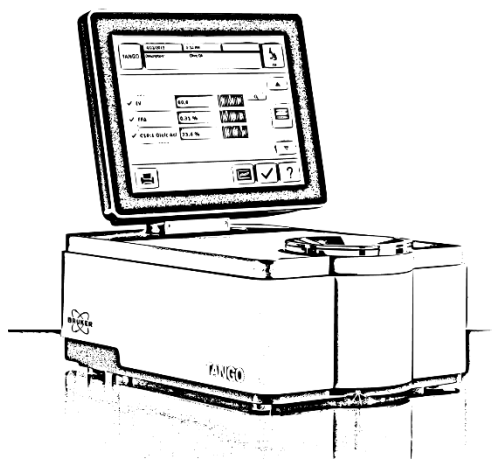


TREBALL DE FI DE GRAU

Universitat Rovira i Virgili

Facultat de Química

Desenvolupament i validació d'un mètode per espectroscòpia d'infraroig proper



Autora: Aina García Casades

Empresa: Clariant Ibérica Producción, S.A.

Curs acadèmic: 2024/2025

Tutora professional:

Dra. Ariadna Campos Carrasco

Tutora acadèmica:

Dra. Maria Besora Bonet

CLARIANT



ÍNDIX DE CONTINGUTS

1.	RESUM.....	1
2.	ABSTRACT.....	1
3.	OBJECTIUS.....	2
4.	INTRODUCCIÓ	3
4.1.	CLARIANT.....	3
4.1.1.	Història.....	3
4.1.2.	Producció i organització interna	3
4.2.	Context i rellevància del projecte	4
5.	FONAMENTS TEÒRICS.....	6
5.1.	Família de productes.....	6
5.2.	Mètodes tradicionals.....	7
5.2.1.	Índex d'hidroxil	7
5.2.2.	Índex de saponificació.....	10
5.3.	Espectroscòpia d'infraroig proper.....	12
5.4.	Anàlisi quimiomètric	15
5.4.1.	Pretractament de les dades.....	16
5.4.2.	Desenvolupament del model de calibratge.....	17
5.4.3.	Validació externa	22
6.	METODOLOGIA EXPERIMENTAL	24
6.1.	Mètodes de referència.....	24
6.1.1.	Índex d'hidroxil	24
6.1.2.	Índex de saponificació.....	26
6.2.	Enregistrament d'espectres <i>NIR</i>	27
6.3.	Desenvolupament del model quimiomètric	29
6.3.1.	Estudi de precisió i robustesa	34
6.3.2.	Implementació del model.....	35
7.	RESULTATS	36
8.	CONCLUSIONS.....	44
10.	ANNEXOS.....	48

1. RESUM

Clariant Tarragona disposa d'un departament de control de qualitat que s'encarrega d'assegurar que els seus productes compleixen els estàndards de qualitat establerts pels clients. Aquesta unitat també s'ocupa de desenvolupar i optimitzar les anàlisis realitzades al laboratori, amb l'objectiu d'incrementar l'eficiència productiva, mantenint la sostenibilitat i la seguretat dels processos.

Per aconseguir-ho, la companyia aposta per tècniques d'espectroscòpia molecular com l'infraroig proper, que permeten construir models quimiomètrics correlacionant espectres *NIR* amb els valors obtinguts de les anàlisis tradicionals de cada mostra, metodologia que proporciona resultats fiables de manera més ràpida que la convencional, minimitzant la generació de residus i l'alteració de la mostra.

En aquest estudi, s'ha creat nous mètodes *NIR*, com a alternativa als actuals, per determinar l'índex d'hidroxil i l'índex de saponificació (mg KOH/g) en productes derivats de l'etoxilació d'un alcohol gras amb òxid d'etilè, mitjançant una tècnica de calibratge multivariant basada en els Mínims Quadrats Parcial (PLS).

2. ABSTRACT

Clariant Tarragona has a quality control department responsible for ensuring that its products meet the quality standards established by customers. This unit is also responsible for developing and optimizing the analyses carried out in the laboratory to increase production efficiency, while maintaining process sustainability and safety.

To achieve this, the company employs molecular spectroscopy techniques such as Near-Infrared, which enable the construction of chemometric models by correlating NIR spectra with the values obtained from traditional analyses of each sample. This methodology provides reliable results faster than conventional methods, minimizing waste generation and sample alteration.

In this study, new NIR methods have been developed, as an alternative to the current ones, to determine the hydroxyl index and the saponification index (mg KOH/g) in products derived from the ethoxylation of a fatty alcohol with ethylene oxide, using a multivariate calibration technique based on Partial Least Squares (PLS).

3. OBJECTIUS

L'objectiu d'aquest Treball de Fi de Grau consisteix a determinar l'índex d'hidroxil (mg KOH/g) i l'índex de saponificació (mg KOH/g) per a una família de productes, mitjançant l'ús d'una tècnica d'espectroscòpia d'infraroig, concretament l'espectroscòpia d'infraroig proper (*Near-Infrared, NIR*).

L'ús d'aquesta tècnica permet reduir considerablement el temps d'anàlisi, fet que contribueix a millorar l'eficiència dels processos dels departaments de control de qualitat i de producció. A partir d'aquesta eina, es pretenen desenvolupar models de calibratge multivariant lineal (*Partial Least Squares, PLS*) per a què, a llarg termini, puguin ser implementats a l'empresa com a mètodes d'anàlisi rutinària.

D'altra banda, aquest projecte té com a propòsit personal consolidar i aprofundir els coneixements adquirits durant el grau, així com desenvolupar-ne de nous en els camps de la quimiometria i l'espectroscòpia d'infraroig proper, mitjançant la creació i validació d'aquests models quimiomètrics.

4. INTRODUCCIÓ

4.1. CLARIANT

Abans d'endinsar-nos en el Treball de Fi de Grau, és necessari contextualitzar l'àmbit en el qual s'ha desenvolupat aquest projecte: l'empresa Clariant Ibérica Producción S.A.

4.1.1. Història

L'empresa va ser fundada l'any 1995, com a resultat de l'escissió de la Divisió Química de Sandoz, i dos anys més tard va establir la seva seu a Muttentz, Suïssa, des d'on va impulsar el seu creixement a través d'una estratègia d'expansió basada en un conjunt d'adquisicions clau. [1]

Aquest procés va començar el 1997 amb la incorporació de *Hoechst Specialty Chemicals*, seguida de la compra de l'empresa britànica *BTP plc* l'any 2000, operacions que li van permetre ampliar de manera significativa el seu abast en el sector dels químics especialitzats. Posteriorment, el 2011, Clariant va reforçar la seva posició en el camp dels catalitzadors i dels adsorbents amb l'adquisició de *Süd-Chemie*, una de les operacions més rellevants de la seva història. [1]

Durant l'any 2016, la companyia va ampliar la seva presència en el sector energètic amb la integració de *Kel-Tech Inc.* i *X-Chem LLC*, empreses especialitzades en el desenvolupament de productes químics per a l'explotació de camps petrolífers. Més recentment, el 2021, l'empresa suïssa va reforçar la seva aposta pel medi ambient i el progrés tecnològic amb l'adquisició de la divisió d'ingredients naturals de Beraca, enfocada en el sector de la cura personal. [1]

Al llarg de la seva trajectòria, Clariant ha destacat per la seva capacitat d'evolució i adaptació constant, impulsada per la innovació i la sostenibilitat, valors que l'han convertida en una referent internacional en el sector dels productes químics especialitzats. Aquesta aposta per un creixement diversificat li ha permès consolidar-se com a una multinacional líder, oferint solucions pioneres i respectuoses amb el medi ambients adaptades a una àmplia varietat d'indústries.

4.1.2. Producció i organització interna

En l'actualitat, Clariant és considerada una companyia amb una àmplia presència global, ja que disposa de diverses plantes de producció distribuïdes en els cinc continents, fet que li permet atendre una gran diversitat de mercats i adaptar-se a les necessitats específiques de cada regió. Paral·lelament, en l'àmbit nacional, Clariant forma part de l'Associació Química Empresarial de Tarragona (*AEQT*), una de les agrupacions industrials més rellevants del sector químic a Espanya.

La producció i activitat de Clariant s'estructura en tres grans àrees de negoci: *Care Chemicals*, *Catalysts* i *Adsorbents & Additives* [2], les quals disposen d'un ampli catàleg de productes en el que s'inclouen des de matèries primeres per a la formulació de productes destinats a la cura personal, fins a aplicacions

industrials en camps com la metal·lúrgia, la mineria, la indústria petroquímica, el sector domèstic, l'agricultura i la indústria farmacèutica.

D'altra banda, pel que fa a l'organització interna del *site* de Tarragona, aquest es divideix en dues àrees molt pròximes entre elles. En una d'elles, s'ubiquen les quatre plantes de producció: *OLIVE*, *MPP*, *ETHOX* i *AMPS*, on es produeixen una àmplia gamma de productes. A més a més, en aquesta mateixa localització es troba el laboratori de control de qualitat, responsable del seguiment analític de les mostres de procés per garantir una bona qualitat dels lots fabricats en les respectives plantes.

L'altra zona que forma aquest recinte es situa l'edifici social de l'empresa. Aquest espai inclou les oficines, el laboratori de producte acabat, on es duen a terme les anàlisis finals abans de la comercialització d'aquests productes.



Figura 1. Edifici social del site de Tarragona (Clariant).

Així mateix, el projecte presentat en aquest treball de fi de grau s'ha desenvolupat principalment al Departament de Control de Qualitat (QC), amb el suport del Departament d'Enginyeria i Tecnologia de Processos (EPT). La primera unitat (QC) compta amb un organigrama

clarament estructurat, liderat per una cap de laboratori, la qual coordina dues seccions diferents. D'una banda, tenim la secció dedicada a les anàlisis rutinàries, amb una supervisora al capdavant i un equip d'analistes que treballen tant en torns rotatius com en horari central. D'altra banda, hi ha una segona secció formada per dues tècniques de laboratori, responsables del manteniment i la validació dels equips i mètodes analítics, així com de la posada en marxa d'equips de laboratori i del desenvolupament de nous mètodes.

Pel que fa al departament d'EPT, aquest és un grup global de Clariant responsable de l'optimització i la digitalització dels processos productius, mitjançant la integració de sistemes automatitzats i el desenvolupament de solucions digitals adaptades a les necessitats específiques de cada planta. Aquesta unitat s'organitza en tres àrees de treball: Enginyeria, Tecnologia de Processos i Solucions Digitals, les quals es divideixen en diversos subgrups distribuïts per regions, com ara Europa, Amèrica i APAC (Àsia-Pacífic).

4.2. Context i rellevància del projecte

Abans de presentar el contingut d'aquest projecte, és pertinent destacar la seva importància, tant per l'empresa Clariant com per a mi, l'estudiant, ja que aquest treball no només m'ha permès adquirir nous coneixements i formar-me com a professional, sinó que també contribuir de manera significativa a l'empresa.

Clariant disposa d'un departament de control i assegurament de qualitat, el qual s'encarrega de garantir que els productes fabricats a planta compleixen amb els estàndards de qualitat establerts. Aquest departament coopera estretament amb el de producció, amb l'objectiu d'assegurar la qualitat del producte com l'eficiència en el procés productiu.

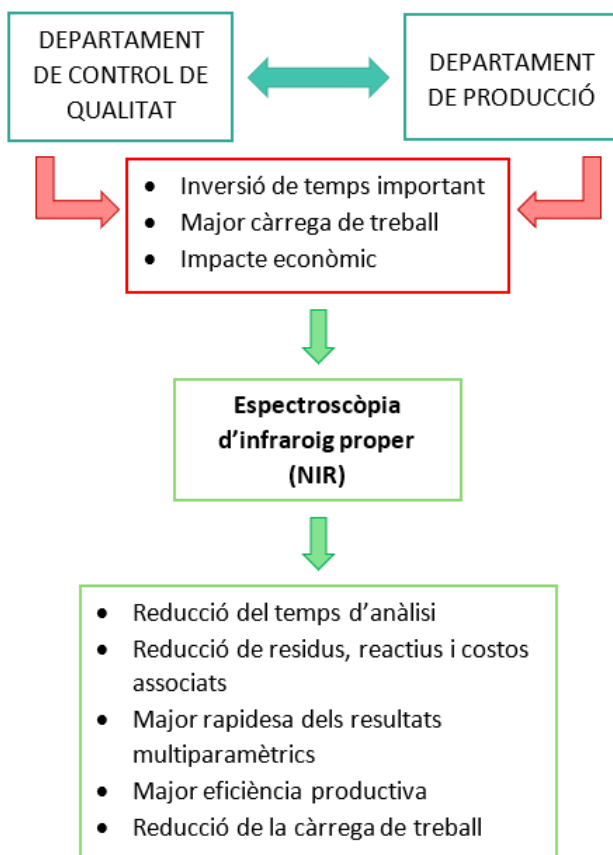
No obstant això, un dels principals inconvenients que sorgeix durant la producció d'aquests productes és la pèrdua de temps derivada de la necessitat d'aturar el procés per tal d'extreure mostres i analitzar-les al laboratori. Aquesta metodologia es realitza per tal d'evitar possibles desviacions en els diferents lots fabricats que podrien comprometre la qualitat del producte final.

La problemàtica rau en el fet que moltes d'aquestes anàlisis són de llarga durada, cosa que provoca un endarreriment en el procés de síntesi, amb l'impacte econòmic que això comporta en l'àmbit energètic i, conseqüentment, en el cost del producte, fet que repercuteix negativament en la capacitat productiva de les plantes.

D'aquesta manera, per reduir aquest temps d'espera, Clariant actualment està apostant per tècniques d'anàlisi més ràpides com l'espectroscòpia molecular, en concret, l'espectroscòpia d'infraroig proper (NIR), una tècnica que, mitjançant la creació de models quimiomètrics, permet establir una correlació entre els resultats dels mètodes tradicionals, els quals s'obtenen a partir del mètode analític de referència, i els espectres enregistrats mitjançant l'infraroig proper (NIR).

L'avantatge d'aquesta tecnologia és que generalment permet conèixer els resultats sense la necessitat de realitzar cap pretractament de la mostra ni d'utilitzar reactius, fet que redueix significativament la generació de residus, convertint a aquesta tècnica en una alternativa més eficient i sostenible.

A més, en comparació amb els mètodes convencionals, redueix notablement el temps d'anàlisi i els respectius costos associats, fet que permet lliurar els resultats al departament de producció amb major rapidesa, cosa que contribueix a millorar l'eficiència productiva i a reduir la càrrega de treball dels analistes.



Esquema 1. Beneficis de la implementació de l'espectroscòpia NIR a la indústria química.

Per últim, cal destacar una altra possibilitat que ofereix aquesta tècnica espectroscòpica. Tot i que no s'ha implementat en aquest projecte, existeix l'opció d'integrar un sistema *NIR online* que permeti el monitoratge directe dels reactors o altres equips utilitzats durant la síntesi dels productes, mitjançant una fibra òptica, una metodologia molt útil per a l'automatització d'aquests processos.

5. FONAMENTS TEÒRICS

A continuació, es presentaran els fonaments teòrics en els quals s'ha basat aquest treball, estructurats en quatre blocs temàtics. En primer lloc, es descriurà la família de productes utilitzada com a objecte d'estudi. Seguidament, s'exposaran els mètodes analítics tradicionals o de referència que es pretenen substituir. Posteriorment, s'explicarà la tècnica d'espectroscòpia d'infraroig proper (*NIR*) i finalment, s'aprofundirà en el procés de construcció de les calibracions d'aquests models, centrant-nos en l'anàlisi quimiomètric necessari per al seu desenvolupament.

5.1. Família de productes

El camp d'aplicació d'aquest treball de fi de grau s'ha centrat en els productes *Emulsogen HCO 040*®, *Emulsogen HCO 060*® i *Emulsogen HCW 049*®, els quals pertanyen a la família de tensioactius no iònics derivats dels alcohols grassos etoxilats. Aquests compostos s'obtenen mitjançant una reacció d'etoxilació, en la qual un alcohol gras, habitualment de cadena alquílica C16-C18, reacciona amb òxid d'etilè en presència d'un catalitzador bàsic. D'aquesta manera, aquest procés facilita la incorporació de grups etoxilats a la molècula original, modificant-ne les propietats fisicoquímiques i atribuint-li una àmplia versatilitat funcional.

En termes generals, l'estructura química d'aquest tipus de tensioactius correspon a la presentada a la *Figura 2*.



Figura 2. Estructura molecular simplificada dels productes *Emulsogen HCO 040*®, *Emulsogen HCO 060*® i *Emulsogen HCW 049*®.

En aquesta representació de l'estructura d'un alcohol gras etoxilat:

- **R** = Representa la cadena alquílica característica de cada producte, que correspon a la part hidròfoba de la molècula.
- **-(CH₂-CH₂-O)_n-** = Són les unitats d'òxid d'etilè repetides, que formen la part hidròfila del compost.

- **n** = Indica el grau d'etoxilació, és a dir el nombre mitjà de grups etoxilats. En el cas, de l'*Emulsogen HCO 040*[®], aquest valor és 4, mentre que per l'*Emulsogen HCO 060*[®] el grau d'etoxilació correspon a unes 6 unitats.

Com es pot comprovar, el grup funcional principal d'aquesta família de productes és l'èter (R-O-R'), específicament el grup etoxi (-O-CH₂-CH₂-O-), fet que els hi atribueix a aquestes molècules un caràcter amfipàtic que fa d'aquests compostos uns tensioactius d'elevat rendiment en múltiples aplicacions industrials.

Així mateix, la família d'*Emulsogen HCO*[®] són utilitzats generalment com a emulsionants, especialment en sistemes d'oli en aigua, gràcies a la seva excel·lent capacitat per estabilitzar emulsions. A més, també actuen com a agents humectants, degut a la seva ràpida penetració en superfícies, així com a detergents industrials, ja que es caracteritzen per oferir una formació d'escuma controlable i moderada.

La diferència entre aquests productes es basa en el grau d'etoxilació d'aquests, és a dir, en el nombre de mols d'òxid d'etilè amb els que ha reaccionat l'alcohol gras de partida. Aquesta variació influeix directament en el balanç hidròfil-lipòfil (*HLB*), paràmetre clau que representa l'afinitat d'un tensioactiu cap a les fases hidròfiles o hidròfobes.

D'una banda, l'*Emulsogen HCO 040*[®] disposa d'un valor *HLB* més baix (aproximadament 4), fet que indica una major afinitat per les fases hidròfobes (lipofíliques). Per aquest motiu, aquest producte és més soluble en olis i té millor capacitat per formar emulsions d'aigua en oli. D'altra banda, l'*Emulsogen HCO 060*[®] amb un valor *HLB* superior (aproximadament 6), mostra una major afinitat per les fases hidròfiles, tot i que continua sent relativament lipofílic, fet que li proporciona propietats intermèdies.

En conclusió, les diferències observades en els valors *HLB* s'atribueixen a la longitud de la cadena d'òxid d'etilè present en l'estructura molecular de cada producte. Com major sigui aquest paràmetre, el compost disposa d'un major grau d'etoxilació, cosa que augmenta la hidrofília del producte.

5.2. Mètodes tradicionals

Per poder desenvolupar aquests tipus de models *NIR*, és necessari realitzar l'anàlisi o la recopilació de resultats dels mètodes tradicionals, que en aquest projecte han estat: l'índex d'hidroxil [3] i l'índex de saponificació [4]. Aquests procediments es caracteritzen per la seva llarga durada, ja que la determinació del primer paràmetre requereix aproximadament dues hores, mentre que el segon pot allargar-se fins a tres hores. Per aquest motiu, s'ha proposat implementar aquesta tècnica com a alternativa per reduir significativament el temps d'anàlisi i disminuir la càrrega de treball al laboratori de control de qualitat.

5.2.1. Índex d'hidroxil

L'índex d'hidroxil és un mètode analític utilitzat per determinar el contingut de grups hidroxil presents en una mostra, i s'expressa en mil·ligrams d'hidròxid potàssic (KOH) per gram de producte.

La determinació d'aquest índex es fonamenta en una reacció d'acetilació dels grups hidroxils primaris o secundaris continguts en la mostra (excloent els terciaris). Per poder realitzar-la, es barreja una determinada quantitat de producte amb un excés d'anhídrid ftàlic o acètic, segons les característiques del producte a analitzar, dissolt en piridina, formant així la mescla d'acetilació. Aquest pas es realitza per triplicat per garantir la precisió dels resultats. Paral·lelament, es preparen dos blancs als quals se'ls aplica el mateix procediment.

Tot seguit, aquesta mescla se sotmet a un procés de reflux fins a completar la reacció. Un cop la mescla es refreda fins a temperatura ambient, s'hidrolitza l'excés d'anhídrid mitjançant l'addició d'aigua i es valora l'àcid format amb una solució de NaOH, fet que permet quantificar de forma indirecta els grups hidroxils presents en el producte mitjançant la següent equació.

$$I(OH) = \frac{(V_b - V_a) \times N \times f \times 56.11}{P} (+IA) \text{ o } (-IB) \quad (1)$$

On:

V_b = Volum de NaOH consumits pel blanc (ml)

V_a = Volum de NaOH consumits per la mostra (ml)

N = Concentració de la solució valorant de NaOH = 1 N

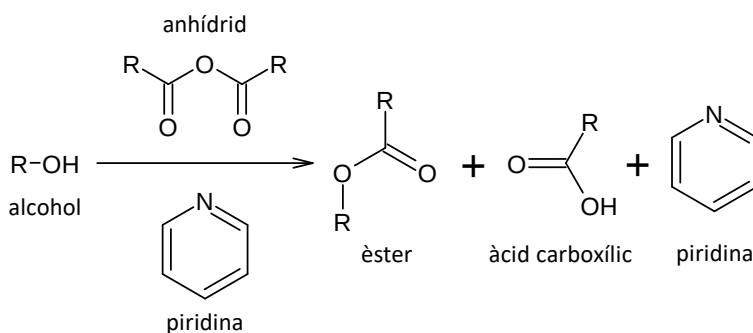
f = factor de la solució valorant de NaOH

P = Pes de la mostra (g)

IA/IB = Índex d'acidesa / Índex de basicitat del producte

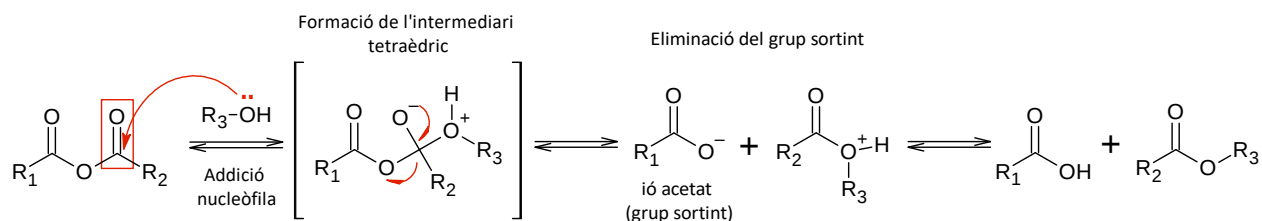
Tanmateix, aquesta anàlisi es desenvolupa en tres etapes, en les quals es produeixen les següents reaccions.

Etapla 1: Acetilació



Esquema 2. Reacció d'acetilació que succeeix en el mètode analític per determinar l'índex d'hidroxil d'un producte.

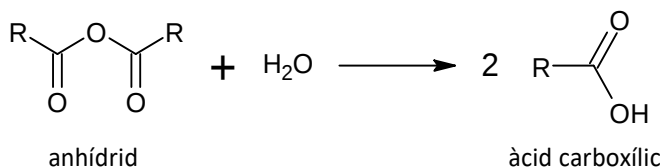
Aquest tipus de reacció d'acilació té lloc entre els grups hidroxil continguts en la mostra i l'anhídrid corresponent en presència de piridina i es produeix durant l'etapa de reflux esmentada en el procediment experimental. El mecanisme d'aquesta reacció es pot observar en l'*Esquema 3* i consisteix en el següent.



Esquema 3. Mecanisme general d'una reacció d'acetilació.

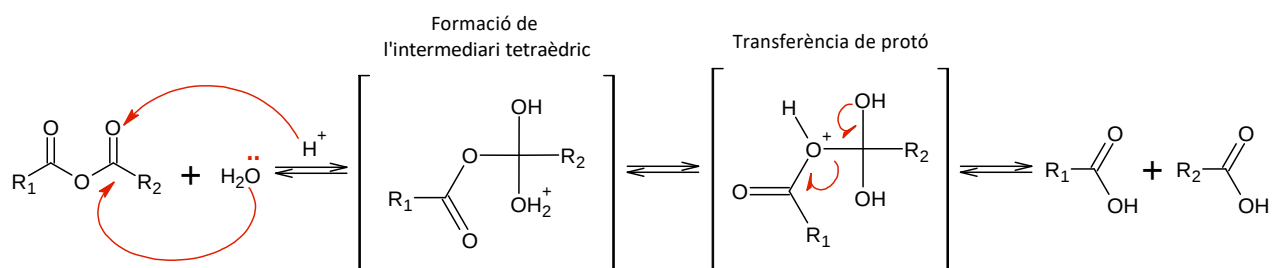
Primer de tot, es produeix l'activació electrònica de l'anhidrid pertinent mitjançant l'ús d'una base de *Lewis*, en aquest cas la piridina. Un cop activat, es produeix l'addició nucleòfila, en la que el parell d'electrons no enllaçants de l'oxigen dels grups hidroxils ataca a un dels carbonis carbonílics de l'anhidrid que actua com a centre electròfil, ja que és parcialment positiu (δ^+). En conseqüència, es forma un intermediari tetraèdric inestable, que es reorganitza mitjançant l'eliminació de l'ió acetat ($R-COO^-$). Aquest pas es troba altament afavorit, ja que aquest anió carboxilat és un excel·lent grup sortint estabilitzat per ressonància. Finalment, es produeix simultàniament la formació de l'èster i de l'àcid corresponent, segons anhidrid utilitzat.

Etapa 2: Hidròlisi del reactiu en excés



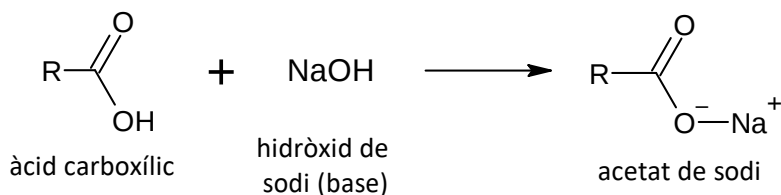
Esquema 4. Hidròlisi de l'anhidrid en excés en el mètode convencional per determinar l'índex d'hidroxil.

La hidròlisi en aquest mètode es produeix quan l'excés d'anhidrid que no ha reaccionat amb els grups hidroxils es fa reaccionar posteriorment amb aigua, donant lloc a una quantitat addicional d'àcid. A continuació, es desenvoluparà el mecanisme de reacció característic d'aquest tipus de reaccions.



Esquema 5. Mecanisme de la hidròlisi d'un anhidrid.

Primerament, es produeix l'atac nucleòfil de l'aigua en el que el seu respectiu oxigen, mitjançant el seu parell d'electrons no enllaçants, ataca el carboni carbonílic (centre electròfil) d'un dels dos grups acils de l'anhidrid. Això genera un intermediari tetraèdric, el qual patirà una transferència de protó que donarà lloc a dues molècules d'àcid.

Etapa 3: Neutralització*Esquema 6. Neutralització de l'àcid format a la primera i segona etapa mitjançant l'ús d'una base.*

Finalment, es produeix la reacció de neutralització. En la que tot l'àcid format, tant el procedent de la reacció d'acetilació com de la hidròlisi de l'excés d'anhidrid, es neutralitza mitjançant una valoració amb una solució bàsica de NaOH.

5.2.2. Índex de saponificació

L'índex de saponificació s'obté mitjançant un mètode convencional que permet determinar la quantitat de grups èster presents en una mostra. Aquest paràmetre s'expressa com el nombre de mil·ligrams d'hidròxid potàssic necessaris per saponificar un gram de producte.

Aquesta anàlisi es basa en una reacció de saponificació entre els èsters continguts en la mostra i una solució de potassa etanòlica, en presència d'isobutanol. Per dur-la a terme, una determinada quantitat de mostra es barreja amb la solució esmentada i es porta a reflux, fet necessari per garantir l'eficiència i la repetibilitat del mètode, ja que la temperatura i la concentració dels reactius són factors determinants que poden afectar el rendiment de la reacció. Així mateix, aquest pas es realitza per triplicat, i, de forma paral·lela es preparen dos blancs per tal de corregir possibles desviacions experimentals que puguin afectar el resultat.

Un cop completada la reacció, l'excés de KOH que no ha reaccionat amb els èsters presents a la mostra es determina mitjançant una valoració amb àcid clorhídric, la qual permet determinar amb precisió la quantitat de potassa etanòlica que ha reaccionat efectivament amb els grups èster. D'aquesta manera, es pot calcular l'índex de saponificació mitjançant l'equació següent.

$$Ind. Sapo. = \frac{(V_b - V_a) \times N \times f \times 56.11}{P} \quad (2)$$

On:

V_b = Volum de HCl consumits pel blanc (ml)

V_a = Volum de HCl consumits per la mostra (ml)

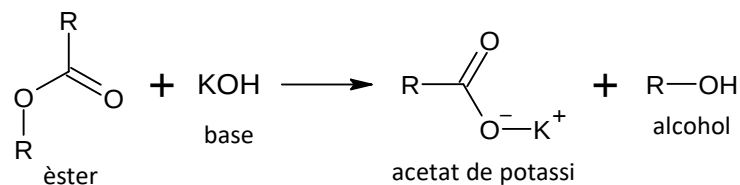
N = Concentració de la solució valorant de HCl = 0.5 N

f = factor de la solució valorant de HCl

P = Pes de la mostra (g)

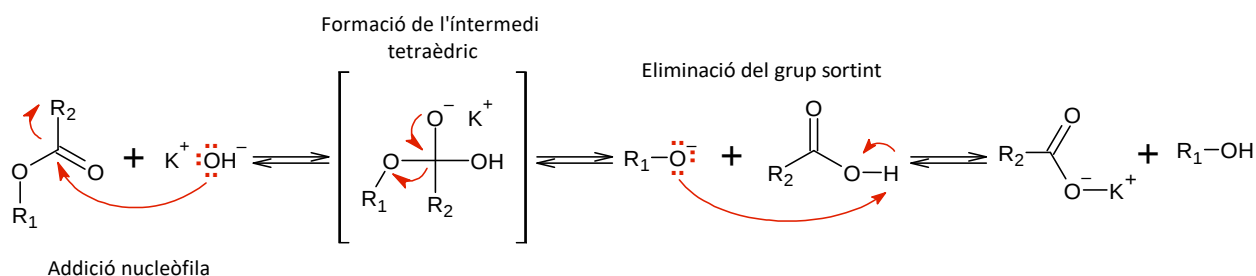
Des del punt de vista químic, aquest mètode consta de dues etapes, en les quals es troben implicades unes determinades reaccions.

Etapa 1: Saponificació



Esquema 7. Reacció de saponificació que ocorre en el mètode analític de l'índex de saponificació.

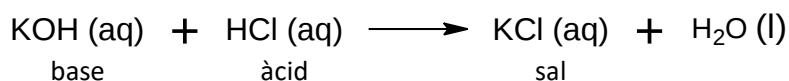
Aquest tipus de reacció té lloc quan els èsters continguts en la mostra entren en contacte amb la solució de potassa etanòlica, fet que comporta que aquests grups es transformin en sabons i s'alliberi l'alcohol corresponent. El mecanisme d'aquesta reacció es desenvolupa a l'*Esquema 8*.



Esquema 8. Mecanisme general d'una reacció de saponificació.

La reacció s'inicia amb l'atac nucleòfil de l'ió hidròxid (OH^-) al carboni carbonílic de l'èster, que actua com a centre electròfil, donant lloc a un intermedi tetraèdric. Aquest intermedi, altament inestable, es reorganitza per expulsar el grup alcoxi (RO^-) com a grup sortint, regenerant el doble enllaç carbonil i formant un àcid carboxílic. A continuació, el grup alcoxi, ric en electrons, actua com a base desprotonant l'àcid recentment format generant així l'acetat de potassi i l'alcohol corresponent.

Etapa 2: Neutralització



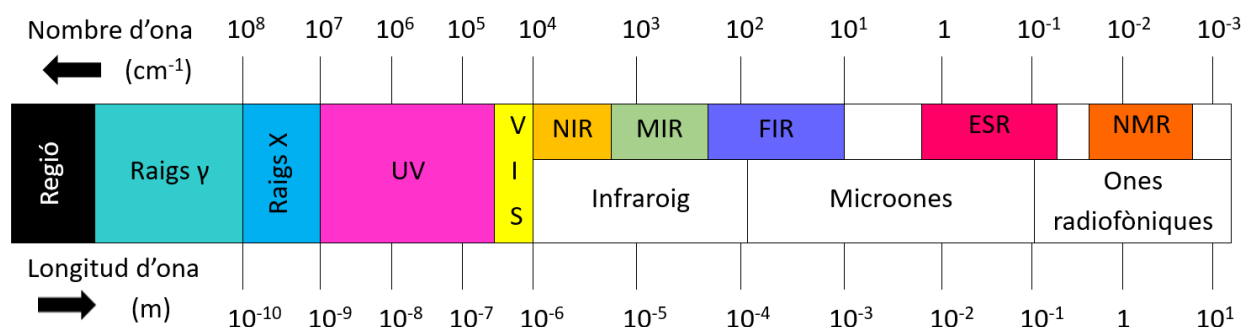
Esquema 9. Neutralització de l'excés de la solució de potassa etanòlica mitjançant àcid clorhídric.

Finalment, en aquesta darrera etapa es neutralitza l'excés de la base emprada amb l'ús d'una solució valorant d'àcid clorhídric, fet que dona lloc a la formació de la respectiva sal i aigua.

5.3. Espectroscòpia d'infraroig proper

L'espectroscòpia d'infraroig proper, també coneguda com a *NIR* (de l'anglès *Near-Infrared*), és una metodologia que permet enregistrar espectres de diverses mostres de forma ràpida i no destructiva. Aquest enregistrament és crucial per al desenvolupament de models quimiomètrics, ja que les dades espectrals obtingudes es poden tractar estadísticament per tal d'optimitzar mètodes analítics rutinaris, reduint significativament el seu respectiu temps d'anàlisi, fet que desencadena una millora de l'eficiència productiva.

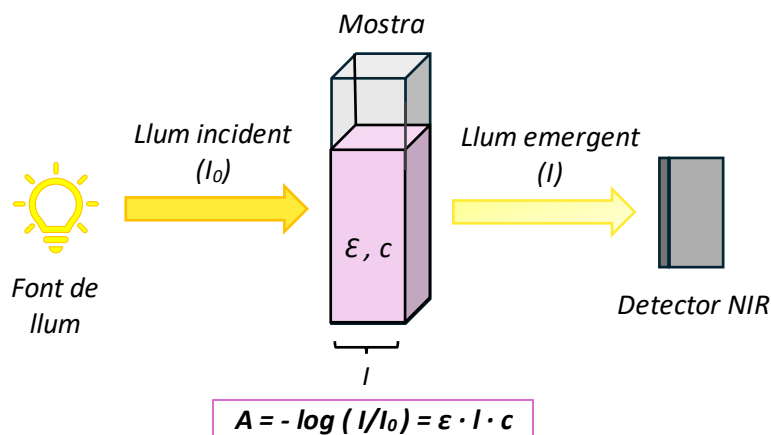
Aquesta tècnica espectroscòpica, que treballa en la regió de l'espectre electromagnètic compresa entre 12.500 i 4.000 cm^{-1} , veure *Esquema 10*, és considerada una de les eines més potents en l'actualitat, ja que permet analitzar mostres de compostos orgànics en estat líquid, sòlid i gasós, tant qualitativament, mitjançant l'ús d'espectres que tenen en compte les regions de les bandes combinades i sobretons, com quantitativament, mitjançant la Llei de *Lambert-Beer*.



Esquema 10. Representació esquemàtica de l'espectre electromagnètic.

El funcionament d'aquesta eina es basa principalment en la interacció de la radiació infraroja amb la matèria. En altres paraules, quan la llum infraroja incideix en una mostra (llum incident), part d'aquesta llum és absorbida per les molècules en major o menor quantitat segons els tipus d'enllaços orgànics que contenen, mentre que la resta continua amb la seva trajectòria fins a arribar al detector de l'equip. En conseqüència, aquesta absorció provoca una disminució de la intensitat de la llum emergent respecte de la incident. [5]

D'aquesta manera, per tal de quantificar la concentració d'un anàlit en una mostra, s'utilitza la Llei de *Lambert-Beer*, la qual expressa que la quantitat d'energia absorbida, és a dir, la diferència d'intensitat entre la llum incident (I_0) i la llum emergent (I), és directament proporcional a la concentració d'anàlit present en la mostra. Aquesta relació consisteix en el que es pot observar a l'*Esquema 11*.

Llei de Lambert-Beer

On:

- A correspon a l'absorbància.
- ϵ representa el coeficient d'absortivitat molar (característic de cada compost).
- l és la longitud del camí òptic (distància que recorre la llum).
- c correspon a la concentració de l'anàlit en la mostra.

Esquema 11. Representació de la Llei de Lambert-Beer.

Un altre aspecte a considerar, són els fenòmens interns que succeeixen a conseqüència de l'absorció d'aquesta radiació infraroja: les vibracions moleculars. Perquè aquest tipus de vibració es produeixi, la freqüència dels fotons que formen aquesta radiació incident ha de permetre la transició entre els estats vibracionals de la molècula, és a dir, ha de ser igual o superior a la freqüència que els separa. [6]

El nombre i tipus d'aquestes vibracions depenen de la quantitat i la disposició espacial dels àtoms que compon la molècula. En aquest context, cal tenir en compte dos conceptes fisicoquímics per comprendre aquest fet: els graus de llibertat i els modes vibracionals.

D'una banda, els graus de llibertat d'una molècula representen totes els diferents tipus de moviment que poden presentar els àtoms que la formen. Aquests es classifiquen en tres tipus: els moviments translacionals, que correspon al moviment en les tres direccions de l'espai (x , y , z), els rotacionals, que representen els girs al voltant dels eixos, i els vibracionals, moviments periòdics dels àtoms respecte a la seva posició d'equilibri. Tanmateix, una molècula formada per N àtoms disposarà de $3N$ graus de llibertat, ja que cada àtom pot moure's en tres dimensions.

D'altra banda, els modes vibracionals corresponen a un dels tres tipus de moviment que pot realitzar una molècula, els quals es distribueixen en dos tipus: vibracions de tensió, que allarguen i escurcen els enllaços, i les de flexió, que els dobleguen en diferents direccions. Per determinar el nombre de modes vibracionals d'una molècula, es resten els moviments translacionals i rotacionals al nombre de graus de llibertat d'aquesta. No obstant això, és important a destacar que no totes les molècules disposaran dels mateixos moviments rotacionals, ja que aquests varien en funció de si la molècula és lineal o no (2 per molècules lineals i 3 per no lineals). Per tant, en molècules lineals, el nombre de modes vibracionals són $3N-5$, mentre que en molècules no lineals són $3N-6$.

L'espectroscòpia d'infraroig proper presenta, com moltes altres tècniques espectroscòpiques, una sèrie d'avantatges i desavantatges que cal tenir en compte. [7]

Entre els principals avantatges d'aquesta tècnica destaca, en primer lloc, la reducció considerable del temps d'anàlisi, ja que proporciona resultats en qüestió de pocs minuts. A més, permet l'anàlisi multiparamètrica, fet que facilita la detecció simultània de diversos paràmetres mitjançant l'enregistrament d'un sol espectre d'una mostra, incrementant així l'eficiència analítica.

Un altre aspecte important a considerar, és que es tracta d'una tècnica no destructiva, que no requereix l'ús de reactius ni la preparació prèvia de la mostra, preservant-ne així la seva integritat. Finalment, cal destacar la facilitat d'integració i automatització d'aquesta eina en processos productius, la qual cosa en facilita l'ús en entorns industrials.

Tanmateix, l'espectroscòpia *NIR* també presenta algunes desavantatges a considerar. Una de les més destacables és la interferència causada per l'aigua, ja que aquesta absorbeix fortament en la regió de l'infraroig proper, cosa que pot afectar la precisió de l'anàlisi en mostres amb contingut variable o elevat d'humitat.

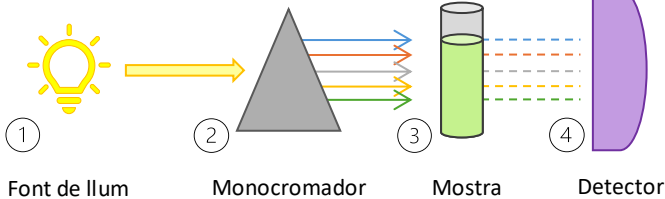
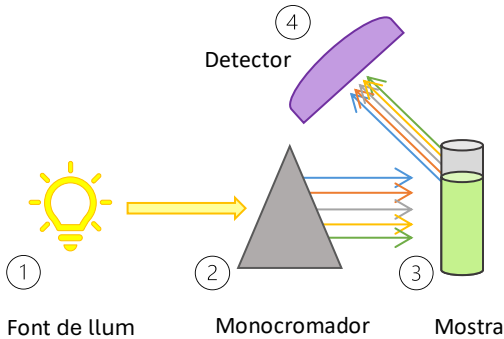
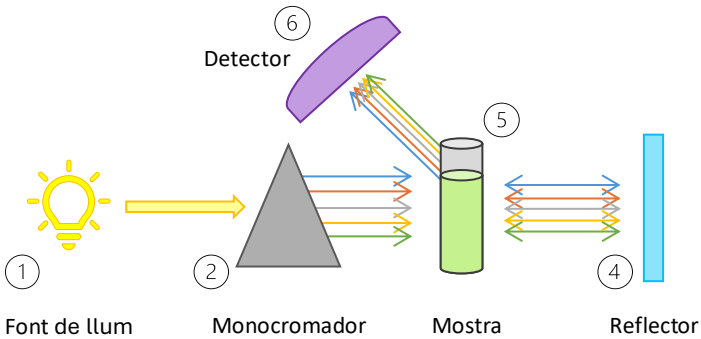
També cal destacar que presenta una sensibilitat menor en comparació amb altres espectroscòpies com l'infraroig mitjà (*MIR*) o l'*UV-Visible*, fet que pot dificultar la detecció d'analits en concentracions molt baixes. A més a més, aquesta tècnica es caracteritza per mostrar una forta dependència dels models predictius, ja que per obtenir resultats fiables, cal desenvolupar i validar models estadístics específics per cada anàlisi o tipus de mostra.

Per últim, aquesta tècnica requereix una homogeneïtat elevada en les mostres, perquè si la mostra no és homogènia, diferents zones poden tenir concentracions variables d'aquests anàlits, fent que l'absorció sigui diferent segons la zona que es mesuri, fet que pot comportar errors en la quantificació dels compostos presents.

No obstant aquestes limitacions mencionades, actualment l'espectroscòpia *NIR* continua sent considerada una tècnica extremadament potent per la seva versatilitat, rapidesa, i caràcter no destructiu, fet que la converteix en una eina indispensable en una gran varietat de sectors.

Per acabar de presentar l'espectroscòpia d'infraroig proper, cal comentar els diferents tipus de mesura que inclou aquesta tècnica, els quals varien en funció de l'instrument i el tipus de mostra que s'utilitzi. Existeixen tres tipus de mesura: la transmissió, reflexió i transfectància, veure *Taula 1*. En aquest projecte, com que s'ha treballat amb productes líquids i homogenis a temperatura ambient o bé a la temperatura que opera l'equip (70 °C) s'ha utilitzat el mètode de mesura de transmissió pel *NIR*.

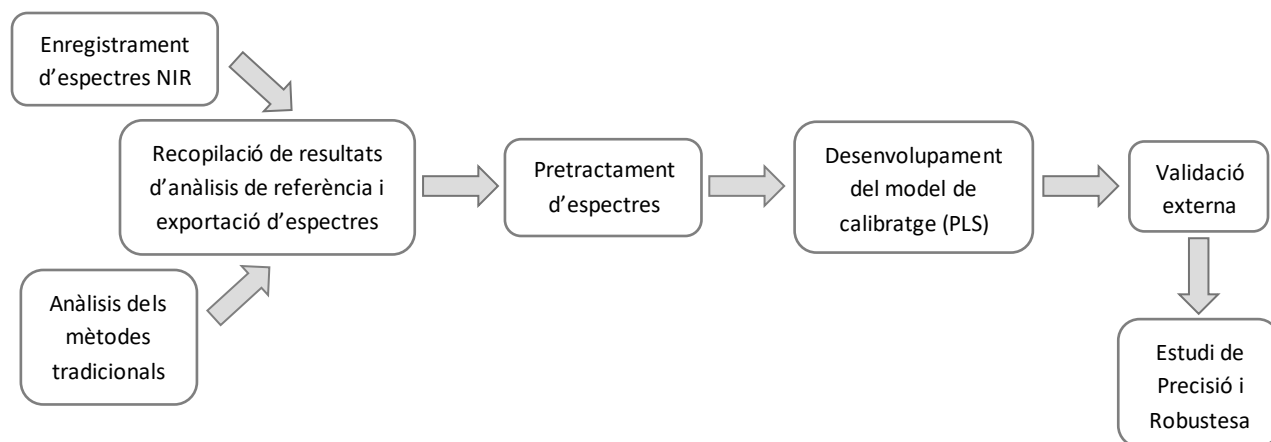
Taula 1. Diferents tipus de mesures de l'espectroscòpia d'infraroig proper.

TIPUS DE MESURA	REPRESENTACIÓ
TRANSMISSIÓ	 Font de llum Monocromador Mostra Detector
REFLECTÀNCIA	 Font de llum Monocromador Mostra Detector
TRANSFLECTÀNCIA	 Font de llum Monocromador Mostra Reflector Detector

5.4. Anàlisi quimiomètric

Per crear aquests tipus de models estadístics, coneguts també com a mètodes *NIR*, és imprescindible l'ús d'una branca específica de la Química, la Quimiometria.

Tal i com s'ha mencionat anteriorment, l'espectroscòpia d'infraroig proper s'utilitza per enregistrar espectres de les mostres, amb l'objectiu de construir una biblioteca espectral que serveixi de base per al desenvolupament d'aquests models. Posteriorment, aquesta disciplina derivada de la Química s'encarrega d'extreure el màxim d'informació possible a partir de les dades espectrals obtingudes durant aquest enregistrament. A més a més, el seu paper és fonamental en totes les etapes necessàries per al desenvolupament i validació dels models de predicció. A continuació, a l'*Esquema 12*, es mostra una representació del diagrama de flux a seguir per crear aquests tipus de model.



Esquema 12. Diagrama de flux per desenvolupar un model quimiomètric.

5.4.1. Pretractament de les dades

Els espectres *NIR* enregistrats durant l'etapa de calibratge sovint poden veure's afectats per diversos factors que comprometen la qualitat del senyal espectral, fet que pot afectar a la fiabilitat del respectiu model quimiomètric. Entre aquests elements, es troben el soroll de fons de l'espectre, les desviacions causades per característiques físiques de la mostra com la mida de partícula o la dispersió de la llum, i les variacions en la longitud del camí òptic, les quals poden provocar alteracions que afectin directament la intensitat dels senyals obtinguts.

Per minimitzar aquestes interferències i millorar la qualitat del senyal, és molt comú l'ús de l'aplicació de pretractaments matemàtics als espectres. Aquest procés és essencial per al desenvolupament d'aquest tipus de models, ja que facilita la interpretació de les dades eliminant possibles distorsions que poden afectar a la precisió del model. Entre els preprocessats més habituals destaquen els mètodes de correcció de dispersió, les tècniques de normalització com el *Standard Normal Variate (SNV)*, i els mètodes basats en derivades, com la primera derivada. Cal destacar que aquests pretractaments poden utilitzar-se tant de forma individual, és a dir, aplicant únicament un d'aquests, com combinant-los. Les dues últimes tècniques mencionades s'explicaran a continuació. [8]

5.4.1.1. *Standard Normal Variate (SNV)*

El *SNV* s'utilitza habitualment per corregir diferències d'escala i desplaçament entre espectres, alteracions que poden ser causades per efectes físics que no estan relacionats amb la composició química de la mostra, però que poden interferir en l'anàlisi de les dades espectrals. Aquest preprocessat s'aplica a cada espectre de forma individual i no necessita un espectre mitjà per obtenir una escala comuna per a tots els espectres.

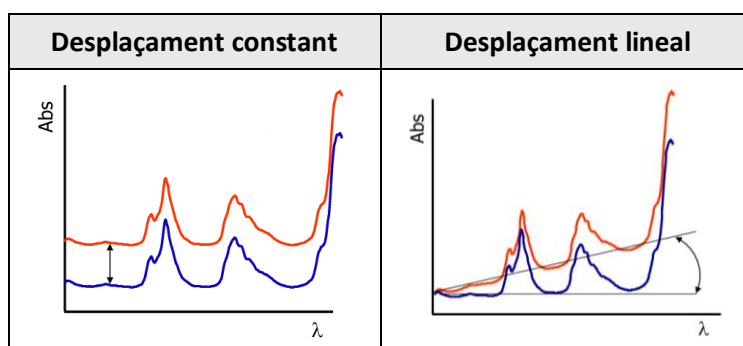
El funcionament d'aquesta tècnica consisteix en el següent. Primer de tot, per a cada espectre es calcula la mitjana de totes les absorbàncies que el formen. Tot seguit, a cada valor d'absorbància, segons longitud d'ona, se li resta aquesta mitjana d'absorbàncies obtinguda, fet que centra l'espectre, és a dir, fa que la

seva mitjana sigui zero. Posteriorment, es divideix cada valor centrat per la desviació estàndard de l'espectre, procés conegut com a autoescalat per files, ja que cada espectre correspon a una de les files d'aquesta matriu de dades espectrals (X). Com a resultat, tots els espectres sotmesos a aquest pretractament queden centrats i normalitzats a una mateixa escala, sense eliminar cap classe d'informació rellevant de les dades espectrals, facilitant així la seva interpretació.

5.4.1.2. Primera derivada

La primera derivada es caracteritza per ser un pretractament espectral utilitzat per corregir la línia base dels espectres. En altres paraules, aquest preprocessat permet alinear les dades espectrals de manera que totes parteixin d'una línia base comuna, eliminant així possibles alteracions que podrien afectar la qualitat del model. Així mateix, pel que fa a la variació entre espectres, és a dir, al desplaçament que poden tenir entre ells, aquest pot ser de tipus constant o lineal.

Taula 2. Tipus de desplaçaments en espectres.



Per una banda, un desplaçament constant es produeix quan la diferència entre els espectres és la mateixa en tot el rang de longituds d'ona. És a dir, aquesta separació no varia de forma progressiva al llarg de l'interval de longituds d'ona, sinó que es manté constant, veure Taula 2. Aquest fenomen sol produir-se per factors que afecten a tots els punts de l'espectre per igual, com ara petites diferències en el posicionament de les mostres dins de l'equip o bé canvis de temperatura o fluctuacions en la font de llum.

En canvi, el desplaçament lineal succeeix quan la variació entre els espectres augmenta o disminueix progressivament al llarg del rang espectral. Aquest comportament se sol associar a canvis sistemàtics en les condicions de mesura que no afecten uniformement l'espectre, com la dispersió de la llum o els canvis en el camí òptic efectiu.

5.4.2. Desenvolupament del model de calibratge

Com que l'espectroscòpia d'infraroig proper no es pot considerar una tècnica analítica primària, és imprescindible l'ús de mètodes de referència per determinar la propietat d'interès que es vol estudiar. D'aquesta manera, aquests mètodes tradicionals proporcionen els valors de referència necessaris per al desenvolupament de models quantitius a partir dels seus espectres *NIR*.

Per a construir i validar aquest tipus de models, es requereixen diversos conjunts de mostres, un primer conjunt corresponent al de calibratge, que s'utilitza per construir el model inicial i un segon conjunt, el de validació, per avaluar la capacitat del model per predir mostres desconegudes. [9]

5.4.2.1. Calibratge multivariant lineal

La construcció de models quimiomètrics requereix l'aplicació d'unes tècniques específiques anomenades tècniques de calibratge, les quals poden ser de tipus univariant o multivariant lineal. En aquest treball, s'ha optat per la tècnica de calibratge multivariant lineal, la qual s'explicarà a continuació.

Aquesta tècnica de calibratge es caracteritza per considerar totes les absorbències en un rang ampli de longituds d'ona al mateix temps, a diferència del calibratge univariant, que només té en compte una única longitud d'ona per predir un únic resultat. Aquesta tècnica multivariant permet analitzar diverses variables independents de forma simultània aportant així informació més rellevant, millorant la discriminació entre mostres, i incrementant la robustesa del model davant de possibles interferències o canvis en les condicions experimentals.

Així mateix, cal tenir en compte que no totes les variables incloses en el model tenen la mateixa importància, de fet, algunes poden resultar irrelevantes, i no aportar gairebé informació útil per a la predicció. Per aquest motiu, és essencial identificar quines regions de l'espectre contribueixen significativament al resultat, fet que ens ho faciliten els coeficients b .

Els coeficients (b) que s'obtenen mitjançant aquest mètode de calibratge poden entendre's com direccions dins de l'espai espectral que capten la informació més rellevant per a la predicció. Això permet reduir la dimensionalitat de les dades espectrals i alhora maximitzar la correlació entre aquestes i els valors de referència [9]. Cada coeficient indica quina influència exerceix una determinada longitud d'ona (o una combinació d'aquestes) sobre la propietat que es pretén modelar. D'aquesta manera, un valor elevat del coeficient implica un major pes d'aquella regió espectral en la predicció, mentre que un valor baix o nul d'aquest correspon a una influència mínima o inexistent en el model. Per poder predir les propietats d'interès estudiades el model utilitza la següent equació, la qual es representa visualment en l'*Esquema 13*.

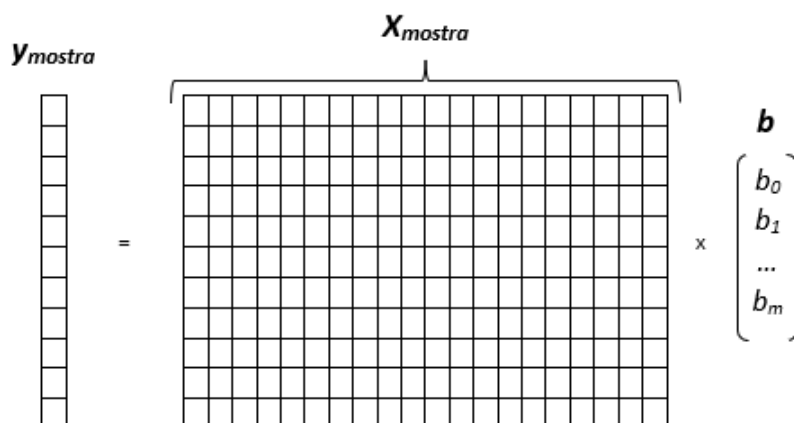
$$y_{mostra} = x_{mostra} \cdot b \leftrightarrow y_{mostra} = x_{mostra} \cdot (b\lambda_0 + b\lambda_1 + \dots + b\lambda_m) \quad (3)$$

On:

y_{mostra} = Resultats de referència (vector), format per tantes files com mostres usades en l'etapa de calibratge.

x_{mostra} = Espectres enregistrats de les mostres (matriu), constituïda per tantes files com mostres, columnes com longituds d'ona.

b = Factor que aconseguirà trobar una correlació (vector) entre la matriu x_{mostra} i el vector y_{mostra} , format per tantes columnes com longituds d'ona.



Esquema 13. Representació esquemàtica d'una calibració multivariant lineal.

Hi ha diverses tècniques de calibratge multivariant lineal, entre d'elles destaquen el *PCR* (*Principal Component Regression*) i *PLS* (*Partial Least Squares*). Per desenvolupar els models quimiomètrics d'aquest projecte, s'ha emprat la Regressió de Mínims Quadrats Parcial (Partial Least Squares, *PLS*), tècnica explicada en el següent apartat.

5.4.2.2. Construcció d'un model de Regressió de Mínims Quadrats Parcial (PLS)

El mètode de calibratge multivariant *PLS*, té com a objectiu establir una relació lineal entre un vector Y , format pels valors de referència de la propietat d'objecte d'estudi i una matriu X corresponent a les dades espectrals.

A diferència d'altres mètodes de regressió, aquesta tècnica modela simultàniament les variables que constitueixen el model per tal d'identificar els components de les dades espectrals (X) que millor expliquen la variabilitat existent en les concentracions del vector Y [9]. Per dur-ho a terme, es transformen les dades originals en un conjunt de noves variables anomenades variables latents (*Latent Variables, LV*), que corresponen a combinacions lineals dels valors originals, però més simplificades.

Així mateix, una vegada s'ha construït el model *PLS*, és imprescindible validar-ne internament la seva capacitat predictiva. Per a això, habitualment s'utilitza la tècnica de validació creuada (*Cross-validation*).

Validació creuada (Cross Validation)

La validació creuada és una tècnica que consisteix a dividir el conjunt de mostres emprades durant l'etapa de calibratge en diversos grups, per tal de poder avaluar com n'és de bona la capacitat predictiva del model. La principal idea és que, en cada cicle, es deixin fora temporalment una part d'aquestes mostres o objectes, s'entreni el model amb la resta i posteriorment es comprovi si aquest pot predir correctament les dades que s'han apartat prèviament. Aquest mètode, permet comprovar si el model és realment fiable i reduir la possibilitat que estigui massa ajustat a les dades d'entrenament (sobreajustament, de l'anglès *overfitting*), sense la necessitat de disposar de mostres addicionals.

Existeixen diferents tipus de validació creuada, però les més comunes són el mètode *leave-one-out*, on es deixa fora un sol objecte en cada iteració i es generen tants submodels com mostres hi ha, i la validació segmentada, en la qual a cada pas s'exclou un grup de mostres en comptes d'una sola. Tot i així, encara que la validació creuada és una bona alternativa a la validació externa per la seva simplicitat, té una limitació important, ja que no garanteix del tot com es comportarà el model amb mostres completament noves que no hagin estat utilitzades en cap fase del respectiu procés de construcció. Per això, sempre que sigui possible, el més recomanable és complementar la validació creuada amb una validació externa.

Així mateix, per obtenir un model *PLS* que funcioni correctament, és clau tenir en compte diferents paràmetres de qualitat [11,12] que ens permetin decidir si el model és adequat per al que es pretén aconseguir.

- **RMSECV: Error quadràtic mitjà de la validació creuada**

El *RMSECV* és un paràmetre que indica com n'és de precisa la predicció el model de calibratge durant la validació interna. Per poder determinar aquest valor es comparen els valors predits pel model amb els valors reals, obtinguts amb els mètodes analítics de referència, calculant la diferència mitjana entre tots dos. Seguidament, es realitza l'arrel quadrada, veure *Equació 4*, de la mitjana dels errors quadràtics obtinguts en cada iteració durant la validació.

$$RMSECV = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (Y_i^{meas} - Y_i^{pred})^2} \quad (4)$$

On:

Y_i^{pred} = valors predits per el NIR.

Y_i^{meas} = resultats obtinguts a partir del mètode de referència.

n = número de mostres.

Tanmateix, aquest error ens ajuda a decidir quantes variables latents són les més adequades per construir un model fiable. Tenint en compte aquest fet, com més baix sigui el *RMSECV*, millor serà la capacitat predictiva del model.

- **Coefficient de correlació (r)**

Aquest paràmetre de qualitat indica com n'és de bona la correlació entre els valors predits pel model i els valors reals obtinguts a partir del mètode de referència. És a dir, proporciona el percentatge de variació en els valors dels components que és recollit en la predicció. Com major sigui aquest coeficient, més acurada és la correlació entre els valors de referència i les dades espectrals.

D'altra banda, un valor de r baix pot ser conseqüència d'una precisió insuficient en els resultats de referència, de la selecció de paràmetres inapropiats per al model, o de la presència de *outliers* que no han estat eliminats en el model.

D'aquesta manera, en funció del coeficient de correlació obtingut, el model es pot aplicar com a alternativa en determinades situacions, veure *Taula 3*.

Taula 3. Avaluació segons l'empresa Clariant del coeficient de correlació d'un model NIR. [11, 12]

r	Classificació del model	Interpretació del model
< 0,25	Molt pobre	No utilitzable
0,26 – 0,49	Pobre	Correlació feble
0,50 – 0,64	Imparcial	Aplicacions de proves preliminars
0,65 – 0,82	Raonable	Acceptable, adequat per aplicacions de proves preliminars
0,83 – 0,90	Bo	Utilitzar-ho amb compte en el control de qualitat de mostres de procés (QC)
0,91 – 0,96	Molt bo	Apte per usar-lo en assegurement de qualitat de producte acabat (QA)
0,97 – 0,998	Excel·lent	Ús per a qualsevol tipus d'aplicació
> 0,998	Superior	Tant bo com el mètode de referència

- **Variables Latents (LV)**

Les variables latents són unes combinacions lineals de les dades originals que el model utilitza per calcular les prediccions, les quals tenen la capacitat de captar l'estructura subjacent de les dades. Aquestes variables en el model *PLS*, tenen com a objectiu identificar els components principals que maximitzin la correlació entre la matriu X , formada per les dades espectrals, i el vector Y , constituït pels resultats de referència.

Així mateix, durant la construcció d'aquests models és fonamental seleccionar correctament el nombre de *LVs*. Generalment, és recomanable utilitzar el menor nombre possible d'aquestes variables que permeti mantenir una capacitat predictiva del model adequada, per tal d'evitar el sobreajustament (*overfitting*), fenomen que succeeix quan el model s'adapta massa a les dades d'entrenament i no generalitza correctament dades noves.

- **Diferència percentual relativa (RPD)**

Aquest paràmetre és un indicador estadístic que es calcula mitjançant la divisió entre la desviació estàndard del valor de referència del conjunt de calibratge (SD_{ref}) i l'error estàndard de predicció (SEP), veure *Equació 5*.

$$RPD = \frac{SD_{ref}}{SEP} \quad (5)$$

El *RPD* (*Relative Percent Difference*) correspon a la mesura quantitativa de l'avaluació de la capacitat predictiva del model durant l'etapa de validació. En general, com major sigui el valor de *RPD* millor serà la qualitat del calibratge, ja que aquest paràmetre també proporciona informació sobre la distribució de les mostres dins del rang de concentracions del calibratge. En aquest context, un valor elevat d'aquest paràmetre indica que en aquest model les mostres que el formen estan distribuïdes de forma homogènia al llarg de l'interval de concentracions. La *Taula 4* recull el criteri que l'empresa Clariant segueix conforme a l'aplicació dels mètodes desenvolupats segons el valor de *RPD* obtinguts.

Taula 4. Interpretació del paràmetre *RPD* segons Clariant. [11,12]

RPD	Classificació del model	Interpretació del model
< 1,0	Molt pobre	No recomanable
1,0 – 2,4	Pobre	No recomanable
2,5 – 2,9	Imparcial	Projecció aproximada
3,0 – 3,9	Reanable	Cribratge o control
4,0 – 5,9	Bo	Control de qualitat de mostres de procés (QC)
6,0 – 7,9	Molt bo	Assegurament de qualitat de producte acabat (QA)
8,0 – 10,0	Excel·lent	Qualsevol aplicació
> 10,0	Superior	Tant bo com el mètode de referència

- **Distància de Mahalanobis (*MD*)**

La distància de Mahalanobis es defineix com la variació entre l'espectre mitjà i la mitjana de tots els espectres enregistrats durant l'etapa de calibratge, els quals formen la matriu espectral (*X*). Quan aquesta diferència augmenta, també ho fa respectivament la distància de Mahalanobis. Aquest increment és causat generalment per factors externs com fluctuacions en la temperatura durant l'enregistrament dels espectres o possibles contaminacions de les mostres, aspectes que desencadenen una alteració de la simetria del pic espectral.

Tanmateix, aquesta distància s'utilitza com un indicador numèric per avaluar la fiabilitat de l'anàlisi, ja que ajuda a identificar tant punts discrepants com aquelles mostres amb valors de concentració fora de l'interval establert com a vàlid. D'aquesta manera, les mostres que es trobin per sobre de la distància de Mahalanobis no poden ser predites amb precisió i, per tant, es consideren valors atípics.

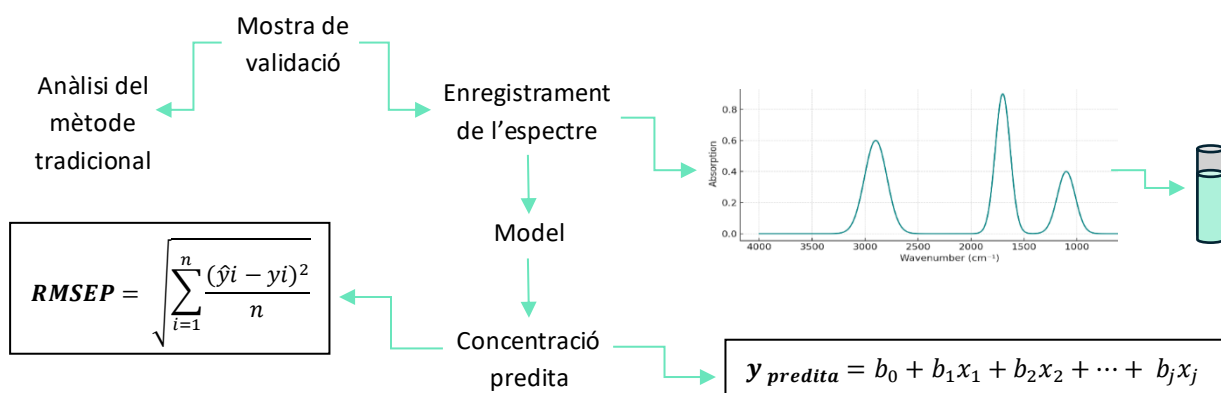
5.4.3. Validació externa

En el desenvolupament d'un model predictiu, la validació externa és una etapa clau per demostrar la seva capacitat per predir noves mostres. Aquesta tècnica serveix per garantir que el model no només s'ajusta

correctament a les dades emprades durant la seva construcció, sinó que també és capaç de generalitzar de forma totalment fiable altres conjunts de dades.

Per avaluar la qualitat dels models, s'utilitzen diferents estratègies, segons la disponibilitat de mostres representatives. Generalment, quan no es disposa d'un conjunt de mostres externes per validar el model, s'opta per mètodes com la validació interna, explicada anteriorment, entre les quals destaca la validació creuada (*Cross-Validation*). No obstant això, la millor manera de validar un model és mitjançant l'ús d'un conjunt de validació independent, conegut com a validació externa, metodologia que proporciona una estimació més realista de la precisió del model. [10]

Durant la validació externa, el model prediu el valor de la propietat d'interès per a cada mostra mitjançant la multiplicació de la suma ponderada de les dades espectrals amb els coeficients de regressió (b) del model, donant com a resultat el valor predit, tal i com es mostra en l'*Esquema 14*.



Esquema 14. Representació del procediment a seguir durant la validació externa d'un model quimiomètric.

D'aquesta manera, per quantificar el comportament predictiu d'aquests models, es calcula la suma de residus quadràtics predictiva (*PRESS*), que correspon a la suma dels quadrats de les diferències entre els valors reals (y_i) i els predits ($\hat{y}_i$). A partir d'aquest valor, es calcula l'error quadràtic mitjà de predicció (*RMSEP*) [10], paràmetre que proporciona una mesura quantitativa de la precisió amb la qual una mostra és predita durant la validació externa. Un *RMSEP* és considerat acceptable quan el seu valor és proper al *RMSECV*, error obtingut durant la validació creuada, sent recomanable que aquest es trobi inclòs en l'interval d'acceptació ($RMSEP \approx RMSECV \pm 0.7 \times RMSECV$). [11,12]

Un cop completada la validació externa, és important assegurar l'adaptació i el manteniment del model, ja que amb el pas del temps poden aparèixer noves fonts de variabilitat (canvis en matèries primeres, condicions experimentals, etc.). Per aquest motiu, és recomanable actualitzar el model periòdicament incorporant noves mostres al conjunt de calibratge. [9]

6. METODOLOGIA EXPERIMENTAL

Un cop explicat el fonament teòric en el qual s'ha basat aquest Treball de Fi de Grau, es presentarà la metodologia experimental emprada en aquest projecte per al desenvolupament dels models *NIR*.

Aquesta part experimental del projecte s'inicia paral·lelament amb dues tasques: d'una banda, l'enregistrament dels espectres de les mostres de calibratge mitjançant l'espectroscòpia d'infraroig proper, i de l'altra, la realització de les anàlisis tradicionals, que en aquest cas són l'índex d'hidroxil i l'índex de saponificació. Posteriorment, aquests espectres s'exporten al *software* adient conjuntament amb els resultats d'aquests mètodes de referència realitzats.

Per completar el model de calibratge, es recuperen els resultats dels mètodes analítics de referència de mostres tant de procés com de producte acabat, extrets de dues bases de dades internes de Clariant. Per seleccionar les mostres adequades, es tenen en compte dos criteris: disposar del corresponent espectre *NIR* i dels seus resultats de referència. A més, se seleccionen les mostres que tenen concentracions que es trobin als extrems de la recta de calibratge per tal d'ampliar el rang de concentracions i augmentar la capacitat predictiva del model.

Quan tots els espectres són carregats al programari, s'aplica el pretractament més adequat per millorar la qualitat del senyal. Tot seguit, es procedeix a la creació del model de calibratge mitjançant la tècnica estadística de *Partial Least Squares (PLS)*. Durant la construcció d'aquest tipus de models, s'eliminen punts discrepants i es generen diversos models per identificar quina és la regió espectral que millora la predicció del model, fet que també facilita la identificació dels grups funcionals continguts a les mostres amb els pics espectrals.

Finalment, amb el model de calibratge ja establert, es duu a terme la seva validació externa per comprovar-ne la fiabilitat del model, seguit de la realització d'un estudi de precisió i robustesa per assegurar que el model manté la seva capacitat predictiva estable en diverses condicions experimentals.

6.1. Mètodes de referència

Per a desenvolupar aquest tipus de models, es requereix l'ús de mètodes analítics de referència, ja que aquests proporcionen els resultats adients amb els quals es correlacionaran els espectres *NIR* enregistrats. En aquest context, aquest projecte se centra en la modelització de dos mètodes tradicionals: l'índex d'hidroxil i l'índex de saponificació, els procediments experimentals dels quals es descriuen als següents apartats.

6.1.1. Índex d'hidroxil

Un cop explicat el fonament teòric en què es basa aquest mètode (*apartat 5.2.1.*), s'explicarà tant el procés operatiu a seguir com el material i reactius necessaris per a la seva determinació, veure *Taula 5*.

Taula 5. Reactius i material necessari per a la determinació d'índex d'hidroxil. [3] (N = normalitat)

REACTIUS	MATERIAL
• Piridina • Anhídrid ftàlic • Solució de NaOH 1 N • Solució de NaOH 0.1 N • Fenolftaleïna 0.1 % en etanol • Aigua destil·lada • Mostra d'Emulsogen HCO 040® o HCO 060® o HCW 049®	• Matrassos erlenmeyers • Pipeta Pasteur • Pipeta aforada • Pera • Balança analítica • Vasos de precipitats (150 ml) • Imants • Agitador magnètic • Mantes calefactores • Material necessari per reflux (condensadors, tubs de plàstic, porcellana porosa,...) • Provetes graduades • Titrador potenciomètric automàtic • Titrador manual amb control a distància • Unitats intercanviables de NaOH 1N i 0.1N

Procediment operatiu

En primer lloc, es prepara la mescla d'acetilació barrejant quantitats precises d'anhídrid ftàlic i piridina en una ampolla de color topazi. Aquests reactius es pesen en un granetari sota l'ús d'una campana de fum, ja que la piridina és un reactiu categoritzat com a CMR. Seguidament, la solució s'agita amb l'ajuda d'un imant i un agitador magnètic fins a la seva dissolució completa, moment en el qual ja pot ser utilitzada.

En erlenmeyers, es pesa la mostra mitjançant una balança analítica, en funció de l'índex d'hidroxil esperat, i es treballa per triplicat per garantir la precisió dels resultats. Paral·lelament, també es preparen dos erlenmeyers addicionals corresponents als blancs (B₁ i B₂). Tot seguit, es distribueix un volum determinat de la mescla d'acetilació a cada erlenmeyer amb l'ajuda d'una pipeta aforada i una pera, i es porta a reflux durant un temps establert.

Un cop transcorregut aquest període i les mostres ja es troben a temperatura ambient, s'hi afegeix una quantitat determinada d'aigua a cada erlenmeyer, mitjançant una proveta, i s'agita correctament per afavorir la hidròlisi de l'anhídrid en excés. A continuació, s'introdueix piridina, seguint el mateix procediment, per tal de facilitar la dissolució dels components restants.



Figura 3. Addició de la mescla d'acetilació a la respectiva mostra.

Per últim, s'afegeix fenolftaleïna com a indicador a cada matràs i aquests es valoren amb una solució valorant de NaOH 1 N, valoració duta a terme amb un valorador manual amb control a distància. Tot seguit, es registren els respectius volums consumits per l'agent valorant i es realitzen els càlculs pertinents (*Equació 1, apartat 5.2.1*) als quals se suma l'índex d'acidesa, paràmetre determinat en paral·lel a partir d'una valoració d'un pes determinat de mostra dissolta en aigua i piridina. [3]



Figura 4. Valoració de la mostra amb una solució de NaOH 1 N.

6.1.2. Índex de saponificació

D'altra banda, pel que fa a l'índex de saponificació, del qual s'han exposat els fonaments teòrics a l'*apartat 5.2.2*, a continuació es descriuen el mètode operatiu a seguir, així com els materials i reactius requerits per a la determinació d'aquest paràmetre, veure *Taula 6*.

Taula 6. Reactius i material necessari per a la determinació d'índex de saponificació. [4] (N = normalitat)

REACTIUS	MATERIAL
- Solució de KOH 0.5 N metanòlica - Isobutanol - Solució de HCl 0.5 N - Fenolftaleïna 0.1 % en etanol - Mostra d'Emulsogen HCO 040® o HCO 060® o HCW 049®	- Matrassos erlenmeyers - Pipeta Pasteur - Balança analítica - Imants - Mantes calefactores - Material necessari per reflux - Titrador potenciomètric automàtic - Titrador manual amb control a distància - Unitats intercanviables de KOH 0.5 N i HCl 0.5 N

Procediment operatiu

Primer de tot, en erlenmeyers, es pesa una quantitat determinada de mostra amb una balança analítica segons l'índex de saponificació esperat, pas que es realitza per triplicat. Addicionalment, es preparen també dos erlenmeyers que actuen com a blancs (B_1 i B_2), els quals no contenen mostra i serveixen per corregir possibles errors sistemàtics derivats del procediment experimental.



Figura 5. Reflux per determinar l'índex de saponificació.

A continuació, s'hi afegeix un volum conegut d'isobutanol i d'una solució de potassa etanòlica (KOH en etanol) a cada matràs, i es porten a reflux durant un temps determinat, per tal que es produeixi la reacció de saponificació.

Un cop ja completada la reacció i les mostres ja es troben a temperatura ambient, s'hi introdueix a cada erlenmeyer fenolftaleïna en etanol com a indicador, i es valora l'excés de KOH amb una solució d'àcid clorhídric 0.5 N. Aquesta valoració es duu a terme amb un valorador manual amb control a distància, i s'anoten els respectius volums consumits de la solució valorant. Finalment, per obtenir l'índex de saponificació es realitzen els càlculs pertinents segons l'Equació 2 que es troba a l'apartat 5.2.2 d'aquest document. [4]



Figura 6. Valoració de la mostra amb una solució de HCl 0.5 N.

6.2. Enregistrament d'espectres NIR

L'enregistrament d'espectres NIR és un dels passos principals en el desenvolupament de models quimiomètrics, ja que aquesta tècnica proporciona informació espectral, de forma ràpida i no destructiva, sobre la composició química dels productes, la qual pot ser correlacionada posteriorment amb la propietat química d'interès.

En aquest treball, s'han gravat els espectres mitjançant l'equip NIR – Antaris II de la marca Thermo Fisher, veure Figura 7. Així mateix, per a l'enregistrament d'aquests, s'han utilitzat els reactius i materials que s'observen a la Taula 7 i s'ha seguit el següent procediment operatiu.



Figura 7. NIR - Antaris II, Thermo Fisher.

Taula 7. Reactius i materials necessaris per a l'enregistrament d'espectres NIR.

REACTIUS	MATERIAL
Mostra d'Emulsogen HCO 040® o HCO 060® o HCW 049®	- Pipeta Pasteur - Cubeta de vidre transparent - Tap de plàstic - NIR – Antaris II, Thermo Fisher

Procediment operatiu

En primer lloc, abans de procedir a enregistrar els espectres, cal seleccionar el diagrama de flux adequat al programari que utilitza l'equip, tal com es mostra a la Figura 8.



Figura 8. Icona per seleccionar el "workflow" adient en el programari de l'equip.

Tot i que, prèviament, cal assegurar-se que, en el diagrama de flux seleccionat, l'equip treballa amb les condicions de mesura següents, en cas contrari, és necessari modificar-les perquè corresponguin a les que s'indiquen a continuació.

Taula 8. Condicions de mesura necessàries per a gravar espectres NIR.

CONDICIONS DE MESURA			
Temperatura (°C)	Temps (Scans)	Resolució (cm ⁻¹)	N ° de mesures per mostra
70	16	8	1

Un cop configurades les condicions establertes, es prepara la mostra per al corresponent enregistrament de l'espectre. En una cubeta de vidre, s'introdueix la mostra amb l'ajuda d'una pipeta *Pasteur* i, un cop omplerta, es tapa amb un tap de plàstic. Cal destacar que la cubeta no s'ha d'omplir del tot, sinó que cal deixar un petit espai perquè el tap pugui exercir correctament la seva funció.

Aquest pas és crucial per realitzar un enregistrament espectral fiable, és per aquest motiu que no només s'ha d'anar amb compte per evitar possibles contaminacions de la mostra, intentant no embrutar la cubeta amb substàncies que puguin interferir en l'adquisició de dades espectrals, sinó que també és important assegurar-se que la mostra sigui homogènia, ja que, en cas contrari, l'espectre obtingut pot ser imprecís o poc representatiu.

Seguidament, cal prémer el botó de l'equip indicat a la *Figura 9* per tal d'obrir el compartiment que conté l'estructura de plàstic on s'ha d'introduir la cubeta amb la mostra per iniciar l'enregistrament. Aquesta estructura disposa d'una tapa que s'ha de retirar per poder col·locar la cubeta dins de la cavitat.

Un cop preparada la mostra i introduïda a l'equip, clicar a la icona "Ir" per iniciar l'enregistrament de l'espectre. Automàticament, s'obrirà una finestra emergent que demana introduir el nom de la mostra. En el cas de Clariant, en aquest camp s'introdueix generalment l'*ID LIMS*, un codi numèric proporcionat pel sistema de gestió de dades *LIMS* que permet identificar diferents lots o operacions associades a diversos productes.

A continuació, verificar aquest codi clicant "Ok", acció que obrirà una nova pestanya on se sol·licita introduir el nom de la persona que realitza l'enregistrament, i, tot seguit, confirmar-ho de nou amb "Ok". En aquest precís moment, s'inicia la gravació de l'espectre de la mostra introduïda, procés que finalitza quan en la pantalla principal del programari apareix l'espectre obtingut.



Figura 9. Compartiment per introduir la mostra a l'equip.

6.3. Desenvolupament del model quimiomètric

Construcció del model de calibratge

Per desenvolupar els models *NIR*, és necessari utilitzar programaris estadístics per dur a terme una anàlisi correcta de les dades introduïdes. En el cas d'aquest projecte, s'ha emprat el *software TQ Analyst*, de la marca *Thermo Fisher*, ja que, com s'ha comentat, s'usa l'equip *NIR – Antaris II*, d'aquesta mateixa marca, per enregistrar els espectres. A continuació, es detalla la metodologia a seguir per crear aquest tipus de models quimiomètrics [13]. Cal destacar que aquest procediment s'ha seguit tant pel model *NIR* de l'índex d'hidroxil com per al de l'índex de saponificació, però s'explicarà de forma individual.

Primer de tot, s'exporten tots els espectres a l'ordinador que s'utilitza per crear el model, de manera que quedin tots guardats en una única carpeta. D'altra banda, també es recopilen tots els resultats dels mètodes de referència per mostra en un full de càlcul, el qual és recomanable que es trobi en una altra carpeta que es localitzi fàcilment.

Seguidament, s'obre el programari *TQ Analyst* i simultàniament apareix la pantalla principal. Com que es vol crear un mètode de calibratge mitjançant la tècnica *PLS*, en primer lloc, cal dirigir-se a la pestanya de "*Description*" i seleccionar aquest mètode, tal com s'indica en la *Figura 10*. A més, també és necessari anomenar-lo com el corresponent mètode analític que es pretén modelar.

Figura 10. Pestanya "*Description*".

A continuació, per introduir les dades necessàries per a la construcció del model, cal accedir a la finestra "*Components*", precisament a l'apartat "*Components Table*". Aquesta pestanya s'utilitza per afegir la variable sobre la qual s'aplicarà el mètode de predicció, així com la seva unitat corresponent entre altres paràmetres requerits, tal com s'observa a la *Figura 11*.

Index	Component	Abbrev.	Unit	Digits	Analysis Low	Analysis High	Uncertainty Limit	Result
1	Índex d'hidroxil	I(OH)	mgKOH/g	2	0.00	0.00	10.00	Compute and report
2								Compute and report

Figura 11. "*Components Table*" de la pestanya "*Components*".

Un cop establert el component a partir del qual es crea el model, adreçar-se a la pestanya "*Standards*", on es carreguen els espectres de les mostres de calibratge enregistrats a l'equip *NIR*. Per exportar-los, cal clicar sobre la icona "*Open Standards*", i seleccionar la carpeta on es troben. Quan tots els espectres ja estan introduïts, es genera de forma

Index	Select	Spectrum Title	Usage	I(OH)
1		Sample_SIN METODO_Desarrollo.0	Calibration	71.10
2		Sample_SIN METODO_Desarrollo.0	Calibration	64.38
3		Sample_EMULSOGEN_Emulsogen HCl Calibration		104.10

Figura 12. "*Standards Table*" de la finestra "*Standards*".

automàtica una columna a la taula “Standards Table”, en la qual cal afegir manualment els resultats obtinguts del mètode analític de referència per a cada mostra.

Seguidament, per seleccionar l'interval de longituds d'ona amb el qual es vol treballar, accedir a la pestanya “Regions” i omplir els diferents apartats de la taula “Regions Table”, procediment visible a la *Figura 13*. Així mateix, habitualment es recomana seleccionar un rang més ampli abans de calibrar el mètode, i reduir-lo posteriorment si és necessari.

Figura 13. Apartats a omplir a la taula “Regions Table”.

Index	Region Type	Location	% or Ht	Baseline Type	Point 1	Point 2	Offset
1	Spectrum Range	4,000.00		None			
		9,000.00					

Un cop configurada la regió espectral de treball, cal clicar a la icona “Calibrate”, que apareix a la finestra principal, per calibrar el model estadístic. Aquest pas comporta que de forma automàtica aparegui una advertència per validar el mètode estadístic mitjançant una *Cross-Validation*. En aquesta nova finestra emergent, seleccionar la primera opció, la qual es pot observar a la *Figura 14*, i dirigir-se a “Continue”.

Per comprovar que aquest pas s’ha realitzat correctament, observar que l’indicador de calibratge aparegui de color verd, veure *Figura 15*.

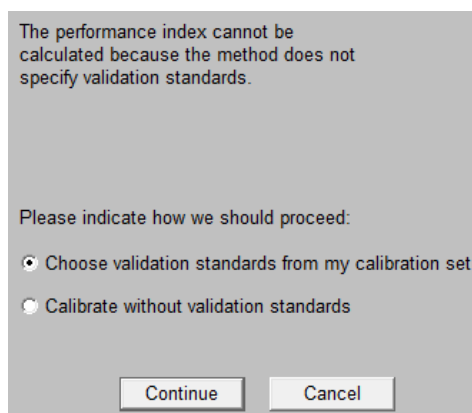


Figura 14. Advertència per realitzar la CV.

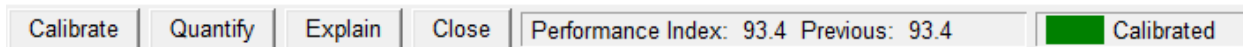


Figura 15. Indicador per comprovar si el model s’ha calibrat correctament.

Optimització del model quimiomètric

Quan el mètode ja es troba calibrat de forma adequada, és el moment d’analitzar els resultats obtinguts i començar amb el procés de millora del model.

Per una banda, un cop construït el model, si s’observa que la regió espectral seleccionada prèviament es pot ajustar de manera més acurada per tal d’enfocar l’anàlisi en les zones més informatives del senyal, millorant així la qualitat de les prediccions, aquesta es pot modificar des de la finestra “Regions” mitjançant l’opció “Edit Regions”. En clicar aquesta icona, s’obre una finestra emergent, la qual permet tornar a seleccionar el rang de longituds d’ona adequat.

D’altra banda, en cas que s’observi que l’ús de pretractaments dels espectres és requerit per tal de corregir desviacions en el senyal espectral que puguin afectar a la capacitat predictiva del model, es poden seguir dos procediments diferents, en funció del tipus de pretractament que es vulgui utilitzar i de la pestanya del programari *TQ Analyst* que se seleccioni.

En cas de voler aplicar els pretractaments primera o segona derivada i el suavitzat (*smoothing*), cal dirigir-se a la pestanya “Spectra” i, dins de l’apartat “Data Format” seleccionar el mètode basat en derivades desitjat així com a l’opció “Smoothing” triar el tipus de suavitzat que es vol aplicar, veure *Figura 16*. Un cop seleccionat, clicar a la icona “Calibrate” i comprovar que l’indicador apareix en color verd.

Figura 16. Aplicació de pretractaments a la pestanya “Spectra”.

Així mateix, quan es vulgui fer servir pretractaments relacionats amb la longitud de pas, els quals es poden observar la *Figura 17*, accedir a la pestanya “Pathlength” i a l’opció “Pathlength Type” escollir el preprocessat que es vol utilitzar. Un cop escollit, també s’ha de calibrar clicant a la icona “Calibrate” i corroborant que l’indicador es troba de color verd.

Figura 17. Aplicació de pretractaments a la pestanya “Pathlength”.

Mitjançant l’aplicació d’aquests pretractaments, es creen diferents models amb el propòsit d’identificar quin preprocessat funciona millor pels espectres que constitueixen el model, seguint com a criteri de selecció aquell que ofereix un nombre de variables latents menor, per evitar el sobreajustament de les dades, i un error de validació creuada (*RMSECV*) més baix.

Després d’haver aplicat els pretractaments espectrals adequats i ajustat la regió de treball, és hora de continuar amb les darreres millores per tal d’aconseguir un mètode estadístic molt més robust i complet. Per dur-les a terme, cal accedir a la barra d’eines i seleccionar l’opció “Diagnostics”. En fer-ho, es desplega una finestra emergent, en la qual s’ha d’escollir els paràmetres que es vulguin modificar, indicats tot seguit.

- **PRESS (Variables Latents)**

Per determinar el nombre de factors (variables latents) més adequat per al nostre model estadístic, cal dirigir-se a l’apartat “PRESS”. En aquest pas, el programari sol·licita introduir el nombre de mostres que es volen excloure en cada cicle del procés de validació interna. Un cop especificat aquest valor, s’obrirà una nova finestra en la qual apareix el gràfic que s’observa a la *Figura 18*.

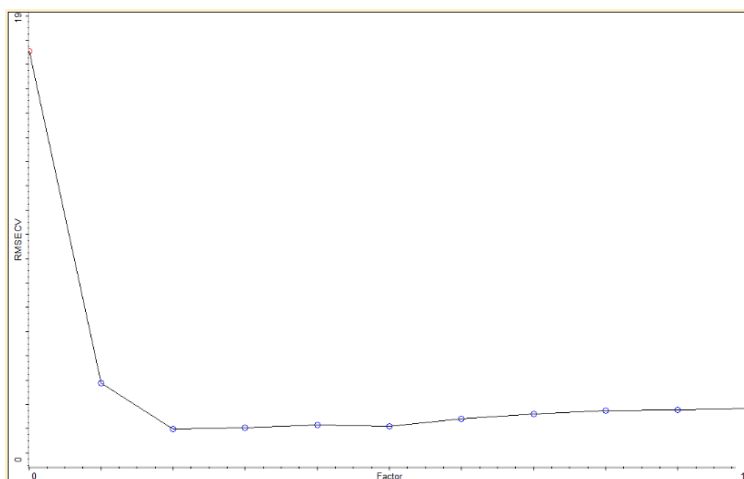


Figura 18. Gràfic *RMSECV* vs Factor (LV).

Aquest procés es realitza amb el propòsit de disposar de la major informació explicada pel model emprant el menor nombre possible de variables latents. No obstant això, cal destacar que la selecció del nombre òptim de variables latents és una decisió força subjectiva, ja que pot haver-hi més d'una opció vàlida, sempre que es pugui justificar adequadament. Així mateix, un cop definides les variables latents a utilitzar, es calibra de nou el model, seguint el procediment explicat anteriorment.

El següent pas a seguir, consisteix a identificar la presència de punts discrepants que poden influir en la capacitat predictiva del model, amb l'objectiu de valorar si eliminar-los contribueix a millorar la qualitat d'aquest. Per detectar possibles outliers s'utilitzen els següents gràfics com a criteri.

- **Residual vs Leverage**

Una de les maneres de comprovar si en el nostre model hi ha valors atípics és mitjançant el gràfic que compara el residual amb el *leverage*. Per poder accedir-hi, clicar, dintre de la finestra desplegada “*Diagnostics*”, a l'opció “*Leverage...*”. En aquest gràfic, si es detecta algun punt a la part superior o inferior dreta, com s'assenyala a la Figura 19, és a dir, amb un residual i *leverage* elevats, tant en valor positiu com negatiu, es tractaran de possibles *outliers*, ja que aquest fet indica que presenten característiques diferents en comparació a la resta de mostres de calibratge i, per tant, es pot considerar que aquestes mostres no s'ajusten correctament al model creat fins al moment.

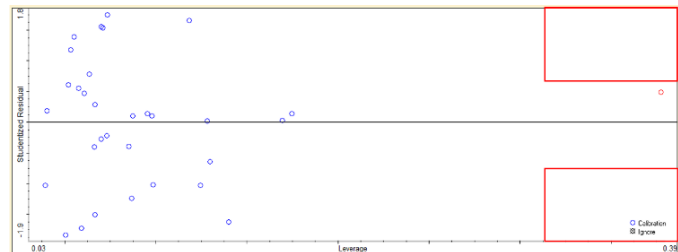


Figura 19. Gràfic Residual vs Leverage.

Com es pot observar en el gràfic proporcionat com a exemple, hi ha una mostra que podria considerar-se un possible *outlier*. Per assegurar-se'n d'aquest fet, abans de decidir no incloure-la en el model de calibratge, cal dirigir-se a l'apartat “*Spectrum Outlier*”.

- **Spectrum Outlier (Distància de Mahalanobis)**

Una altra manera de comprovar l'existència d'*outliers* és l'ús de l'opció “*Spectrum Outlier*”, disponible també dins l'apartat “*Diagnostics*”. Aquest gràfic es basa en la diferència entre cada espectre del conjunt de calibratge i l'espectre mitjà del conjunt d'entrenament, anomenada distància de Mahalanobis. Com s'ha mencionat anteriorment, aquest paràmetre permet quantificar la desviació de cada mostra respecte al comportament general del conjunt de calibratge. Per tant, com més gran és aquesta distància, més probable és que sigui un punt discrepant.

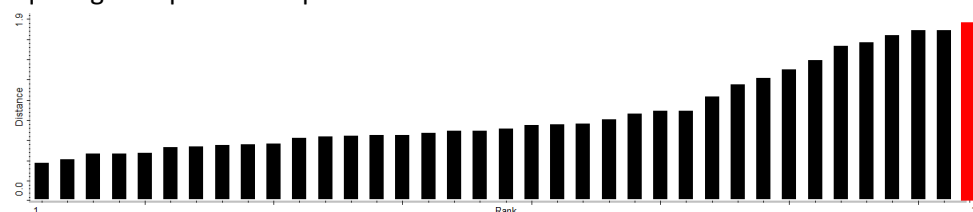


Figura 20. Gràfic Distància de Mahalanobis vs Rank.

Tanmateix, cal tenir en compte que el fet que una mostra sigui assenyalada com a possible valor atípic per alguns dels mètodes de detecció no implica que hi sigui realment. Per aquest motiu, per assegurar-se de què veritablement es tracta de punts discrepants, cal analitzar si eliminant les mostres candidates a *outliers* el model de calibratge millora o no, i, en cas d'excloure la mostra, que aquesta eliminació tingui una coherent justificació.

Validació externa

Un cop el model s'ha optimitzat de forma adequada i s'ha completat la validació interna, per tal de poder donar per finalitzat el desenvolupament d'aquest, es realitza la validació externa.

Per poder dur-la a terme, és necessari guardar correctament el mètode accedint "*File*", icona que es troba a la barra d'eines del programari, i seleccionar l'opció "*Save Method As...*", que permet desar el mètode amb el nom que es vulgui.

Tal com s'explica a l'*apartat 5.4.3*, per començar amb aquesta etapa primer de tot es recopilen un conjunt de mostres que no han estat utilitzades durant l'etapa de calibratge del model. Paral·lelament, s'enregistren els espectres corresponents, els quals s'han d'exportar en una carpeta accessible en l'ordinador que utilitzi, i es duen a terme les anàlisis dels mètodes de referència, resultats que es recopilen en un full de càlcul.

Seguidament, per tal de predir els resultats mitjançant el model creat, cal accedir a l'*apartat "Diagnostics"* del programari *TQ Analyst*. Un cop dins, dirigir-se a l'opció "*Multiple Summary – Select Files...*". De forma simultània, s'obre una finestra emergent, veure *Figura 21*, en la qual cal carregar els espectres enregistrats del conjunt de validació seleccionant la carpeta corresponent i clicant a la icona "*Open*". Aquest pas permet generar un full de càlcul amb els resultats predits per a cada mostra de validació a partir dels espectres proporcionats.

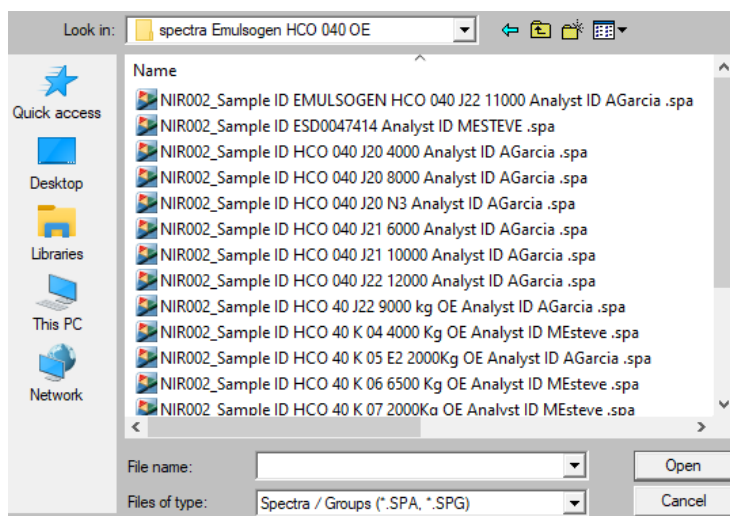


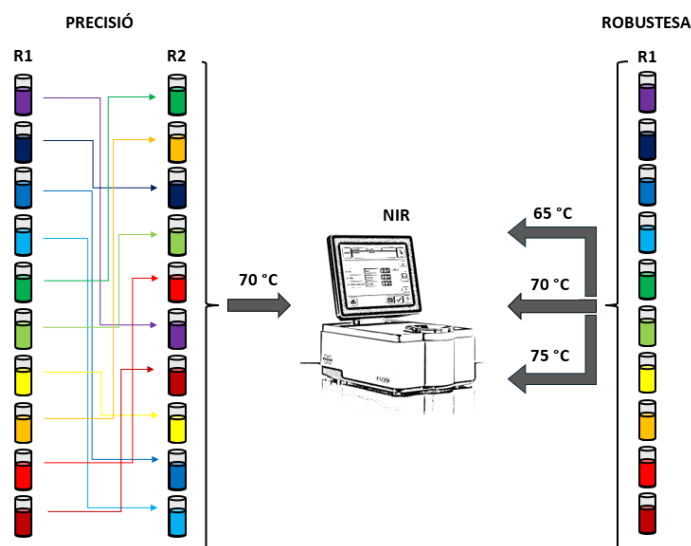
Figura 21. Finestra per seleccionar els espectres del conjunt de validació.

Per tal de poder calcular el *RMSEP* i comparar-lo amb el *RMSECV* obtingut durant la validació interna del model, és necessari crear un nou excel i calcular els paràmetres adients. Un cop determinats, si el *RMSEP* és semblant al *RMSECV*, es considera que el model prediu correctament mostres noves i, per tant, es pot donar per finalitzada la validació.

6.3.1. Estudi de precisió i robustesa

Un cop completada la validació externa, per acabar de confirmar la fiabilitat dels models creats, l'empresa Clariant considera necessari realitzar un estudi per avaluar-ne la precisió i la robustesa d'aquest, veure *Esquema 15*.

Aquests paràmetres de qualitat es determinen emprant una mateixa mostra del producte Emulsogen HCO 040®, la qual és seleccionada de forma que la seva concentració se situï als extrems de la recta de calibratge que constitueix el model. Així mateix, aquest estudi es duu a terme amb el



Esquema 15. Representació de l'estudi de precisió i robustesa.

propòsit de comprovar si es manté estable la capacitat predictiva davant possibles variacions en les condicions experimentals que poden ocórrer de forma rutinària en un laboratori.

Per una banda, per avaluar la precisió del model, es preparen 10 cubetes d'aquesta mostra, numerades de l'1 al 10. D'aquesta manera, cada cubeta es mesura dues vegades seguint un ordre a l'atzar i a una temperatura constant de 70 °C, per tal de minimitzar possibles errors sistemàtics o acumulatius.

Per altra banda, pel que fa a la robustesa, s'utilitzen les mateixes cubetes que en l'estudi de la precisió. En aquest cas, s'avalua l'efecte de la temperatura en el model, i, per a això, es realitzen dues mesures addicionals a temperatures lleugerament desviades respecte de l'estàndard (70 °C), és a dir, s'enregistren dues lectures a 65 °C i 75 °C. Com en el cas anterior, aquestes també s'efectuen aleatòriament.

Una vegada registrats tots els espectres, aquests s'introdueixen pel *software TQ Analyst* per predir-ne la concentració, seguint el mateix procediment emprat en l'obtenció dels valors estimats en l'etapa de la validació externa. A partir d'aquests resultats, es calcula la diferència màxima entre valors, la desviació estàndard i la desviació estàndard relativa (*RSD*), aquest últim paràmetre es determina a partir de la següent equació.

$$RSD(\%) = \frac{\text{desv. est.}}{\text{promig}} \times 100 \quad (6)$$

Segons els criteris establerts per Clariant, si aquest paràmetre estadístic és inferior al 2 %, es pot concloure que els resultats són satisfactoris i que, per tant, el model presenta una bona precisió, robustesa i capacitat predictiva, fet que ens indica que el mètode està preparat per a la seva implementació.

6.3.2. Implementació del model

Abans de procedir a la implementació del model, és imprescindible avaluar si els errors obtinguts (*RMSECV* i *RMSEP*) són acceptables i comparables. Per a això, cal contrastar-los amb les especificacions establertes pels clients dels productes analitzats. Si aquests valors es troben inclosos en l'interval d'acceptació, definit per aquestes especificacions, es poden considerar errors assumibles. Aquest fet indicaria que el model presenta una precisió suficient per ser utilitzat com a tècnica alternativa en control de qualitat tant de producte acabat com de mostres de procés.

Per exemple, si els productes estudiats disposen d'un interval d'especificació d'una amplitud de 10 unitats, i els errors obtinguts són comparables entre ells, tenint en compte el criteri utilitzat a Clariant, el qual diu que el *RMSEP* assolit durant la validació externa ha de trobar-se dins de l'interval mostrat a l'Equació 7 perquè es consideri que el model desenvolupat prediu correctament mostres noves.

$$RMSEP = RMSECV \pm 0.7 \times RMSECV \quad (7)$$

En cas que ambdós errors (*RMSECV* i *RMSEP*) superin aquestes 10 unitats, tot i ser comparables entre ells, no es podrien acceptar, ja que això implicaria que els resultats proporcionats pel model per a les diferents mostres presentin un error associat suficientment elevat com per a situar-les fora d'especificació.

Així mateix, si es determina que els errors del model són tolerables, cal redactar també tot un seguit de documentació necessària per a la seva posterior implementació. Aquesta consta de dos arxius principals: el mètode analític i l'informe quimiomètric. En aquests documents es descriu detalladament el procediment emprat per al desenvolupament del model, incloent-hi tota classe d'informació necessària per a la seva construcció, com ara els espectres utilitzats, els valors de referència i els resultats predits. Tanmateix, també és necessari que en aquest arxiu s'incorporin els gràfics generats pel programari *TQ Analyst*, les característiques més significatives del model i els resultats obtinguts en les diferents etapes prèvies: l'etapa de calibratge, la validació externa i l'estudi de precisió i robustesa.

A més, per complementar aquesta documentació, també és fonamental dur a terme una xerrada informativa de caràcter breu, la qual s'adreça habitualment als analistes i tècnics de laboratori que formen part del departament de qualitat, amb l'objectiu de comunicar que el nou mètode finalment ha estat implementat i explicar, de forma clara, el nou procediment a seguir a partir de la seva posada en marxa.

Per últim, s'actualitza el programari de l'equip, substituint el model utilitzat durant la fase de desenvolupament, que l'única funció que exercia fins aquell moment era l'enregistrament dels espectres, pel model definitiu, el qual permet la quantificació i proporciona resultats fiables de la propietat d'interès.

7. RESULTATS

En aquest apartat es detallen els resultats obtinguts en els models d'índex d'hidroxil i d'índex de saponificació. Abans de presentar-los de manera individual, es procedirà a realitzar un anàlisi dels espectres obtinguts durant l'etapa d'enregistrament, pas essencial per al desenvolupament d'aquests models *NIR* que es poden observar a la *Figura 22*.

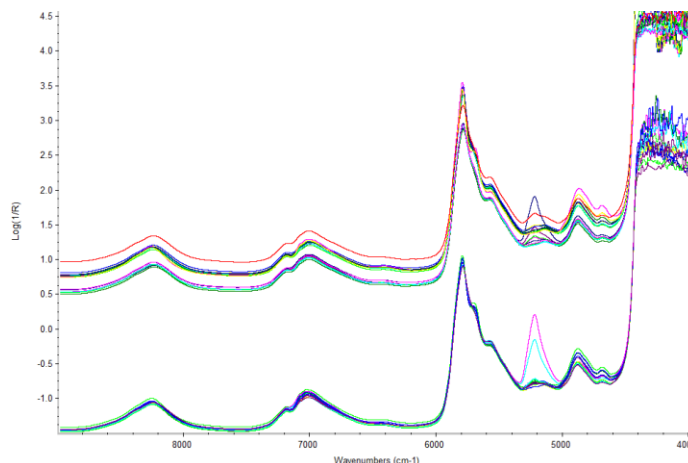


Figura 22. Espectres NIR continguts en ambdós models.

Així mateix, es presenta també una taula d'assignació de bandes *NIR* (*Taula 9*), amb l'objectiu de facilitar la posterior interpretació de les regions seleccionades en la construcció d'ambdós models. Aquesta assignació ha estat possible amb el suport de l'*Annex 1*. [11]

Avaluant el conjunt d'espectres que constitueixen aquest models s'ha detectat una zona saturada a partir dels 4500 cm^{-1} , la qual és deguda a la presència d'aigua en les mostres analitzades, ja que l'aigua absorbeix fortament en aquesta regió de l'espectre *NIR*. Aquesta condició impedeix l'ús d'aquesta franja espectral, ja que la informació que és proporcionada es veu distorsionada i podria afectar a la fiabilitat de les prediccions del model.

Malgrat aquesta limitació, la resta de l'espectre presenta pics ben definits i amb una resolució adequada, cosa que ens indica que qualsevol regió espectral que no es trobi saturada pot ser utilitzada com a regió de treball en la construcció model quimiomètric, ja que conté la informació rellevant suficient per assegurar la viabilitat de continuar amb el desenvolupament d'ambdós models.

Taula 9. Assignacions dels pics dels espectres NIR. [11]

Nº d'ona (cm ⁻¹)	Tipus de banda	Assignació	Observacions
8500 – 8000	2n sobretò	-CH	8440 – 8370 cm ⁻¹
		-CH ₂	8330 – 8260 cm ⁻¹ Stretching simètric
		-CH ₃	8400 – 8330 cm ⁻¹ Stretching simètric
7250 – 7150	1r sobretò	-OH	7170 – 7020 cm ⁻¹
7100 – 6850	1r sobretò	-OH	6970 – 6760 cm ⁻¹
	Banda combinada	H ₂ O	6900 – 6850 cm ⁻¹
		-CH ₂	7090 – 7040 cm ⁻¹ 2xCH-Stretching + 3xCH-Bending
5875 – 5760	1r sobretò	-CH ₃	5850 – 5780 cm ⁻¹ Stretching antisimètric
5760 – 5600	1r sobretò	-CH	5700 – 5630 cm ⁻¹
		-CH ₂	5760 – 5710 cm ⁻¹ Stretching antisimètric
		-CH ₃	5650 – 5600 cm ⁻¹ Stretching simètric
5450 – 5000	2n sobretò	-COOR	5180 – 5130 cm ⁻¹ 2xC=O Stretching
	Banda combinada	H ₂ O	5180 – 5150 cm ⁻¹ OH-Stretching + OH-Bending
4950 – 4500	Banda combinada	-OH	4850 – 4780 cm ⁻¹ OH-Stretching + OH-Bending
		H ₂ O	4550 – 4450 cm ⁻¹ OH-Stretching + OH-Bending

Resultats del model d'Índex d'hidroxil

Abans d'endinsar-se en els resultats d'aquest model, cal proporcionar la informació adient per comprendre'ls adequadament. Tanmateix, per a la construcció del model d'índex d'hidroxil, es van utilitzar 41 mostres: 39 mostres del producte *Emulsogen HCO 040®* i 2 d'*Emulsogen HCO 060®*. Aquestes mostres van fer possible el desenvolupament del model quimiomètric gràcies a l'obtenció dels corresponents espectres i a la recopilació dels valors de referència. Les característiques d'aquest conjunt d'entrenament s'observen a la *Taula 10*.

Taula 10. Tipus de mostres emprades en el desenvolupament del model d'Índex d'hidroxil.

Procés	Producte acabat	Punts discrepants	Total
35	2	4	41

Així mateix, l'interval de concentracions, és a dir, els valors d'índex d'hidroxil de les mostres de calibratge que constitueixen el model, abasta des de 58.50 a 131.74 mg KOH/g. Aquest rang força extens resulta especialment positiu, ja que incrementa la precisió en l'estimació d'aquests índexs en mostres noves. D'aquesta manera, com major és la variabilitat de concentracions utilitzades durant l'entrenament, més diversitat de valors es té en compte i més acurades poden ser les seves prediccions.

Un cop feta aquesta breu introducció, ens disposarem a comentar els resultats obtinguts d'aquest model de calibratge. Durant l'etapa de desenvolupament, havent emprat els mètodes de detecció d'*outliers* esmentats a l'*apartat 6.3*, es van identificar 4 punts discrepants que finalment es va decidir excloure'ls, ja que l'eliminació d'aquests comportava millores significatives en el model. Havent descartat aquests punts, els resultats estadístics del model han estat els que s'observen a la *Taula 11*.

Taula 11. Resultats estadístics obtinguts del model de calibratge d'Índex d'hidroxil.

Mètode analític	RMSECV	Variables Latents	Regio (cm ⁻¹)	r	RPD	Pretractament
Índex d'hidroxil	1.950	2	7800-8600	0.994	9.461	Primera derivada

En aquesta taula es mostren diversos paràmetres estadístics que permeten avaluar la qualitat del model desenvolupat. Considerant els criteris esmentats anteriorment a l'*apartat 5.4.2.2*, els valors obtinguts classifiquen el model com a excel·lent i apte per utilitzar-lo en qualsevol tipus d'aplicació.

Pel que fa al valor de *RPD*, aquesta se situa entre 8-10 unitats (*Taula 4*), cosa que confirma l'elevada capacitat predictiva del model. D'altra banda, el nombre de variables latents d'aquest model ha estat prou baix, un fet positiu, ja que s'evita el fenomen de sobreajustament i el model resulta més generalitzable.

Així mateix, el coeficient de correlació obtingut (r) mostra una bona correlació entre els valors predits i els de referència, tal com es pot corroborar a la *Figura 23* del model definitiu d'índex d'hidroxil.

Un altre aspecte destacable dels resultats obtinguts és la regió de treball seleccionada.

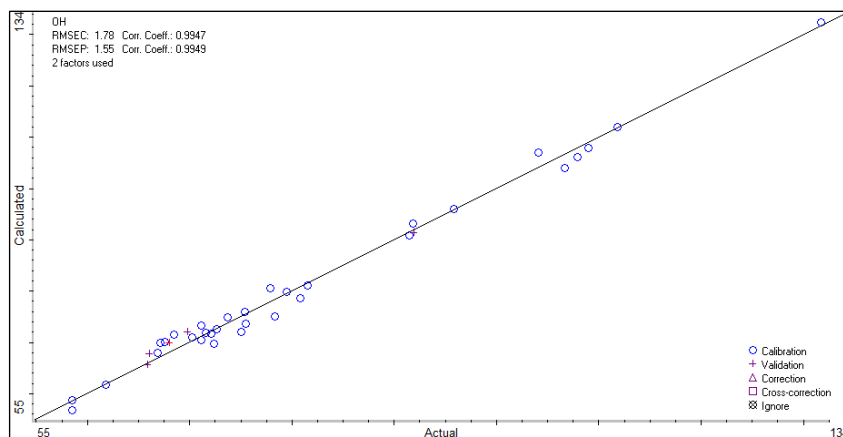


Figura 23. Gràfic de la correlació obtinguda entre els valors d'índex d'hidroxil predits (Calculated) i els de referència (Actual).

Aquesta franja, compresa entre $7800\text{--}8600\text{ cm}^{-1}$, s'assigna als grups funcionals que constitueixen la cadena alquílica dels productes analitzats, tal com es pot observar a la *Taula 9*, fet peculiar ja que aquest model ha estat desenvolupat per predir valors d'índex d'hidroxil i, teòricament, es podria esperar la selecció de bandes associades a aquests grups funcionals.

Tanmateix, la tria d'aquesta regió queda completament justificada pel seu bon comportament predictiu, ja que aquesta permet una predicció més ajustada de les mostres, fet que pot ser degut a que correspon a una franja menys afectada per altres grups funcionals, com ara l'aigua. En resum, tot i que s'observin altres franjes espectrals amb senyals més intenses, aquestes es troben més interferides, i, per tant, no contribueixen de forma positiva al model.

Un cop creat el model, es va dur a terme la validació externa per tal d'avaluar la capacitat predictiva mitjançant mostres no incloses en el seu desenvolupament, les quals presenten valors d'índex d'hidroxil compresos entre 57.47 i 81.51 mg KOH/g . Així mateix, els resultats d'aquesta fase es mostren la *Taula 12*, en la qual s'observen diversos indicadors que evidencien el bon comportament del model.

Taula 12. Resultats obtinguts durant la validació externa del model d'índex d'hidroxil.

Mètode analític	N ° mostres	Mostres de HCO 040 ®	Mostres de HCO 060 ®	Dif. Màx.	RMSEP
Índex d'hidroxil	11	9	2	1.255	0.746

En primer lloc, durant aquesta etapa s'ha detectat una diferència màxima entre els valors predits pel model i els de referència de tan sols 1.255 mg KOH/g , una discrepància força reduïda que reflecteix una elevada capacitat per reproduir els resultats obtinguts mitjançant el mètode tradicional amb molt poca desviació.

Un altre aspecte clau a comentar és l'error mitjà de predicció (*RMSEP*) que segons el criteri establert per Clariant, aquest valor es considera acceptable si el *RMSEP* es troba dins del següent interval: $RMSECV \pm$

$0.7 \times RMSECV$ (Equació 7, apartat 6.3.2). En aquest cas, la validació externa del model presenta un $RMSEP$ de 0.746, mentre que el valor de $RMSECV$ és de 1.950. Per tant, l'interval d'acceptació corresponent és 1.950 ± 1.365 (0.7×1.950), és a dir, entre 0.585 i 3.315. Com que el $RMSEP$ obtingut està inclòs en aquest rang, es pot afirmar que el model prediu correctament tant mostres noves com mostres incloses en el conjunt d'entrenament.

Així mateix, cal destacar que ambdós valors d'errors ($RMSEP$ i $RMSECV$) són considerats assumibles, ja que es troben dins de les especificacions establertes de valors d'índex d'hidroxil pels clients. Per tant, els últims resultats a comprovar abans de la implementació del model són els de l'estudi de precisió i robustesa.

Taula 13. Ordre establert per enregistrar les diferents cubetes pel NIR.

HCO 040®	R1 70 °C	R2 70 °C	65 °C	75 °C
1	7	3	8	6
2	6	8	4	3
3	10	1	7	9
4	4	9	5	10
5	9	10	3	1
6	2	7	1	4
7	1	2	10	8
8	5	4	9	2
9	8	5	6	5
10	3	6	2	7

Per últim, es va dur a terme l'estudi de precisió i robustesa, explicat en l'apartat 6.3.1. Aquest estudi va consistir en l'anàlisi de 10 cubetes d'una mateixa contramostra del producte *Emulsogen HCO 040®*, enregistrades seguint un ordre aleatori i en un mateix dia. Així mateix, la seqüència establerta tant per avaluar la precisió com la robustesa del model es pot consultar a la Taula 13, així com els resultats obtinguts d'aquest estudi, que es mostren a la Taula 14.

Fixant-nos en la primera part de la darrera taula mencionada (Taula 14), corresponent a l'estudi de precisió, es pot concloure que la precisió de l'equip *NIR Thermo* és prou bona tenint en compte els següents factors. En aquesta anàlisi s'ha obtingut un RSD (%) d'un 0.476%, molt per sota del límit del 2%, establert com a criteri d'acceptació per l'empresa Clariant. Segons aquesta condició, una desviació estàndard relativa inferior a aquest valor indica que el model desenvolupat té la capacitat de proporcionar resultats fiables i

Taula 14. Resultats estudi de precisió i robustesa pel model d'índex d'hidroxil.

I(OH)	ESTUDI DE PRECISIÓ			ESTUDI DE ROBUSTESA		
Repetició	70°C (R1)	70°C (R2)	Mitjana	70°C	65°C	75°C
1	67.100	67.960	67.530	67.530	66.800	66.500
2	68.040	67.700	67.870	67.870	67.920	67.070
3	68.150	67.640	67.895	67.895	68.530	68.840
4	67.730	68.960	68.345	68.345	68.330	68.280
5	68.700	68.240	68.470	68.470	67.990	67.080
6	68.110	69.010	68.560	68.560	67.440	67.560
7	67.960	68.290	68.125	68.125	68.260	67.560
8	68.880	68.190	68.535	68.535	69.190	67.550
9	68.370	68.510	68.440	68.440	68.620	68.370
10	67.830	68.440	68.135	68.135	67.200	68.500
Mín.	67.100	67.640	67.530	67.530	66.800	66.500
Màx.	68.880	69.010	68.560	68.560	69.190	68.840
Màx. - Mín.	1.780	1.370	1.030	1.030	2.390	2.340
Mitjana	68.087	68.294	68.191	68.191	68.028	67.731
Desv. Est.	0.477	0.439	0.325	0.325	0.682	0.707
CV	0.007	0.006	0.005	0.005	0.010	0.010
RSD (%)	0.701	0.643	0.476	0.476	1.002	1.044

repetibles, fet que reflecteix una elevada precisió en les prediccions.

A més, també s'ha observat una diferència màxima entre el valor màxim i el mínim de 1.780 mg KOH/g, fet que reforça la idea que el model presenta una estabilitat adequada. És a dir, que els resultats no es veuen afectats de forma significativa per possibles variacions en les condicions de mesura ni per l'ordre d'enregistrament de les cubetes.

D'altra banda, en aquest estudi de robustesa s'ha avaluat l'efecte de la temperatura en la predicció dels resultats. Aquest paràmetre pot influir significativament en els espectres *NIR* enregistrats, ja que la temperatura afecta la vibració dels enllaços continguts en les mostres, alterant així la resposta espectral i, com a conseqüència, els resultats. Per comprovar que el model és robust en aquest aspecte, es van realitzar mesures a tres temperatures diferents: 65 °C, 70 °C i 75 °C, és a dir, amb una variació de ± 5 °C respecte a l'estàndard. S'ha escollit aquesta diferència, ja que aquesta reflecteix situacions reals que poden produir-se en un ambient rutinari de laboratori.

En aquest cas, els valors de *RSD* (%) obtinguts, han estat lleugerament superiors als observats en les condicions de temperatura establertes (70 °C), fet que afirma la sensibilitat d'aquesta classe d'equips a la temperatura. Tot i això, aquestes desviacions relatives continuen sent inferiors al valor del 2%, i la diferència màxima observada ha estat de 2.390 mg KOH/g.

Per tant, a partir dels resultats esmentats, es pot concloure que el model es manté estable davant fluctuacions de temperatura i disposa d'una precisió adequada en la predicció de les mostres. És per aquest motiu que, una vegada redactada la documentació adient, es considera que aquest model està llest per la seva implementació al laboratori de control de qualitat com a alternativa del mètode analític convencional per determinar l'índex d'hidroxil.

Resultats del model d'índex de saponificació

D'altra banda, en aquest projecte es va construir un altre model *NIR* com a alternativa per substituir el mètode convencional per determinar l'índex de saponificació. En aquest cas, per al desenvolupament d'aquest model, es va fer-ne ús de 43 mostres: 41 mostres del producte *Emulsogen HCO 040*® i 2 d'*Emulsogen HCW 049*®. El tipus de mostres que formen aquest conjunt de calibratge es mostren a la Taula 15.

Taula 15. Tipus de mostres emprades en la construcció del model d'índex de saponificació.

Procés	Outliers	Total
39	4	43

Tanmateix, l'interval d'índexs de saponificació de les mostres utilitzades durant la creació d'aquest model s'estén de 50.80 a 123.98 mg KOH/g. Tal i com succeeix en el model d'índex d'hidroxil, aquest rang també és força ampli, fet que es considera positiu, ja que, com més gran és la variabilitat de concentracions emprades durant l'entrenament, més diversitat de valors són considerats, afavorint així unes prediccions més precises.

També cal destacar que, durant la fase de calibratge del model, es van detectar 4 *outliers*, utilitzant com a criteri diversos mètodes de detecció de punts discrepants. Com que l'exclusió d'aquests punts comportava un augment del rendiment del model, es va optar per eliminar-los. Un cop descartats els *outliers*, es van obtenir els següents valors estadístics del model.

Taula 16. Resultats estadístics obtinguts del model de calibratge d'índex de saponificació.

Mètode analític	RMSECV	Variables Latents	Regio (cm^{-1})	r	RPD	Pretractament
Índex de saponificació	1.400	1	8000-8400	0.997	13.260	Primera derivada

A la *Taula 16* es presenten certs paràmetres estadístics que s'utilitzen freqüentment per avaluar la qualitat d'aquests tipus de models *NIR*. Tenint en compte els criteris destacats prèviament (*apartat 5.4.2.2*), els valors obtinguts classifiquen la qualitat del model com a superior, fet que indica que aquesta alternativa es pot considerar tan bona com el mètode de referència.

Pel que fa al valor de *RPD*, aquest supera les 10 unitats (*Taula 4*), resultat que reforça la idea que el model creat podria substituir al mètode analític convencional, ja que aquest presenta una elevada capacitat per realitzar prediccions fiables i precises. Així mateix, el coeficient de correlació obtingut mostra una relació

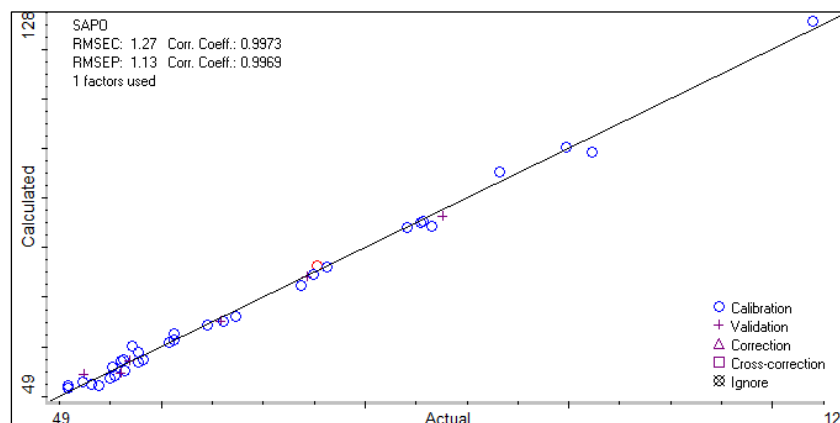


Figura 24. Gràfic de la correlació obtinguda entre els valors d'índex de saponificació predits (*Calculated*) i els de referència (*Actual*).

lineal molt bona entre els resultats proporcionats pel model *NIR* i els valors de referència, fet que es reflecteix visualment a la *Figura 24*.

Un altre paràmetre estadístic fonamental en el qual fixar-se és el nombre de variables latents. En aquest cas, aquest ha estat molt reduït, només una unitat, aspecte que resulta extremadament positiu, ja que redueix el risc de *overfitting* i contribueix a crear un model caracteritzat per una notable capacitat de generalització i estabilitat davant noves dades.

D'altra banda, un element rellevant a destacar dels resultats obtinguts és la regió espectral que ha estat seleccionada en la fase de desenvolupament del model. Concretament, s'ha escollit l'interval comprès entre 8000 i 8400 cm^{-1} , el qual s'associa als grups funcionals presents a la cadena alquílica dels compostos analitzats, tal com s'indica a la *Taula 9*. Aquesta elecció pot causar certa incertesa en un principi, ja que, en tractar-se d'un model destinat a predir l'índex de saponificació, seria raonable l'ús de bandes relacionades amb els grups èster, situades típicament en la regió dels 5450 i 5000 cm^{-1} . Tot i això, aquesta

última regió no va ser seleccionada perquè es trobava interferida per altres grups funcionals que afectaven el comportament predictiu del model.

Una vegada desenvolupat el model, es va procedir a realitzar la validació externa mitjançant l'ús de mostres no incloses en l'etapa de calibratge que mostraven valors d'índex de saponificació compresos entre 44.61 i 60.79 mg KOH/g. Tanmateix, a la següent taula, es poden observar els resultats obtinguts que reflecteixen una bona resposta del model.

Taula 17. Resultats obtinguts durant la validació externa del model d'índex de saponificació.

Mètode analític	N ° mostres	Mostres de HCO 040 ®	Mostres de HCO 060 ®	Dif. Màx.	RMSEP
Índex de saponificació	11	9	2	0.980	0.568

Tal com s'observa a la *Taula 17*, en aquesta etapa la diferència màxima obtinguda entre els valors predits i els de referència ha estat només de 0.980 mg KOH/g. Aquesta desviació tan reduïda indica que el model disposa d'un rendiment elevat per realitzar estimacions de mostres noves molt properes als resultats mitjançant el mètode convencional, aspecte corroborat mitjançant l'error mitjà de predicció (*RMSEP*). Aquest paràmetre és considerat acceptable a l'empresa Clariant si es troba inclòs en l'interval esmentat anteriorment ($RMSECV \pm 0.7 \times RMSECV$).

En el cas d'aquest model, el *RMSEP* obtingut és de 0.568 mentre que el *RMSECV* calculat durant la validació interna creuada ha estat d'1.400. Considerant aquests darrers errors l'interval d'acceptació obtingut per aquest model és de 1.400 ± 0.980 (0.7×1.400), en altres paraules, entre 0.420 i 2.380. Per aquest motiu, tenint en compte que *RMSEP* se situa dins aquest rang, es pot confirmar que el model prediu correctament tota menada mostres d'aquests productes, incloses o no en el model.

A més, cal destacar que tant el *RMSEP* com el *RMSECV* es troben dins l'amplitud dels límits d'especificació dels valors d'índex de saponificació establerts pels clients i, per tant, es poden considerar errors assumibles.

Tanmateix, els darrers paràmetres a validar abans de la implementació del model corresponen als de l'estudi de precisió i robustesa. Aquest s'ha realitzat seguint el mateix procediment utilitzat per al model d'índex d'hidroxil, amb les mateixes condicions experimentals i emprant la mateixa mostra del producte *Emulsogen HCO 040®*, seguint exactament l'ordre aleatori especificat a *Taula 13*. Els resultats obtinguts en aquest estudi es poden observar a la següent taula.

D'una banda, l'anàlisi de l'estudi de precisió realitzat evidencia que la precisió de l'equip *NIR* és notablement superior en el model de predicció de l'índex de saponificació en comparació amb el d'índex d'hidroxil. Aquesta conclusió s'extreu del fet que, en aquest cas, s'ha obtingut un valor de *RSD* encara més baix, en concret, de 0.140%, valor que es troba significativament per sota del llindar del 2% establert com

a límit segons Clariant. A més, la diferència màxima observada entre el valor més alt i el més baix és de 1.780 mg KOH/g, una variació que, al ser tant reduïda, indica que el model és capaç de predir resultats repetibles per a una mateixa mostra, i que aquests no es veuen distorsionats per possibles variacions en les condicions experimentals.

D'altra banda, els resultats obtinguts en l'avaluació de la robustesa del model d'índex de saponificació han estat molt positius. De fet, els valors de desviació estàndard relativa obtinguts en les anàlisis a diferents temperatures, amb una variació de $\pm 5^\circ\text{C}$ respecte a l'estàndard, no

Taula 18. Resultats estudi de precisió i robustesa pel model d'índex de saponificació.

Índex de saponificació	ESTUDI DE PRECISIÓ			ESTUDI DE ROBUSTESA		
Repetició	70°C (R1)	70°C (R2)	Mitjana	70°C	65°C	75°C
1	53.150	52.960	53.055	53.055	53.250	53.390
2	52.920	52.860	52.890	52.890	53.190	52.820
3	52.810	52.820	52.815	52.815	53.200	53.050
4	52.750	53.060	52.905	52.905	53.280	53.000
5	52.920	52.930	52.925	52.925	53.070	52.890
6	52.820	52.940	52.880	52.880	53.050	52.950
7	52.720	52.840	52.780	52.780	53.180	52.980
8	52.980	52.720	52.850	52.850	53.360	52.910
9	52.850	52.950	52.900	52.900	53.310	53.110
10	52.710	52.900	52.805	52.805	53.240	53.080
Mín.	52.710	52.720	52.780	52.780	53.050	52.820
Màx.	53.150	53.060	53.055	53.055	53.360	53.390
Màx. - Mín.	0.440	0.340	0.275	0.275	0.310	0.570
Mitjana	52.863	52.898	52.881	52.881	53.213	53.018
Desv. Est.	0.128	0.088	0.074	0.074	0.093	0.150
CV	0.002	0.002	0.001	0.001	0.002	0.003
RSD (%)	0.242	0.167	0.140	0.140	0.175	0.283

s'han allunyat notablement del *RSD* corresponent a la temperatura estàndard. Així mateix, en aquest estudi totes les desviacions estàndard relatives continuen sent notablement inferiors al límit del 2%, i la diferència màxima observada entre valors s'ha produït en l'enregistrament realitzat a 75°C , amb un valor de 0.570 mg KOH/g. Aquest resultat era esperable, ja que aquesta la resposta espectral d'aquesta tècnica espectroscòpia pot veure's afectada per fluctuacions de temperatura durant la gravació d'espectres *NIR*.

En conclusió, considerats els resultats destacats anteriorment, es pot afirmar que aquest últim model ofereix una bona precisió en les seves prediccions i mostra una notable estabilitat davant possibles variacions de temperatura. Per tant, un cop es disposi de la documentació necessària, aquest model també estarà preparat per ser implementat en el procés de control de qualitat de la família de productes de l'*Emulsogen HCO®*, com a substitut del mètode analític tradicional utilitzat per a la determinació de l'índex de saponificació.

8. CONCLUSIONS

Un cop finalitzat aquest Treball de Fi de Grau, s'han extret les següents conclusions:

- S'han adquirit habilitats pràctiques en l'anàlisi dels productes estudiats mitjançant els mètodes convencionals d'índex d'hidroxil i l'índex de saponificació, fet que ha permès familiaritzar-se amb els equips i procediments habituals en un laboratori de control de qualitat.
- S'ha aprofundit tant en el funcionament de l'espectroscòpia d'infraroig proper com del programari estadístic *TQ Analyst*, adquirint d'aquesta manera una àmplia gamma de coneixements quimiomètrics. A més, s'ha après a seleccionar els models multivariants més adequats en funció de la seva capacitat predictiva i fiabilitat.
- S'ha realitzat la construcció, la validació externa i el posterior estudi de precisió i robustesa dels models *NIR* per determinar l'índex d'hidroxil i l'índex de saponificació, per a productes obtinguts mitjançant l'etoxilació d'un alcohol gras amb òxid d'etilè. Tanmateix, durant el desenvolupament d'ambdós models, no es va poder utilitzar la regió espectral pròpia dels grups hidroxil ni dels grups èster, ja que aquesta es trobava interferida per altres grups funcionals, optant finalment per la franja espectral associada la cadena alquílica d'aquests productes.
- Els dos models multivariants desenvolupats han demostrat una bona qualitat de predicció, tal com indiquen els diversos paràmetres estadístics obtinguts. Així mateix, els valors de *RMSECV* i *RMSEP* assolits són comparables i assumibles, ja que aquests es troben inclosos en els intervals d'especificació establerts pels clients segons producte. Aquests resultats permeten definir-los com a models precisos, robustos i aplicables als tres productes analitzats i, per tant, es considera que ambdós models estan preparats per la seva implementació en processos de control de qualitat.

Tot i l'èxit obtingut en el desenvolupament d'aquests models *NIR*, es plantegen diverses propostes de millora per dur a terme en un futur:

- Una de les propostes de futur pensades consisteix a incorporar diversos productes de química similar, amb el propòsit d'augmentar la representativitat de les dades espectrals dels models i poder treure'n el màxim rendiment d'aquests.
- Així mateix, també seria una opció recomanable incloure una major quantitat de mostres, tant de producte acabat com de procés, en la validació externa d'ambdós models. L'objectiu d'aquesta iniciativa seria posar a prova el model amb un conjunt de validació més divers, afavorint així la generalització i la robustesa dels resultats proporcionats pels models *NIR*.
- Entre altres propostes, es podria estudiar la possibilitat d'incloure altres paràmetres continguts dins del pla de control d'aquests productes, especialment aquells que impliquin un temps d'anàlisi i l'ús de reactius elevats, ja que aquest fet permetria escurçar la durada d'aquestes anàlisis, incrementar l'eficiència del procés productiu i minimitzar la generació de residus, contribuint així a un procés més sostenible.

CONCLUSIONS

Once this bachelor's degree final project has been completed, the following conclusions have been drawn:

- Practical skills have been acquired in the analysis of the products studied using the conventional methods of hydroxyl index and saponification index, which allowed me to become familiar with the usual equipment and procedures in a quality control laboratory.
- Both the operation of near-infrared spectroscopy and the TQ Analyst statistical software have been deepened, thus acquiring a wide range of chemometric knowledge. In addition, I have learned to select the most appropriate multivariate models based on their predictive capacity and reliability.
- The construction, external validation, and subsequent study of precision and robustness of the NIR models have been carried out to determine the hydroxyl index and the saponification index for products obtained by the ethoxylation of a fatty alcohol with ethylene oxide. However, during the development of both models, it was not possible to use the spectral region of the hydroxyl groups or the ester groups since it was interfered with other functional groups, finally opting for the spectral band associated with the alkyl chain of these products.
- The two multivariate models developed have demonstrated good prediction quality, as indicated by the various statistical parameters obtained. Likewise, the RMSECV and RMSEP values achieved are comparable and acceptable since they are included in the specification intervals established by customers according to product. These results allow them to be defined as precise and robust models applicable to the three products analyzed. Therefore, it is considered that both models are ready for implementation in quality control processes.

Despite the success obtained in the development of these NIR models, some proposals for improvement are planned to be carried out in the future:

- One of proposed future actions is to incorporate several products of similar chemistry, with the aim of increasing the representativeness of the spectral data of the models and being able to get the most out of them.
- Likewise, it would also be an advisable option to include a greater number of samples, both finished product and process, in the external validation of the two models. The objective of this initiative would be to test the model with a more diverse validation set, thus favoring the generalization and robustness of the results provided by the NIR models.
- Among other proposals, the possibility of including other parameters contained in the control plan for these products could also be studied, especially those that involve a high analysis time and the use of reagents, since this would shorten the duration of these analyses, increase the efficiency of the production process and minimize the generation of waste, thus contributing to a more sustainable process.

9. BIBLIOGRAFIA

- [1] Clariant Ltd. Basel. (2024, 15 de març). *History*. Clariant Ltd. <https://www.clariant.com/es/Company/Clariant-at-a-glance/History> (Accedit 30/04/2025).
- [2] Clariant Ltd. Basel. (2023, 21 de desembre). *Solutions for industries*. Clariant Ltd. <https://www.clariant.com/es/Solutions/Industries> (Accedit 30/04/2025).
- [3] Clariant Tarragona. *AC-TH035 – Método interno QA/QC*; Clariant Tarragona: Tarragona, 2018; Versión 6. Basado en DIN 53240; UNE 55-730; UNE 55-731. (Accedit 07/05/2025).
- [4] Clariant Tarragona. *AC-TH007 – Método interno QA/QC*; Clariant Tarragona: Tarragona, 2018; Versión 4. Basado en ISO 3657; DIN EN ISO 3681. (Accedit 07/05/2025).
- [5] Owen-Reece, H.; Smith, M; Elwell, C.E.; Goldstone, J.C. *Near Infrared Spectroscopy*. British Journal of Anaesthesia. 1999, 82 (3), 418–426. <https://doi.org/10.1093/bja/82.3.418> (Accedit 08/05/2025).
- [6] Silicann Systems. *The difference between VIS spectroscopy and NIR spectroscopy*. Silicann Blog, September 22, 2022. [The difference between VIS spectroscopy and NIR spectroscopy](#) (Accedit 08/05/2025).
- [7] Cozzolino, D. *Advantages and limitations of using near infrared spectroscopy in plant phenomics applications*. Comput. Electron. Agric. 2023, 212, 108078. <https://doi.org/10.1016/j.compag.2023.108078> (Accedit 08/05/2025).
- [8] Rinnan, Å.; van den Berg, F.; Engelsen, S. B. *Review of the Most Common Pre-Processing Techniques for Near-Infrared Spectra*. TrAC Trends Anal. Chem. 2009, 28 (10), 1201–1222. <https://doi.org/10.1016/j.trac.2009.07.007> (Accedit 13/05/2025).
- [9] Roggo, Y.; Chalus, P.; Maurer, L.; Lema-Martinez, C.; Edmond, A.; Jent, N. *A Review of Near Infrared Spectroscopy and Chemometrics in Pharmaceutical Technologies*. J. Pharm. Biomed. Anal. 2007, 44 (3), 683–700. <https://doi.org/10.1016/j.jpba.2007.03.023> (Accedit 13/05/2025).
- [10] Wold, S.; Sjöström, M.; Eriksson, L. *PLS-regression: A basic tool of chemometrics*. Chemometrics and Intelligent Laboratory Systems. 2001 58(2), 109–130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1) (Accedit 13/05/2025).
- [11] Conzen, J.P. *Multivariate Calibration – A Practical Guide for Developing Methods in the Quantitative Analytical Chemistry*, 3rd ed.; Bruker Optik GmbH. Ettlingen, Germany, 2014. (Accedit 19/05/2025).

[12] European Medicines Agency. *Guideline on the Use of Near Infrared Spectroscopy by the Pharmaceutical Industry and the Data Requirements for New Submissions and Variations*. Science Medicine Health. London, 2014. (Accedit 19/05/2025).

[13] Clariant Tarragona. *TA-INCE-34A – Instrucción de trabajo interna QA/QC*; Clariant Tarragona: Tarragona, 2024, Versión 1. (Accedit 23/05/25).

10. ANNEXOS

10.1. Taula d'assignació de bandes *NIR*

