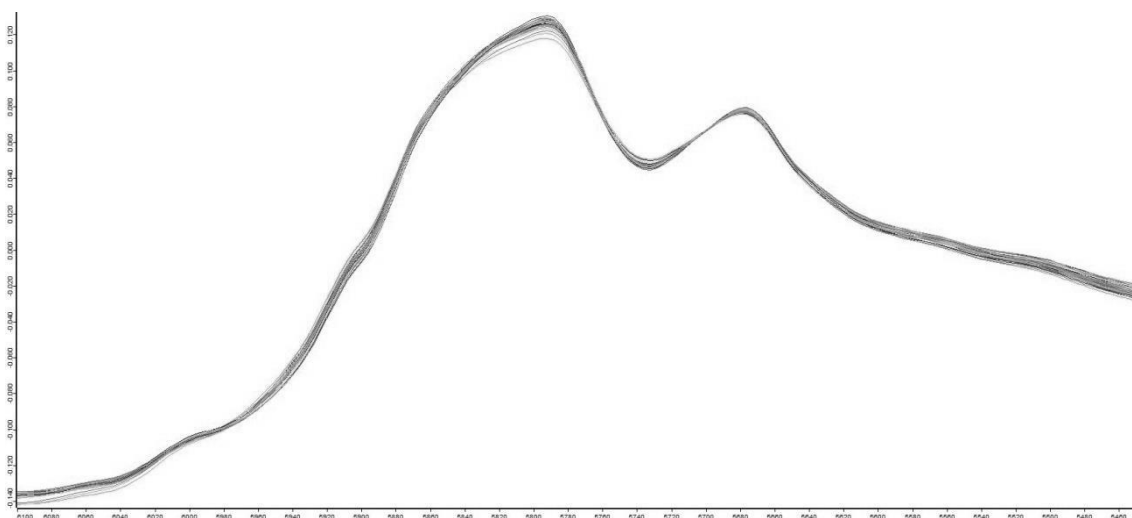


Figura 20. Regió entre 6100 - 5448 cm^{-1} i fent una normalització vectorial com a pretractament (àcid linoleic)

Per a cada mètode s'ha realitzat un pretractament espectral de normalització vectorial. En aquest tipus de normalització primer es realitza un centrat dels espectres. Tot seguit es calcula el sumatori de tots els quadrats de tots els valors Y, i els respectius espectres es divideixen entre l'arrel quadrada d'aquest sumatori. El vector de normalització dels espectres resultants és sempre 1.

6.5. Resultats de validació i exclusió d'*outliers*

Un cop ajustats els paràmetres i feta la validació creuada s'han obtingut els resultats per a cada model que s'observen a la Taula 5.

Model	RMSECV	PC	R^2	RPD
Humitat	0,161	3	99,34	12,3
Matèria Grassa	0,970	5	97,03	5,81
Àcid Linoleic	0,537	8	99,28	11,8

Taula 5. Resultats de validació per a cada model

En els tres models hi havia la presència d'un *outlier*. Per millorar els models s'han exclòs els *outliers* i s'ha tornat a fer la validació. Els nous resultats es poden veure a la Taula 6.

Model	RMSECV	PC	R ²	RPD
Humitat	0,128	3	99,56	15,1
Matèria Grassa	0,888	5	97,58	6,43
Àcid Linoleic	0,391	8	99,63	16,4

Taula 6. Resultats de validació per a cada model sense *outliers*

Observant els resultats de validació per a cada model es pot veure que hi ha una clara millora un cop s'han exclòs els *outliers*. En tots ells ha disminuït l'error (RMSECV) significativament i han augmentat tant el coeficient de determinació (R²) com la Desviació de Predicció Residual (RPD).

6.6. Predicció de noves mostres

Quan s'ha finalitzat l'optimització dels models i s'han agrupat en un mètode, s'ha fet la predicció de 5 noves mostres problema, de diferents productes i processos, i s'han comparat els valors reals amb els de la predicció. Els valors reals de les mostres noves estan presents a la Taula 7, els valors de la predicció a la Taula 8, la diferència dels reals amb els predits a la Taula 9 i el percentatge de la diferència a la Taula 10.

Mostres Noves	Humitat Real (%)	M. Grassa Real (%)	À. Linoleic Real (%)
Pasta Avellana	0,92	69,37	12,03
Ametlla Natural	5,33	55,10	19,40
Ametlla Repelada	4,35	58,02	22,47
Avellana Torrada	1,65	65,61	11,17
Pasta Ametlla	1,19	62,61	22,03

Taula 7. Valors reals de les mostres noves

Mostres Noves	Humitat Predita (%)	M. Grassa Predita (%)	À. Linoleic Predita (%)
Pasta Avellana	0,82	69,86	11,93
Ametlla Natural	5,32	55,12	17,81
Ametlla Repelada	4,47	57,53	22,88
Avellana Torrada	1,52	65,91	12,50
Pasta Ametlla	1,37	61,29	22,30

Taula 8. Valors predits de les mostres noves

Mostres Noves	Diferència Humitat (%)	Diferència M. Grassa (%)	Diferència À. Linoleic (%)
Pasta Avellana	0,10	-0,49	0,10
Ametlla Natural	0,01	-0,02	1,59
Ametlla Repelada	-0,12	0,49	-0,41
Avellana Torrada	0,13	-0,30	-1,33
Pasta Ametlla	-0,18	1,32	-0,27

Taula 9. Diferència dels valors reals amb els predits

Mostres Noves	Diferència Humitat (%)	Diferència M. Grassa (%)	Diferència À. Linoleic (%)
Pasta Avellana	10,87	-0,71	0,83
Ametlla Natural	0,19	-0,04	8,20
Ametlla Repelada	-2,76	0,84	-1,82
Avellana Torrada	7,88	-0,46	-11,91
Pasta Ametlla	-15,13	2,11	-1,23

Taula 10. Percentatge de la diferència de valors reals amb predits

Els valors d'aquesta última Taula 10 s'han calculat a partir de la següent fórmula:

$$\% \text{ Diferència} = \frac{\text{Valor real} - \text{Valor predit}}{\text{Valor real}} \times 100$$

Analitzant els resultats de les prediccions en relació amb els valors de referència es pot veure que els models de calibració creats són bastant fiables i que presenten errors de predicció bastant baixos.

7. CONCLUSIONS

When the experimental project is finished, several conclusions are obtained. They are described below:

- The sample calibration bench has to be representative, considering the parameter values of the samples that later on are analysed.
- Reference methods have to give very accurate results, because the calibration depends on these values.
- The spectral acquisition should be made without making changes that could affect the values determined with the reference methods. Special care has been taken in order to not vary the moisture % of the samples between the reference value acquisition and the spectral acquisition, because in case of variation the assignment between the value and the spectrum wouldn't be correct and therefore the subsequent predictions would be wrong.
- The reading of the samples with the NIR should be made in the same way that later on the operators will do.
- It's very important not being in a hurry when the method is being optimised and to be sure that the required checks to know if the created model is correct are made.
- The results can be considered correct while the values of each parameter are in the straight calibration, because in multivariate calibration the extrapolation is not allowed.
- Observing validation results and the new samples prediction it can be said that the optimized model is correct and fulfils expectations. It can be used by the company in order to predict new samples.

8. BIBLIOGRAFIA

- [1] Groenewald, T.; Köster, H. Espectroscopia de Infrarrojo Cercano (NIRs) - La técnica de análisis rápidos del futuro. *Engormix* [En línea] 2006. <http://www.engormix.com/MA-balanceados/formulacion/articulos/espectroscopia-infrarrojo-cercano-nirs-t577/800-p0.htm> (data de consulta 03 març 2015)
- [2] Osborne, B. G.; Fearn, T.; Hindle, P. H. *Practical NIR Spectroscopy: with applications in food and beverage analysis*, 2^a edició; Longman Scientific & Technical: Harlow, 1993
- [3] Cairós, C. J. *Espectroscopia de imagen en el infrarrojo próximo*. Tesi Doctoral, Universitat Autònoma de Barcelona. 2009
- [4] Urrego Novoa, J. R.; Díaz, G. J. Aflatoxinas: mecanismos de toxicidad en la etiología de cáncer hepático celular. *rev.fac.med.* [En línea] 2006, v.54, n.2. (data de consulta 18 març 2015)
- [5] Segura, R. Composición nutricional de los frutos secos. <http://www.sabormediterraneo.com/salud/frutossecos.htm> (data de consulta 17 març 2015)
- [6] Castillo, M. *Aplicación de la Espectroscopia NIR al Control Analítico de Procesos de la Industria Química*. Tesi Doctoral, Universitat Autònoma de Barcelona. 2007
- [7] Conzen, J. P. *Multivariate Calibration: A practical guide for developing methods in the quantitative analytical chemistry*. 3^a edició; Bruker Optik GmbH: Ettlingen, 2014

ANNEX 1: Procediment per la determinació de la humitat

Per a la determinació de la humitat de les mostres amb l'Analitzador Halogen d'Humitat utilitzem el programa bàsic, el qual treballa a 115 °C. Introduïm uns 6 grams de la mostra, prèviament molturada, i automàticament ens anirà calculant la pèrdua de pes de la mostra. El procés finalitza quan la mostra ja no perd més pes. Finalment ens calcula la humitat que hi havia a la mostra.

Per la determinació de la humitat de les pastes mitjançant el mètode tradicional pesem uns 10 grams de mostra en una placa petri de vidre i l'introduïm a l'estufa a 105 °C. Com en el cas de l'analitzador donarem per finalitzat l'assecat quan ja no s'observi pèrdua de pes i procedirem a calcular la humitat amb la següent fórmula:

$$\% \text{ Humitat} = \frac{\text{Pes inicial} - \text{Pes final}}{\text{Pes inicial}} \times 100$$

ANNEX 2: Procediment per la determinació de la matèria grassa

Per la determinació de la matèria grassa amb l'aparell de ressonància primerament hem de crear una recta de calibració amb un nombre suficient de patrons, en el nostre cas són 10 patrons. Aquests patrons es preparen amb diferents quantitats d'oli d'oliva, entre 3 i 10 grams, introduïts en vials de vidre. Per afavorir la dispersió de l'oli, dins dels vials introduïm rotllos de paper de filtre, de 4 cm d'amplada. Un cop feta la recta de calibratge, introduïm la mostra a analitzar, uns 14 grams de mostra molturada i seca, i l'aparell ens dóna automàticament el contingut de matèria grassa.

Per la determinació en pastes introduïm en un erlenmeyer uns 10 grams de mostra, 150 mL d'èter de petroli i ho posem en agitació fins a la seva completa dissolució. Tot seguit filtrem al buit amb una placa filtrant, fent les rentades necessàries, i transvasem el contingut del Kitasato a un matràs de fons rodó. Evaporem el dissolvent amb el rotavapor a uns 50 °C i posem el matràs a l'estufa per assegurar-nos l'evaporació total del dissolvent. Finalment pesem el contingut d'oli resultant i calculem el contingut de matèria grassa amb la següent fórmula:

$$\% \text{ Matèria Grassa} = \frac{\text{Pes pasta} - \text{Pes oli extret}}{\text{Pes pasta}} \times 100$$

ANNEX 3: Procediment per la determinació de l'àcid linoleic

Primer hem de preparar les mostres per a la seva posterior injecció al cromatògraf. En primer lloc introduïm uns 150 mg d'oli de la mostra a analitzar en un tub de plàstic. Afegim 2 mL d'heptà i agitem. Tot seguit afegim 0.2 mL d'una dissolució de KOH 2M, agitem i deixem decantar. Un cop les dues fases s'han separat correctament introduïm la fase apolar (superior) en un vial per a la seva posterior injecció al cromatògraf de gasos.

Els paràmetres i condicions utilitzats pel cromatògraf venen descrits a la Taula 11.

Columna	HP-88 Agilent (88% cianopropil)-aril-polisiloxà, 60 m x 0.25 mm x 0.20 µm
Gas portador	Heli. Flux de 40 cm/s
Injector	Split/Splitless en mode Splitless a 250°C
Detector	FID a 300°C
Rampa de temperatura	50 °C durant 2 min; 50-180 °C a 10 °C/min; 180 °C durant 5 min; 180-240 °C a 5 °C/min

Taula 11. Paràmetres i condicions cromatògraf