

**AUTENTIFICACIÓ D'OLIS D'OLIVA DE LES
DOP SIURANA I LES GARRIGUES
MITJANÇANT MESURES DE
FLUORESCÈNCIA ACOBLADES A
TRACTAMENTS QUIMIOMÈTRICS**

TREBALL DE FI DE GRAU

JUNY 2021

Alumne: Albert Miravent Sans

Tutora: Itziar Ruisánchez Capelastegui

AGRAÏMENTS

En primer lloc, agrair a la tutora del treball, la Dra. Itziar Ruisánchez Capelastegui, per haver-me donat la oportunitat de realitzar aquest estudi amb tot el seu grup del Departament de Química Analítica i Química Orgànica i per la seva constant atenció i ajuda al llarg d'aquests mesos.

També donar les gràcies a la Facultat de Química de la Universitat Rovira i Virgili (URV), per poder permetre la realització del treball i tota la seva investigació de manera presencial amb totes les mesures de precaució presents després de tots aquests mesos de pandèmia mundial sent conscient de que no era una situació gens fàcil.

Agrair al Tauler Oficial de Tast d'Olis d'Oliva Verge de Govern de Catalunya amb seu a Reus per haver subministrat totes les mostres d'oli d'oliva utilitzades al llarg del treball.

I finalment donar les gràcies a la meva família i als meus amics, que han estat en tot moment al meu costat, s'han preocupat i han donat tot tipus de suport en qualsevol moment de necessitat.

Moltes gràcies.

ÍNDEX

1. RESUM.....	1
2. OBJECTIUS	3
3. INTRODUCCIÓ.....	4
3.1. Estudis fins l'actualitat	5
4. MARC TEÒRIC	7
4.1. Tècnica instrumental.....	7
4.2. Pretractament de dades.....	11
4.2.1. Centrat (Mean Center)	11
4.3. Tècniques quimiomètriques.....	12
4.3.1. Representació de dades	12
4.3.1.1. Anàlisi de Components Principals (PCA)	13
4.3.2. Tècniques de classificació.....	14
4.3.2.1. Modelat suau independent d'analogies de classes (SIMCA)	15
4.3.2.2. Anàlisi discriminant per mínims quadrats parcials (PLS-DA).....	17
4.4. Paràmetres de qualitat.....	19
4.4.1. Corba característica de rendiment (PCC)	20
5. PART EXPERIMENTAL	23
5.1. Descripció de les mostres.....	23
5.2. Condicions de mesura	23
5.3. Tractament de dades	24
6. RESULTATS.....	25
6.1. Anàlisi dels espectres de les mostres d'oli d'oliva	25
6.2. Anàlisi dels espectres de les mostres d'oli de gira-sol	26
6.3. Anàlisi per components principals (PCA)	27
6.4. Models SIMCA	29
6.4.1. Càlcul dels paràmetres de qualitat.....	31
6.4.2. Corba PCC i paràmetres $CC\alpha$, $CC\beta$ i regió de no fiabilitat	32
6.5. Models PLS-DA	35
6.5.1. Càlcul dels paràmetres de qualitat.....	37
6.5.2. Corba PCC i paràmetres $CC\alpha$, $CC\beta$ i regió de no fiabilitat	38
6.6. Comparativa SIMCA i PLS-DA	40
7. CONCLUSIONS	41
8. BIBLIOGRAFIA	43
9. ANNEXOS.....	45

1. RESUM

Degut a que els preus de l'oli d'oliva tant del verge extra (EVOO) com el verge (VOO) són realment alts, i també per els valors nutricionals positius que proporcionen, es duen a terme moltes vegades diverses adulteracions. Aquest fet és actualment de molt interès per diversos motius, com per exemple amb l'objectiu de disminuir el preu de la producció i consegüentment obtenir més benefici econòmic en la venda o obtenir més quantitat d'oli sense tenir en compte la qualitat. Aquestes adulteracions es realitzen a partir d'altres olis. En podrien ser d'exemple l'oli de gira-sol o de soja. De fet, el Parlament Europeu cita l'oli d'oliva com un dels primers en la llista d'aliments que pateixen activitat fraudulentas en la resolució de Gener de 2014 [referència 2013/2091 (INI)].

L'objectiu principal del treball consisteix en classificar i diferenciar mostres adulterades d'oli d'oliva, tant verge extra (EVOO) com verge (VOO) de la varietat Arbequina de dues denominacions d'origen catalanes (Les Garrigues i Siurana), i posteriorment obtenir paràmetres de qualitat d'aquests models de classificació. Les adulteracions s'han realitzat amb oli de gira-sol refinat de la marca *Borges*. Les mostres han estat analitzades als laboratoris de la Facultat de Química de la Universitat Rovira i Virgili (URV) dins del departament de Química Analítica i Química Orgànica (QAIQO) mitjançant un espectrofotòmetre de fluorescència.

Les tècniques de classificació emprades han sigut *Soft Independent Modelling of Class Analogy* (model SIMCA) i *Partial Least Squares – Discriminant Analysis* (model PLS-DA). Per a obtenir els diferents paràmetres de qualitat semi-quantitatius, concretament $CC\alpha$, $CC\beta$ i la regió d'incertesa, s'han utilitzat les *performance characteristic curves* (corba PCC). La validació dels models SIMCA s'ha realitzat mitjançant un conjunt d'entrenament (*training set*) i un conjunt de prova (*test set*). Per al segon cas s'ha utilitzat un altre grup de mostres. També s'han utilitzat el càlcul de diversos paràmetres de qualitat: sensibilitat i especificitat.

Due to the high prices of the extra virgin olive oils (EVOO) and the virgin olive oils (VOO) and the positive effects of their nutritional values, it is widely known that some adulterations are carried out. There is an interest about it for several reasons, for example to decrease the producer price and get more profit in the sales or obtain more quantity of oil without considering the quality. These adulterations are produced by blending with other oils. Some examples can be sunflower or soybean oils. In fact, the European Parliament cites olive oil as one of the first on the list of foods that suffer from fraudulent activity in its January 2014 resolution [reference 2013/2091 (INI)].

The main objective of this work is to classify and differentiate adulterated samples of olive oils, both extra virgin (EVOO) and virgin (VOO) from the Arbequina variety of two Catalan certificates of origin (Les Garrigues and Siurana) and subsequently, to obtain quality parameters of these classification models. Adulterations were produced with refined sunflower oil from the brand of *Borges*. Samples have been analysed in the laboratory of the Chemistry Faculty of the Rovira i Virgili University (URV) in the Analytical and Organic Department (QAIQO) using fluorescence spectroscopy as experimental technique.

The classification methods have been Soft Independent Modelling of Class Analogy (SIMCA) and Partial Least Squares – Discriminant Analysis (model PLS-DA). To obtain some quality semi-quantitative parameters, specifically $CC\alpha$, $CC\beta$ and the unreliability region, performance characteristic curves (curve PCC) have been used. To the first models, the validation was done with a training and a test set. To the second models, another group of samples was used. Also some quality parameters were calculated such as sensibility and specificity for the purpose of validating.

2. **OBJECTIUS**

El principal objectiu d'aquest treball consisteix en realitzar models de classificació per a poder determinar adulteracions en olis d'oliva de la varietat arbequina, tant verge com extra verge, i per tant tenir la capacitat de comprovar les possibles activitats fraudulentas i l'autenticació de qualsevol mostra. L'anàlisi de les mostres es realitza mitjançant espectroscòpia de fluorescència i el tractament de dades es realitza amb el programa Matlab i el PLS Toolbox. Dintre d'aquest objectiu prioritari, es poden enunciar els següents objectius:

- Realització d'una investigació experimental i aprenentatge dels diversos problemes que sorgeixen al desenvolupar i ficar a punt un mètode d'anàlisi, fet que al llarg del grau de Química no es pot arribar a realitzar amb aquesta extensió.
- Possibilitat de treballar més profundament amb Matlab i més concretament amb el PLS Toolbox, que és un programa amb el qual no es treballa durant el grau de Química.
- Càlcul dels paràmetres de qualitat dels models de classificació. A banda dels paràmetres acceptats i més coneguts (sensibilitat, especificitat i eficiència) s'ha establert la sistemàtica per estimar paràmetres relacionats amb la concentració d'adulterant ($CC\alpha$, $CC\beta$ i regió de confiança).

3. INTRODUCCIÓ

Tal i com és conegut, l'oli d'oliva és un producte alimentari bàsic que té una gran importància en la dieta Mediterrània. Aquest gran ús és degut a que proporciona uns grans valors nutricionals per la salut ja que està format per compostos que són realment importants per al consum humà. La composició de l'oli d'oliva és principalment triacilglicèrids (98% del total) i la fracció restant està formada per una gran varietat de compostos, com per exemple fenols, tocoferols, fosfolípids, àcids grassos lliures, hidrocarburs o compostos volàtils, entre altres¹. Una altra motiu important pel qual l'oli d'oliva està tant sol·licitat, concretament l'extra verge, és que la seva producció consisteix en una única extracció mitjançant un premsat mecànic en fred de les olives sense la necessitat d'afegir solvents ni cap tractament tèrmic ni químic posterior².

La IOC (*International Olive Council*) juntament amb la Comissió Europea han publicat regulacions i estàndards amb les quals defineixen diversos tipus d'olis d'oliva segons el seu grau de qualitat. Principalment existeixen quatre categories d'oli d'oliva: extra verge (EVOO), verge (VOO), refinat (ROO) i d'orujo (POO). Però el primer i el segon es poden considerar dins de la mateixa categoria: olis d'oliva verges. La principal diferència entre aquests dos és la composició d'àcid, expressat com a àcid oleic. Per a l'extra verge aquesta quantitat no supera els 0.80 grams per 100 grams, i en canvi per al verge no supera el valor de 2 grams per 100 grams. L'oli d'oliva refinat és obtingut a partir d'oli d'oliva verge a partir de processos de refinament, tal i com indica el seu nom, però que no provoquen alteracions en la estructura inicial de triacilglicèrids. Igual que en els casos anteriors, també està marcat per la quantitat d'àcid oleic i en aquest cas no supera els 0.30 grams per 100 grams. I per últim l'oli d'orujo correspon a l'oli obtingut mitjançant tractaments físics o amb solvents de l'orujo de l'oliva. En aquest cas hi ha diversos tipus, com per exemple cru, refinat o una mescla amb olis d'oliva verge³.

Per aquests motius, l'oli d'oliva verge i sobretot l'extra verge són molt apreciats a causa del seu gust i els grans beneficis que proporciona per a la salut. Però a vegades es pot modificar la seva qualitat mitjançant frau durant els processos de producció i de comercialització. Per això hi ha un gran interès en les possibles activitats fraudulentas que es poden produir al oli d'oliva. Un exemple actual són les adulteracions. Aquestes es poden produir amb olis més barats, com per exemple de palma, blat de moro, soja, girasol, refinats, etc. D'aquesta manera els olis no compleixen els límits marcats per la IOC

però encara apareixen etiquetats com extra verges o verges i en canvi el productor obté un gran benefici venent mescles de menor qualitat a preus més alts.

Per tant, és important per als consumidors i minoristes que els olis d'oliva compleixin la qualitat assignada. Quan es venen realment aquestes adulteracions, no es pot identificar a primera vista si es tracte d'un oli original o d'una mescla, ja que tenen pràcticament les mateixes qualitats físiques (color, sabor, etc.). Aleshores, com es poden diferenciar? Aquí és on la química analítica té un paper fonamental.

Mitjançant aquest estudi s'intentarà detectar adulteracions d'olis d'oliva, tant extra verge com verge, de la varietat Arbequina i de les denominacions d'origen Les Garrigues i Siurana mitjançant l'espectroscòpia de fluorescència acoblades a tècniques quimiomètriques.

3.1. Estudis fins l'actualitat

Amb l'objectiu d'obtenir informació de com s'han detectat les diverses activitats fraudulentades de l'oli d'oliva, tant adulteracions o autenticacions, s'ha procedit a realitzar una recerca bibliogràfica dels últims anys. S'ha marcat un interval dels últims 5 o 6 anys, és a dir, amb antiguitat inferior al 2015.

La conclusió principal que s'extreu és que hi ha una àmplia varietat d'estudis amb un gran ventall de tècniques d'anàlisi i tècniques quimiomètriques emprades.

Pel que fa a les tècniques d'anàlisi, principalment s'han utilitzat una gran diversitat de tècniques espectroscòpiques, però no s'ha trobat que ninguna sigui més utilitzada que una altra. En primer lloc s'ha observat l'ús de l'espectroscòpia infraroja, ja sigui propera (NIR)⁴ o mitjana (MIR)⁵. També una altra tècnica bastant utilitzada ha sigut la espectroscòpia UV-Visible⁵. I finalment, fent referència a la tècnica d'interès en aquest estudi, també s'ha empleat en moltes ocasions l'espectroscòpia de fluorescència. Concretament s'ha utilitzat la fluorescència molecular⁶, consisteix en fixar la longitud d'ona d'excitació i enregistrar l'espectre d'emissió. Un altre mètode de mesura és la fluorescència síncrona^{7,8}, en la qual tant els monocromadors d'excitació com els d'emissió escanegen simultàniament de manera sincronitzada aconseguint que es mantingui un interval constant ben definit entre les longituds d'ona d'excitació i

d'emissió ($\Delta\lambda = \lambda_{\text{em}} - \lambda_{\text{ex}}$). Fins i tot s'han arribat a realitzar diversos estudis on es realitza una comparació d'aquestes tècniques⁹.

Per una altra banda, s'utilitza la quimiometria per a l'estudi de les dades recollides (espectres). En aquest cas s'ha trobat un predomini clar d'algunes tècniques. Primer de tot cal diferenciar que hi ha tècniques de classificació i de calibració, en funció de si es realitza un anàlisi qualitatiu o quantitatiu. La majoria d'estudis realitzen prèviament un anàlisi de dades mitjançant la tècnica PCA amb l'objectiu de reduir la dimensionalitat de les dades, veure quin és el comportament de les mostres analitzades i poder treure alguna conclusió, com per exemple si hi ha alguns grups de mostres o no, si hi ha alguna diferent a la resta, etc. A posteriori, les tècniques de classificació més utilitzades han sigut el PLS-DA^{7,10} i el SIMCA^{4,5,10}, depenent de si l'objectiu és modelar una classe o més d'una. Després s'han empleat altres tècniques amb menys freqüència, com per exemple, KNN, SVM, OPLS-DA, UNEQ, etc¹¹.

4. MARC TEÒRIC

4.1. Tècnica instrumental

L'espectroscòpia de fluorescència¹² és una tècnica òptica, dintre del terme *fotoluminescència*, que consisteix en realitzar una excitació de molècules mitjançant una absorció de fotons amb l'objectiu de generar un espectre d'emissió que proporciona informació per a un posterior anàlisi qualitatiu i/o quantitatiu. Normalment, aquestes emissions es produeixen a longituds d'ona superiors que les de la radiació que s'utilitza per a l'excitació. En canvi, en alguns casos pot ser que la radiació absorbida es torni a emetre sense cap canvi de freqüència. Aquest fenomen es coneix com *fluorescència de ressonància*.

En la figura 1 es pot observar un diagrama parcial de nivells d'energia per a una molècula anomenat *diagrama de Jablonski*. Tal i com es pot veure, es pot donar la promoció d'una molècula per arribar a qualsevol nivell vibracional durant el procés d'excitació electrònica. El temps de vida mitjà d'una molècula excitada en forma vibracional es inferior a 10^{-12} s i per tant la relaxació vibracional és molt eficaç. Degut a aquest fet, la fluorescència d'una dissolució sempre implica una transició des del nivell vibracional més baix de l'estat electrònic excitat.

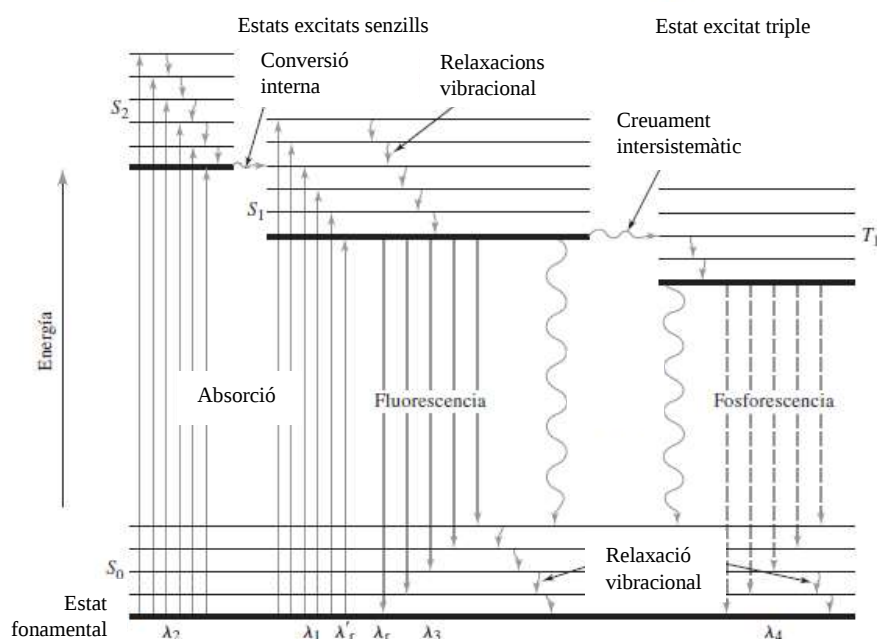


Figura 1. Diagrama de Jablonski per a una molècula fotoluminiscent

En aquest cas, les transicions energètiques electròniques que es donen no canvien l'espí de l'electró i això implica que els estats excitats on hi ha fluorescència són de curta durada ($< 10^{-5}$ s). L'estat electrònic fonamental d'una molècula és aquell en el qual tots els espins dels electrons estan aparellats i per tant no es veuen afectats quan s'exposen a un camp magnètic. Hi ha altres molècules, com els radicals lliures, que el seu estat fonamental és un estat doble. Això és degut a que l'electró té la possibilitat d'adoptar dues orientacions en presència d'un camp magnètic. Quan un d'aquests electrons és excitat passa a un nivell superior d'energia i es forma l'estat electrònic excitat senzill o triple. Tal i com es pot observar en la figura 2, en el primer cas l'espí de l'electró excitat continua aparellat amb l'electró de l'estat fonamental. En canvi, en l'estat triple els electrons ja no estan aparellats i són paral·lels ja que tenen la mateixa direcció. Per tant, es pot arribar a la conclusió que les propietats d'una molècula en l'estat triple excitat no seran les mateixes que les d'una molècula en l'estat senzill excitat.

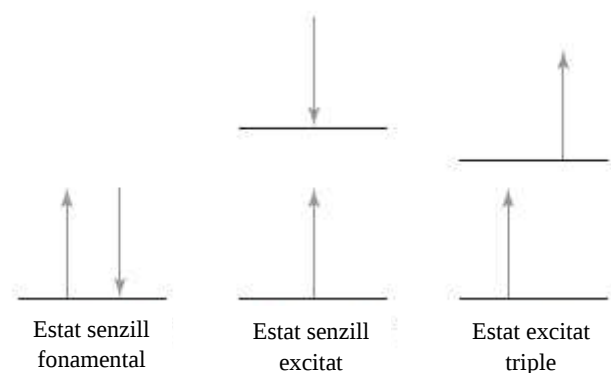


Figura 2. Estats electrònics dels espins de les molècules

La instrumentació utilitzada per a mesurar la fluorescència s'anomenen fluorímetres o espectrofluorímetres. En la següent figura es pot observar un esquema de les diferents components que formen aquests tipus d'instruments. Casi tots utilitzen un equip òptic de doble feix amb l'objectiu de compensar les possibles fluctuacions en la potencia radiant.

El primer feix passa per un filtre o monocromador, que s'encarrega de seleccionar la longitud d'ona d'excitació i transporta la radiació que causa la fluorescència. Aquesta es propaga en totes les direccions, però la observació més adequada és aquella que forma un angle recte amb el feix d'excitació ja que redueix considerablement les contribucions de la dispersió i de la radiació intensa de la font. A continuació, la radiació emesa travessa un altre filtre o monocromador, però en aquest cas selecciona la longitud d'ona d'emissió.

La radiació resultant aïllada xoca amb un fototransductor el qual la transforma en senyal elèctrica per a la seva posterior mesura.

El feix de referència travessa un atenuador amb l'objectiu de reduir la seva potència a aproximadament la de la fluorescència (disminució en un factor de 100 o més). A continuació, igual que en el cas anterior xoca contra un altre fototransductor per a obtenir una senyal elèctrica. Un sistema de dades computat processen les senyals i determinen la relació d'intensitats de la emissió de fluorescència i de la font d'excitació, donant lloc l'espectre resultant amb les dades de longitud d'ona.

La diferència entre els fluorímetres i els espectrofluorímetres resideix en els selectors de longitud d'ona. En el primer cas s'utilitzen filtres mentre que en el segon són monocromadors. També n'hi ha que són mixtes ja que utilitzen un filtre per a seleccionar la longitud d'ona d'excitació i un monocromador per a seleccionar la longitud d'ona d'emissió.

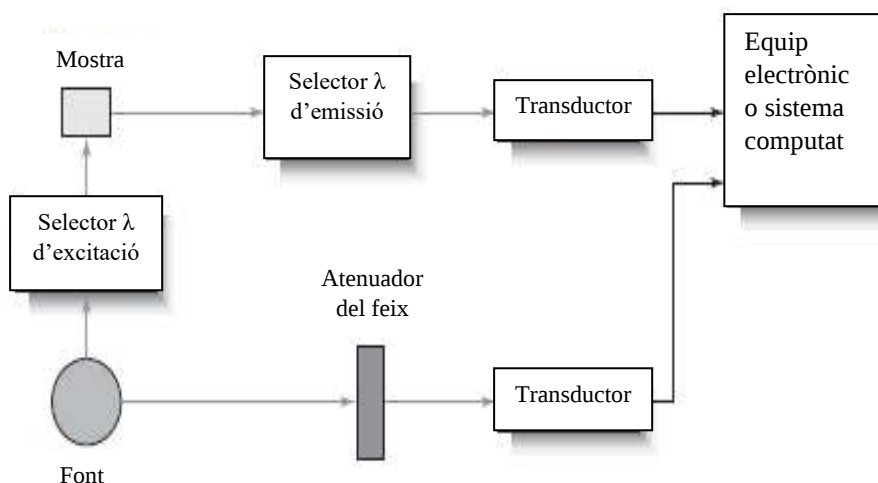


Figura 3. Parts d'un fluorímetre o espectrofluorímetre

Els components que formen aquests instruments són els següents:

- Fonts:

A diferència dels mètodes d'absorció que utilitzen làmpades de tungstè o deuteri, en aquest cas es fan servir fonts més intenses degut a que la magnitud de la senyal de sortida en aquest tipus de mesures, és a dir, la sensibilitat, és directament proporcional a la potència de la font radiant. N'hi ha de diversos tipus:

- Làmpades: Per als fluorímetres la làmpada que s'utilitza més habitualment és la de vapor de mercuri a baixa pressió amb una finestra de sílice fosa. En canvi per als espectrofluorímetres s'utilitza normalment una làmpada de xenó d'alta pressió degut a que es requereix una font de radiació continua. Per als dos instruments, en algunes situacions també s'utilitzen díodes emissors de llum blava.
- Raigs làser: Són d'un gran interès els raigs làser sintonitzables i de colorant els quals utilitzen un sistema de bombeig de raig làser de nitrogen polsant o un làser de Nd-YAG.
- Filtres i monocromadors:
Tal i com s'ha comentat anteriorment, els filtres són utilitzats per a seleccionar la longitud d'ona d'excitació com la d'emissió en els fluorímetres. En canvi, els espectrofluorímetres tenen com a mínim un monocromador i moltes vegades fins a dos, amb la mateixa funció que els filtres.
- Transductors:
Aquests components són necessaris ja que les senyals d'emissió són de baixa intensitat. Els més comuns són els tubs fotomultiplicadors. Un altre exemple és els dispositius de transferència de càrrega.
- Cubetes i compartiments per a cubetes:
Per a les mesures s'utilitzen cubetes on s'introdueix la mostra, tant cilíndriques com rectangulars, fabricades amb vidre, sílice o quars. A l'hora de dissenyar el compartiment de la cubeta s'ha de tenir en compte la reducció de la quantitat de radiació dispersada que arriba fins al detector. També s'ha d'evitar deixar empremtes dactilars a les cubetes ja que la grassa de la pell manifesta fluorescència.
- Sistema computat:
Els actuals instruments contenen diferents programes de computació per tal de gestionar les dades. Algunes funcions que inclouen aquests tipus de programes són la realització de la senyal del blanc, la correcció d'espectres d'excitació i emissió, detecció de màxims i processament, entre altres.

En aquest estudi s'utilitza per a realitzar els diversos anàlisis de les mostres d'oli d'oliva un espectrofluorímetre mitjançant la fixació de la longitud d'ona d'excitació i variant la d'emissió. Aquest tipus d'anàlisi s'anomena *fluorescència molecular*.

4.2. Pretractament de dades

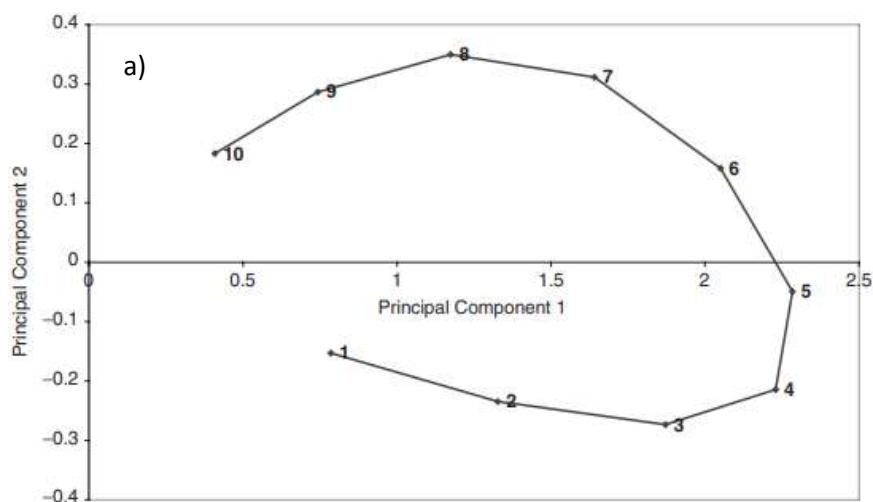
En general, abans de construir qualsevol model matemàtic o d'analitzar un conjunt de dades pot ser necessari realitzar alguna modificació d'aquestes. Aquest fet és el que s'anomena *pretractament de dades*. L'objectiu d'aquest procediment consisteix en eliminar les fonts estranyes que no són d'interès en l'anàlisi. Si no es realitza aquest pretractament per a eliminar o disminuir els errors aleatoris presents en qualsevol experiment, el model no serà correcte i serà necessari un treball bastant més complex per a obtenir la informació¹³. A continuació es comenta el pretractament de dades que s'utilitza en aquest estudi.

4.2.1. Centrat (Mean Center)

Aquest pretractament consisteix en calcular la mitjana de cada columna de la matriu de dades i restar-li a cada valor de la columna. D'aquesta manera el que s'aconsegueix es traslladar els eixos de coordenades del sistema cap al centre de les dades i es poden observar només les diferències que hi ha amb la mitjana de les dades originals¹³.

$$x_i^{centrat} = x_i^{original} - \bar{x}_i \quad \text{Equació 1}$$

Mitjançant l'equació 1 es pot observar l'expressió matemàtica del que es duu a terme mitjançant aquest pretractament. Aquest pretractament es realitza amb l'objectiu de que la màxima variància possible de les dades sigui explicada pel primer PC.



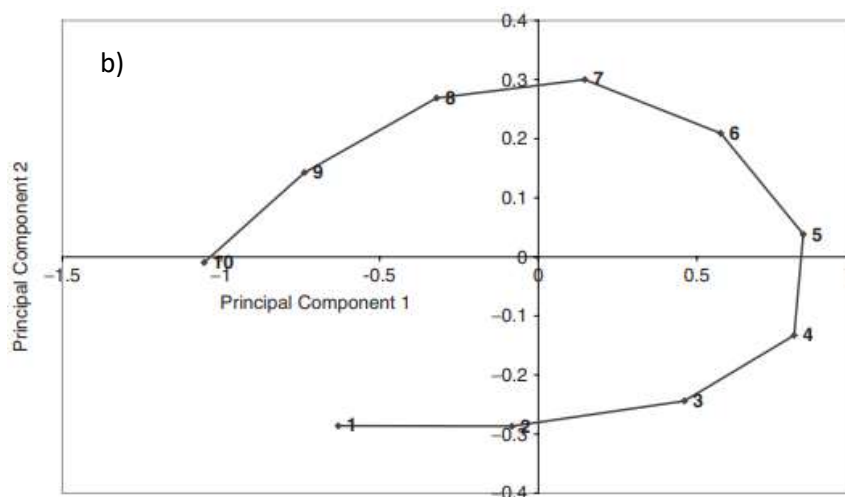


Figura 4. Gràfic de scores d'un model PCA. a) Sense pretractament; b) Amb el centrat

4.3. Tècniques quimiomètriques

Què és la quimiometria? És la disciplina química que utilitza la lògica formal, l'estadística i les matemàtiques amb l'objectiu de dissenyar procediments experimentals, obtenir la màxima informació química possible a través d'un anàlisi de dades i obtenir coneixements dels sistemes químics¹³. Aquestes tècniques s'utilitzen quan les mostres són descrites per moltes variables, com és el cas d'un espectre. Dintre d'aquesta disciplina hi ha una gran varietat de tècniques amb les quals es pot aconseguir els objectius descrits, però en aquest estudi s'utilitzen les tècniques d'anàlisi multivariant que es comenten a continuació.

4.3.1. Representació de dades

Per a representar les mostres es necessita tantes dimensions com variables s'han mesurat. I per tant, tal i com és conegut, en quan aquestes dimensions són superiors a tres, és impossible realitzar una representació gràfica per a extreure qualsevol tipus d'informació visual. És per aquest motiu que aquestes tècniques proporcionen mètodes que permeten representar aquestes dades multivariants¹⁴.

Les tècniques de representació de dades són les tècniques quimiomètriques que tenen com a objectiu mostrar quin és el comportament de les dades, és a dir, de les mostres. Mitjançant un conjunt de gràfics es pot observar quina és la distribució de les mostres en

funció de les variables mesurades, com si hi ha certes agrupacions, tendències o si hi ha alguna mostra totalment diferent a la resta i per tant poden ser considerades punts discrepants (*outliers*) i eliminades. A continuació es descriu la tècnica de representació de dades utilitzada en aquest estudi, anomenada *Principal Components Analysis* (PCA).

4.3.1.1. Anàlisi de Components Principals (PCA)

El PCA és una tècnica quimiomètrica que s'utilitza normalment com a primer pas en qualsevol anàlisi de dades per tal de detectar alguns tipus de patrons en les mostres. El que realitza aquesta tècnica és una reducció de la dimensionalitat i permet retenir de manera visual la màxima informació possible de les dades. El que es produeix és una transformació de les variables originals en unes de noves variables no correlacionades entre si anomenades *principal components* (PCs). Cada PC correspon a una combinació lineal de les variables originals. Aquesta tècnica proporciona un conjunt d'eixos ortogonals que representen les direccions de màxima variància de les dades. Una característica important d'aquestes variables és que no tots els PCs expliquen la mateixa quantitat de informació i que estan organitzats de forma jeràrquica. Els primers són els que contenen la major variància de les dades i per tant la informació més important, però en canvi els darrers corresponen als que no tenen informació tant rellevant ja que poden estar relacionats, per exemple, amb l'error experimental o el soroll de fons de l'instrument. Per tant, es pot aconseguir una gran reducció del número de variables¹⁵. En resum les noves variables són les següents:

- PC₁: conté la màxima variància del total.
- PC₂: conté la màxima variància del total que no és explicada pel PC₁. Ha de ser ortogonal al PC₁.
- PC_a: conté la màxima variància del total que no és explicada pel PC₁, PC₂, ..., i PC_{a-1}. Ha de ser ortogonal a PC₁, PC₂, ... i PC_{a-1} (on *a* és el nombre de PCs totals).

En aquesta tècnica la matriu de les variables mesurades (*X*) es descompon en 3 matrius segons l'equació 2:

$$X = TP^T + E \quad \text{Equació 2}$$

on X és la matriu de dades a tractar, T és la matriu de *scores* per a cada PC, P^T és la matriu transposada dels *loadings* i E és una matriu d'error que conté la variació residual de X que no és explicada pel model amb a PCs. Els *scores* són els valors de les mostres en el nou sistema de coordenades definit pels PCs. Els *loadings* són els coeficients que defineixen els PCs i indica quina és la contribució de les variables a cada PC¹². En la figura 5 es pot veure l'esquema de l'equació 2 on P' correspon a P^T , n és el nombre de mostres, m és el nombre de variables originals i a el nombre de PCs,

El diagrama mostra l'equació de PCA: X (matriu $n \times m$) és igual a T (matriu $n \times a$) multiplicada per P' (matriu $a \times m$) més E (matriu $n \times m$). La matriu P' està representada amb una línia horitzontal i una línia vertical que la divideixen.

Figura 5. Esquema de matrius de la tècnica PCA

Mitjançant la representació gràfica de PCs es poden observar les diferents interrelacions entre les diferents variables i detectar grups, similituds o diferències. Hi ha diversos gràfics que s'utilitzen en aquesta tècnica, però els més empleats són dos: el de *scores* i el de *loadings*. El primer té normalment el PC1 i el PC2 com a eixos de coordenades i mostren les similituds entre les mostres. En canvi, el gràfic de *loadings* és la representació de les variables originals en l'espai dels PCs i indica quina és la relació entre variables i entre variables i mostres.

4.3.2. Tècniques de classificació

Les tècniques de classificació són aquelles en les quals s'estableix una funció matemàtica amb l'objectiu de classificar mostres tenint en compte una propietat específica mitjançant mesures analítiques. És a dir, aquestes tècniques permeten realitzar assignacions d'una mostra a una classe en funció de les variables mesurades. Hi ha dos tipus de tècniques de classificació en funció de la regla de classificació que s'utilitzi: de modelatge i discriminants. Les tècniques de classificació de modelatge són aquelles que busquen similitud dins d'una classe per establir un model matemàtic per a cada classe. Per tant, la classificació dels objectes es produeix en aquella classe a la que millor s'ajusten al seu model. En canvi, les tècniques de classificació discriminants estableixen regles les quals separen les classes en zones en l'espai i assignen objectes desconeguts a alguna d'elles.

Les mostres que tenen propietats comunes entre elles estaran situades en la mateixa regió, mentre que dues mostres amb característiques diferents estaran posicionades en diferents regions. Hi ha diverses tècniques de classificació que es poden arribar a realitzar, però en aquest estudi les que s'utilitzen són el *Soft Independent Modeling by Class Analogy (SIMCA)* i el *Partial Least Squares-Discriminant Analysis (PLS-DA)*.

4.3.2.1. Modelat suau independent d'analogies de classes (SIMCA)

El SIMCA correspon a una tècnica de classificació de modelatge basada en l'anàlisi per components principals (PCA). Aquest model és el mètode més conegut per a descriure l'estructura d'una classe d'un conjunt de dades. En aquesta tècnica el que es realitza és un anàlisi PCA per a cada classe i es determina el nombre suficient de PCs que retenen la majoria de la variància per a cada classe¹⁶. Aquest nombre de PCs no té perquè ser el mateix per a cada classe.

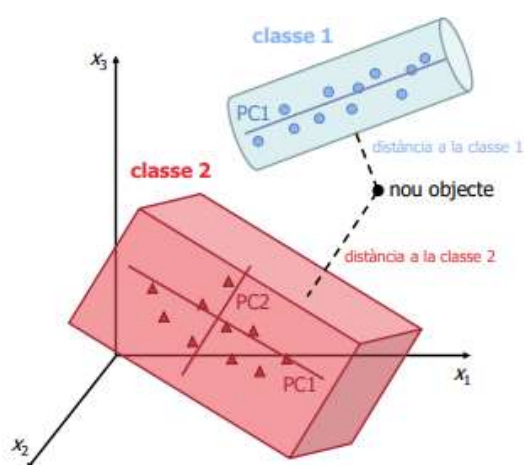


Figura 6. Realització model SIMCA per a dues classes diferents

El procediment a seguir per a la realització d'un model de classificació SIMCA és el següent:

1. Fer el PCA de cada classe amb un conjunt de mostres de classe coneguda.
2. Determinar el número de PCs òptim per a cada classe.

Tal i com es mostra en la figura 6, per a la classe 1 (blava) és suficient realitzar el model amb només un PC però en canvi per a la classe 2 (vermella) es necessiten dos PCs. Com es pot observar, depenent del nombre de PCs la regió del model de cada classe té una

forma geomètrica o una altra (en aquest cas un cilindre i un ortoedre). La selecció del nombre de PCs es realitza observant els gràfics de la variància explicada vs el número de PCs. Es pot veure en la figura 7 que arriba un moment el qual a partir d'un PC (PC_a) la variància no canvia significativament, i per tant es selecciona el nombre de PCs fins aquest menys un (PC_{a-1}).

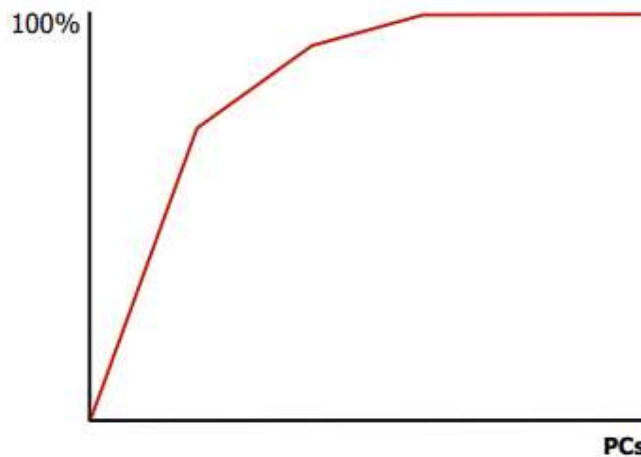


Figura 7. Gràfic variància explicada respecte PCs

3. Establir el límits de cada classe. Es calculen dos paràmetres estadístics que són els límits de cada model, és a dir, els límits a partir dels quals les mostres desconegudes es classificaran dintre les classes o no. Aquests paràmetres són la distància residual o Q_{lim} i la T^2_{lim} de Hotelling. La T^2_{lim} fa referència a la variància retinguda i la Q_{lim} correspon a la variància residual que el model no ha retingut. Per a establir aquests límits, s'utilitza un interval de confiança. Aquest estudi es realitza amb un 95%.
4. Validació del model. Hi ha dues maneres per a realitzar la validació d'un model de classificació: mitjançant la validació creuada (*cross-validation*) o mitjançant un conjunt d'entrenament (*training set*) i un altra de prova (*test set*). El primer cas consisteix en realitzar models deixant cada cop una mostra fora de la matriu de dades i a continuació predir aquesta amb els models construïts. En canvi, en la realització d'un conjunt d'entrenament i un de prova es divideix el conjunt de dades en dos grups. El *training set* s'utilitza per a establir el model i posteriorment es prediuen les mostres del *test set* per a validar-los. Com que són de la mateixa classe que les del model, la classificació hauria de ser correcta al 100%

teòricament i s'observa si el model és suficientment bo. En aquest estudi s'utilitza la segona opció de validació.

5. Predicció de mostres desconegudes. Quan es realitza aquesta predicció, s'obtenen els paràmetres T^2_i i Q_i de cada mostra. La distància Q_i correspon a la distància ortogonal de l'objecte al model. En canvi, la distància T^2_i de Hotelling mesura la distància ponderada d'un objecte al centre de la classe. Una mostra i pertany a una classe si es compleixen alhora les següents condicions:

- $T^2_i < T^2_{\text{limit, classe}}$
- $Q_i < Q_{\text{limit, classe}}$

6. L'assignació es visualitza a través del gràfic de membres (*membership plot*). Aquest gràfic fa referència a una sola classe. Es representen els valors de Q i T^2 per a cada mostra i es senyalen els límits de cada classe (Q_{lim} i T^2_{lim}). Els objectes que queden per sota de ambdós límits pertanyen a la classe corresponent.

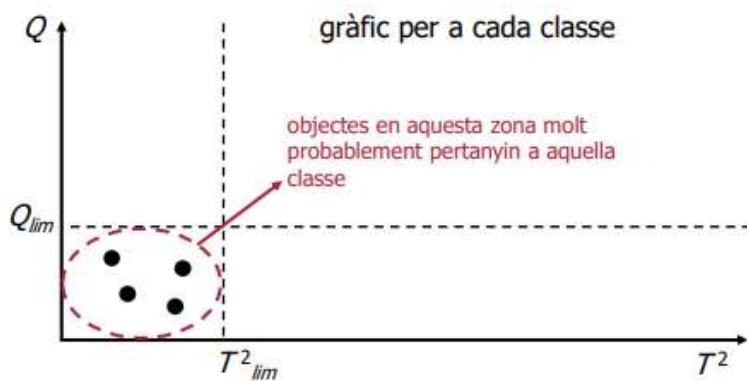


Figura 8. Membership plot del model SIMCA

4.3.2.2. Anàlisi discriminant per mínims quadrats parcials (PLS-DA)

Un model PLS-DA¹⁷ correspon a una tècnica de classificació discriminant que realitza un anàlisi basat en la regressió per mínims quadrats parcials. La regressió s'estableix entre una matriu X , la qual conté les variables independents, i una matriu Y que correspongui a les variables dependents. Per aquest estudi, la matriu X correspon a les dades espectrals de fluorescència obtingudes i la matriu Y fa referència a quina classe pertany cada mostra

utilitzant un codi binari (0,1) per a classificar. En el cas a estudi, la matriu X conté els espectres de dues classes diferents i per tant a la matriu Y s'assigna el 1 a una classe i el 0 a l'altra. A continuació es realitza un model PLS, amb el qual s'obté un model de components principals per a la matriu X tenint en compte la informació que proporciona la matriu Y.

Els passos a seguir per a realitzar un model PLS-DA són els següents:

1. El primer que es realitza és la tria del nombre de variables latents (LVs). Per a poder fer la tria, es pot mirar el gràfic de la variància explicada vs LVs. El procediment és el mateix que per a la tria de PCs en el SIMCA: el gràfic té una forma similar a la figura 7 i es selecciona el nombre de LVs més petit possible però que contingui un percentatge suficientment important d'informació (variància explicada). Normalment, a partir d'un valor de LVs la variància comença a ser pràcticament constant i aquest és el valor òptim, ja que si s'agafen més LVs s'està produint un sobre ajust al model.
2. S'estableix un valor límit entre les classes. Aquest valor separa les classes entre elles. Aquest límit de classificació és pot obtenir mitjançant el teorema de Bayes, que consisteix en realitzar la millor discriminació possible entre les classes amb la probabilitat més petita d'errors de classificació en les prediccions.
3. Validació del model. Tal i com s'ha comentat en l'apartat del SIMCA hi ha dues maneres de validar els models. A diferència que pel SIMCA, en aquest cas hem utilitzat la validació creuada (*cross-validation*).
4. Predicció de mostres de pertinença desconeguda. Quan es realitza aquesta predicció, cada mostra obté un valor de Y predita. Si aquest valor supera el valor límit de la classe comentat en l'apartat 2, aleshores serà assignada a la classe. En canvi, si la Y predita té un valor inferior a aquest límit es considerarà que no pertany a la classe. En el cas particular de tindre només 2 classes, si una mostra no es classifica a una classe queda assignada en l'altre classe.

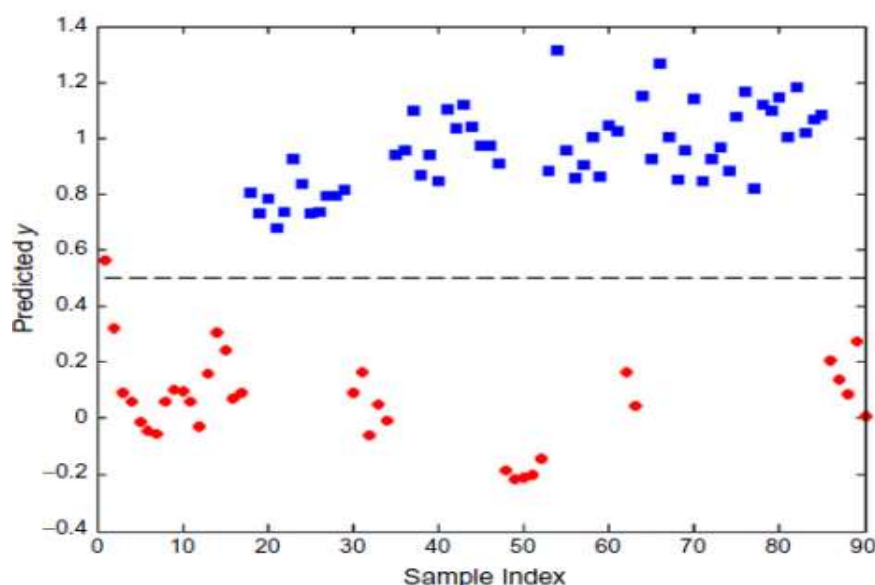


Figura 9. Exemple d'un gràfic de predicció d'un model PLS-DA

4.4. Paràmetres de qualitat

Els paràmetres de qualitat són valors que es calculen amb l'objectiu de validar els models i treure conclusions referent a la seva qualitat. En aquest estudi s'han utilitzat la sensibilitat i l'especificitat, que són descrites a continuació breument:

- Sensibilitat: aquest paràmetre indica la capacitat que té el model d'assignar correctament mostres a la seva classe. La següent equació mostra l'expressió matemàtica utilitzada per al seu càlcul:

$$\text{Sensibilitat}_j = \frac{n^{\circ} \text{ mostres de la classe } j \text{ assignades a la classe } j}{n^{\circ} \text{ mostres totals de la classe } j} \quad \text{Equació 3}$$

- Especificitat: fa referència a la capacitat del model de reconèixer mostres que realment no són d'aquella classe, és a dir, l'habilitat del model de detectar correctament mostres d'altres classes. Es pot calcular mitjançant la següent equació:

$$\text{Especificitat}_j = 1 - \left(\frac{n^{\circ} \text{ mostres que no són de la classe } j \text{ assignades a la classe } j}{n^{\circ} \text{ mostres totals que no són de la classe } j} \right) \quad \text{Equació 4}$$

4.4.1. Corba característica de rendiment (PCC)

La corba característica de rendiment¹⁸ (PCC), o *performance characteristic curve* en anglès, és una corba que és utilitzada per a estimar paràmetres addicionals de qualitat de models de classificació amb els quals es pot obtenir informació semi-quantitativa a través d'un mètode analític qualitatiu.

Per a poder explicar com es representa aquesta corba, es considera un exemple de la presència d'un contaminant que està regulat per la legislació. Per tant, aquest contaminant no pot ser superior a partir d'un valor de concentració (per exemple 5 mg/L). Es defineixen dues classes: classe no contaminada i la classe contaminada. Quan es fa la predicció de les mostres, la resposta del model de classificació només pot ser binària: resposta positiva si la mostra està contaminada i resposta negativa si no està contaminada. Aquesta assignació de positiu/negatiu s'assigna per conveni, però pot ser al revés. Per a la realització de la corba s'analitzen mostres amb concentracions al voltant d'aquest valor (per exemple, 3, 4, 6, 7 i 8 mg/L). La probabilitat d'obtenir respostes positives, $P(x)$, correspon a la freqüència d'aquestes senyals a cada concentració analitzada. Idealment, aquesta $P(x)$ hauria de ser 100% quan la concentració sigui superior a 5 mg/L i 0% quan sigui igual o inferior a aquest valor, i per tant la corba resultant seria la de la figura 10.

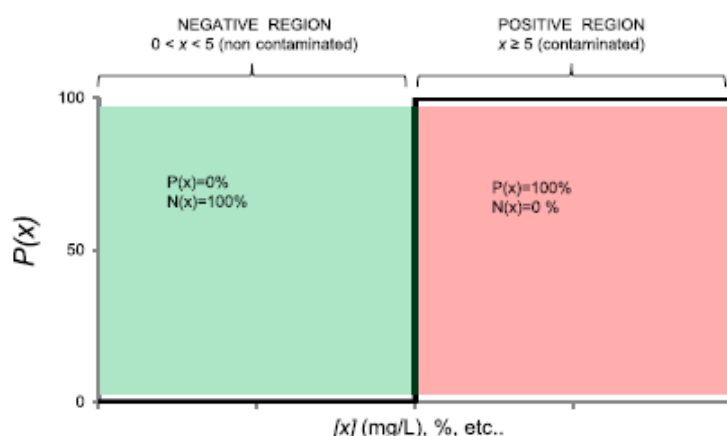


Figura 10. Corba PCC en una teòrica situació ideal

Però, com en tota experimentació, la realitat mai és igual que la idealitat. El que en realitat succeeix és que aquesta corba s'ajusta a una funció sigmoïdal de totes les possibles minimitzant l'error quadràtic dels residuals. Un exemple de funció sigmoïdal seria la següent equació:

$$p(x) = 1 - e^{-\left(\frac{x}{a}\right)^b} \quad \text{Equació 5}$$

Els diferents paràmetres d'aquesta funció són els següents: $p(x)$ correspon a la probabilitat de tenir una senyal positiva, x és la concentració d'analit en aquest exemple, i a i b corresponen als coeficients de regressió que han estat ajustats per tal de minimitzar l'error quadràtic dels residuals. El primer paràmetre correspon a l'amplitud de la corba i el segon a la pendent. Mitjançant aquest ajust, el que s'obté és una corba d'aquest tipus:

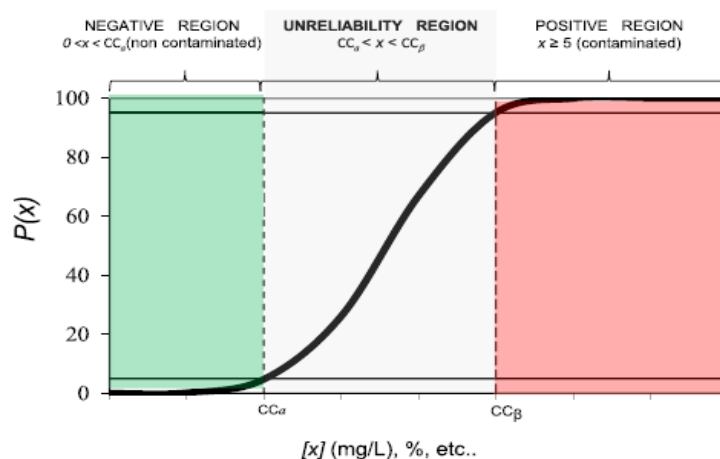


Figura 11. Corba PCC en una teòrica situació real

Tal i com es pot observar, es poden obtenir les següents concentracions límit. Per a obtenir-les s'han de traçar dues línies horitzontals al gràfic. Aquestes normalment es situen quan $P(x) = 95\%$ i 5% .

- CC_{α} : anomenat *decision limit*, és a dir, el nivell de concentració més petit que es pot detectar en una mostra amb una probabilitat de cometre un error de fals positiu (*FP: false positive*) del 5%. És a dir, considerar una mostra amb contaminant a 5 mg/L o més quan en realitat està per sota. S'obté on es creuen la corba sigmoïdal i la línia horitzontal realitzada a $P(x) = 5\%$.
- CC_{β} : anomenat *detection capability*, és a dir, és la mínima quantitat d'analit que es pot detectar en una mostra amb una probabilitat de cometre un error de fals negatiu (*FN: false negative*) del 5%. És a dir, en aquest exemple aquest error consisteix en considerar una mostra que no té contaminant a 5 mg/L o superior quan en realitat si supera aquesta quantitat. S'obté on es creuen la corba sigmoïdal i la línia horitzontal realitzada a $P(x) = 95\%$.
- *Unreliability region*: correspon a la regió compresa entre els dos anteriors paràmetres. És a dir, en aquesta zona hi ha possibilitats d'obtenir falsos resultats tant falsos positius com negatius.

Per tant, la zona amb valors inferiors a CC_α es considera com a zona de confiança ja que hi ha un 95% de probabilitat de detectar mostres que no superen el valor de contaminant legislatiu i que sigui així de cert. I per tant de manera similar, la zona amb valors superiors a CC_β també es considera com a zona de confiança ja que la probabilitat de detectar mostres que superen el valor de contaminant legislatiu és superior al 95%.

5. PART EXPERIMENTAL

5.1. Descripció de les mostres

Per aquest estudi s'ha disposat d'un total de 60 mostres d'oli d'oliva arbequina, tant extra verge com verge subministrades pel Tauler Oficial de Tast d'Olis d'Oliva Verge de Govern de Catalunya amb seu a Reus. Aquestes estaven dividides en dos grups segons la denominació d'origen catalana a la que pertanyen. En aquest cas, Siurana i Les Garrigues. Degut a que el nombre de mostres no és tant elevat com el desitjat, s'han creat noves mostres que corresponen a mescles. Per tant, finalment s'han obtingut per ambdues classes 42 mostres per a cadascuna (30 mostres pures + 12 mescles). Es pot veure la nomenclatura de les mostres en l'Annex 1.

5.2. Condicions de mesura

Les mesures de les diferents mostres s'han realitzat mitjançant la tècnica de fluorescència molecular. Els diferents espectres han sigut obtinguts amb l'equip Shimadzu RF-5301PC. Les condicions de mesura han sigut les següents:

- Longitud d'ona d'excitació: 350 nm
- Rang de mesura de l'espectre d'emissió: 360-800 nm
- Temps d'integració: 0.1 s
- Interval de longitud d'ona: 10 nm

Per a la mesura de les diferents mostres, un cop s'ha omplert la cubeta és important assegurar que no es formen petites bombolles, ja que s'ha pogut observar que afecten clarament als espectres.

A l'hora de dur a terme preparació de les mostres adulterades amb oli de gira-sol, s'ha seleccionat aleatòriament 15 mostres de cada classe i s'han adulterat als següents nivells (% w/w) per ambdues classes: 0.5, 1, 2, 5, 7.5, 10, 12.5, 15, 17.5, 20, 25, 30 i 35.

5.3. Tractament de dades

Un cop registrats els espectres, les dades han sigut exportades i tractades amb el programa Matlab, versió Matlab R2012b i el seu complement PLS Toolbox. També s'ha utilitzat el programa Excel per a realitzar els ajustos sigmoïdals necessaris per a obtenir la corba PCC, mitjançant l'eina Solver i optimitzant els paràmetres de l'equació sigmoïdal utilitzada fent mínim el sumatori dels residuals al quadrat.

6. RESULTATS

6.1. Anàlisi dels espectres de les mostres d'oli d'oliva

El primer que s'ha realitzat és la representació del conjunt de dades, tant de la denominació de Siurana com la de Les Garrigues conjuntament, amb l'objectiu d'observar els espectres.

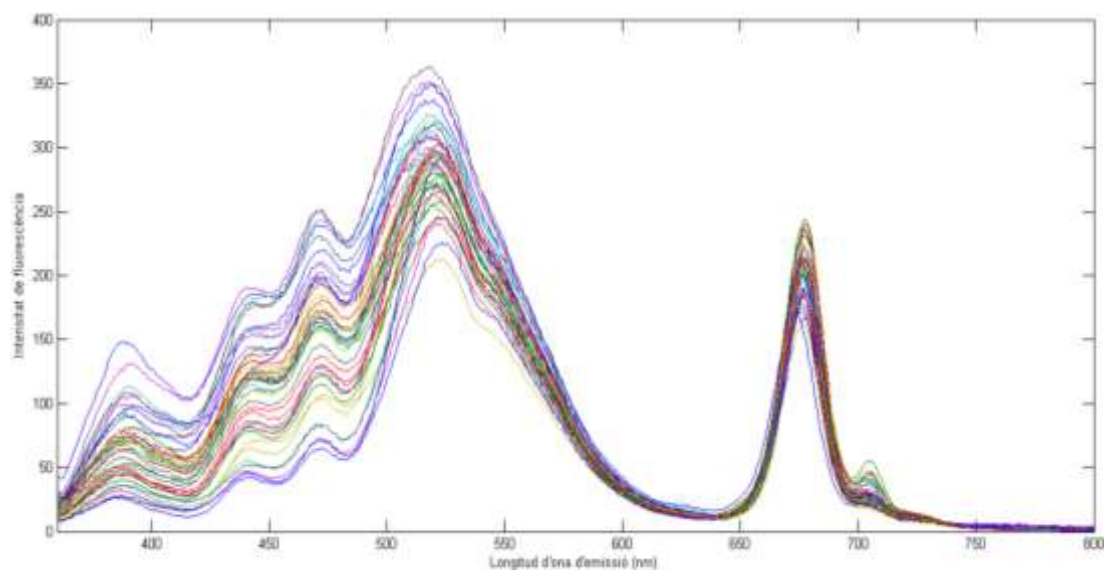


Figura 12. Espectres de fluorescència d'emissió de les 42 mostres d'oli d'oliva DO Siurana

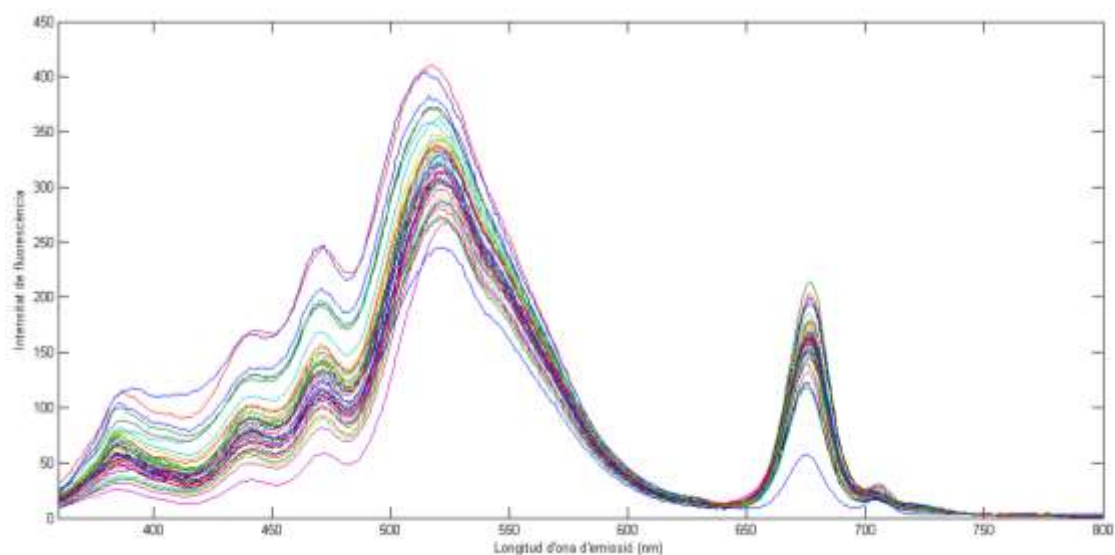


Figura 13. Espectres de fluorescència d'emissió de les 42 mostres d'oli d'oliva DO Les Garrigues

Tal i com es pot observar en les figures 12 i 13, es poden diferenciar 5 bandes en tot el rang de l'espectre. Les primeres corresponen a intensitats baixes, la quarta té la intensitat més gran de tot l'espectre i la cinquena és la que té una intensitat inferior a l'anterior però superior a totes les primeres. El primer pic cau sobre 380-390 nm, el segon a uns 440 nm, el tercer a uns 470 nm, el quart a uns 520 nm i el darrer a uns 675 nm.

Les bandes de l'espectre de fluorescència dels olis d'oliva són degudes a la presència en l'oli de la clorofil·la, els carotenoides i els productes d'oxidació dels àcids grassos. També són rics en compostos fenòlics i tocoferols. Les bandes de baixa intensitat en el rang de 440-470 nm corresponen als hidroperòxids conjugats i a les diverses oxidacions dels àcids grassos. Els dobles i triples enllaços no conjugats dels lípids passen a ser conjugats (diens i triens) degut a aquestes oxidacions. La banda a 520 nm fa referència a la presència de carotenoides (β -carotè i luteïna) que actuen com a antioxidants i prevenen l'oxidació de l'oli. I per acabar, l'últim pic a 675 nm correspon a la clorofil·la. És un pigment fotosintètic que conté magnesi. En els processos de refinació d'oli, la quantitat de clorofil·la disminueix aproximadament un 80% en els diversos processos físics i mecànics⁶.

6.2. Anàlisi dels espectres de les mostres d'oli de gira-sol

El primer que s'ha fet ha sigut analitzar el oli de gira-sol pur per tal de veure quines regions presenten la màxima intensitat de fluorescència i d'aquesta manera es pot relacionar amb com hauria d'afectar a mesura que s'augmenta el seu percentatge d'adulteració. En la següent figura es mostra l'espectre:

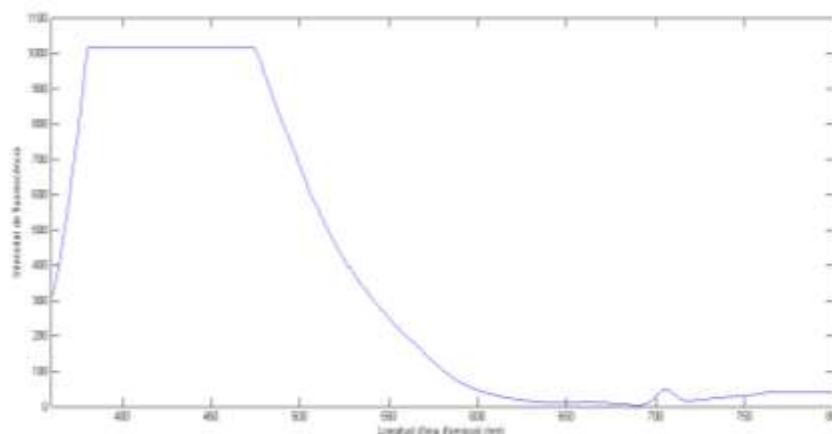


Figura 14. Espectre de fluorescència d'emissió d'oli de gira-sol

Es pot observar que hi ha una regió inicial on l'oli de gira-sol proporciona una gran intensitat (satura el detector) de fluorescència. Això implica que aquesta regió és la que principalment hauria d'anar variant a mesura que s'augmenti el percentatge d'adulteració en les mostres. Aquest fet es pot observar en l'exemple de la figura 15. Els espectres corresponen tots a la mateixa mostra però a diferents nivells d'adulteració. S'observa un augment de la intensitat en aquesta regió a mesura que el percentatge és superior.

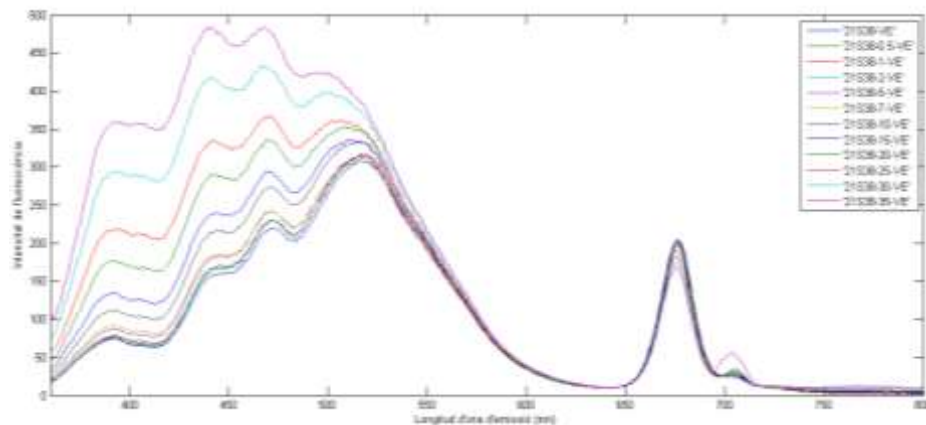


Figura 15. Espectres de la mostra 38 de Siurana: pura i diferents nivells d'adulteració

6.3. Anàlisi per components principals (PCA)

Abans d'establir els models de classificació, es fa un anàlisi per components principals (PCA) per tal de veure possibles agrupacions, tendències o mostres discrepants (*outliers*), en funció de l'espectre de les mostres.

La figura 16 mostra els gràfics dels *scores* per cada classe i per les 2 classes juntes i es poden treure les següents conclusions.

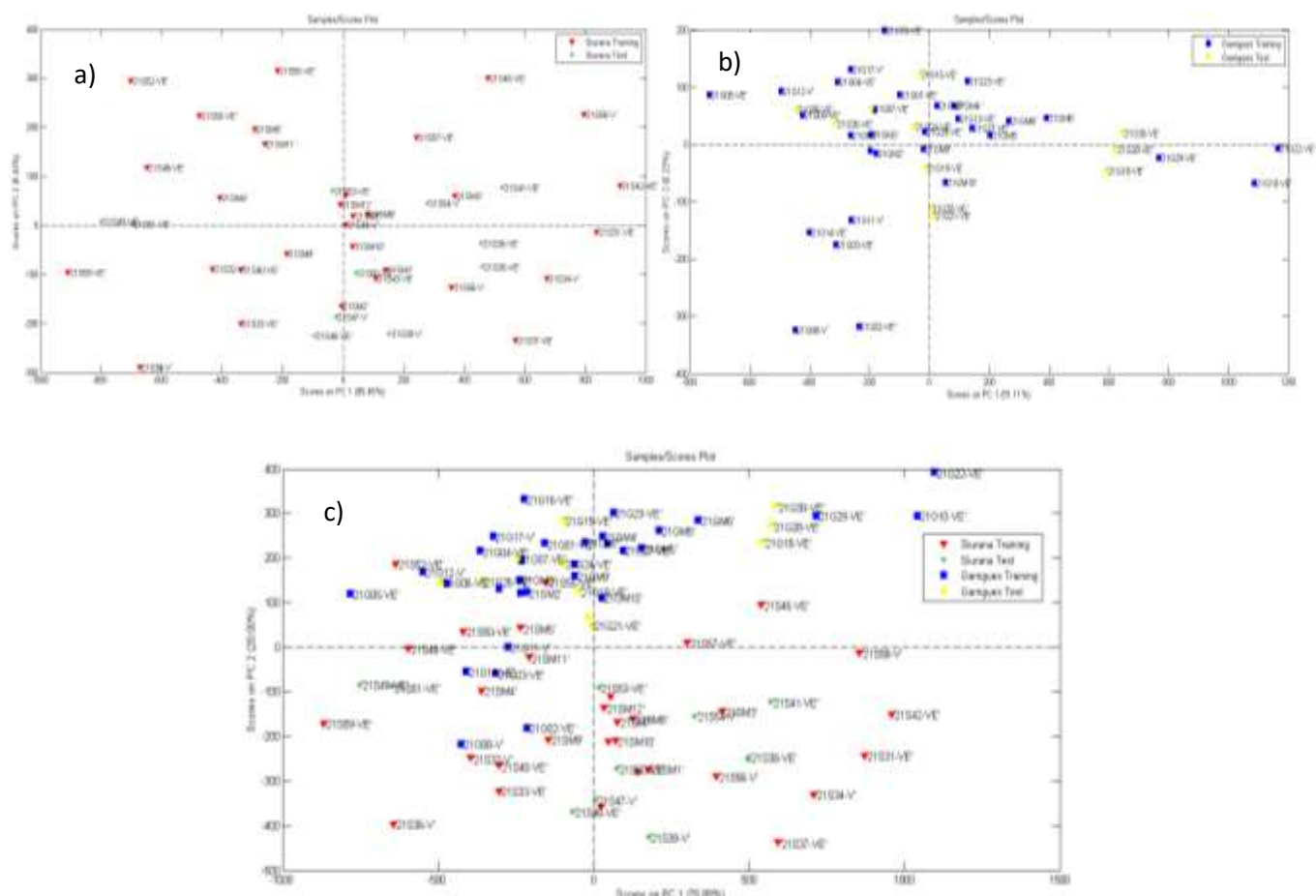


Figura 16. Gràfic de scores del PCA de: a) Siurana; b) Les Garrigues; c) Siurana + Les Garrigues

En tots 3 gràfics es pot observar una distribució homogènia de les mostres. Es pot destacar que al PCA de les Garrigues (Fig. 16b), hi ha algunes mostres que s'allunyen lleugerament de la resta: dos (2 i 8) amb valors més grans i negatius al PC2 i altres dos mostres (10 i 22) amb valors grans i positius al PC1.

Al gràfic de scores de mostres de les dues classes juntes (Fig. 16c), encara que no hi ha una separació clara de classes, les mostres s'agrupen al llarg del PC2. Majoritàriament les mostres de les Garrigues tenen un valor positiu del score al PC2 i per les mostres de Siurana és negatiu.

6.4. Models SIMCA

Amb el model SIMCA es proposa fer una aproximació *one-class* (una classe), el que vol dir que només s'estableix 1 model per la classe no adulterada. Es fa la predicció de les mostres adulterades en aquest model, que haurien de quedar fora del model (no pertany).

Per a la creació dels models SIMCA, primer s'han de triar el nombre de PCs a partir del gràfic de la variància explicada, tan per a cadascuna de les classes (Siurana i Garrigues) com per a les dues juntes. Per als models individuals de cada classe, s'han utilitzat 4 PCs ja que, com es pot observar com a exemple en la figura 17 que correspon a la classe de Siurana, 4 PCs expliquen el 99.29% de la variància. Amb un cinquè PC, aquest explica el 0.34% més i per tant no canvia significativament. I per a la classe Les Garrigues, 4 PCs expliquen el 99.61% de la variància. Amb un cinquè PC, aquest explica el 0.12% més i per tant no canvia significativament. En canvi, per al model realitzat amb les dues classes juntes amb 3PCs ja n'hi ha suficient, ja que expliquen el 98.06% de la variància però hi ha una tendència clara a estabilitzar-se al augmentar els PCs a partir d'aquest.

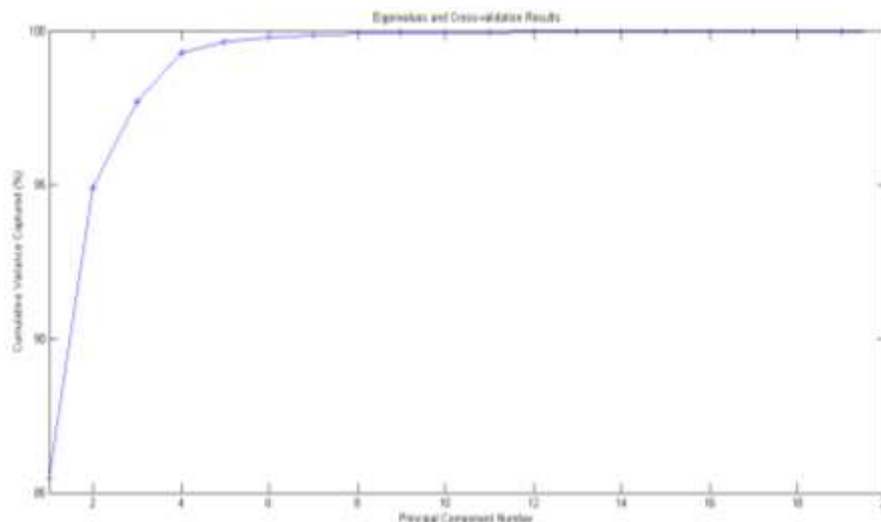


Figura 17. Gràfic variància explicada vs PCs de la classe Siurana

A continuació s'ha procedit a realitzar els models SIMCA. Aquests models es realitzen mitjançant un conjunt d'entrenament (*training set*) i un conjunt de prova (*test set*). La selecció de les mostres que formen part d'un grup o l'altre es realitza mitjançant el gràfic de *scores* de PC1 vs PC2. S'escull un nombre suficientment representatiu per al primer grup, conclouent que el *training set* tindrà 31 mostres i el *test set* 11. Les mostres del

training set estan distribuïdes homogèniament per tot el gràfic de *scores*, havent-hi d'aquesta manera una bona variabilitat escollint mostres del centre i també dels extrems. En canvi per al model realitzat amb les dues classes conjuntament, de les 84 mostres totals disponibles, s'han utilitzat 54 mostres per al *training set* i 30 mostres per al *test set*. Aquestes 30 són les mateixes que s'utilitzen per als diferents nivells d'adulteració (les 15 de cada classe). En les figures 16a, 16b i 16c les mostres seleccionades pels *test sets* es mostren encerclades. Es poden observar quines són les mostres utilitzades per a cada grup en l'Annex 1.

La classificació/assignació de les mostres a la classe modelada es fa a través del gràfic Hotelling- T^2 vs Q (figura 18, a) i b) pels models establerts per cada classe individual i c) per al model de les dues classes conjuntes). Les mostres que tenen uns valors de Hotelling- T^2 i Q inferiors als límits del model (línies blaves) queden assignades a la classe.

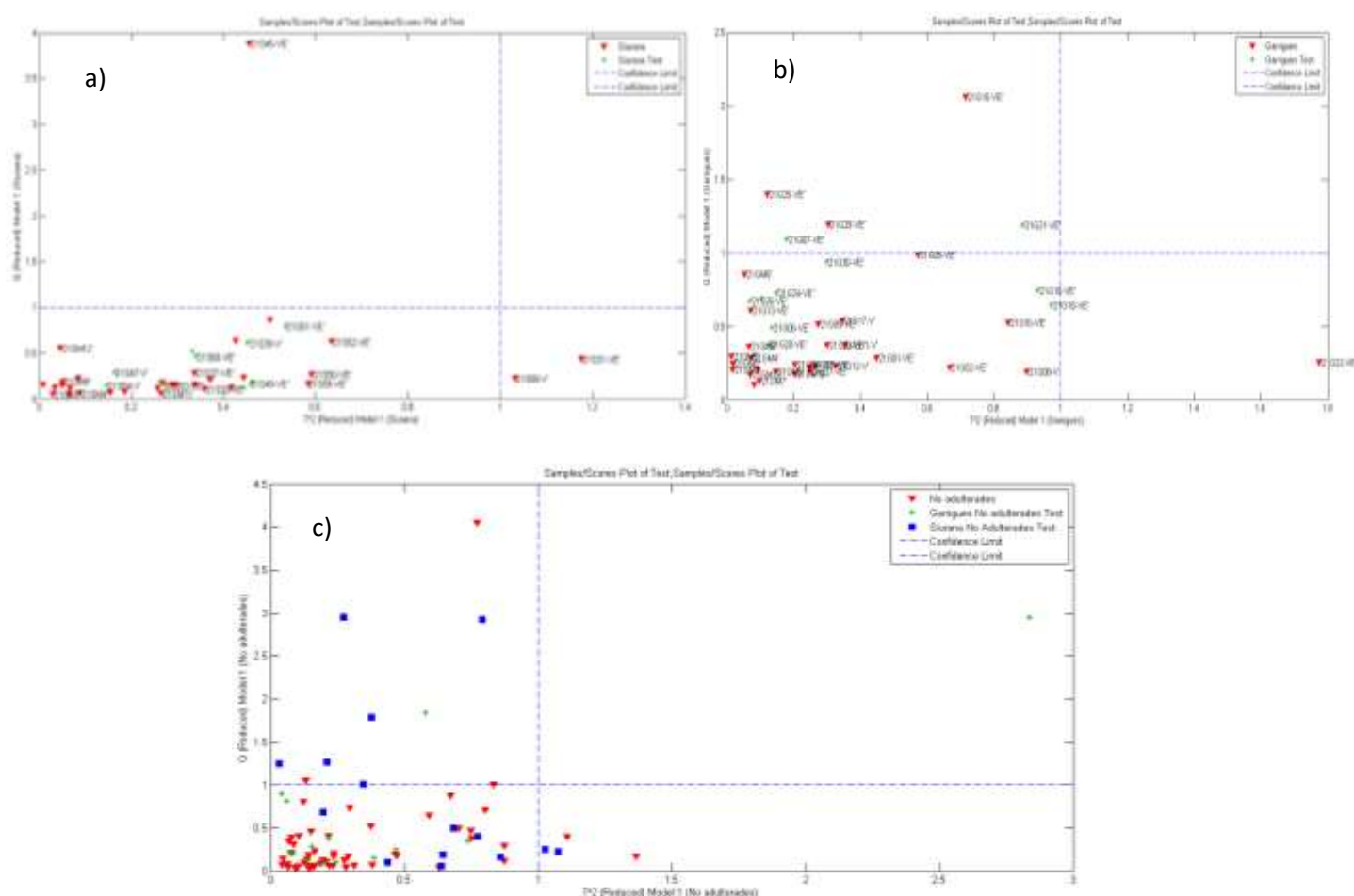


Figura 18. Gràfic T^2_i vs Q_i del training i test set del SIMCA de: a) Siurana; b) Garrigues; c) Siurana + Les Garrigues

Es pot observar com en el cas del model per a la classe de Siurana (fig. 18a) les mostres del *training set* (triangle vermell) queden totes ben assignades exceptuant 3 mostres i les del *test set* (creu verda) són totes classificades dintre. Per una altra banda, pel cas del model per a la classe de Les Garrigues les mostres del *training set* queden totes ben assignades excepte 4 mostres i les del *test set* són totes classificades dintre excepte 2 mostres. I finalment per al model de Siurana i Les Garrigues conjuntament queden classificades fora del model 5 mostres del *training set* i 10 mostres del *test set*.

6.4.1. Càlcul dels paràmetres de qualitat

Pel càlcul de la sensibilitat, també es fa la predicció del conjunt de prova (*test set*) per als diferents models segons l'equació 3, que queden mostrats en la següent taula:

Taula 1. Sensibilitats dels models SIMCA

Paràmetres de qualitat	Models SIMCA		
	Siurana	Les Garrigues	Siurana + Les Garrigues
Sensibilitat <i>training set</i> (%)	87.10	83.87	87.04
Sensibilitat <i>test set</i> (%)	100	63.64	66.67

Pel càlcul de l'especificitat, es fa la predicció del conjunt de mostres adulterades a cada nivell. Per a cada predicció s'obté el gràfic de T_i^2 vs Q_i on es pot observar l'assignació a la classe “No Adulterada” si es situen per sota dels límits ($Q_i < 1$ i $T_i^2 < 1$). En canvi si es situen per damunt l'assignació correspon a “Adulterat”. Un exemple d'aquests és el següent:

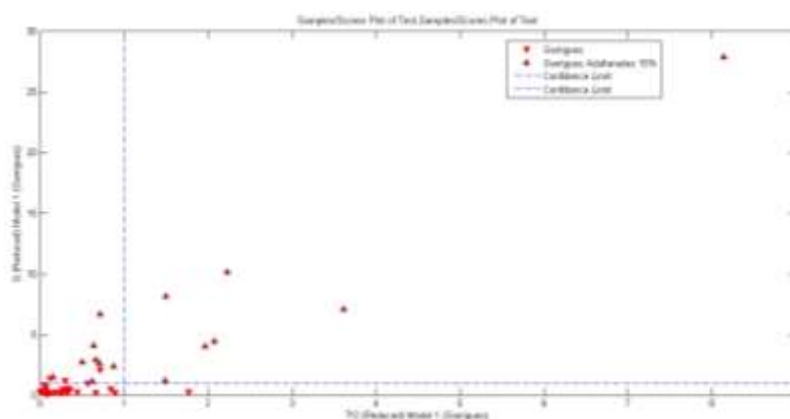


Figura 19. Gràfic de T_i^2 vs Q_i de la predicció de mostres adulterades al 15% de Les Garrigues

Posteriorment s'han extret les dades que proporcionen aquests gràfics. És a dir, s'obtenen els valors de Q i T^2 de cada mostra en un arxiu Excel i es produeix l'assignació definitiva dintre o fora de la classe “No Adulterat” per a cada mostra.

Taula 2. Especificitats de cada nivell d'adulteració per a cada model SIMCA

	Models SIMCA		
Nivells d'adulteració (%)	Siurana	Les Garrigues	Siurana + Les Garrigues
0	0.00	36.36	33.33
0.5	20.00	80.00	36.67
1	20.00	46.67	33.33
2	20.00	46.67	40.00
5	40.00	66.67	40.00
7.5	40.00	100	50.00
10	53.33	93.33	50.00
12.5	73.33	100	86.67
15	73.33	100	63.33
17.5	73.33	100	83.33
20	80.00	100	73.33
25	93.33	100	100
30	100	100	100
35	100	100	100

6.4.2. Corba PCC i paràmetres $CC\alpha$, $CC\beta$ i regió de no fiabilitat

Es considera el valor de l'especificitat o la probabilitat de resposta positiva $P(x)$ (és a dir, adulterat) per a cada nivell d'adulteració. Per al 0% (no adulterat), el valor de $P(x)$ correspon al valor de 100-sensibilitat del model. La corba PCC es calcula mitjançant la realització d'un ajust sigmoidal. Aquest tipus de corba es pot crear a partir de diverses equacions, però l'escollida per aquest estudi és la de l'equació 5, on a i b són els paràmetres a optimitzar i la següent taula mostra aquests valors:

Taula 3. Paràmetres a i b de l'equació 5 un cop realitzat l'ajust sigmoidal dels models SIMCA

	Models SIMCA		
Paràmetres optimitzats	Siurana	Les Garrigues	Siurana + Les Garrigues
a	10.97	1.19	8.06
b	0.89	0.40	0.56

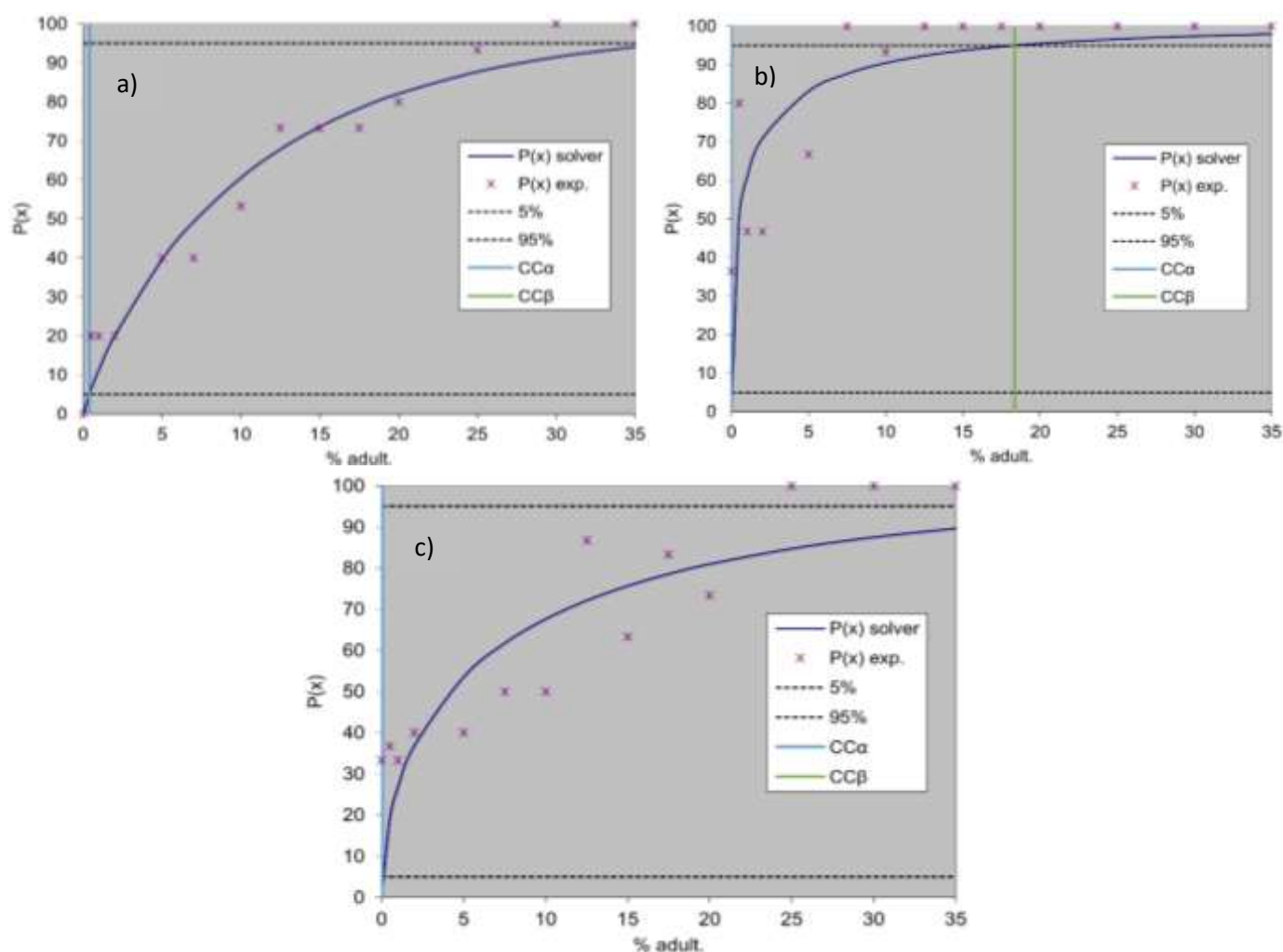


Figura 20. Corba PCC obtinguda del model SIMCA de: a) Siurana; b) Garrigues; c) Siurana + Les Garrigues

Tal i com mostra la figura 20, no es pot arribar a observar una corba sigmoidal, sinó que en realitat es produeix una corba exponencial. O el que és el mateix, el resultat que dona és la meitat superior de la corba sigmoidal teòrica. Aquest fet es produeix perquè el model correspon a la classe no adulterada (0%) i ràpidament a l'inici augmenta el valor de $P(x)$ ja que no tenim una sensibilitat del 100% i a nivells d'adulteració molt petits, la $P(X)$ és molt elevada.

A partir de les corbes, es calculen els paràmetres $CC\alpha$ i $CC\beta$, que corresponen quan $P(x)$ equival a 5% i 95% respectivament, però pel segon cas s'ha de realitzar una extrapolació de l'ajust per als casos dels models de Siurana i Siurana + Garrigues, ja que queda fora de l'interval dels nivells d'adulteració utilitzats per aquest estudi. El càlcul d'aquests valors es realitza amb la següent fórmula, que no és més que l'equació 6 però aïllant la x enlloc de $P(x)$:

$$x = a \cdot [-\ln(1 - p)]^{1/b} \quad \text{Equació 6}$$

En la següent taula es mostren els valors d'aquests paràmetres obtingut en les corbes PCC dels diferents models SIMCA:

Taula 4. $CC\alpha$, $CC\beta$ i regió d'incertesa dels models SIMCA

	Models SIMCA		
Paràmetres corba PCC	Siurana	Les Garrigues	Siurana + Les Garrigues
$CC\alpha$	0	0	0
$CC\beta$	37	18	58
Regió d'incertesa	0 - 18	0 - 18	0 - 58

Al tractar-se d'un adulterant, és molt important poder conèixer com es comporta el model no adulterat enfront a diferents percentatges d'adulteració, informació que es pot obtenir de la corba PCC i dels límits $CC\alpha$ i $CC\beta$.

Tal i com es poden observar aquests valors conjuntament amb els gràfics de les corbes representades en l'apartat anterior, amb els models SIMCA es produeix un augment molt ràpid de $P(x)$ a nivells d'adulteració molt propers a 0 i per tant s'obté $CC\alpha = 0$. Aquest valor ens està dient que a un percentatge molt petit d'adulterant, el podem diferenciar d'una mostra no adulterada però només en un 5% dels casos, la majoria (95%) les identificarà erròniament com no adulterades. Per exemple pel cas de Siurana (Fig. 20a) de cada 100 mostres adulterades al 0.5% al voltant de 80 les donaria com no adulterades, ja que la $P(X)$ és 20%. I també es pot observar que concretament pels models realitzats amb la classe Siurana i amb les dues classes conjuntes el valor de $CC\beta$ són molt grans i queden fora de l'interval utilitzat per a realitzar les corbes. El valor més gran amb el que s'ha realitzat la corba va ser d'una adulteració al 35%. Per tant només la corba establerta per Les Garrigues (Fig. 20b) es pot afirmar que de cada 100 mostres adulterades al 18% només 5 les donaria erròniament com no adulterades.

Cal recordar que aquestes corbes donen informació sobre mostres adulterades. Sobre les mostres no adulterades, la informació l'obtenim directament a partir de la sensibilitat del model desenvolupat.

6.5. Models PLS-DA

A l'hora de dur a terme el model PLS-DA es proposa fer una aproximació per a dues classes, aquestes dues seran “No Adulterat” i “Adulterat”. Però com tenim diferents nivells d'adulteració, quin de tots ells s'utilitza per a realitzar el model? S'ha determinat utilitzar el nivell d'adulteració al 15% ja que és el que millor s'assembla la corba PCC obtinguda a una forma sigmoïdal. Per als models individuals de Siurana i Les Garrigues s'han utilitzat 42 mostres per a la classe “No Adulterat” i 15 mostres per a la classe “Adulterat” (les del 15% d'adulteració). En canvi pel model de les dues classes conjuntes s'ha tingut una disponibilitat de 54 mostres per a la classe “No Adulterat” i 30 mostres per a la classe “Adulterat” (les del 15% d'adulteració de cada classe).

El primer pas ha consistit en seleccionar el nombre de LVs. Es realitza de la mateixa manera que per a la selecció del nombre de PCs en els models SIMCA. S'observa el gràfic de la variància explicada vs LVs i s'escull el valor més petit possible de LVs que contingui la màxima variància possible i que comenci a tenir una tendència estabilitzar-se.

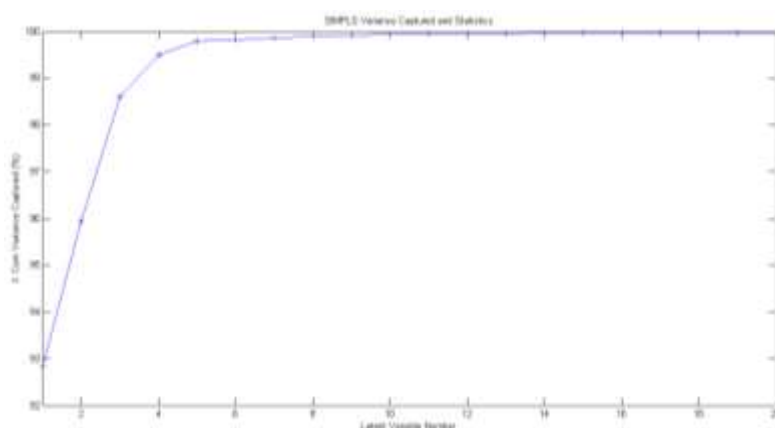


Figura 21. Gràfic variància explicada vs LVs de la classe Les Garrigues

En aquest cas, que correspon al model PLS-DA per a la classe Les Garrigues, s'han seleccionat 4 LVs sent la variància explicada de 99.49%. El mateix succeeix per als

models de Siurana i de Siurana + Garrigues. En aquestes s'explica una variància del 99.19% i del 98.44% respectivament.

A continuació s'ha procedit a realitzar els models PLS-DA i s'obté el valor límit o Y límit per cada classe. Una mostra queda assignada a la classe si el seu valor de Y predit és més gran que la Y límit. La figura 22 mostra el valor límit per la classe no adulterada i els valors de Y predits per les mostres no adulterades (triangle vermell) i les mostres adulterades (creus verdes).

S'obtenen els següents gràfics per a cada model, on es pot observar quin és la Y límit i es representa cadascuna de les mostres segons la Y predita corresponent.

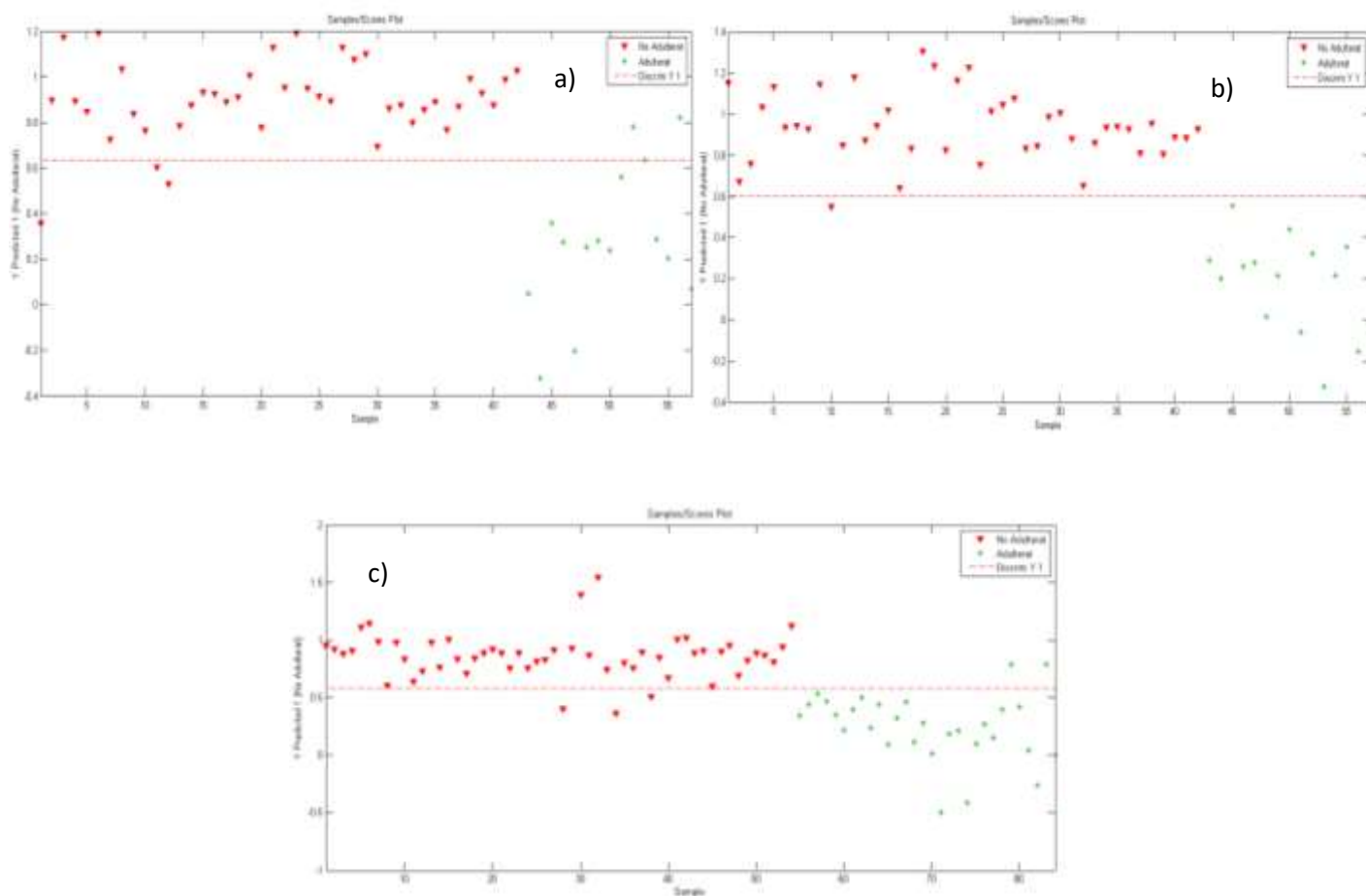


Figura 22. Gràfic de Y predita del PLS-DA de a) Siurana; b) Garrigues; c) Siurana + Les Garrigues

6.5.1. Càlcul dels paràmetres de qualitat

De la mateixa manera que s'ha realitzat amb els models SIMCA, es pot observar com aquests models PLS-DA classifica les diferents mostres tant de la seva classe com de l'altra i calcular posteriorment la sensibilitat (Equació 3) i l'especificitat (Equació 4) de cada model. En aquest estudi també es realitza sobre la classe "No Adulterat". Tant pel cas del model de Siurana com pel fet conjuntament amb les dues classes es pot observar que 3 mostres estan per sota del límit i per tant les classifica com a "Adulterat". En canvi per al model fet amb Les Garrigues només hi ha 1 mostra que no l'assigna correctament.

Taula 5. Sensibilitat dels models PLS-DA

Paràmetres de qualitat	Models PLS-DA		
	Siurana	Les Garrigues	Siurana + Les Garrigues
Sensibilitat (%)	92.86	97.62	94.44

Després d'haver realitzat el model, s'han predit totes les mostres adulterades i tots els nivells d'adulteració. Per a cada nivell també s'obté el gràfic de la Y predita, com per exemple el de la següent figura:

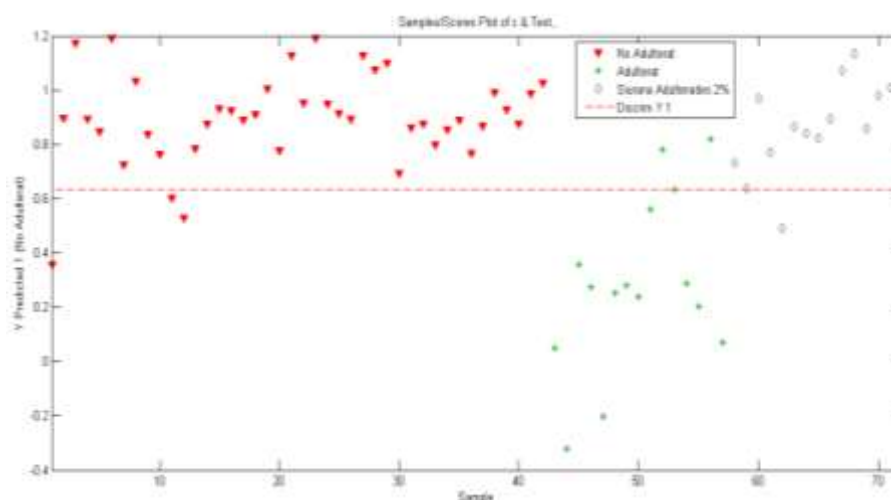


Figura 23. Gràfic de Y predita del PLS-DA de la predicció de mostres adulterades al 2% de Siurana

En aquest exemple es pot veure que la majoria de les mostres de la predicció queden per damunt de la Y límit. Es pot considerar lògic ja que el model ha estat fet amb mostres no adulterades i adulterades al 15%, i com que les de la predicció són adulterades al 2%

aleshores es pot pensar que seran més semblants a les no adulterades. Igual que pels models SIMCA, es calculen els valors d'especificitat per a cada nivell d'adulteració.

Taula 6. Especificitats de cada nivell d'adulteració per a cada model PLS-DA

Nivells d'adulteració (%)	Models PLS-DA		
	Siurana	Les Garrigues	Siurana + Les Garrigues
0	7.14	2.38	3.33
0.5	6.67	46.67	26.67
1	6.67	20.00	13.33
2	13.33	26.67	13.33
5	20.00	40.00	30.00
7.5	20.00	100	60.00
10	73.33	100	73.33
12.5	66.67	100	83.33
15	80.00	100	93.33
17.5	80.00	100	93.33
20	93.33	100	96.67
25	100	100	100
30	100	100	100
35	100	100	100

6.5.2. Corba PCC i paràmetres CC_{α} , CC_{β} i regió de no fiabilitat

Per acabar s'ha realitzat l'exportació de les dades de les Y predites de cada mostra a un arxiu compatible amb l'Excel amb l'objectiu de poder calcular les probabilitats de positius ($P(x)$) per tal de formar la corba PCC i trobar CC_{α} , CC_{β} i la regió d'incertesa. De la mateixa manera que per als models SIMCA, les corbes s'han obtingut amb l'equació 5 optimitzant els paràmetres a i b obtenint els següents valors:

Taula 7. Paràmetres a i b de l'equació 5 un cop realitzat l'ajust sigmoidal dels models PLS-DA

Paràmetres optimitzats	Models PLS-DA		
	Siurana	Les Garrigues	Les Garrigues + Siurana
a	11.54	4.37	8.25
b	1.72	0.99	1.34

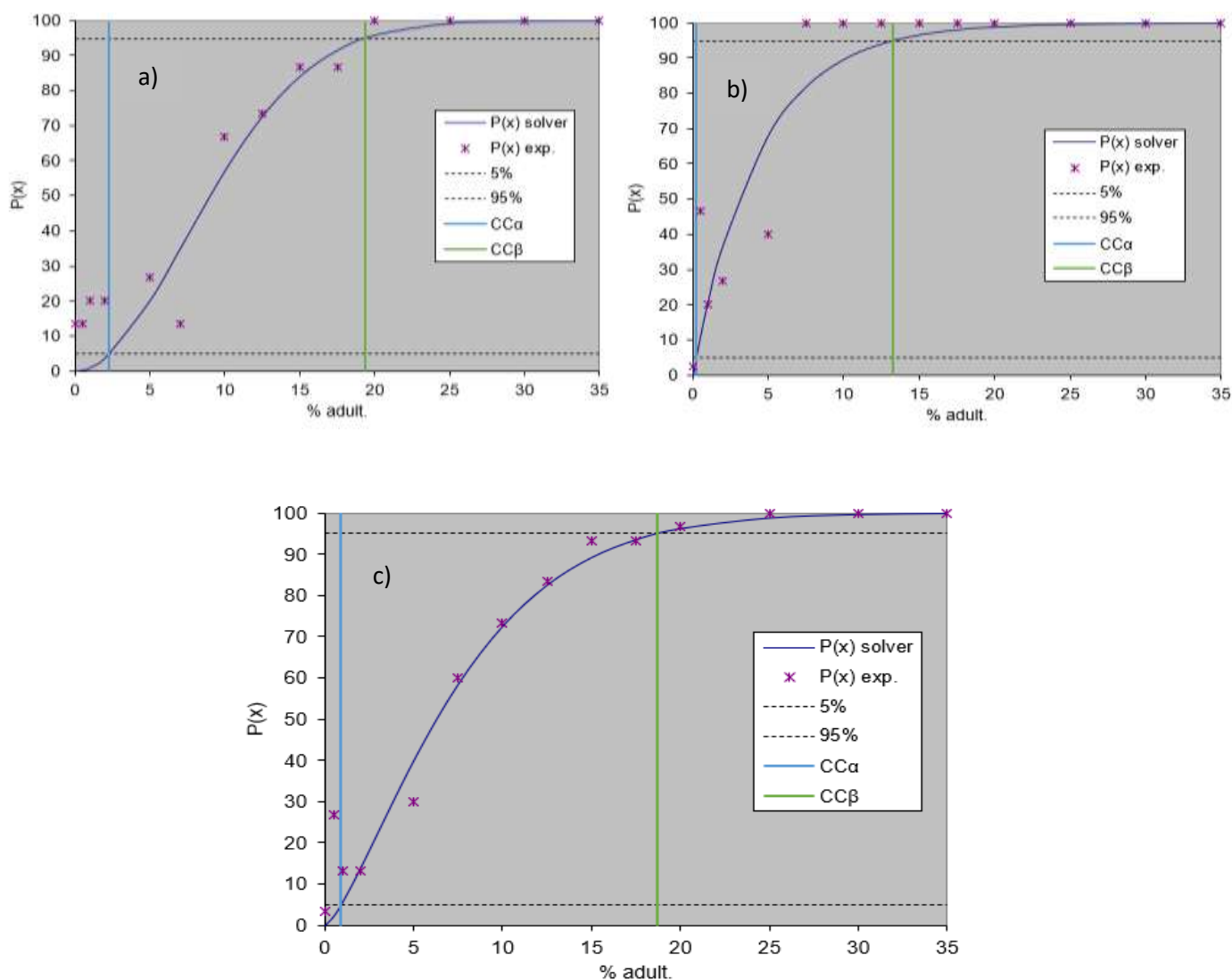


Figura 24. Corba PCC obtinguda del model PLS-DA de: a) Siurana; b) Garrigues; c) Siurana + Garrigues

Es pot arribar a observar unes petites diferències en la forma de la corba comparant-ho amb les dels models SIMCA, sobretot en el de Siurana i en el de Siurana + Garrigues. Aquestes dues corbes s'assemblen més a una forma sigmoïdal. Per a la classe Garrigues continua donant com a resultat una corba exponencial. També es poden calcular els paràmetres $CC\alpha$ i $CC\beta$ mitjançant l'equació 6 igual que s'ha fet amb els models SIMCA, els quals es mostren en la següent taula:

Taula 8. $CC\alpha$, $CC\beta$ i regió d'incertesa dels models PLS-DA

	Models PLS-DA		
Paràmetres corba PCC	Siurana	Les Garrigues	Siurana + Les Garrigues
$CC\alpha$	2	0	1
$CC\beta$	19	13	19
Regió d'incertesa	2 - 19	0 - 13	1 - 19

6.6. Comparativa SIMCA i PLS-DA

Tal i com es pot observar, pel que respecte a la classe no adulterada, amb el model PLS-DA s'obtenen millors valors de sensibilitat (majors que 92%) que amb el model SIMCA (major de 87%). Però cal indicar que la validació del PLS-DA s'hauria de fer amb *training* i *test set* i que el model s'ha establert front la classe adulterada al 15%.

A partir de les corbes PCC s'han obtingut els paràmetres relacionats amb l'adulteració: $CC\alpha$, $CC\beta$ i regió d'incertesa. No s'observa una diferència significativa respecte al valor de $CC\alpha$ (0% amb SIMCA i 2% amb PLS-DA). En canvi, el valor de $CC\beta$ és molt més petit i per tant la regió d'incertesa. Aquest és un canvi significatiu d'establir un model d'una classe respecte a un model de les dues classes.

Si comparem la forma de les corbes ajustades, el model PLS-DA de la classe Les Garrigues s'ha obtingut un comportament similar al dels diferents models SIMCA, tot i que en aquest cas el valor de $CC\beta$ si que queda dintre l'interval. Però per als models realitzats amb la classe Siurana i amb les dues classes conjuntes es pot observar que la corba, tal i com s'ha comentat anteriorment, perd el comportament exponencial per a convertir-se en sigmoidal, fet que fa que la probabilitat de dir que la mostra no està adulterada sigui més gran. És a dir, per al cas del model de la classe Siurana, de cada 100 mostres adulterades al 2% només 5 es reconeixeran com a adulterades. I a partir d'una adulteració del 19% es pot dir que la mostra està adulterada amb un nivell de confiança del 95%.

7. CONCLUSIONS

En primer lloc, un dels objectius d'aquest estudi ha sigut la realització de dos tipus de models de classificació (SIMCA i PLS-DA) amb valors de sensibilitat i especificitat el suficientment bons com per a que siguin acceptats.

Per una altra banda, s'ha establert les corbes PCC per tal d'obtenir els límits $CC\alpha$, $CC\beta$ i la regió d'incertesa. Després de realitzar la comparativa entre els dos tipus de models, s'ha arribat a la conclusió de que per a la tècnica de classificació PLS-DA, com discrimina entre 2 classes, els límits depenen de la classe adulterada modelada. En concret el $CC\beta$ estarà al voltant del valor d'adulteració de la classe modelada.

Les corbes PCC han demostrat ser d'utilitat per establir paràmetres de qualitat qual l'objectiu és un adulterant. La forma exponencial és deguda a que al tractar-se d'un adulterat, el límit inferior és zero. Quan el límit a partir del qual es vol discriminar és major de zero, aleshores pren una forma sigmoïdal.

Per acabar, altres objectius que es van proposar a l'inici d'aquest estudi també s'han aconseguit satisfactòriament. A nivell d'aprenentatge, considero aquests mesos que ha durat aquest estudi com a realment molt positius ja que tal i com s'ha comentat anteriorment, la realització d'una investigació d'aquest tipus no es pot arribar a dur a terme durant el grau. I també la utilització amb més profunditat del programa Matlab i més en concret el PLS Toolbox que inicialment era desconegut han sigut experiències molt enriquidores.

Com a treball de futur s'haurien d'establir els models PLS-DA amb la classe adulterada a diferents percentatges d'adulteració per veure com es modifica la corba. Per una altra banda, s'hauria d'estudiar com optimitzar el llindar del model per tal d'assegurar que cap mostra adulterada sigui predita com no adulterada encara que perdéssim sensibilitat (capacitat del model de predir mostres no adulterades com a tal).

Firstly, one of the main objectives of this study has been to make two types of classifications models (SIMCA and PLS-DA) with sensitivity and specificity values good enough to be accepted.

Besides that, the PCC curves have been established to obtain the limits $CC\alpha$, $CC\beta$ and the uncertainty region. After making the comparison between the two types of models, it has been concluded that the model of the PLS-DA classification technique, which discriminates between 2 classes, its limits depend on the modelled adulterated class.

The PCC curves have been shown to be useful in setting quality parameters when the target is an adulterant. The exponential form is due to the fact that in the case of an adulterator, the lower limit is zero. When the limit from which you want to discriminate is greater than zero, then it takes a sigmoidal form.

Finally, other objectives that were proposed at beginning of this study have also been successfully achieved. In terms of learning, I consider that these months that this study has lasted as very positively because, as it was mentioned above, the research can not be carried out during the degree. And also, the more in-depth use of the Matlab program and more specifically the PLS Toolbox that was initially unknown has been very enriching experiences.

On the one hand, PLS-DA models with adulterated class at different percentages of adulteration should be established as future work to see how the curve changes. On the other hand, we should study how to optimize the model threshold to ensure that no adulterated samples are predicted as unadulterated even if we lose sensitivity (ability of the model to predict non-adulterated samples as such).

8. BIBLIOGRAFIA

- (1) Genovese, A.; Caporaso, N.; Sacchi, R. Flavor Chemistry of Virgin Olive Oil: An Overview. *Appl. Sci.* **2021**, *11* (4), 1639. <https://doi.org/10.3390/app11041639>.
- (2) Abamba Omwange, K.; Al Riza, D. F.; Saito, Y.; Suzuki, T.; Ogawa, Y.; Shiraga, K.; Giametta, F.; Kondo, N. Potential of Front Face Fluorescence Spectroscopy and Fluorescence Imaging in Discriminating Adulterated Extra-Virgin Olive Oil with Virgin Olive Oil. *Food Control* **2021**, *124* (August 2020), 107906. <https://doi.org/10.1016/j.foodcont.2021.107906>.
- (3) International Olive Oil Council. International Trade Standard Applying to Olive Oils and Olive-Residue Oils (COI/T.15/No.1). **1985**, No. 3, 1–17.
- (4) Yan, J.; van Stuijvenberg, L.; van Ruth, S. M. Handheld Near-Infrared Spectroscopy for Distinction of Extra Virgin Olive Oil from Other Olive Oil Grades Substantiated by Compositional Data. *Eur. J. Lipid Sci. Technol.* **2019**, *121* (12), 1–12. <https://doi.org/10.1002/ejlt.201900031>.
- (5) Casale, M.; Oliveri, P.; Casolino, C.; Sinelli, N.; Zunin, P.; Armanino, C.; Forina, M.; Lanteri, S. Characterisation of PDO Olive Oil Chianti Classico by Non-Selective (UV-Visible, NIR and MIR Spectroscopy) and Selective (Fatty Acid Composition) Analytical Techniques. *Anal. Chim. Acta* **2012**, *712*, 56–63. <https://doi.org/10.1016/j.aca.2011.11.015>.
- (6) Ali, H.; Saleem, M.; Anser, M. R.; Khan, S.; Ullah, R.; Bilal, M. Validation of Fluorescence Spectroscopy to Detect Adulteration of Edible Oil in Extra Virgin Olive Oil (EVOO) by Applying Chemometrics. *Appl. Spectrosc.* **2018**, *72* (9), 1371–1379. <https://doi.org/10.1177/0003702818768485>.
- (7) Mabood, F.; Boqué, R.; Folcarelli, R.; Busto, O.; Jabeen, F.; Al-Harrasi, A.; Hussain, J. The Effect of Thermal Treatment on the Enhancement of Detection of Adulteration in Extra Virgin Olive Oils by Synchronous Fluorescence Spectroscopy and Chemometric Analysis. *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* **2016**, *161*, 83–87. <https://doi.org/10.1016/j.saa.2016.02.032>.
- (8) Dankowska, A.; Kowalewski, W. Comparison of Different Classification Methods for Analyzing Fluorescence Spectra to Characterize Type and Freshness of Olive Oils. *Eur. Food Res. Technol.* **2019**, *245* (3), 745–752. <https://doi.org/10.1007/s00217-018-3196-z>.
- (9) Uncu, O.; Ozen, B. A Comparative Study of Mid-Infrared, UV–Visible and Fluorescence Spectroscopy in Combination with Chemometrics for the Detection of Adulteration of Fresh Olive Oils with Old Olive Oils. *Food Control* **2019**, *105* (June), 209–218. <https://doi.org/10.1016/j.foodcont.2019.06.013>.

- (10) Vera, D. N.; Jiménez-Carvelo, A. M.; Cuadros-Rodríguez, L.; Ruisánchez, I.; Callao, M. P. Authentication of the Geographical Origin of Extra-Virgin Olive Oil of the Arbequina Cultivar by Chromatographic Fingerprinting and Chemometrics. *Talanta* **2019**, *203* (February), 194–202. <https://doi.org/10.1016/j.talanta.2019.05.064>.
- (11) Brereton, R. G. *Applied Chemometrics for Scientists*; 2007. <https://doi.org/10.1002/9780470057780>.
- (12) Skoog, D.; Holler, F.; Crouch, S. *Principios de Análisis Fundamental*; 6a edición.
- (13) Lucio Gutiérrez, J. R. Aplicación de Métodos Quimiométricos Para La Caracterización y Control de Calidad de Plantas Medicinales. **2012**, 147.
- (14) D.L. Massart; B.G.M. Vandeginste; L.M.C. Buydens; S. De Jong; P.J. Lewi; J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics*, Elsevier, Volume 20, Part A, **1998**, Pages 1-20. [https://doi.org/10.1016/S0922-3487\(97\)80031-7](https://doi.org/10.1016/S0922-3487(97)80031-7).
- (15) Chen, Y.; Xie, M. Y.; Yan, Y.; Zhu, S. Bin; Nie, S. P.; Li, C.; Wang, Y. X.; Gong, X. F. Discrimination of Ganoderma Lucidum According to Geographical Origin with near Infrared Diffuse Reflectance Spectroscopy and Pattern Recognition Techniques. *Anal. Chim. Acta* **2008**, *618* (2), 121–130. <https://doi.org/10.1016/j.aca.2008.04.055>.
- (16) Gemperline, P. *Practical Guide to Chemometrics*; Second Edition; CRC Press, Boca Raton FL, **2006**. <https://doi.org/10.2307/1269627>.
- (17) Moncayo Martín, S. Desarrollo y Aplicación de Métodos Quimiométricos Para El Estudio de Muestras Mediante Espectroscopia de Ablación Láser (LIBS). **2017**, 40–58.
- (18) López, M. I.; Callao, M. P.; Ruisánchez, I. A Tutorial on the Validation of Qualitative Methods: From the Univariate to the Multivariate Approach. *Anal. Chim. Acta* **2015**, *891*, 62–72. <https://doi.org/10.1016/j.aca.2015.06.032>.

9. ANNEXOS

Annex 1. Taules de nomenclatura de les mostres

NOM MOSTRA	DO
21G01-VE	Les Garrigues
21G02-VE	Les Garrigues
21G03-VE	Les Garrigues
21G04-VE	Les Garrigues
21G05-VE	Les Garrigues
21G06-VE	Les Garrigues
21G07-VE	Les Garrigues
21G08-V	Les Garrigues
21G09-VE	Les Garrigues
21G10-VE	Les Garrigues
21G11-V	Les Garrigues
21G12-V	Les Garrigues
21G13-VE	Les Garrigues
21G14-VE	Les Garrigues
21G15-VE	Les Garrigues
21G16-VE	Les Garrigues
21G17-V	Les Garrigues
21G18-VE	Les Garrigues
21G19-VE	Les Garrigues
21G20-VE	Les Garrigues
21G21-VE	Les Garrigues
21G22-VE	Les Garrigues
21G23-VE	Les Garrigues
21G24-VE	Les Garrigues
21G25-VE	Les Garrigues
21G26-VE	Les Garrigues
21G27-VE	Les Garrigues
21G28-VE	Les Garrigues
21G29-VE	Les Garrigues
21G30-VE	Les Garrigues
21GM1	Les Garrigues
21GM2	Les Garrigues
21GM3	Les Garrigues
21GM4	Les Garrigues
21GM5	Les Garrigues
21GM6	Les Garrigues
21GM7	Les Garrigues
21GM8	Les Garrigues
21GM9	Les Garrigues
21GM10	Les Garrigues
21GM11	Les Garrigues
21GM12	Les Garrigues

NOM MOSTRA	DO
21S31-VE	Siurana
21S32-V	Siurana
21S33-VE	Siurana
21S34-V	Siurana
21S35-VE	Siurana
21S36-V	Siurana
21S37-VE	Siurana
21S38-VE	Siurana
21S39-V	Siurana
21S40-VE	Siurana
21S41-VE	Siurana
21S42-VE	Siurana
21S43-VE	Siurana
21S44-V	Siurana
21S45-VE	Siurana
21S46-VE	Siurana
21S47-V	Siurana
21S48-VE	Siurana
21S49-VE	Siurana
21S50-VE	Siurana
21S51-VE	Siurana
21S52-VE	Siurana
21S53-VE	Siurana
21S54-V	Siurana
21S55-VE	Siurana
21S56-V	Siurana
21S57-VE	Siurana
21S58-V	Siurana
21S59-VE	Siurana
21S60-VE	Siurana
21SM1	Siurana
21SM2	Siurana
21SM3	Siurana
21SM4	Siurana
21SM5	Siurana
21SM6	Siurana
21SM7	Siurana
21SM8	Siurana
21SM9	Siurana
21SM10	Siurana
21SM11	Siurana
21SM12	Siurana

Llegenda:
Training set SIMCA
Test set SIMCA