

Treball de Fi de Grau

Desenvolupament d'un mètode qualitatiu multivariant per la determinació d'adulterants en pastes d'anacards

Departament de Química Analítica i Química Orgànica

Facultat de Química

Alumne: Jair Alfonso Wells Llanos

Directora: María Pilar Callao Lasmarias



**UNIVERSITAT
ROVIRA i VIRGILI**

Agraïments

Dedico els agraïments a la directora del treball, la doctora i professora Maria Pilar Callao, per donar-me l'oportunitat de participar en un dels seus projectes on m'ha ensenyat el món de la quimiometria i, més important, la seva perspectiva de la química on no només es busca el coneixement sinó que també una aportació útil a la societat.

També agraeixo a la Universitat Rovira i Virgili i la Facultat per haver-me dotat aquests anys de les seves instal·lacions i el seu equip humà.

Per últim, agraeixo a la meua companya Glòria Rovira, pels seus coneixements i ajuda imprescindible.

Índex

1. Objectiu	6
2. Introducció	6
2.1. Àmbit de Treball	8
3. Dades prèvies	8
3.1. Mostres	8
3.2. Obtenció de dades	9
3.2.1. Espectroscòpia Infraroja Propera (NIR).....	9
3.2.2. Espectroscòpia IR de Reflectància Total Atenuada (ATR)	10
3.2.3. Software	10
3.3. Tractament de dades	10
3.3.1 Pretractaments.....	10
3.3.2. Anàlisi exploratori de dades: descomposició per components principal.....	13
3.3.3. Tècniques de classificació: Soft Intedependent Modelling of Class Analogy	14
3.3.4. Validació: paràmetres de qualitat del mètode de classificació.....	16
5. Resultats	18
5.1. NIR	18
5.1.1. Espectres	18
5.1.2. Anàlisi exploratori de dades (PCA)	19
5.1.3. Tècnica classificatòria: Soft Independent Modelling of Class Analogy	22
5.2. ATR	24
5.2.1. Espectres	24
5.2.2. Anàlisi exploratori de dades (PCA)	26
5.2.3. Tècnica classificatòria: Soft Independent Modelling of Class Analogy	27
6. Conclusions	29
7. Bibliografia	30

Resum

L'objectiu d'aquest estudi ha estat desenvolupar i validar un mètode qualitatiu multivariant per a determinar l'adulteració en pastes d'anacards a partir de dades obtingudes per dues tècniques espectromètriques NIR i ATR i tècniques de classificació

S'ha fet servir l'anàlisi exploratòria de dades (PCA) per estudiar quines serien les estratègies més útils per desenvolupar la tècnica de classificació. Hi destaca l'estudi de diferents pretractaments.

S'ha desenvolupat un model SIMCA per cada tècnica infraroja. Tots dos s'han desenvolupat a partir d'una sola classe de mostres no adulterades. Els models s'han fet servir per predir la resta de mostres conegudes, tant adulterades com adulterades en un ampli rang de concentracions, per cadascun dels quatre fruits secs que s'han fet.

El model ha estat validat crosvalidació i test set. El resultat ha sigut l'obtenció de dos models classificatoris amb una especificitat del 100%, una sensibilitat al voltant del 93% i una eficàcia aproximada del 95%.

Abstract

The main objective of this study has been to develop and validate a multivariate qualitative method for determining adulteration in cashews using data obtained from both NIR and ATR spectrometric techniques and classification techniques.

Exploratory data analysis (PCA) has been used to study which strategies would be most useful for developing the classification technique. The study of different pre-treatments is notable.

A SIMCA model has been developed for each infrared technique. Both have developed from a One Class of non-adulterated samples. Models have been used to predict the remaining known samples, both adulterated and adulterated in a wide range of concentrations, for each of the four dry fruits that have been made.

The model has been validated by cross-validation and test set. The result has been the acquisition of two classificatory models with a 100% specificity, a sensitivity around 93% and an approximate effectiveness of 95%.

1. Objectiu

Aquest treball té com a objectiu establir un mètode per avaluar la possible l'adulteració de pastes d'anacards procedents de Brasil amb diferents fruits secs mitjançant una anàlisi qualitatiu multivariant a partir de dades espectrals NIR i ATR amb els corresponents tractaments quimiomètrics.

Aquest objectiu incorpora les etapes següents:

- Estudiar diferents pretractaments de dades per optimitzar els resultats.
- Visualitzar el comportament de les mostres aplicant tècniques exploratòries de dades.
- Separar, definir i caracteritzar quines són les classes d'interès a partir de l'ús de tècniques de classificació.
- Validar resultats.
- Predir la possible adulteració en mostres desconegudes.

2. Introducció

El cultiu de fruits secs a Brasil és un dels més grans del món, destaca el conreu de l'anacard amb un ritme de producció de 135 mil tones anualment des de 1985¹ i el del cacauet amb cerca de 580 mil tones produïdes l'any 2019. El valor de l'anacard és molt major al del cacauet a causa de les condicions més limitades en les que pot plantar-se, créixer i recollir-se, això fa que el seu preu de venda al públic sigui més car que altres fruits secs.

Les característiques similars entre els anacards i els cacauets, indiquen que aquest últim podria fer-se servir per substituir el primer en l'elaboració de pastes. Una escassetat del gènere podria provocar una petita adulteració amb un altre producte més barat i més abundant, fent possible mantenir la producció i abaratir costos.

Els motius per provocar una adulteració al menjar poden ser diversos, però el més probable és que es busqui un benefici econòmic, així que és lògic pensar que pot haver-hi barreges en dos productes semblants que tenen valors distints. Sense importar el motiu de la modificació del producte, es tracta d'un frau i un engany al client que compra un producte de major qualitat i en rep un de pitjor, raó suficient per evitar aquesta pràctica, però a més, l'ocultació d'al·lèrgens o altres substàncies poden provocar el desenvolupament de malalties, produir els efectes de l'al·lèrgia o la intolerància.

A Espanya, la llei 28/2015² per la defensa de la qualitat alimentària és la que declara una il·legalitat la pràctica d'adulteració alimentària a més de dictaminar l'obligatorietat de la realització d'anàlisi als productes per tal d'assegurar que no s'hagi comès cap adulteració.

Existeixen una gran quantitat d'adulterants, molts dels quals encara són desconeguts, sent un problema a l'hora de detectar-los. Aquesta presència de múltiples contaminants a la mostra dificulta la seva quantificació, l'alternativa és desenvolupar una anàlisi qualitativa amb respostes binàries, positiva o negativa, que és molt més útil en aquest cas i evita buscar concentracions exactes que realment no interessen.

L'anàlisi multivariant de dades és una estratègia àmpliament emprada per resoldre problemes que no poden ser resolts amb una sola variable, es tracta d'una tècnica quimiomètrica que empra múltiples dades per cada mostra, per tant, permet recollir molta informació. Una manera d'obtenir grans quantitats de dades per mostra és fer servir tècniques espectrofotomètriques

com són el NIR i l'ATR que són capaces de crear un gran volum de dades en poc temps i a costos baixos.

L'anàlisi qualitativa multivariant es basa en l'ús de les tècniques de classificació i és emprat no només per detectar anomalies a les mostres com poden ser les adulteracions sinó que també serveix per detectar propietats globals de la mostra com per exemple la denominació d'origen d'un producte i així reconèixer el seu valor afegit³.

Les tècniques de classificació segueixen el procediment descrit a la figura 1, a partir de les dades originals pertanyents a més d'una classe es pot arribar a crear un model de classificació per predir mostres totalment desconegudes. El procés es realitza amb un conjunt d'entrenament, amb el que es crea un model, i amb un conjunt de prova, que és predit; com que la classe real de les mostres emprades ja és conegut, podem valorar la capacitat de classificar del model creat i avaluar el nivell de fiabilitat.

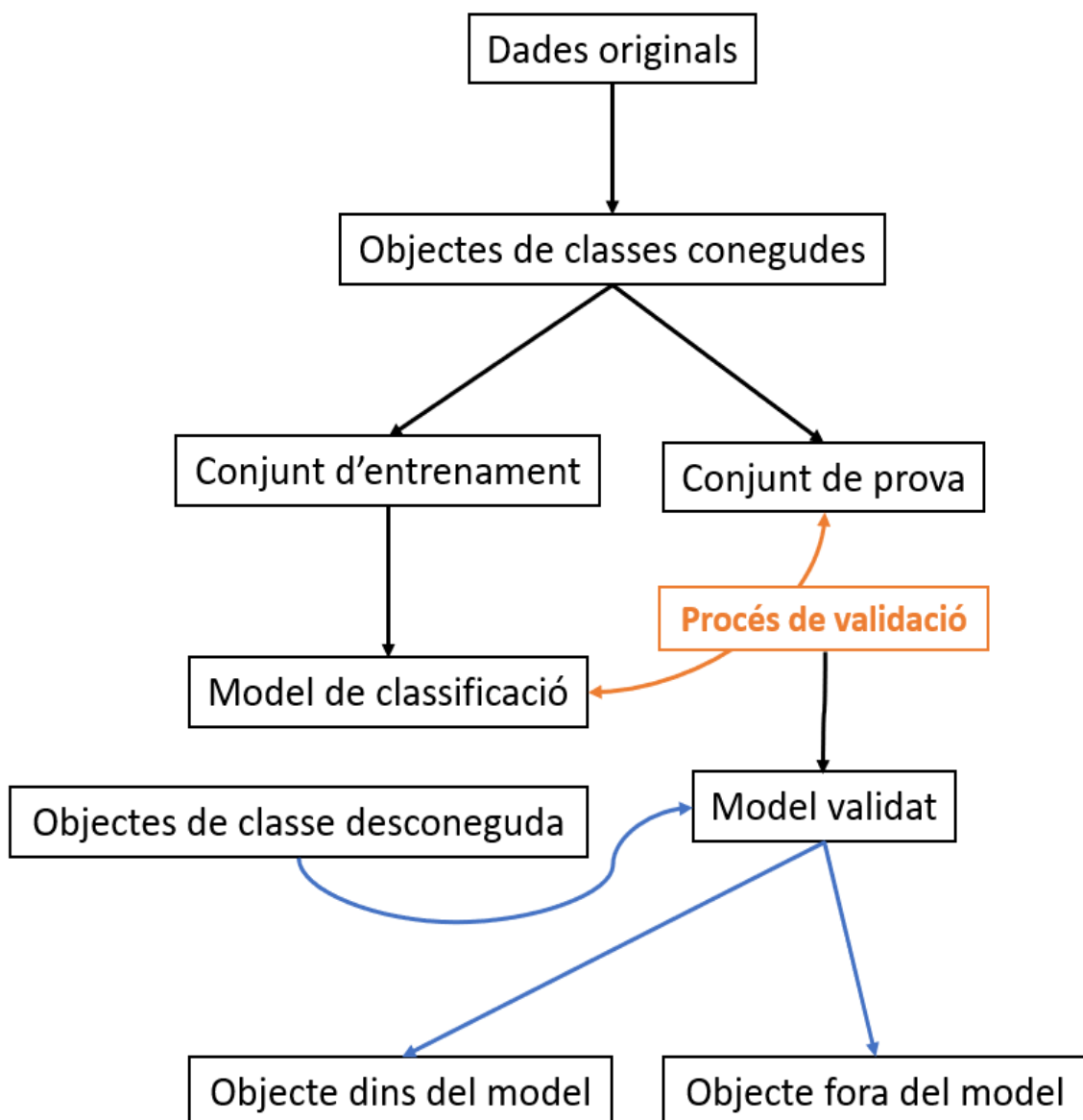


Figura 1. Procediment de classificació.

Una vegada creat el model, aquest es validarà i podrà ser emprat per classificar mostres totalment desconegudes.

Hi ha dos grans tipus de tècniques de classificació:

- Tècniques de modelatge, que busquen les similituds entre les mostres d'una mateixa classe per crear el model corresponent.
- Tècniques discriminants, que defineix una funció que discrimina les mostres de les diferents classes.

2.1. Àmbit de Treball

Aquest estudi està realitzat dins del grup de treball de quimiometria, qualimetria i nanosensors que pertany al departament d'analítica i química orgànica de la Universitat Rovira i Virgili, en col·laboració amb el departament de ciència alimentària de la Universitat Federal de Minas Gerais, Brasil.

Els dos grups tenen diferents responsabilitats, el de Minas Gerais s'ha encarregat de fer les tasques experimentals referents a l'obtenció de mostres representatives i de les dades, mentre que el grup de Tarragona hi treballa en el tractament de dades i en la creació d'un model qualitatiu multivariant.

No és la primera col·laboració que es dona entre ambdues universitats, s'han realitzat treballs publicats als darrers anys amb objectius del mateix àmbit d'aquest treball^{4,5}. Alguns dels temes tractats són la detecció d'adulterants, l'ús de mètodes de classificació multivariant o la selecció de variables⁶. La feina prèvia servirà per donar suport teòric i pràctic a l'hora d'avançar en aquest estudi.

En estudis anteriors tenim una guia sobre la detecció del frau alimentari a partir de mètodes qualitius multivariants⁷, en ell s'informa com a partir de les dades es poden seguir diferents vies per desenvolupar un model:

- Si hi ha un nombre elevat de dades es poden treballar amb un conjunt d'entrenament i un conjunt de prova, en canvi, si tenim un nombre menor de dades això s'hauria d'evitar i emprar una altra estratègia.
- Si la problemàtica requereix saber la concentració d'adulterant caldria trobar paràmetres addicionals.

Avui en dia, encara es podria dir que no hi ha consens sobre els protocols de validació i la terminologia adient pels mètodes qualitius; termes com sensibilitat i especificitat es poden considerar àmpliament acceptats, però encara falten atributs per definir la qualitat d'un mètode que al cap i a la fi, indirectament servirà per analitzar la qualitat de la decisió sobre una mostra d'origen desconegut.

3. Dades prèvies

3.1. Mostres

El material d'anàlisi multivariant d'aquest treball seran les pastes d'anacards brasilers procedents de la universitat col·laboradora on s'ha realitzat, prèviament, el procés experimental mitjançant les tècniques espectromètriques NIR i ATR, entre altres.

Per estudiar l'adulteració de la d'anacards (CN), s'adquireixen de tres proveïdors diferents un seguit de fruits secs que actuaran com adulterants, aquests són: nous pacanes (CNPN), nous de Brasil (CNBN), nous de macadàmia (CNM) i cacauets (CNP), a banda dels propis anacards (no adulterades). Com que l'adulteració s'estudia en pastes, ja que si fossin sòlids seria fàcil

assabentar-se de la presència d'altres fruits, se segueix aquest procediment per homogeneïtzar les mostres:

1. S'aixafen les mostres en un processador.
2. Es barregen.
3. Es dipositen en envasos de polietilè.
4. Se segellen.
5. Es mantenen a temperatura ambient.

Una vegada obtingudes les pastes de fruits secs es procedeix a la preparació de les mostres no adulterades i adulterades tal com indica la figura 2. S'obtenen 280 mostres en total, les quals 56 romandran sense adulterar i la resta seran modificades amb els 4 tipus diferents de fruits secs, fent que finalment hi hagi 56 mostres cada. Després, s'apliquen 8 nivells de concentració pes/pes 7 vegades, amb cadascun dels grups de contaminants.

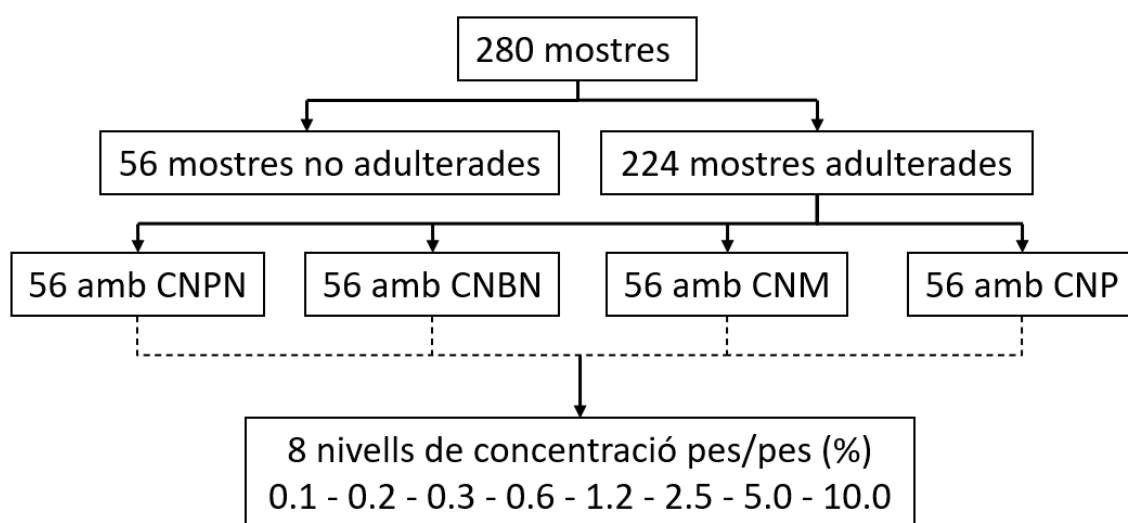


Figura 2. Esquema de preparació de les mostres adulterades.

3.2. Obtenció de dades

3.2.1. Espectroscòpia Infraroja Propera (NIR)

La tècnica de l'infraroig proper es basa en la capacitat de generar un espectre a partir de la vibració dels enllaços moleculars. La majoria de les molècules es troben al seu estat fonamental de vibració a temperatura ambient, els àtoms que participen en els enllaços químics es troben en moviment constant depenent del tipus d'unió en què es trobin, les amplituds d'aquests desplaçaments són nanomètriques i poden ampliar-se si es transfereix energia a la molècula per tal de capturar l'espectre generat.

L'energia (E_p) pot ser transferida per un fotó a una determinada longitud d'ona (λ), mitjançant la següent equació:

$$E_p = h\nu = \frac{hc}{\lambda}$$

On h és la constant de Planck i c és la velocitat de la llum.

L'energia del fotó, transferit per l'espectrofotòmetre provoca l'excitació de l'estat vibracional. La diferència de nivells energètics provoca una resposta selectiva del sistema molecular, això

significa que en un rang de longituds d'ona s'obindrà una sèrie d'absorbàncies que formaran l'espectre final⁸.

Per aquesta anàlisi del NIR, una solució portable de Viavi, l'equipament emprat va ser un MicroNIR® 1700. Una petita porció de la mostra va ser col·locada a una placa Petri (de 3.5 cm de diàmetre x 1.2 cm d'altura) fins a cobrir completament la superfície del recipient. Després, la placa va ser col·locada al MicroNIR, on cada mostra va rebre una lectura de 20 escanejos amb una resolució de 6.25 nm per obtenir un espectre que comprenia la regió de longitud d'ona d'entre els 908 nm i els 1676 nm.

3.2.2. Espectroscòpia IR de Reflectància Total Atenuada (ATR)

L'infraroig de transformada de Fourier de reflectància total atenuada (ATR FTIR) és una tècnica d'anàlisi ràpida que no requereix preparació de la mostra i necessita molt poca quantitat de mostra per funcionar⁹, ja que fa servir una peça de diamant per crear una reflexió on incidir els rajos infrarojos.

Per l'anàlisi es va treballar amb un espectrofotòmetre Frontier de Perkin Elmer (Waltham, MA, EUA) amb un diamant ATR d'accessori. Les mostres van ser prèviament homogeneïtzades i després dipositades en el cristall ATR on van ser pressionades pel cap de metall. Cada mostra va ser analitzada en una lectura de 16 escanejos, l'espectre generat va des dels 4000 cm⁻¹ fins els 650 cm⁻¹ (amb una resolució de 4 cm⁻¹). Les lectures van ser realitzades en ordre aleatori sota les mateixes condicions.

3.2.3. Software

Les dades experimentals rebudes han sigut treballades als programes MATLAB R2012b i Microsoft Excel 2013.

MATLAB és un programa àmpliament emprat en investigacions científiques i matemàtiques. En aquest treball ha sigut utilitzat per dur a terme el tractament de les dades i per crear les figures resultants de l'anàlisi de dades. Per treballar-hi, ha sigut necessària l'addició d'una caixa d'eines anomenada PLS_Toolbox 4.0 que és una expansió de les funcions que permet l'ús de tècniques quimiomètriques emprades en química analítica.

3.3. Tractament de dades

3.3.1 Pretractaments

Les dades obtingudes de les tècniques espectrofotomètriques NIR i ATR formen una matriu, és a dir, una taula de valors formada per cadascuna de les mostres i les seves respectives dades multivariants obtingudes. A títol d'exemple per veure com s'arriba a aquesta matriu està la figura 3 i la figura 4:

- A la figura 3 es pot observar com de cada mostra s'obté un espectre. Aquests espectres són realment una representació gràfica d'una taula numèrica on cada valor X en dona una resposta de valor Y. De cada mostra se'n pot extreure una dada multivariant, la mostra 1, 2 i n tenen espectres diferents, per això, la taula de valors que els formen no són iguals.

- Pel que fa a la figura 4, aquesta és la matriu final obtinguda a l'unir les diferents taules de numèriques de les diferents mostres, aquesta matriu final pot representar-se espectralment amb les diferents matrius superposades o amb la mitjana d'aquestes.

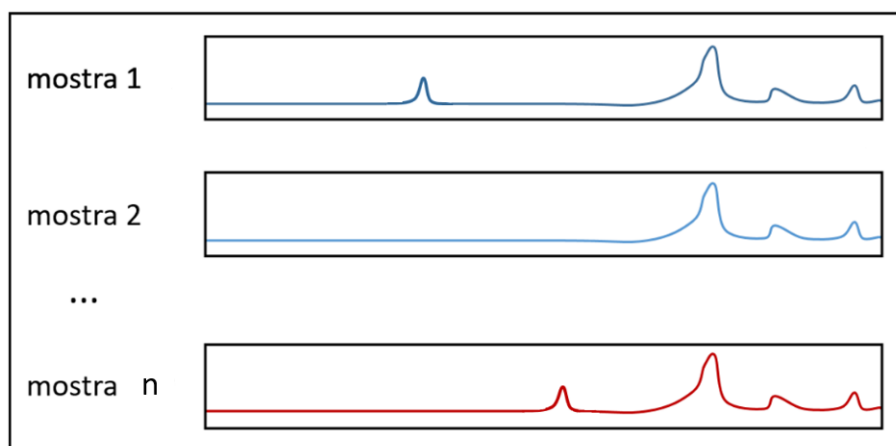


Figura 3. Exemple de les respostes espectrals obtingudes.

		Valors X				variables
		1	2	...	m	
mostra 1		0	1		2	
mostra 2		0	0		2	
...						
mostra n		0	3		2	

Figura 4. Exemple de la matriu numèrica obtinguda.

Les dades obtingudes es poden representar tant espectralment com en una matriu numèrica. El realment important són els valors numèrics que són els que ens aportaran la informació necessària de les mostres per classificar-les i obtenir resultats qualitius.

Per millorar les dades s'aplicaran els pretractaments necessaris amb els objectius següents:

- Eliminar senyals que no depenguin de la mostra com pot ser el soroll spectral provocat per l'instrument.
- Maximitzar la informació spectral.

L'aplicació de pretractaments transformarà la matriu obtinguda, els valors numèrics de l'espectre seran modificats matemàticament. L'elecció de quin pretractament es farà servir hi depèn de:

- La forma que tingui l'espectre.
- L'augment d'informació que aporti.
- Les necessitats del model.

D'entre els pretractaments emprats hi destaquen els següents:

Centrat

El centrat pretén que el valor de les variables estigui a la mateixa escala. Consisteix en el fet que a partir de la matriu, es calculi el valor mitjà de cada columna i després es resta a la variable original, d'aquesta manera, la matriu resultant tindrà columnes amb una alçada mitjana de zero i el nou valor correspondrà a la diferència entre el valor original i la nova mitjana.

$$X \text{ centrada} = X \text{ original} - X \text{ mitjana columna}$$

Autoescalat

L'escalat busca que les diferents variables tinguin la mateixa escala i variabilitat.

La matriu es transforma, primer centrant cada columna i després dividint per la desviació estàndard d'aquella columna. El resultat de cada columna és una mitjana de zero i una desviació estàndard d'1.

$$X \text{ autoescalada} = \frac{X \text{ original} - X \text{ mitjana columna}}{S \text{ columna}}$$

Correcció de la dispersió multiplicativa promig (MSC mean)

La tècnica MSC corregeix la dispersió dels valors de la matriu fent servir una matriu de referència. Aquesta matriu pot ser la mitja de l'espectre o simplement una línia dibuixada.

Suavitzat (Savitsky-Golay)

És un mètode de reducció de soroll, es substitueix cada variable per valors nous tenint en compte el seu entorn.

El mètode de Savitsky-Golay aplica regressions lineals de grau x localitzades a $x+1$ nombre de variables, cada regressió lineal retorna un valor a la matriu calculat a partir del mateix polinomi generat per l'ajust lineal. El mètode genera tantes regressions lineals com variables té la matriu. El resultat és un espectre amb menys soroll.

Derivació (Savitsky-Golay)

La derivació es sol utilitzar per millorar la resolució visual, resoldre la superposició de pics i ressaltar estructures. La derivació per Savitsky-Golay funciona semblant al suavitzat, però, en comptes d'aplicar una regressió lineal s'aplica una derivada.

3.3.2. Anàlisi exploratori de dades: descomposició per components principal

Un component principal (CP) és una combinació lineal de les variables originals. El CP es farà servir per substituir les variables originals a la matriu de dades tal com s'indica a la figura 5.

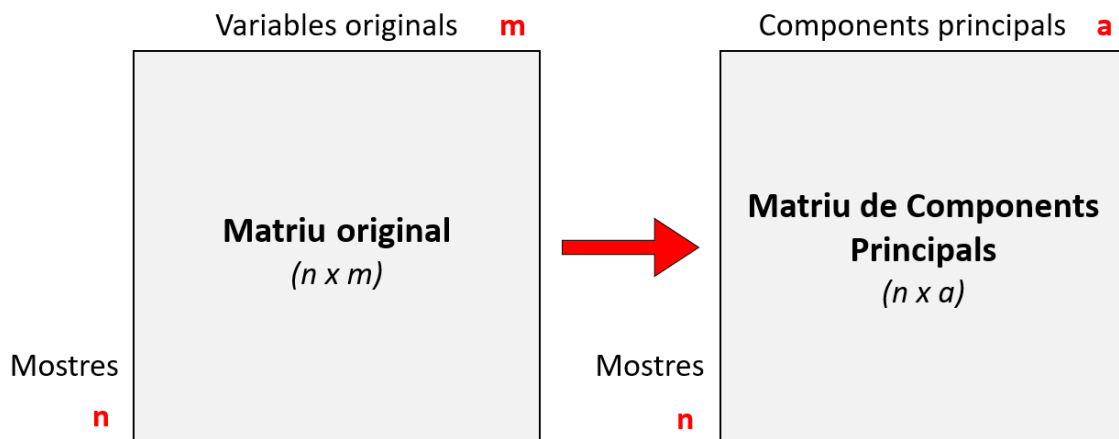


Figura 5. Transformació de la matriu original en la matriu de components principals.

La combinació lineal segueix la direcció de màxima variació a les dades, així retenir la màxima informació continguda a les dades. A la figura 6 es representa la regressió lineal i la projecció del CP resultant.

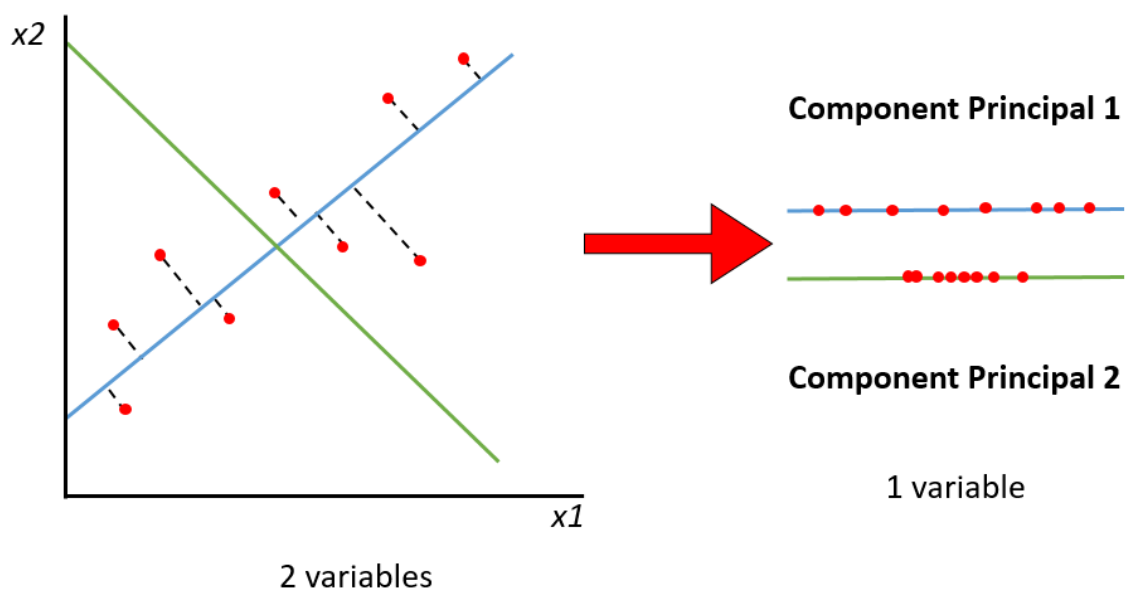


Figura 6. Transformació dels valors originals en un component principal i representació de les projeccions lineals dels components principals 1 i 2.

El CP format només en té una variable anomenada *score*, representada als punts vermells del component principal 1 a la figura X. Per arribar a aquest valor es segueix la següent fórmula:

$$t_1 = p_{11} \cdot x_1 + p_{12} \cdot x_2$$

On:

- El terme t_1 és el valor *score*.
- Els termes p_{11} i p_{12} són els valors del coeficient *loadings*. Els subíndexs representen el número del PC i la variable.
- Els termes x_1 i x_2 són les dades originals.

El primer CP segueix la direcció amb més variació en les dades i, per tant, conserva major informació continguda en aquestes. La següent combinació lineal, el CP 2, serà ortogonal i retindrà informació que la CP 1 no ha agafat¹⁰. Cada component principal expressa un % de la informació (variància) explicada i entre tots els components principals s'explica el 100% on s'inclou la informació útil i el soroll. Per tal que els CP 1 i 2 retinguin informació diferent, han de seguir projeccions lineals distintes, tal com s'indica a la figura 6.

L'anàlisi exploratòria per components principals (PCA) es fonamenta en la seva representació gràfica, ja que permet enfrontar els valors dels 2 primers PCs en una representació 2D o els 3 primers en una representació 3D, que són els components principals amb més variància acumulada. De les gràfiques es pot observar la presència de punts discrepants, de tendències i agrupacions (com és el cas de la figura 7), que serviran per prendre decisions com la eliminació de punts discrepants o aplicar una tècnica de classificació SIMCA basada en les mostres marcades¹¹.

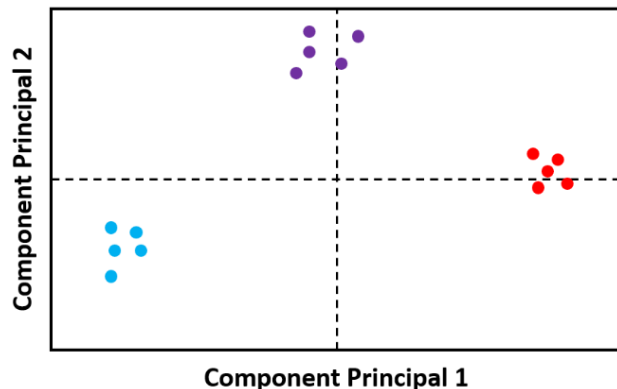


Figura 7. Exemple de representació de dos CP.

3.3.3. Tècniques de classificació: Soft Intedependent Modelling of Class Analogy

La tècnica es basa en el PCA, on cada classe és modelada independentment de la resta. Dins d'aquest model, per a cada mostra s'obtenen dos estadístics: T^2 de Hotelling i Q .

L'estadística Q es defineix com la suma de quadrats dels residuals, per tant, està relacionada amb la quantitat d'informació original de la mostra que no està inclosa al model. El límit de la Q , Q_{lim} , pot ser calculada pel model a un nivell de significança específica (α), usualment 0.05.

L'estadístic de T^2 de Hotelling mesura la informació de cada mostra dintre del model SIMCA, per tant, hi dóna una mesura de com de bé encaixa cada mostra al model. El límit de T^2 , T^2_{lim} , pot ser calculat en una α específica, usualment en 0.05.

Els valors de T^2 de Hotelling reduïda i la Q reduïda poden ser calculats a partir de la ràtio entre l'estadística corresponent de la mostra i el límit corresponent a $\alpha=0.05$.

Segons el "SIMCA original", els límits són definits considerant les estadístiques reduïdes, una mostra amb un valor inferior a 1 en ambdues estadístiques podrà ser considerada dins del model. Per altra banda, una nova versió anomenada "SIMCA restrictiu" pren com a criteri la distància, el límit que considera les mostres dins del model és un semicercle amb radi 1.

$$d_{ij} = \sqrt{T_{red}^2 + Q_{red}}$$

On:

- La d_{ij} representa la distància de la mostra i a la classe j .
- La T_{red}^2 representa la T^2 de Hotelling reduïda.
- La Q_{red} representa la Q reduïda.

A la figura 8 hi trobem una representació gràfica del que seria el SIMCA original i el SIMCA restrictiu. Les línies de guions definirien els límits del model i quines mostres queden dins o fora segons el criteri que s'agafi, mentre la línia blava, que pertany al SIMCA original accepta algunes mostres roges dins del model, la línia verda, que pertany al SIMCA restrictiu, no hi accepta cap.

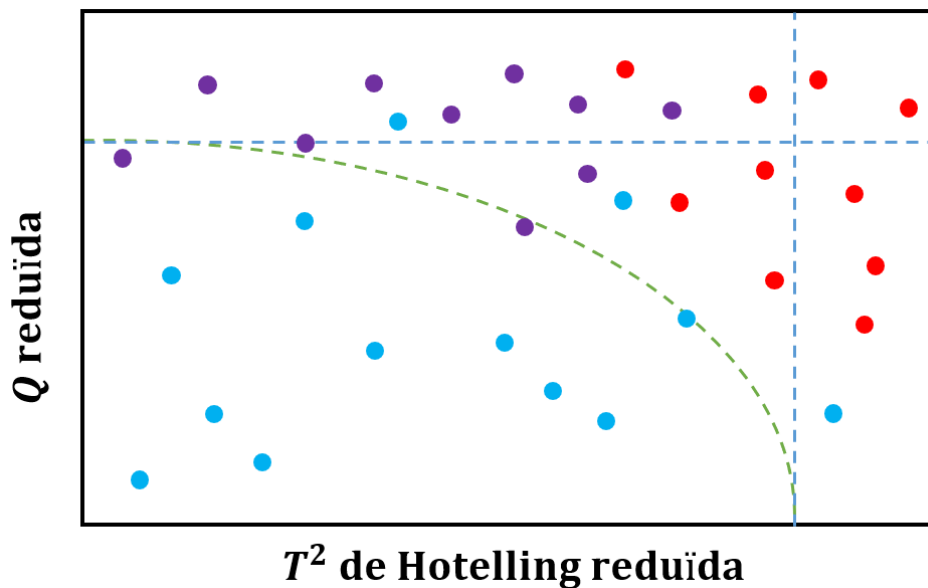


Figura 8. Representació del model SIMCA. Línies blaves de guions, límit del SIMCA original. Línia verda de guions, límit del SIMCA restrictiu.

Com ja s'ha comentat, el model SIMCA es crea a partir dels components principals; cada CP aporta un % de la informació de les dades incloent el soroll instrumental i altres errors. Idealment, el model creat no ha d'acumular el 100% de la informació, només la necessària per ser representatiu.

A l'hora de fer una anàlisi SIMCA es poden seguir dues estratègies per classificar les mostres: la One Class i la Multi Class.

Per una banda, l'estratègia One Class consisteix a definir al model una única classe, de manera que les mostres predites només poden ser classificades com "pertanyents a la classe" o "No pertanyents a la classe", mentre que per l'altra banda, a l'estratègia Multi Class s'especifiquen més d'una classe al model, fent possible que les mostres predites siguin classificades "pertanyents a més d'una classe", "pertanyents a una classe" o "no pertanyents a cap classe".

En una anàlisi qualitatiu d'aquest estudi amb determinar si una mostra ha estat adulterada o no, és suficient, per tant, l'ús de l'estratègia One Class és més que suficient. Tot i això, ambdues estratègies es poden emprar.

3.3.4. Validació: paràmetres de qualitat del mètode de classificació

Una vegada plantejat el SIMCA com a model de classificació de dades cal definir quins seran els paràmetres de qualitat amb els que s'estimarà si un model és vàlid o no.

Respecte a les mostres, s'hi consideren positives aquelles que pertanyen a la classe modelada i negatives aquelles que no formen part de la classe modelada. Quan es fa un nou model SIMCA, s'utilitzen les dades de les mostres positives per crear-lo, mentre que a la predicció es fa tant de mostres positives com negatives.

Segons el tipus de resposta obtinguda pel model SIMCA sumat a la naturalesa real de la mostra, en tenim quatre de qualificacions per cada mostra:

- Vertader Positiu (TP): mostra que realment és positiva i el model la classifica acuradament com a tal.
- Fals Positiu (FP): mostra que realment és negativa i el model la classifica erròniament com positiva.
- Vertader Negatiu (TN): mostra que realment és negativa i el model la classifica acuradament com a tal.
- Fals Negatiu (FN): mostra que realment és positiva i el model la classifica erròniament com negativa.

Si a més, es realitza un model SIMCA de més d'una classe, dos per exemple, es poden obtenir també els següents casos:

- Doble assignació o doble positiu: la mostra ha sigut classificada dins de dos models alhora, donant un resultat inconclús.
- Sense assignació o doble negatiu: la mostra no ha sigut classificada en cap dels dos models, donant un resultat inconclús.

Per al model SIMCA d'una sola classe modelada, en realitzar les prediccions corresponents s'organitzarà una taula com la següent:

Taula 1. Assignacions d'un model SIMCA d'una sola classe.

		Naturalesa real de les mostres	
		Positiva	Negativa
Predicció del model	Positiva	Nº de TP	Nº de FP
	Negativa	Nº de FN	Nº de TN

A partir de la taula s'obtenen els paràmetres de qualitat, que són els següents:

- Sensibilitat: És l'habilitat que té el model per predir correctament les mostres de naturalesa positiva.
- Especificitat: És l'habilitat que té el model per predir correctament les mostres que són de naturalesa negativa.
- Eficiència: Correspon a la capacitat general que té el model per predir correctament les mostres.

Taula 2. Càlcul dels paràmetres de qualitat.

Sensibilitat	$\frac{TP}{TP + FN} \cdot 100$
Especificitat	$\frac{FN}{TP + FN} \cdot 100$
Eficiència	$\frac{TN + TP}{TN + FP + TP + FN} \cdot 100$

Per validar el model, és a dir, per estudiar els seus paràmetres de qualitat, es poden seguir dues estratègies:

- La primera és la validació creuada *leave one out* que consisteix en crear n-1 nombre de models a partir de les n mostres que té la matriu inicial. Cada model té n-1 nombre de mostres, és a dir, es fan servir totes les mostres de la matriu menys una, ja que aquesta darrera serà utilitzada per ser predita.
- La segona, la validació mitjançant *test set*, consisteix en fer servir només una part de les mostres de la classe per crear el model, mostres *training*, i fer servir la part que s'ha deixat fora per predir, mostres *test*. Només es crea un model, a partir de les mostres *training*, i amb aquest, es prediu tant les mostres del test com les mostres que no pertanyen a la classe modelada.

5. Resultats

5.1. NIR

5.1.1. Espectres

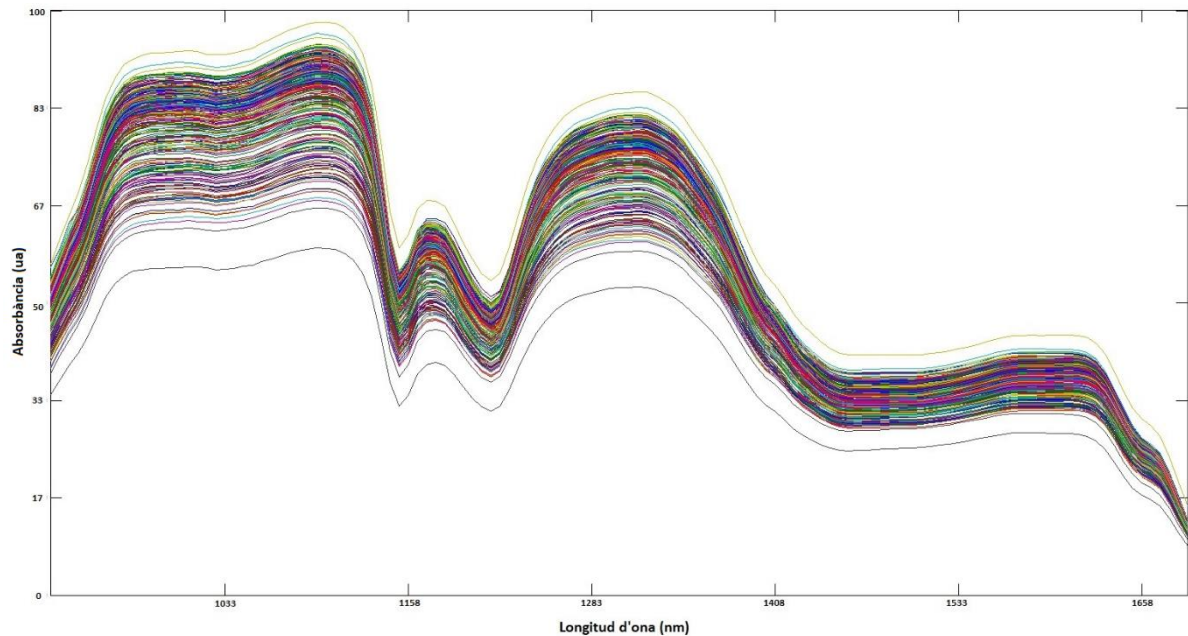


Figura 9. Espectres NIR superposats de les mostres no adulterades i adulterades.

A la figura 9 es representa l'espectre NIR de totes les mostres tant adulterades com no adulterades superposades en un mateix gràfic. Es pot observar, per una banda, cap mostra adulterada presenta cap pic d'absorbància diferent que sigui característic de l'adulterant. Per altra banda, els espectres no tenen la mateixa intensitat indicant que el grau d'adulteració ha pogut influir¹².

L'espectre ens dona la mateixa informació qualitativa, aparentment, la figura 10 presenta mitjanes de cada classe on a simple vista no es poden veure diferències notables. També caldria trobar diferències entre els graus de concentració de contaminant.

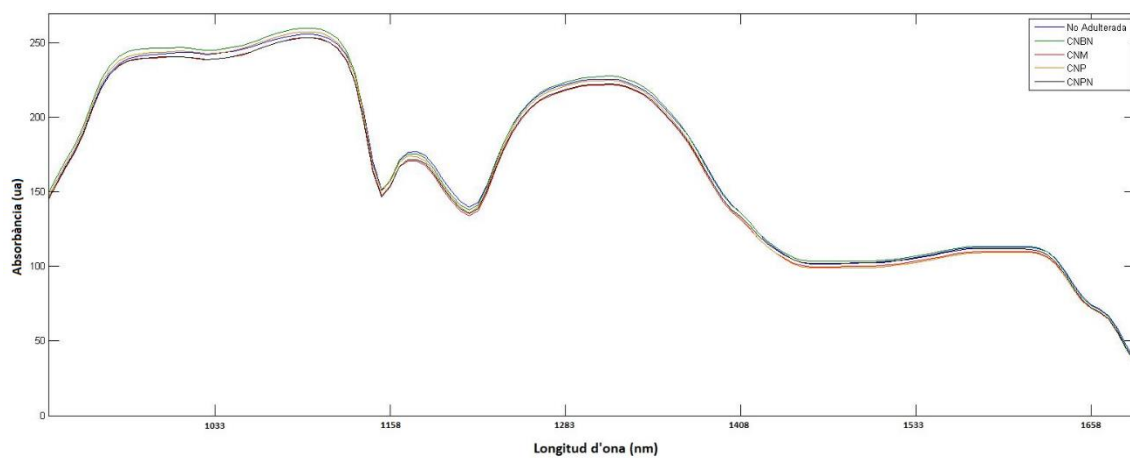


Figura 10. Espectres NIR de les mitjanes de les mostres no adulterades i adulterades.

5.2.1.1. Pretractaments

D'entre les diferents modificacions matemàtiques que s'han emprat per millorar l'estudi de les dades hi destaquen el centrat, l'autoescalat i la primera derivada.

El centrat i l'autoescalat fan la funció d'igualar les escales i la variabilitat de les dades, és interessant emprar-les als gràfics NIR, ja que tant a la figura 10 com 11 es pot observar una gran fluctuació entre els valors més alts i els més baixos.

La primera derivada s'aplica per millorar la resolució, és una tècnica àmpliament utilitzada quan es tracta del NIR¹³, a la figura 11 es pot observar com queda l'espectre NIR després d'aplicar-s'hi.

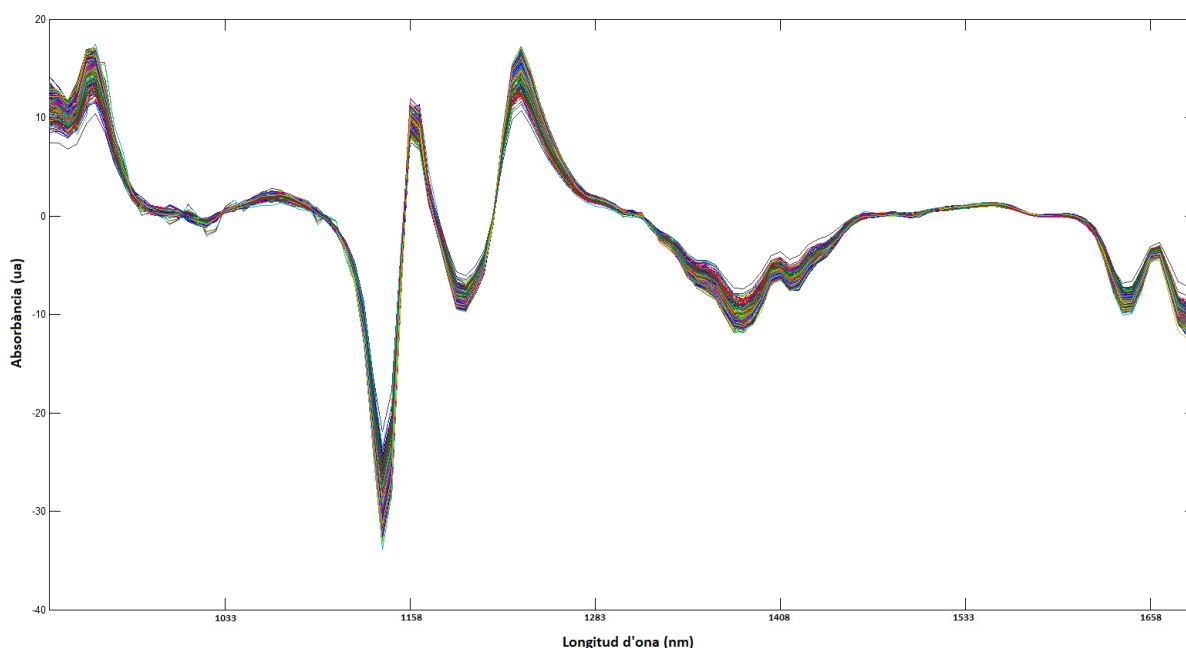


Figura 11. Espectre NIR pretractat amb la primera derivada.

5.1.2. Anàlisi exploratori de dades (PCA)

Per començar a estudiar les dades s'hi fa anàlisi exploratòria de dades al conjunt de les mostres adulterades i no adulterades per observar l'existència d'agrupacions o tendències a les seves dades.

Se li fica un nom a les mostres segons la informació que es té d'elles, aquest nom serà informatiu només per l'exploració de dades, ja que, el programa no sap de quin tipus de mostra es tracta. Les dades procedents de les mostres que no han estat adulterades s'anomenaran "No adulterades", mentre que aquelles que si ho han estat s'anomenaran igual que el seu respectiu adulterant (CNP, CNBN, CNM o CNP). Es comença fent servir aquest criteri perquè és una manera de separar les mostres per la seva estructura química, cada adulterant té una composició distinta a la resta i s'espera que amb l'anàlisi de dades del NIR es trobin diferències. Idealment, es trobarien agrupacions distintes de cadascuna de les 5 marques que s'han fet, però també seria un bon resultat que la PCA en pogues separar les mostres no adulterades de les adulterades.

La primera PCA es crea sense pretractaments, a la figura 12 es pot veure com s'obté una representació de les mostres en la qual no es pot treure cap tipus de conclusió esperable de

l'exploració de les dades. Aquesta PCA justifica l'ús de pretractaments, ja que les dades originals no mostren, almenys algunes, signes de tendència.

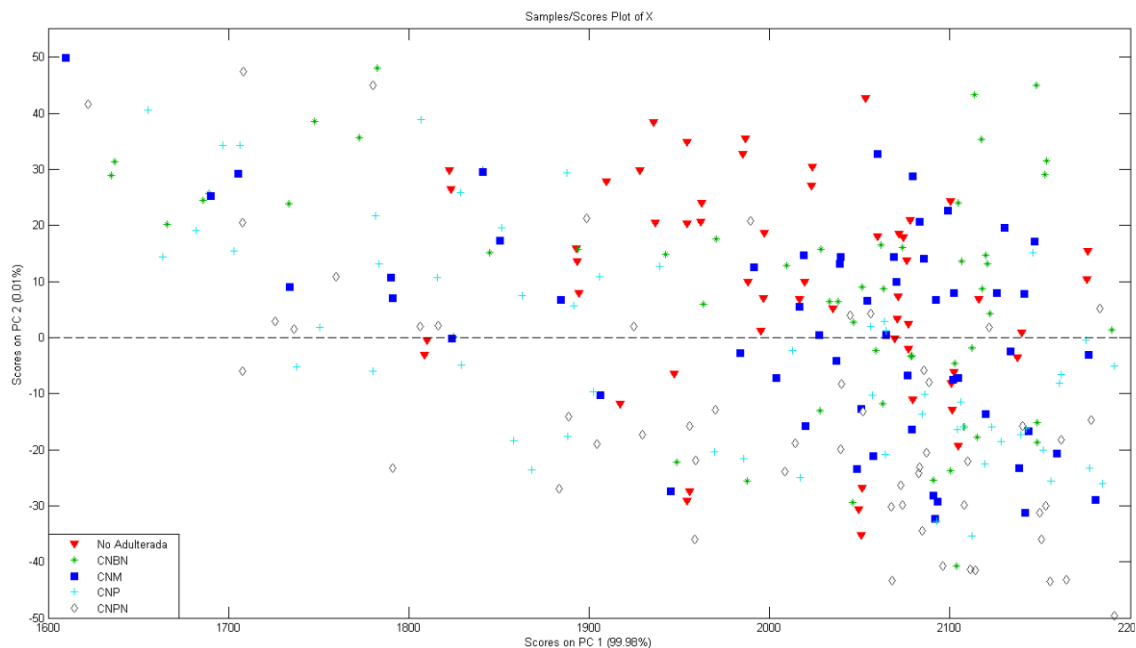


Figura 12. Anàlisi de Components Principals (PCA) de les dades NIR sense pretractar.

D'entre els diferents PCA realitzats, s'exposa a la figura 13 el que està format pels pretractaments de centrat i primera derivada, en ell es poden veure dues clares agrupacions de les classes, la primera és de les mostres no adulterades que queden aïllades de la resta de punts exceptuant-ne 6, la segona la formen les mostres adulterades que encara que no es separin entre elles, sí que s'aconsegueix que s'aïllin de les mostres no adulterades. Per tant, la classe no adulterada queda quasi completament separada de les quatre classes no adulterades, sent un resultat adequat per a fer un model de classificació One Class SIMCA¹⁴.

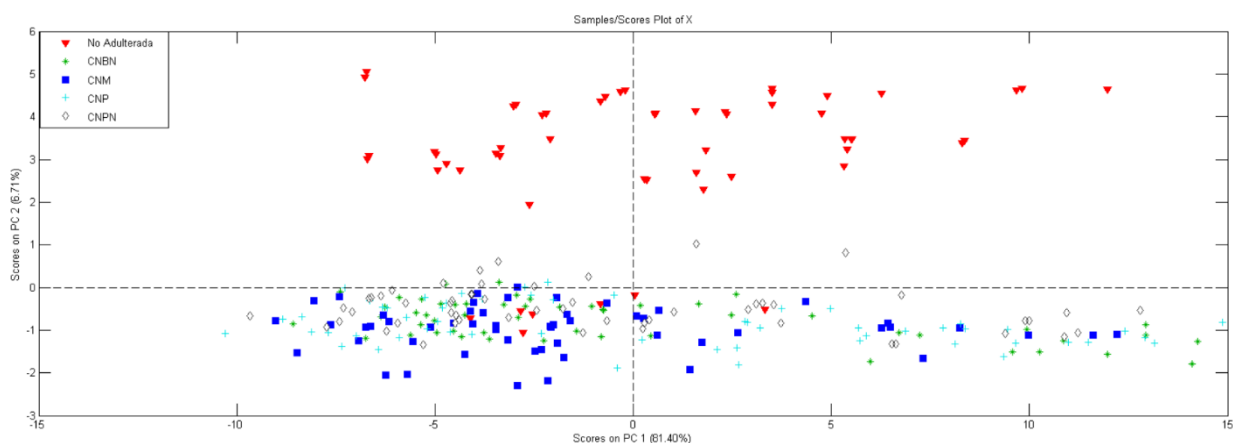


Figura 13. Anàlisi de Components Principals (PCA) de les dades NIR pretractades amb centrat i primera derivada.

Respecte als resultats d'altres anàlisis de components principals amb altres pretractaments, s'aconsegueixen formar agrupacions fent servir pretractaments com l'autoescalat o el centrat per separat, però les tendències no es veuen tan clares ni la variància queda ben repartida entre els primers components principals on el CP1 acumula cerca del 98% de la informació i el CP 2 hi acumula menys d'un 1%.

Les dades originals obtingudes pel NIR també es poden analitzar segons el seu grau de concentració en comptes de per la seva composició química com s'ha fet abans. La nova divisió es podria arribar a visualitzar 9 classes, on tindríem el grup de mostres no adulterades amb un grau de concentració 0 i altres 8 grups amb concentracions que van de 0.1% fins 10%. Així i tot, seria convenient no emprar tantes marques, ja que visualment no s'apreciarien les agrupacions o tendències.

Per veure com es comporten els diferents graus d'adulteració, els 8 percentatges de concentració es marquen 3 grups:

- L1: mostres amb 10.0 i 5.0 % de concentració d'adulterant.
- L2: mostres amb 2.5, 1.0 i 0.6 % de concentració d'adulterant.
- L3: mostres amb 0.3, 0.2 i 0.1 % de concentració d'adulterant.

S'han agrupat els antics graus d'adulteració en trams de concentració, d'aquesta manera hi haurà 4 grups de dades a la PCA i resultarà més fàcil detectar si s'han format tendències o no. Els 3 grups d'adulterants L1, L2 i L3 segueixen una progressió de més adulterades, L1, cap a menys adulterades, L3.

El PCA de la figura 14 s'ha fet amb les mateixes condicions de la PCA de la figura 13, amb els pretractaments de 1ª derivada i centrat. D'aquest canvi, no es pot extreure gaire, no s'aconsegueix diferenciar entre el grau d'adulteració, però sí entre si la mostra ha estat adulterada o no.

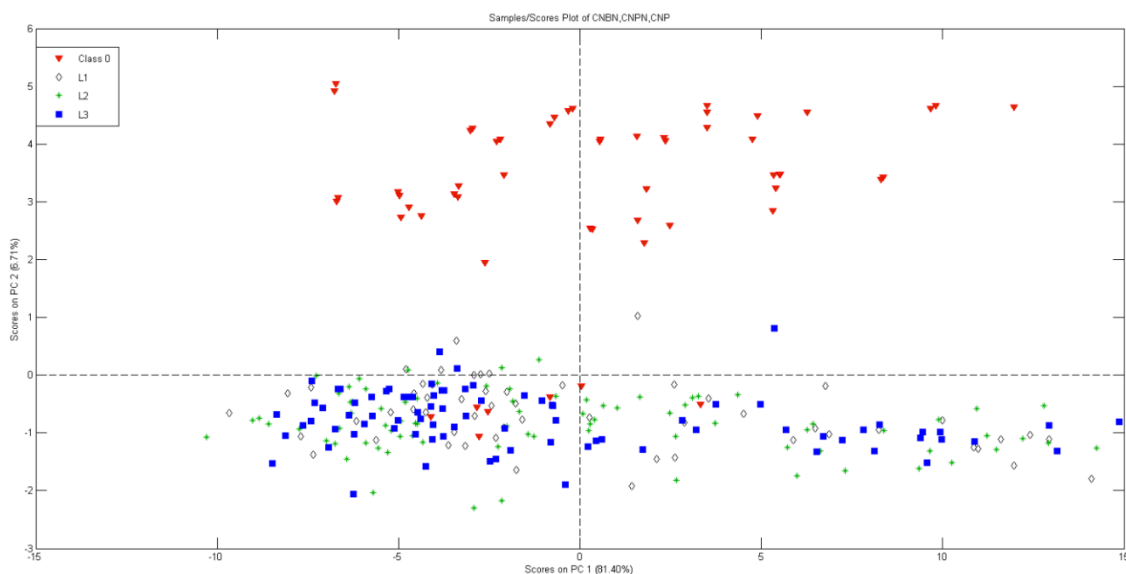


Figura 14. Anàlisi de Components Principals (PCA) de les dades NIR pretractades amb centrat i primera derivada.

Una possible explicació a què totes les mostres adulterades estiguin dispersades a la mateixa zona a pesar de tenir graus de concentració diferents és que el llindar de concentració necessària per trobar diferències entre no adulterades i adulterades és molt baix, és a dir, que es podria arribar a detectar mostres per sota del 0.1% de concentració.

Una altra variant d'aquesta PCA s'aconsegueix emprant la segona derivada en comptes de la primera, els resultats són semblants, les mostres no adulterades es separen però les adulterades no, també cal fer servir 3 components principals en una representació tridimensional per observar les tendències. La PCA sense pretractament no dóna bons resultats.

5.1.3. Tècnica classificatòria: Soft Independent Modelling of Class Analogy

Pel model SIMCA se segueix l'estratègia One Class, el model es fa amb les mostres no adulterades i es fa la predicció dels 4 grups de mostres adulterades, idealment, les mostres adulterades haurien de ser considerades fora del model i les mostres del set training dins. Una altra situació esperable seria que els anacards amb menys percentatge d'adulteració siguin considerats dins del model i progressivament, els anacards amb més contaminació siguin predits cada vegada més mostres fora del model.

L'ús de l'estratègia One Class es fa servir per diverses raons, com s'ha vist a l'anàlisi exploratòria de dades, l'única classe que es diferencia de la resta és la classe no adulterada, així, és ideal que les mostres no adulterades formin el model en què basar l'estudi. Per altra banda, si es volgués seguir una estratègia Multi Class també hi seria possible, però la gran quantitat de fruits secs o compostos químics que podrien funcionar com adulterants no es podrien abastar amb els 4 que hi ha disponible i cabria esperar-se del model més resultats inconclusos de múltiple assignació o de falta d'assignació.

El model creat a continuació s'ha fet seguint l'esquema de la Figura 15, el conjunt de dades es divideixen en 5 classes, una no adulterada i quatre adulterades, després es realitza el SIMCA agafant la classe no adulterada pretractada amb la primera derivada i el centrat per crear el model, mentre les classes adulterades són predites.

Per valorar els resultats es tenen en compte els paràmetres de qualitat que en són la sensibilitat, l'especificitat i l'eficàcia. Els paràmetres han sigut calculats a partir dels valors de la T2, Q i la distància, però en aquest cas només es presenten els resultats del SIMCA restrictiu.

A banda d'aquest esquema, també s'han seguit d'altres on la principal variació la trobem en l'ús de diferents pretractaments abans de muntar el model, però amb resultats menys destacables que els que en trobem aquí.

A la taula 3 estan representats els resultats de les diferents variacions que trobem al crear el model, l'especificitat representada a les taules correspon a la mitja obtinguda per les quatre classes adulterades.. El millor resultat s'obté utilitzant la primera derivada i el centrat com

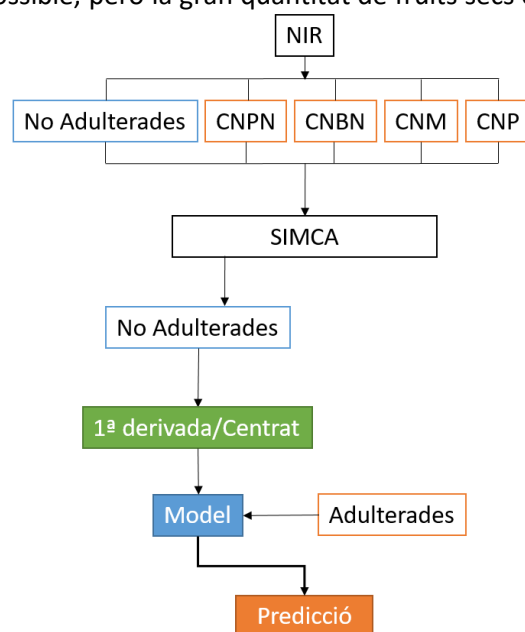


Figura 15. Esquema del model SIMCA.

	1ª derivada	2ª derivada	Centrat	1ª derivada	2ª derivada	Autoescalat
	Centrat	Centrat		Autoescalat	Autoescalat	
Sensibilitat	87,5	83,93	92,86	85,71	85,71	94,64
Especificitat	100	100	35,71	100	100	74,11
Eficàcia	93,75	91,96	64,29	92,86	92,86	84,38

Taula 3. Paràmetres de qualitat dels models SIMCA generats a partir dels espectres NIR.

pretractaments, s'arriba a una sensibilitat del 87,5% i una especificitat del 100%, tenint el valor més alt d'eficàcia d'entre tots els models creats.

Sobre la resta de variacions, s'ha arribat fins a una sensibilitat de 94,64% en el cas de fer servir només l'autoescalat, el que seria un bon resultat si no fos perquè l'especificitat que té el model és una mica baixa, el risc d'emprar aquest model seria obtenir falsos positius amb molta freqüència, empitjorant la capacitat de detectar adulterants. Respecte a l'aplicació de la primera i segona derivada, tant en el centrat com l'autoescalat hi ha una pujada de la sensibilitat fins al 100%, fent que sigui teòricament impossible trobar un fals positiu, l'eficàcia es queda al voltant del 93% sent un valor prou fiable per emprar el model.

Com el valor d'eficàcia obtingut emprant el centrat i la primera derivada és el més alt, es procedeix a aplicar la tècnica de validació mitjançant test set al model. Per fer-ho, la classe no adulterada, composta per 56 mostres, és dividida en 42 pel training i 14 pel test, escollides a l'atzar.

Amb el nou model creat es torna a calcular la sensibilitat, on s'arriba a què el training millora els resultats originals, però prediu amb errors les dades del test i la sensibilitat promig coincideix amb la sensibilitat original. La predicció de les mostres adulterades dóna el mateix resultat, s'obté un 100% d'especificitat (Taula 4).

Sensibilitat training	95,24
Sensibilitat test	42,86
Especificitat	100
Eficàcia	97,62

Taula 4. Paràmetres de qualitat del training-test i la predicció del model generat a partir del model.

La baixa sensibilitat del test i l'alta del training, impliquen que s'hauria de canviar la divisió feta entre les mostres del training i test set, estudiant i aplicant algorismes de selecció més representatius que no s'han abordat en aquest treball.

5.2. ATR

El treball que s'ha portat a terme seguint el mateix esquema que al NIR.

5.2.1. Espectres

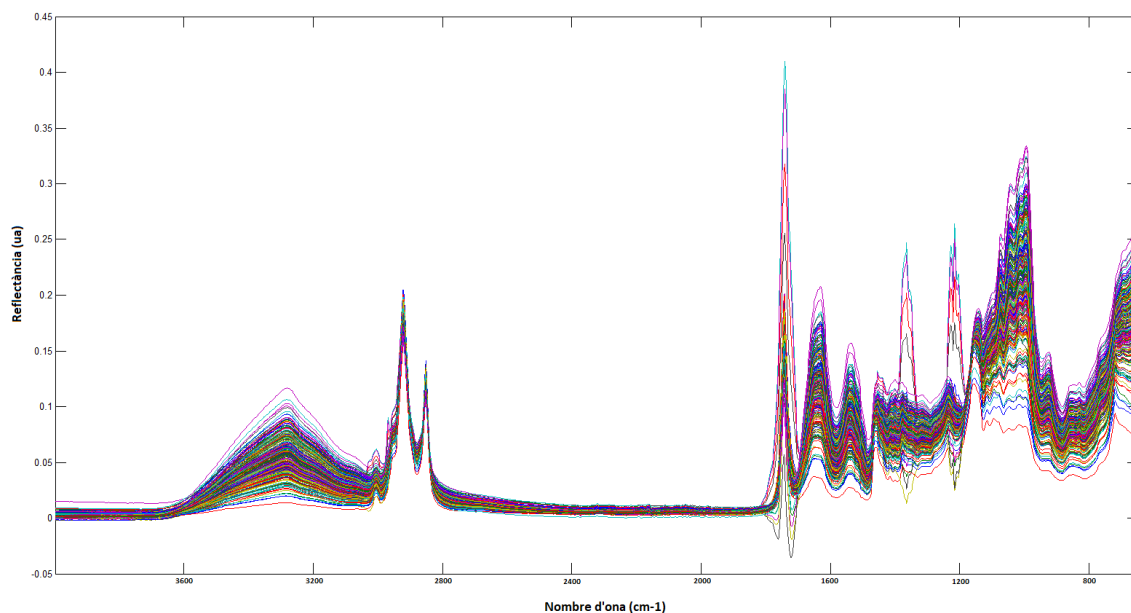


Figura 16. Espectres ATR superposats de les mostres no adulterades i adulterades. (Corregir Escala).

A la figura 16 en tenim els 280 espectres ATR superposats en un mateix pla, la regió que ocupen va dels números d'ona que van des dels 4000 cm^{-1} fins els 650 cm^{-1} amb una resolució de 2 cm^{-1} , donant un total de 1676 variables per cadascuna de les mostres estudiades.

Destaquen dues zones principals, la primera compresa entre els nombres d'ona de 3600 cm^{-1} i 2800 cm^{-1} on s'observen 3 pics a diferent alçada, la segona va dels 1800 cm^{-1} fins els 650 cm^{-1} on es pot veure una àrea amb més pics, més soroll i amb més particularitats entre les respostes instrumentals.

Tal com reflecteix la figura 17, que està creada a partir de les mitjanes de cadascun dels adulterants i la mitjana de la mostra no adulterada, existeix una notable diferència entre les classes d'adulterant i la mateixa matriu sense adulterar, però, a les regions sense pics on l'espectre viatja paral·lel a la línia base, se'n troben comportaments similars.

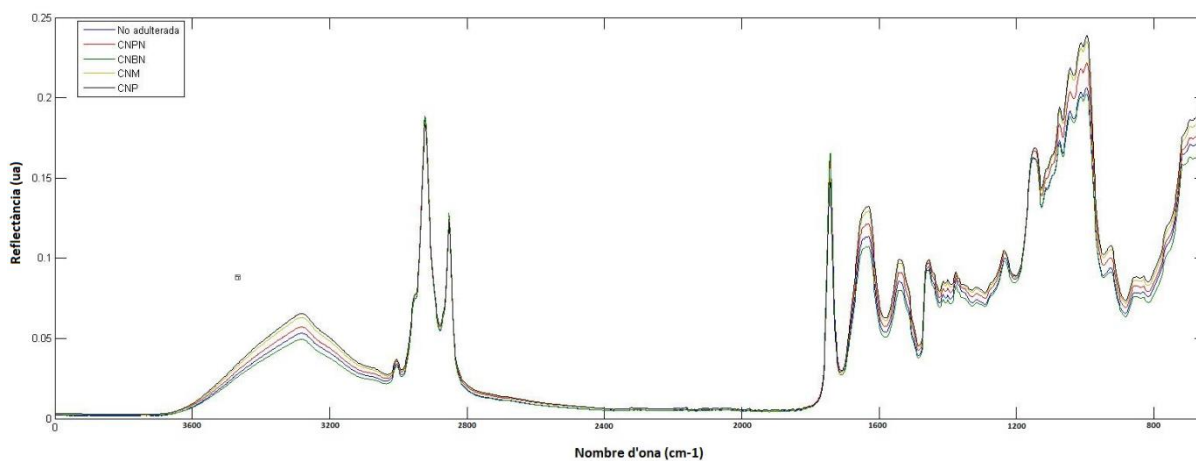


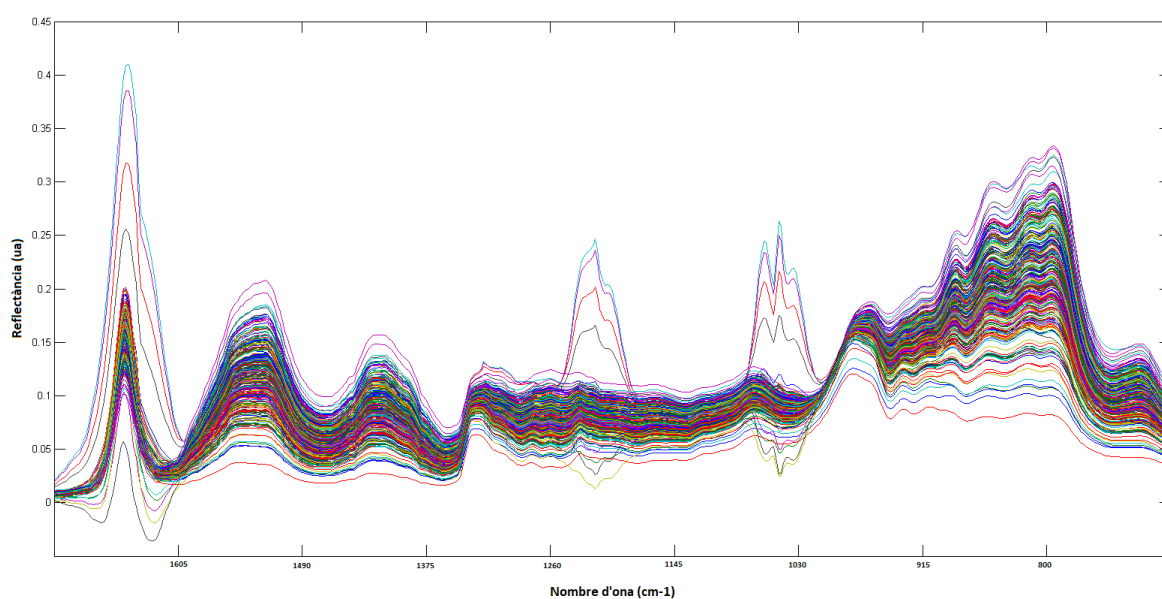
Figura 17. Espectres ATR de les mitjanes de les mostres no adulterades i adulterades. (corregir escala a 4000 cm^{-1})

5.2.1.1. Pretractaments

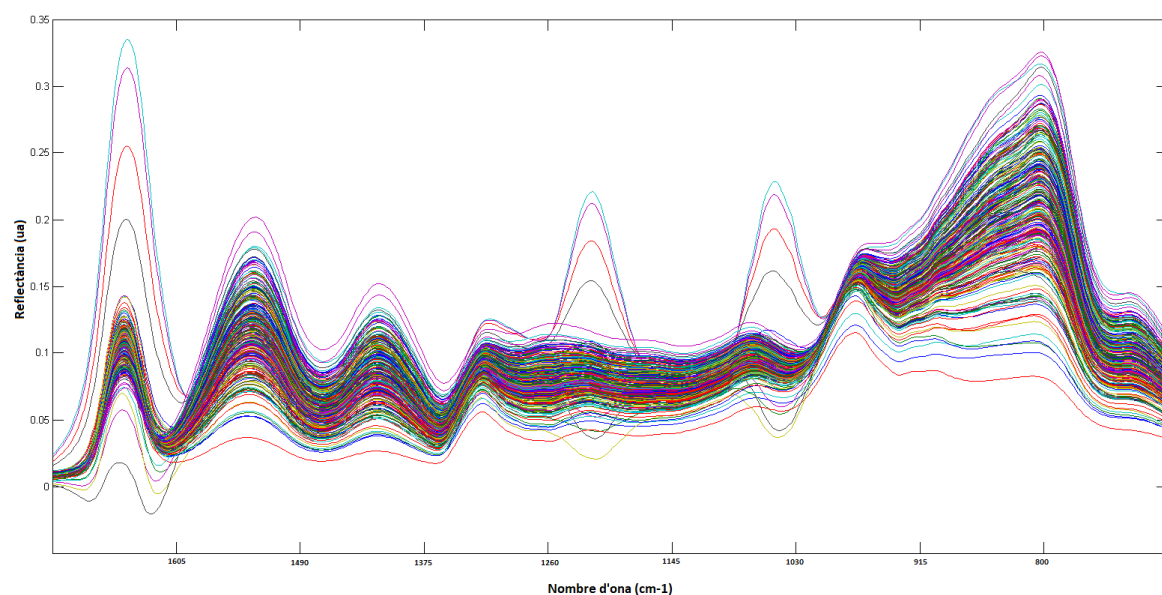
L'autoescalat i el centrat són pretractaments que s'han d'aplicar en quasi la totalitat d'anàlisi degut a la seva capacitat d'igualar escales i fer que les variables siguin més comparables entre elles, s'utilitza en solitari o en acompanyament d'altra tècnica que no centri les dades per si mateixa com pot ser la correcció de la línia base.

Fer servir el suavitzat permetria reduir el soroll de la zona d'entre 1600 cm^{-1} i 650 cm^{-1} fent que la variància que porti cada component principal no sigui tan detallada, però amb més capacitat d'informació quant a que el soroll eliminat no aporta informació útil.

A la figura 18 i 19 es pot observar l'efecte del suavitzat. En primera instància estan les dades de les mostres sense pretractament on s'observen una gran quantitat de pics entre els nombres d'ona d'entre 1600 cm^{-1} i 800 cm^{-1} . A l'aplicar-s'hi el pretractament, les dades se suavitzen considerablement.



Il·lustració 18. Rang d'entre 1700 cm^{-1} i 700 cm^{-1} de l'espectre ATR original.



Il·lustració 19. Rang d'entre 1700 cm^{-1} i 700 cm^{-1} de l'espectre ATR amb el pretractament de suavitzat.

5.2.2. Anàlisi exploratori de dades (PCA)

Un objectiu de les PCA és trobar esquemes de classificació i/o pretractaments candidats per poder crear futurs models SIMCA eficaços, òbviament, els pretractaments que donin resultats poc satisfactoris no seran precursors de la tècnica classificatòria.

L'exploració de dades comença amb l'estudi dels comportaments de la classe no adulterada amb les quatre classes que si ho han estat, igual que amb les dades procedents del NIR, a continuació també es farà una anàlisi exploratòria dividint les mostres segons el seu grau d'adulteració.

D'entre els múltiples PCA realitzats, es destaca el fet amb les tècniques de GLSW, suavitzat i MSC. El resultat ens indica una clara separació de les mostres no adulterades i també una separació parcial entre les mostres no adulterades, sobretot a l'adulterant CNP que és el que millor se separa dels altres adulterants encara que també s'apropa més a l'àrea ocupada per les mostres sense adulterar tal com es veu a la figura 21.

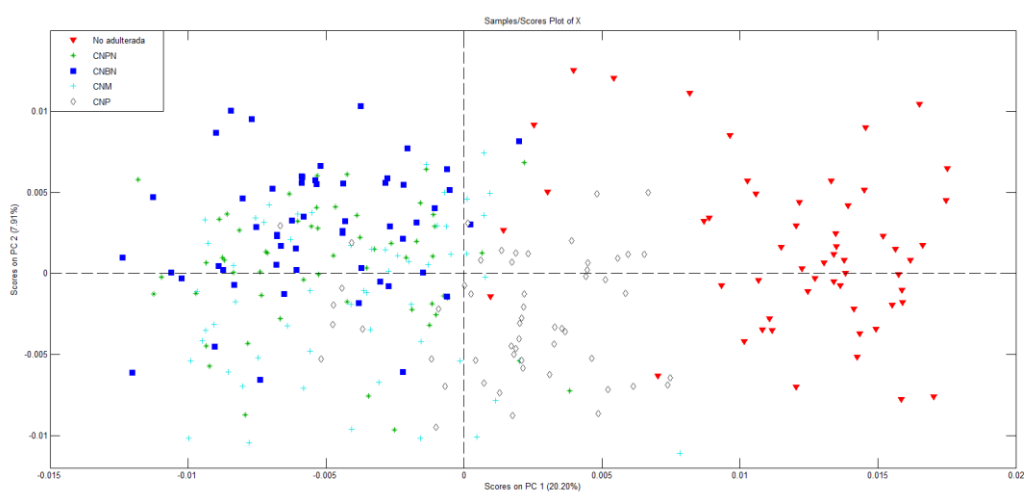


Figura 21. Anàlisi de Components Principals (PCA) de les dades ATR pretractades amb GLSW, suavitzat i MSC.

La resta de variacions emprades per produir gràfics de PCA, donen resultats similars entre si. Respecte a la classificació de les mostres per grau d'adulteració, es va realitzar amb els mateixos pretractaments que el PCA classificat per tipus d'adulterant. Per fer aquest PCA es classifiquen les mostres en L1, L2 i L3 com es va fer anteriorment.

A la figura 22 es pot veure com continua la separació entre mostres no adulterades i adulterades, en canvi, ja no es veuen tendències entre les adulterades, la concentració de les mostres no queda palpable al gràfic.

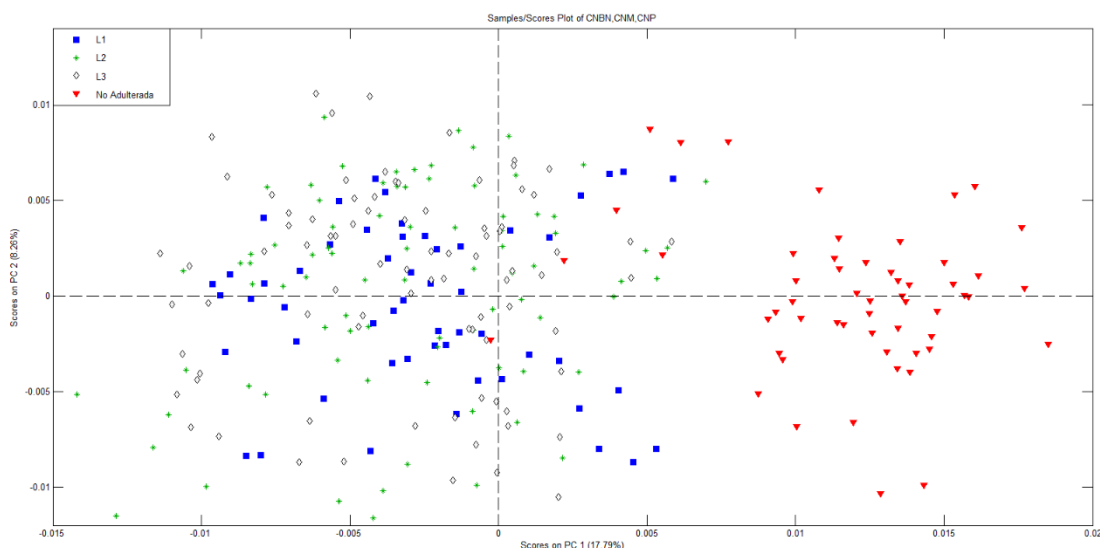


Figura 22. Anàlisi de Components Principals (PCA) de les dades ATR pretractades amb GLSW, suavitzat i MSC.

Encara que sembli que els gràfics aportin la mateixa informació, el primer ens indica que fàcilment les mostres adulterades s'agrupen al mateix espai, sent una opció viable classificar-les a totes com una sola classe adulterada, en canvi, el segon gràfic ens indica que una adulteració de 0.1% es separa igual del 0% que una mostra del 10%. A efectes pràctics, els dos esquemes donarien resultats equivalents en un model SIMCA.

5.2.3. Tècnica classificatòria: Soft Independent Modelling of Class Analogy

Pel model SIMCA es segueix l'estratègia One Class amb l'objectiu de dir si les mostres analitzades són o no adulterades, sense focalitzar en quin tipus d'adulterant s'ha afegit a la mostra, igualment, es podria crear un model Multi Class amb els distints contaminants, però com ja s'ha comentat, no es considera del tot adient seguir aquesta estratègia.

Es segueix un esquema similar amb les dades del ATR, les mostres es tornen a classificar per tipus d'adulterant i varien els pretractaments emprats, que són els mateixos del PCA.

Al buscar el millor model possible, s'han fet servir altres pretractaments indicats a la Taula 5 que resumeix els valors dels paràmetres de qualitat amb SIMCA restrictiu. Entre les variacions que han donat bons resultats en tenim les següents:

	Suavitzat	2ª derivada	1ª derivada	Suavitzat
	Autoescalat	Autoescalat	MSC	GLSW
Sensibilitat	82,14	75,00	77,55	78'5
Especificitat	45,54	89,26	25,89	100
Eficàcia	63,84	82,14	51,72	89,28

Taula 5. Paràmetres de qualitat dels models SIMCA generats a partir dels espectres ATR.

El model més bo és que el que arriba fins al 89,28 % d'eficàcia, que té un 100% d'especificitat, però un 78,5% de sensibilitat, indicant que el model li costa detectar mostres que si pertanyen a la classe no adulterada i tendeix a deixar a valorar com negatives mostres que són realment positives.

Pel que fa a la resta de pretractaments emprats, en tenim la derivació i el suavitzat, dues tècniques que no es poden emprar a l'hora, ja que fan servir el mètode Savitsky-Golay que no és preferible no superposar amb les mateixes dades. Per altra banda, altres tècniques no ensenyades a la taula com poden ser el centrat o el normalitzat, no acabaven de donar valors, en alguns casos, superiors a 20% d'especificitat.

Una vegada analitzats els valors obtinguts mitjançant la validació creuada *leave one out*, es procedeix a realitzar el *training test* per validar el model amb els pretractaments GLSW, suavitzat i MSC. Igual que amb el NIR, s'agafen 42 mostres a l'atzar de la classe no adulterada per fer el training i altres 14 pel test.

Els resultats no varien gaire dels obtinguts prèviament, a la taula 6 es poden observar els paràmetres de qualitat validats, els valors augmenten considerablement com és el cas de la sensibilitat i se segueix obtenint un 100% d'especificitat.

Sensibilitat training	92,86
Sensibilitat test	71,43
Especificitat	100
Eficàcia	96,42

Taula 6. Paràmetres de qualitat del training-test i la predicció del model generat a partir del

La sensibilitat del test no és gaire semblant a la del training, tal com es va comentar amb el NIR, s'hauria de revisar l'algoritme amb els que s'escullen les mostres training i del test set.

Comparant resultats entre NIR i ATR a la taula 7, si no es té en compte la sensibilitat del test, s'observa com els resultats del NIR són superiors als obtinguts per l'ATR, tot i això, tots dos models són bons i haurien de ser capaços de predir mostres desconegudes amb fiabilitat gràcies al valor alt de sensibilitat obtingut al training i al 100% de sensibilitat aconseguit.

	NIR	ATR
Sensibilitat training	95,24	92,86
Sensibilitat test	42,86	71,43
Especificitat	100	100
Eficàcia	97,62	96,42

Taula 7. Paràmetres de qualitat del training-test i la predicció del model generat dels models NIR i ATR.

Respecte a les tècniques, el NIR té l'avantatge de ser molt més simple i assequible instrumentalment, això li donaria avantatge al model generat a partir d'aquesta tècnica per poder ser emprat en un futur. Per altra banda, l'ATR necessita una instrumentació més cara i la seva velocitat d'anàlisi és més lenta ja que maneja un major nombre de variables.

6. Conclusions

S'ha desenvolupat un mètode per determinar l'adulteració en pastes d'anacards del qual s'han pogut extreure dos models validats d'eficàcies del 97.2% i 96.52% en anàlisis classificatòries mitjançant validació creuada.

Amb l'anàlisi exploratòria de dades, s'ha pogut estudiar el comportament de les mostres i estudiar quins pretractaments eren els més idonis per aconseguir separar les mostres segons la seva adulteració, a partir del PCA s'ha pogut definir quines eren les classes d'interès que formarien el model SIMCA. En el cas del NIR, s'ha arribat a la decisió que una derivació i un centrat serien la millor alternativa per modelar, mentre que a l'ATR ha requerit un pretractament previ més complex que incorpora l'aplicació de GLS, suavitzat i MSC.

La tècnica de classificació SIMCA restrictiu ha sigut capaç de caracteritzar les mostres i determinar la seva classe, la validació dels dos models creats ha aconseguit uns paràmetres de qualitat del 100% de sensibilitat i una mitja del 93% de sensibilitat, sent una qualitat dels models són prou bones per a predir mostres desconegudes amb fiabilitat.

Les mostres han estat analitzades a partir de l'ús de tècniques fotomètriques com el NIR i l'ATR, la possibilitat de combinar ambdues tècniques per millorar els resultats va ser estudiada però, finalment no es va profunditzar en aquesta estratègia per diversos motius, entre els quals destaquen, els bons resultats obtinguts a partir de les dades individuals i la dificultat de combinar les 125 variables del NIR enfront de les 1676 de l'ATR (reduïbles per selecció de variables).

El model creat a partir del NIR ha aconseguit ser el millor, té uns paràmetres de qualitat més alts, és una tècnica espectromètrica ràpida d'emprar i de costos baixos, també genera menys variables i això fa que computacionalment sigui més ràpida d'analitzar i econòmica a l'hora d'emmagatzemar la informació.

En general, s'han complert satisfactòriament tots els objectius del treball, l'anàlisi qualitativa multivariant ha demostrat ser eficaç a l'hora de determinar l'adulteració en els aliments i una eina important a l'hora de detectar frauds gràcies a l'ús de diverses tècniques quimiomètriques.

7. Bibliografia

1. Food and Agriculture Organization of the United Nations <http://www.fao.org/faostat/en/#data/QC/> (accés el 8 de Juny, 2021).
2. BOE.es - BOE-A-2015-8563 Ley 28/2015, de 30 de julio, para la defensa de la calidad alimentaria. <https://www.boe.es/buscar/act.php?id=BOE-A-2015-8563> (accés el 9 de juny, 2021).
3. Márquez, C.; López, M. I.; Ruisánchez, I.; Callao, M.P.; FT-Raman and NIR Spectroscopy Data Fusion Strategy for Multivariate Qualitative Analysis of Food Fraud. *Talanta*. **2016**, *161*, 80–86. <https://doi.org/10.1016/j.talanta.2016.08.003>.
4. Miaw, C. S. W.; Santos, P. M.; Silva, A. R. C. S.; Gozzi, A.; Guimarães, N. C. C.; Callao, M. P.; Ruisánchez, I.; Sena, M. M.; de Souza, S. V. C. Comparison of Different Multivariate Classification Methods for the Detection of Adulterations in Grape Nectars by Using Low-Field Nuclear Magnetic Resonance. *Food Analytical Methods*. **2019**, *13* (1), 108–118. <https://doi.org/10.1007/s12161-019-01522-7>.
5. Sheng Whei, C.; Martins, M.; Vitorino, S.; Callao, M. Detection of Adulterants in Grape Nectars by Attenuated Total Reflectance Fourier-Transform Mid-Infrared Spectroscopy and Multivariate Classification Strategies. *Food Chemistry*. **2018**, *266*, 254–261. <https://doi.org/10.1016/j.foodchem.2018.06.006>
6. Miaw, C. S. W.; Sena, M. M.; Souza, S. V. C. de; Ruisánchez, I.; Callao, M. P. Variable Selection for Multivariate Classification Aiming to Detect Individual Adulterants and Their Blends in Grape Nectars. *Talanta*. **2018**, *190*, 55–61. <https://doi.org/10.1016/j.talanta.2018.07.078>.
7. Callao, M. P.; Ruisánchez, I. An Overview of Multivariate Qualitative Methods for Food Fraud Detection. *Food Control*. **2018**, *86*, 283–293. <https://doi.org/10.1016/j.foodcont.2017.11.034>.
8. Pasquini, C.; Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. *Journal of the Brazilian Chemical Society*. **2003**, *14* (2), 198–219. <https://doi.org/10.1590/S0103-50532003000200006>.
9. Guiliano, M. Asia, L. Onoradini, G. Mille, G.; Applications of Diamond Crystal ATR FTIR Spectroscopy to the Characterization of Ambers. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*. **2007**, *67* (5), 1407–1411. <https://doi.org/10.1016/j.saa.2006.10.033>.
10. Miller, J. N.; Miller, J. C. Estadística Y Quimiometría Para Química Analítica; *Pearson*, **2002**, 224-228.
11. M Blanco; CerdáV. Serie Química Analítica: Quimiometría.; Universitat Autònoma De Barcelona: *Bellaterra*, **1988**.

12. Woodcock, T.; Downey, G.; O'Donnell, C. P. Confirmation of Declared Provenance of European Extra Virgin Olive Oil Samples by NIR Spectroscopy. *Journal of Agricultural and Food Chemistry*. **2008**, 56 (23), 11520–11525. <https://doi.org/10.1021/jf802792d>.
13. Skvaril, J.; Kyprianidis, K.; Avelin, A.; Odlare, M.; Dahlquist, E. Utilization of near Infrared (NIR) Spectrometry for Detection of Glass in the Waste-Based Fuel. *Energy Procedia*. **2015**, 75, 734–741. <https://doi.org/10.1016/j.egypro.2015.07.502>.
14. Kang, S.; Cho, S.; Kang, P. Multi-Class Classification via Heterogeneous Ensemble of One-Class Classifiers. *Engineering Applications of Artificial Intelligence*. **2015**, 43, 35–43. <https://doi.org/10.1016/j.engappai.2015.04.003>.

