

APARTAT III

ANÀLISI QUANTITATIVA

RESOLUCIÓ DE PROBLEMES QUANTITATIUS. DETERMINACIÓ DE L'ÍNDEX D'ACIDESA EN OLIS D'OLIVA PER ESPECTROFOTOMETRIA FTIR-ATR^A

Com s'ha esmentat anteriorment, l'estudi d'un conjunt de dades mitjançant eines quimiomètriques ve condicionat fonamentalment pel tipus d'informació que se'n vol obtenir, ja sigui qualitativa o quantitativa.

En espectrofotometria IR, tot i que les bandes que s'originen són força específiques, el procés de calibració univariant pot arribar a ser complicat. En aquells casos en els que es donen solapaments amb altres bandes, o bé la contribució de l'analit al senyal global és petita, els resultats que s'obtenen ja no són tan satisfactoris i convé buscar altres mètodes de calibració.

Els mètodes multivariants, conjuntament amb els pretractaments del senyal són una bona alternativa, ja que permeten la introducció de major nombre de variables, el que equival a més informació, i poden arribar a 'destriar' les petites contribucions de l'analit al senyal global.

Es proposa un mètode alternatiu de determinació de l'índex d'acidesa lliure en olis d'oliva, una determinació relativament senzilla però lenta de resoldre per via humida, fent servir dades espectrofotomètriques en l'Infraroig Mitjà. Al llarg del treball, es veu

^A Treball publicat sota el títol: 'Determination of Olive Oil Free Fatty Acid by Fourier Transform Infrared Spectroscopy' al *Journal of American Oil Chemists' Society*, **76**, 611-, (1999).

que el problema és complex i és necessari seguir una certa sistematicitat a l'hora de seleccionar les mostres per tal d'obtenir resultats satisfactoris.

L'oli d'oliva és un ingredient essencial de la dieta mediterrània. La Unió Europea (UE) ha establert diferents categories d'oli d'oliva atenent a les condicions de producció i/o al seu processat¹. Amb la finalitat de poder emmarcar els diferents productes en les seves categories, es fa necessari realitzar diferents anàlisis, d'entre les quals destaca l'índex d'acidesa lliure per la seva senzillesa en la determinació i la seva popularitat.

Els mètodes oficials, en general, són grans consumidors de temps i de reactius. Alguns d'ells són complicats, altres fan servir moltes hores de mà d'obra especialitzada i molts d'ells generen residus que han de ser tractats. Aquests inconvenients poden ser evitats en part fent servir tècniques espectrofotomètriques, que permeten el tractament de gran número de mostres per part de personal amb uns coneixements mínims i fent servir un volum mínim (o nul en alguns casos) de reactius. Aquests aspectes, principalment, han estat els impulsors de les tècniques espectrofotomètriques com a tècniques alternatives i han fet que es popularitzessin enormement. En el camp de l'oli d'oliva se n'han desenvolupat d'orientades a la detecció d'adulteracions², a la preparació de sistemes de classificació³, i a la determinació de l'origen geogràfic⁴. També s'han desenvolupat anàlisis quantitatives basades en tècniques espectrofotomètriques, sense fer servir separacions prèvies. En aquesta òrbita, l'espectroscopia infrarroja de transformada de Fourier, (FTIR), ha esdevingut una alternativa vàlida per la simplicitat en la preparació i tractament de la mostra. Prova d'això n'han estat les aplicacions desenvolupades per a l'anàlisi de lípids, com la monitorització de l'oxidació d'olis comestibles⁵, la determinació del contingut en enllaços cis- i trans-⁶, l'índex de peròxids⁷ i l'índex d'acidesa lliure⁸.

La urgència en l'aplicació d'aquestes tècniques ha comportat, en molts casos, el fet de prestar poca atenció al procés de calibració en si. Dels diferents aspectes implicats, el que ha sortit més mal parat ha estat el de la selecció de mostres per a la preparació dels conjunts de calibració i validació. Com s'ha esmentat en apartats anteriors, aquesta part del procés és fonamental si es volen obtenir equacions de calibració que resultin representatives i amb una bona capacitat predictiva per a totes (o gairebé totes) les mostres de la població. D'aquesta manera, el desenvolupament de mètodes d'anàlisi en aquest context necessita una certa cura a l'hora d'utilitzar les principals tècniques d'anàlisi multivariable, fent servir mostres de predicció externes amb valors de concentració coneguts com a referents per a avaluar la consistència dels resultats^{9,10}.

L'objectiu concret ha estat desenvolupar un mètode d'anàlisi de l'índex d'acidesa lliure en olis d'oliva de diferents categories i orígens, fent servir espectres FTIR-ATR. Com a objectiu de fons, la comprovació de la viabilitat d'un mètode espectrofotomètric a l'hora de substituir un mètode oficial quan el número de mostres a analitzar és molt elevat i amb la idea de realitzar una determinació multiparamètrica a partir d'una única mesura espectral.

Metodologia Experimental

En el que fa referència a les **mostres**, es va intentar que l'estudi fos el més representatiu possible. Amb aquesta intenció es van seleccionar mostres d'olis d'oliva comercials de diferents orígens i categories. Es van analitzar un total de 87 mostres, de les quals 36 eren d'oli d'oliva verge, 40 d'oli pur d'oliva i 11 d'oli de sansa d'oliva. Els seus continguts en acidesa lliure oscil·laven entre el 0.1% i el 2.1%. Tot i que l'índex d'acidesa lliure dels olis verges extra acostuma a estar entre el 0% i l'1%, per a expandir l'interval de calibració, algunes d'aquestes mostres varen ser enriquides amb quantitats conegudes d'àcid oleic. Amb la intenció de disposar d'una calibració el més universal possible es van incorporar mostres amb diferents nivells d'estrès tèrmic, 13 mostres addicionals varen ser preparades per l'escalfament a 250°C durant 2-3 h de diferents olis d'oliva.

El contingut en àcids grassos lliures es va determinar fent servir el mètode de referència proposat per la normativa europea¹. Aquest consisteix en una volumetria en medi no aquós (éter dietílic – etanol 1:1) fent servir com a agent valorant una solució etanòlica de KOH 0.1 M.

Els espectres d'infraroig mitjà varen ser enregistrats amb un instrument de transformada de Fourier Perkin-Elmer 16PC, fent servir com a dispositiu de mesura un sistema ATR de canal dissenyat específicament per a aquest instrument per Graseby-Specac. Aquest dispositiu és adequat per a fluïds viscosos i fa servir un cristall de ZnSe tallat a 45° que proporciona 6 reflexions. Per a l'adquisició i processat de les dades espectrals es va fer servir el software de control i tractament *IR Data Manager* v. 3.34, també de Perkin-Elmer.

Per a obtenir l'espectre de cada mostra, es va dispersar per la superfície del cristall entre 1 i 1.5 mL d'oli, de manera que es cobris de forma uniforme tot el canal de mesura. Es va registrar un espectre per mostra, fent servir l'aire com a referència. Cadascun d'ells va ser el resultat de promitjar 50 escombrats al llarg de l'interval de números d'ona de 4000 a 650 cm^{-1} , fent servir una resolució de 2 cm^{-1} . Abans

d'enregistrar cada espectre, el cristall va ser netejat convenientment fent servir una solució aquosa a l'1% de Triton X-100 com a detergent. Seguidament va ser esbandit amb aigua destil·lada i alcohol isopropilic. A la figura AIII.1 es pot veure un exemple d'espectre FTIR-ATR obtingut mitjançant aquest procediment, acotat a l'interval 1775-775 cm^{-1} .

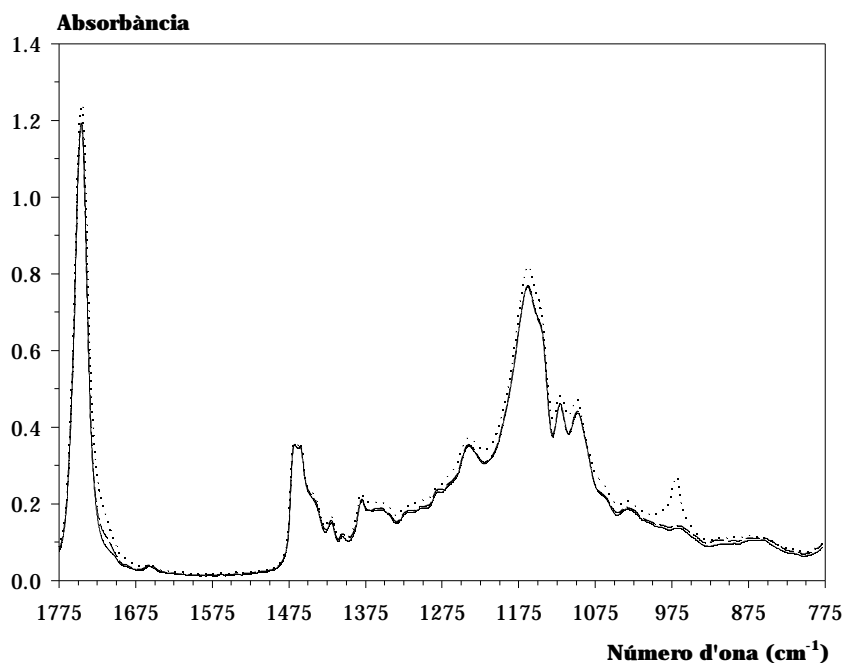


Figura AIII.1. Espectre FTIR-ATR d'absorbància per a olis d'oliva verges amb diferents continguts en àcids grassos lliures: baix (línia contínua), alt (línia segmentada) i estressades tèrmicament (línia de punts).

Per a **processar** les dades, es van assajar diferents tractaments espectrals amb la intenció d'eliminar els petits desplaçaments de la línia de base. Així, se'n van seleccionar tres: la primera i segona derivades i SNV (Standard Normal Variate)^{11,12}. Les equacions de calibració es van calcular fent servir la regressió parcial per mínims quadrats (PLSR)^{13,14} partint de dades autoescalades. Per a calcular aquestes equacions, així com els tractaments de primera i segona derivades, es va fer servir el paquet de software per a anàlisi multivariant i disseny experimental *The Unscrambler* v. 6.1., produït per Camo ASA.

Per a la determinació de les condicions de calibració que proporcionaven una millor capacitat predictiva es va fer servir el mètode '*cross validation*' de validació seguint el

procediment *'leave one out'* com a mètode de selecció de les mostres de validació¹⁵. La selecció del número de components PLS adequats es va fer tenint en compte l'estadístic que reflecteix valor de l'error mitjà de predicció (MSEP)^B. En el cas particular en el que es calcula per a reflectir l'error de validació en un procés de validació creuat rep el nom de *Mean Square Error of Cross Validation*, MSECv.

La representació de MSECv envers el número de components PLS va ser asimptòtica en la majoria dels casos comprovats. Amb la intenció d'evitar possibles sobreajustos de la calibració, es va seleccionar el número de components PLS aplicant una variant del criteri de selecció de variables 'cap endarrera' als valors de MSECv per a cada component^C. Com a interval de confiança per a acceptar o rebutjar el valor de l'estadístic calculat es va seleccionar $\alpha=0.25$. Per a determinar quin era el mode espectral més adequat, així com l'interval de números d'ona, les diferents equacions de calibració es van aplicar a un conjunt de dades de predicció, no emprades en el procés de calibració. Per a comparar els resultats obtinguts, es va fer servir el valor de RMSE, calculat com

$$\text{RMSE} = \sqrt{\frac{1}{I} \sum_{i=1}^I (y_{i\text{calc}} - y_{i\text{ref}})^2}$$

On $y_{i\text{calc}}$ equival a l'índex d'acidesa lliure calculat mitjançant PLSR, $y_{i\text{ref}}$ és el valor de referència i I equival al total de mostres del conjunt de predicció.

La selecció de l'interval de números d'ona es va fer en base al coneixement químic previ que es tenia del sistema. Es varen estudiar dos intervals espectrals:

- El primer (1775-1689 cm^{-1}) presenta una banda d'absorció intensa a 1748 cm^{-1} , corresponent a l'absorció de l'enllaç carbonílic del grup éster (COOR'); i una altra a 1710 cm^{-1} , deguda a l'absorció dels grups carbonílics dels àcids grassos lliures (en forma COOH), que es troba fortament solapada amb la primera.
- Amb la intenció de complementar la informació espectral i comprovar quin podia ser el seu efecte en la capacitat predictiva dels models de calibració preparats es va afegir un interval espectral al primer, i que comprenia el senyal

^B Veure l'apartat 1.3.1.5. del Rerafons Teòric.

^C En aquesta variant s'ha calculat l'estadístic F de dos components PLS consecutius, fent servir el valor de MSECv, començant des del que presenta el valor mínim. Successivament, es calcula el valor de F per a parells de components PLS consecutius fins que deixa de complir-se la hipòtesi nul·la (H_0).

d'absorció IR entre 1480 i 1050 cm^{-1} . Aquesta zona de l'espectre proporciona informació sobre el moviment d'estirament asimètric en grups metil i metilènic (1465 cm^{-1}), sobre l'estirament de l'enllaç C-O dels ésters alifàtics (1160 cm^{-1}) i l'absorció deguda als grups metilènic alifàtics [C-(CH₂)_n-C, 721 cm^{-1}].

La confirmació en la selecció d'ambdós intervals es va aconseguir tenint en compte el valor dels vectors de *loading* PLS per a cadascuna de les variables espectrals.

Com s'ha dit al començament, per a que una equació de calibració presenti una capacitat predictiva apropiada, aquesta ha de ser calculada tenint en compte una sèrie de dades que siguin representatives de la població total. De fet, tot es redueix a una adequada **selecció de mostres**.

Per a fer-ho, en el nostre cas, es va decidir agrupar-les primer per subconjunts que tinguessin en compte la seva similitud. Així, es va realitzar una anàlisi d'aglomeracions (*clusters*) i es van seleccionar mostres de cadascun d'ells per a ser introduïdes al conjunt de mostres de calibració¹⁶. Per a calcular aquesta anàlisi es va recórrer a un paquet estadístic estàndard com SPSS per a Windows v.7.5.2. i fent servir com a dades d'entrada els valors de *Score* obtinguts al fer un PCA de totes les dades.

Per això, es va calcular PCA dels espectres tractats mitjançant SNV de 65 mostres al llarg de l'interval 1400-700 cm^{-1} , que conté la zona anomenada 'd'empremta digital', i que és la que mostra les diferències més marcades entre les diferents mostres, tal i com es pot veure a la figura AIII.1.

Presentació i discussió dels resultats

De les dades SNV tractades mitjançant PCA varen ser seleccionats els primers 8 components principals obtinguts, que en conjunt descrivien el 85.9% de la variança total continguda en el conjunt d'espectres. Es va fixar aquesta frontera amb la intenció de no incorporar un número de components superior amb poca variança explicada per component. Aquestes dades es van sotmetre posteriorment a una anàlisi de *clusters* fent servir *complete linkage* com a criteri d'aglomeració

Tal i com es pot veure a la figura AIII.2, les mostres formen 7 aglomeracions diferents a una distància d'onze unitats. Aquestes aglomeracions reflecteixen les diferents categories d'oli d'oliva (verge, pur i de sansa) i també les diferències degudes a les pròpies característiques de les mostres (terbolesa no corregida,

procedència geogràfica, etc.). Aquestes diferències entre elles, lluny de ser un inconvenient, presenten l'avantatge d'incorporar el màxim de variabilitat no lligada amb la concentració que és esperable en mostres d'aquestes característiques. Fent servir aquesta informació es va mirar de construir conjunts de calibració i predicció amb fonts de variació el més homogènies possible, de tal forma que incorporessin totes les possibles fonts de variabilitat.

A partir de les aglomeracions obtingudes, es va construir un conjunt de calibració C_1 amb 37 mostres (16 d'oli d'oliva verge, 17 d'oli pur d'oliva i 4 d'oli de sansa d'oliva), que cobrien de forma uniforme l'interval d'acideses lliures entre 0.1 i 2.12%. De la mateixa manera, es va construir el conjunt P_1 de mostres de predicció amb 22 mostres. La finalitat d'aquest últim va ser la d'avaluar la capacitat predictiva de les diferents equacions de calibració obtingudes, al fer servir els diferents modes espectrals i intervals de número d'ona.

Interval	Mode Espectral	PC	RMSEC₁	RMSEP₁
1775-1689 cm ⁻¹	Absorbància	4	0.044	0.072
	Primera Derivada	3	0.052	0.071
	Segona Derivada	3	0.069	0.108
	SNV	3	0.041	0.111
1775-1689 + 1480-1050 cm ⁻¹	Absorbància	5	0.053	0.100
	Primera Derivada	4	0.048	0.121
	Segona Derivada	5	0.044	0.181
	SNV	3	0.067	0.138

Taula AIII.1. Valors de RMSEC₁ i de RMSEP₁ proporcionats per diferents equacions de calibració.

La taula AIII.1 presenta els resultats obtinguts pels conjunts C_1 i P_1 pels intervals de número d'ona estudiats. Tal i com es pot veure a partir dels resultats de RMSE pels dos conjunts preparats, els errors de predicció (RMSEP₁) varen ser, en general, més alts que els errors de calibració (RMSEC₁). Aquest fet pot ser indicatiu d'un possible sobreajust de l'equació de calibració. La inclusió del segon interval de números d'ona (1480-1050 cm⁻¹) va incrementar els errors de predicció. Aquest efecte pot relacionar-se amb la incorporació d'informació espectral corresponent a d'altres espècies presents a la mostra, que presenten una variabilitat espectral similar a la dels àcids grassos lliures. És per això que no va semblar recomanable incorporar aquesta informació en calibracions successives.

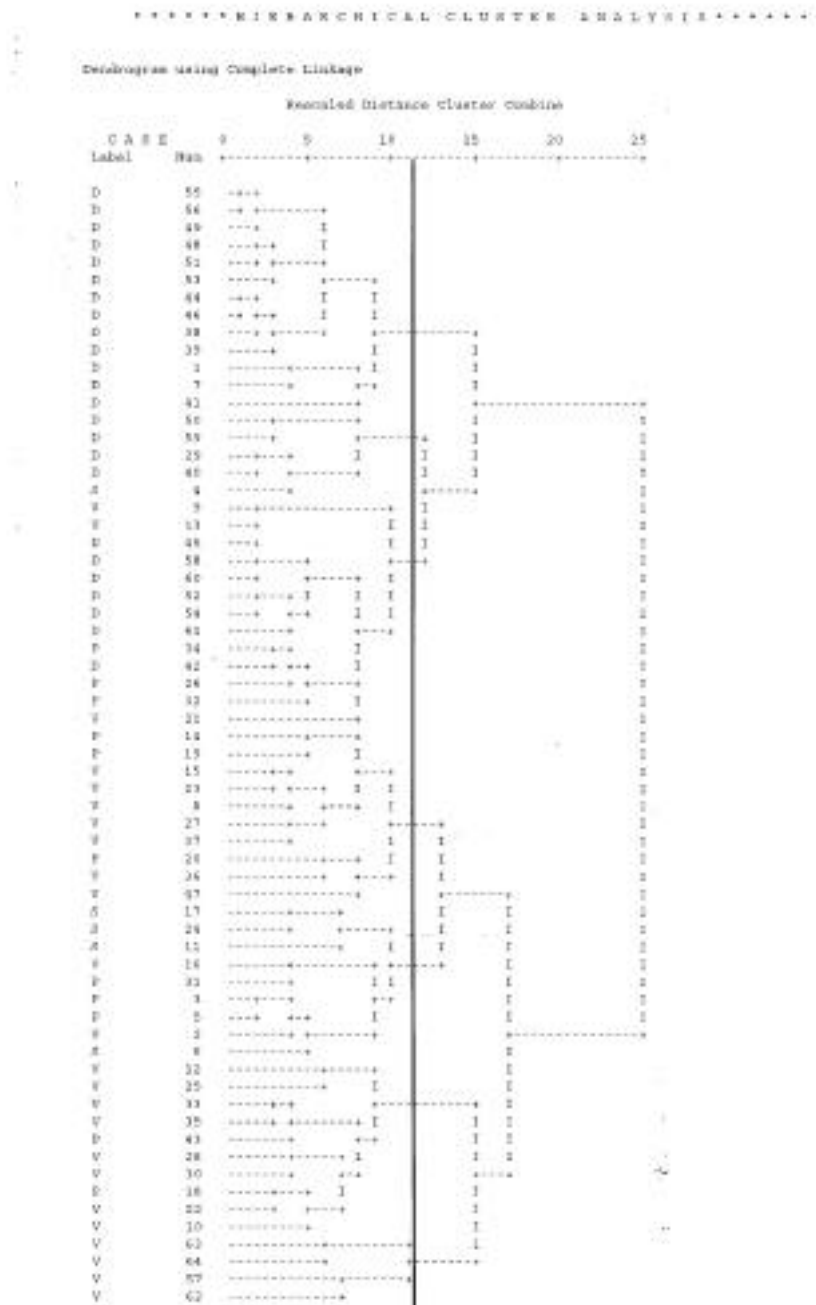


Figura AIII.2. Selecció de mostres. Dendrograma obtingut fent servir *complete linkage* com a tècnica d'aglomeració. Mostres: D (olis d'oliva dopats amb àcid oleic), V (olis d'oliva verge), P (olis purs d'oliva) i S (olis de sansa d'oliva).

Els resultats predictius obtinguts per a la calibració preparada amb dades d'absorbància, que presenten una bona capacitat predictiva i una baixa tendència al sobreajust, van fer que es seleccionés aquest mode espectral com a òptim. La

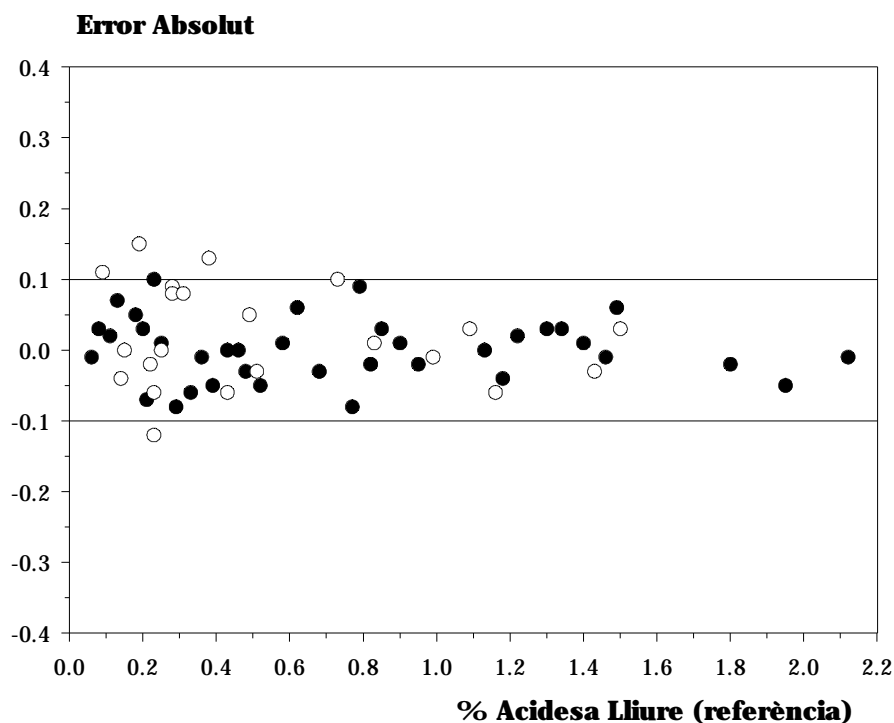


Figura AIII.3. Errors absoluts de les mostres de calibració (○) i de predicció (●) al fer servir dades d'absorbància en l'interval 1775-1689 cm^{-1} .

figura AIII.3 en presenta la variació en l'error absolut en funció de la concentració pels conjunts de calibració (C_1) i de predicció (P_1), a l'interval de números d'ona 1775-1689 cm^{-1} . Les dues fronteres d'error dibuixades a ± 0.1 representen l'error després d'aplicar el mètode de referència. Per a calcular l'error del mètode, es van fer servir els resultats de 57 volumetries en 19 mostres diferents i es van expressar com a unitats de desviació estàndard tenint en compte la norma ISO¹⁷. El resultat obtingut va ser del 0.09%. Tal i com es pot veure a la figura, la dispersió dels resultats obtinguts fent PLSR no va ser homogènea. De fet, va ser lleugerament superior per a valors d'índex d'acidesa inferiors al 0.5%. Es va comprovar que aquesta dispersió podia venir donada per la influència que exercien les mostres amb índexs d'acidesa alts. El seu efecte palanca en l'equació de calibració resultava suficientment important com per a que es produís un desajust en les mostres de baixes concentracions.

Amb la intenció de millorar la capacitat predictiva, i tenint en compte la disposició dels residuals, es va decidir fraccionar la calibració en dos intervals de concentració (per sobre i per sota de 0.5%). La intenció va ser la d'obtenir dues equacions de calibració, una per a cada interval de concentracions, amb una millora en la capacitat predictiva al tram baix d'índexs d'acidesa¹⁸.

Contingut inferior al 0.5%

Es van preparar novament dos conjunts de calibració i predicció amb continguts en àcids grassos lliures inferiors a 0.5%. La selecció de mostres per a cadascun dels dos conjunts es va fer seguint el mateix criteri que a l'apartat anterior, de tal manera que a cadascun dels dos conjunts apareguessin mostres de les diferents aglomeracions. D'aquesta manera es va disposar de dos conjunts, C₂ i P₂, amb 36 i 14 mostres respectivament.

La taula AIII.2 presenta els valors de RMSE obtinguts per a cadascun dels dos conjunts fent servir els diferents modes espectrals assajats, en l'interval 1775-1689 cm⁻¹. Tal i com es pot veure, l'ús de dades de primera i segona derivada varen proporcionar prediccions més dolentes que en els altres dos casos. L'aplicació de la derivada segona als espectres fa augmentar la rellevància del soroll, generant un increment del número de components PLS necessaris a l'equació de calibració i disminuint la capacitat predictiva.

Interval	Mode Espectral	PC	RMSEC₂	RMSEP₂
1775-1689 cm ⁻¹	Absorbància	4	0.042	0.050
	Primera Derivada	3	0.054	0.068
	Segona Derivada	4	0.053	0.115
	SNV	3	0.044	0.057

Taula AIII.2. Valors de RMSEC₂ i de RMSEP₂ proporcionats per diferents equacions de calibració en l'interval d'acideses lliures 0.1 – 0.5 %.

Els errors de quantificació obtinguts varen ser relativament similars en la resta de casos (Absorbància i SNV), però les equacions calculades al fer servir dades SNV eren més simples, ja que únicament feien servir 3 components PLS. A l'hora de comparar l'exactitud del mètode espectrofotomètric, es va fer una representació dels valors d'acidesa lliure calculats a partir de dades SNV envers els valors de referència. Els intervals de confiança en els paràmetres es van calcular amb un nivell de significació P=0.05. El resultat va ser una recta de pendent 0.89 ± 0.11 , amb una ordenada a l'origen 0.03 ± 0.03 i un coeficient de regressió $r=0.943$. Al fer un test t de diferències entre els valors de referència i els calculats pel conjunt

de predicció (P_2), es va confirmar l'absència de diferències significatives entre ambdós mètodes ($t_{exp} = 0.39 < t_{critica}$ per a un nivell de significació $P=0.05$). La disposició dels residuals calculats per a ambdós conjunts (C_2 i P_2) es mostra a la figura AIII.4. Tal i com es pot veure, el valor dels residuals és inferior a ± 0.1 per a totes les mostres.

Contingut superior al 0.5%

Fent servir els mateixos criteris que en el cas anterior, en el que fa referència a la representativitat, es van preparar dos conjunts més de calibració i predicció (C_3 i P_3) respectivament. Es va cobrir en aquest cas el tram alt de concentracions, compès entre 0.5 i 2.1 %, amb 25 mostres de calibració i 8 de predicció.

A la taula AIII.3 apareixen els valors de RMSE per a ambdós conjunts ($RMSEC_3$ i $RMSEP_3$) pels diferents modes espectrals comprovats a l'interval 1775-1689 cm^{-1} . Paral·lelament al cas anterior, les dades espectrals tractades amb SNV proporcionen resultats predictius comparables als obtinguts per absorbància, però fent servir

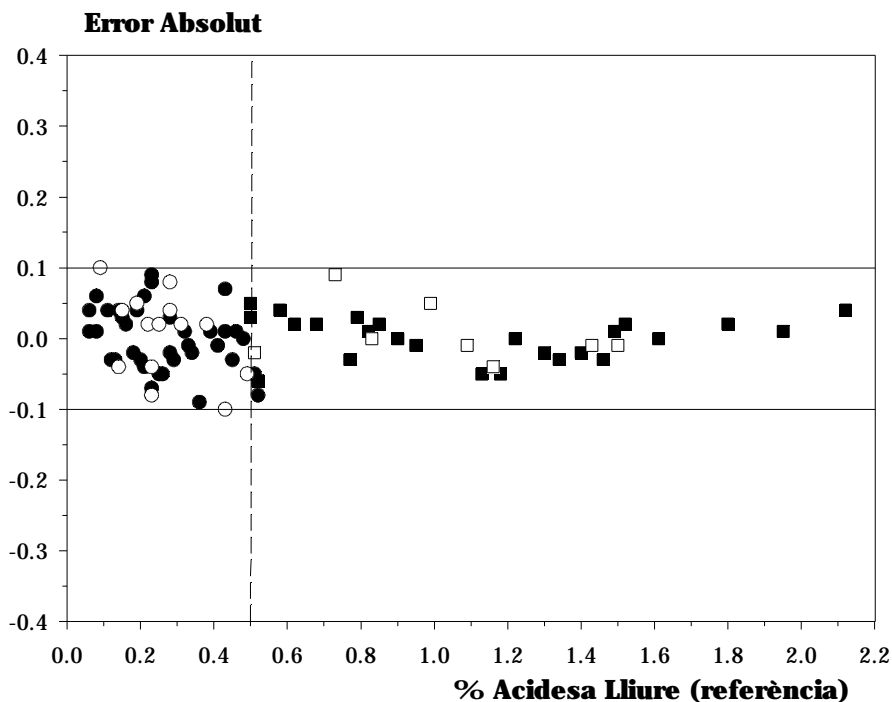


Figura AIII.4. Errors absoluts que s'obtenen al fer servir dades tractades amb SNV en l'interval 1775-1689 cm^{-1} . Tram baix de concentracions: calibració, C_2 (○) i predicció, P_2 (●). Tram alt de concentracions: calibració, C_3 (□) i predicció, P_3 (■).

equacions més simples. La representació dels valors d'índex d'acidesa calculats envers els valors de referència proporciona una recta amb pendent 1.0 ± 0.03 , una ordenada a l'origen de 0.0 ± 0.03 i un coeficient de regressió $r=0.998$ per a un nivell de confiança $P=0.05$. Com en el cas anterior, es va aplicar el test t als resultats de predicció. El valor de t_{exp} calculat va ser menor que el valor de t_{critic} pel mateix nivell de confiança. A la mateixa figura AIII.4 es mostren els residuals per a ambdós conjunts (C_3 , calibració i P_3 , predicció).

Predicció externa

Un cop es van seleccionar les condicions idònies per a la determinació de l'índex d'acidesa a cadascun dels dos trams de concentració, es va comprovar la capacitat predictiva de les equacions de calibració obtingudes sobre 22 mostres que no havien estat emprades (conjunt P). Cadascuna de les mostres va ser primerament quantificada fent servir l'equació de calibració per al tram complet de concentracions (0.1-2.1%). El resultat obtingut va constituir una estimació que va

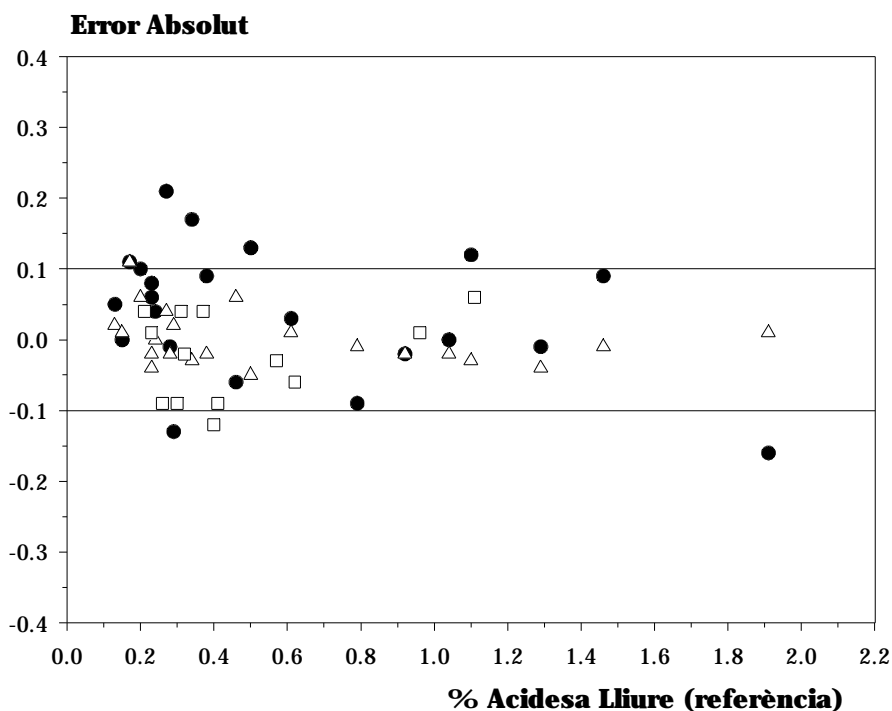


Figura AIII.5. Errors absoluts a la quantificació de les mostres de predicció (P), obtinguts al fer servir l'equació de calibració d'interval complet (□). Errors absoluts per a les mostres de predicció, (P) (△), i per a les mostres estressades tèrmicament (●) fent servir intervals fraccionats.

permetre seleccionar l'equació de calibració més adequada per a cadascuna d'elles, amb la intenció d'obtenir prediccions més acurades.

El valor obtingut de RMSE de les 22 mostres, a l'aplicar l'equació que contemplava el tram complet d'índexs d'acidesa va ser de 0.100. Aquest valor era molt més gran que l'obtingut pel mateix model quan es feien servir els conjunts C_1 i P_1 . Quan es va fer servir l'equació de calibració per al tram 0.1-0.5% per a quantificar les mostres que havien estat identificades com a pertanyents a aquest interval, el valor de RMSE va ser de 0.046. Aquest valor de RMSE sí que va ser comparable a l'obtingut per les mostres de calibració-predicció pertanyents a aquest tram de concentracions i emprades en el procés de calibració ($RMSEC_2$ i $RMSEP_2$). De la mateixa manera, les mostres que varen resultar classificades en el segon interval de concentracions (0.5-2.1%) van proporcionar un valor de RMSE de 0.019, similar a l'obtingut per a les mostres dels conjunts C_3 i P_3 . La figura AIII.5 mostra els residuals d'aquest conjunt de mostres P, quan van ser quantificades per l'equació que contemplava l'interval complet de concentracions així com pels dos subinterval. Els darrers van proporcionar resultats satisfactoris, amb residuals inferiors a $\pm 0.1\%$.

La capacitat predictiva de les equacions obtingudes a partir d'espectres corregits per SNV d'olis d'oliva de diferents categories, es va comprovar també per a un conjunt de 13 mostres d'oli d'oliva estressat tèrmicament. Aquest pretractament va ser aplicat al senyal espectral de mostres d'aquest tipus per a corregir els efectes deguts a la terbolesa observada. A la figura AIII.5 es mostren els residuals de predicció del seu índex d'acidesa. Es va poder observar com les mateixes equacions de calibració podien ser aplicades per igual a mostres fresques i estressades tèrmicament, cosa que no havia estat possible fent servir espectres d'absorbància.

Conclusions

En aquest treball es va mostrar com la conjunció d'una tècnica espectrofotomètrica relativament estàndard, com ATR-FTIR, amb un mètode quimiomètric de regressió permetia obtenir un mètode alternatiu a l'oficial en la determinació de l'acidesa lliure en olis d'oliva. A l'hora d'organitzar la calibració, el fraccionament de l'interval de concentracions en dos trams diferents va permetre l'obtenció d'equacions amb una major capacitat predictiva envers mostres externes. L'ús d'alguns pretractaments com SNV va proporcionar resultats predictius molt similars als obtinguts a partir d'espectres d'absorbància. La principal diferència va provenir de la major simplicitat de les equacions de calibració calculades a partir de SNV així com la capacitat d'aquestes darreres equacions de predir el valor de l'acidesa lliure

de mostres estressades tèrmicament. La determinació de l'índex d'acidesa en aquestes condicions obre les portes a determinacions simultànies de diferents paràmetres d'interès sobre espectres obtinguts en un mateix enregistrament.

- ¹ Directiva UE, **2568/91**, No L 248 (1991).
- ² Y. W. Lai, E.K. Kemsley i R.H. Wilson, *Food Chem.*, **53**, 95, (1995).
- ³ K.M. Bewig, A.D. Clarke, C. Roberts i N. Unklesbay, *J. Am. Oil Chem. Soc.*, **71**, 195, (1994).
- ⁴ M. Tsimidou, K.X. Karakostas, *J. Sci. Food Agric.*, **62**, 253, (1993).
- ⁵ F.R. van de Voort, A.A. Ismail, J. Sedman i G. Emo, *J. Am. Oil Chem. Soc.*, **71**, 243, (1994).
- ⁶ F.R. van de Voort, A.A. Ismail i J. Sedman, *J. Am. Oil Chem. Soc.*, **72**, 873, (1995).
- ⁷ F.R. van de Voort, A.A. Ismail, J. Sedman, J. Dubois i T. Nicodemo, *J. Am. Oil Chem. Soc.*, **71**, 921, (1994).
- ⁸ A.A. Ismail, F.R. van de Voort, G. Emo i J. Sedman, *J. Am. Oil Chem. Soc.*, **70**, 335, (1993).
- ⁹ M. Blanco, J. Coello, H. Iturriaga, S. Maspocho, C. de la Pezuela, *Anal. Chim. Acta.*, **333**, 147, (1996).
- ¹⁰ M. Blanco, J. Coello, H. Iturriaga, S. Maspocho i C. de la Pezuela, *Analyst*, **122**, 761, (1997).
- ¹¹ R.J. Barnes, M.S. Dhanoa i S.J. Lister, *Appl. Spectrosc.*, **43**, 772, (1989).
- ¹² M.S. Dhanoa, S.J. Lister, R. Sanderson i R.J. Barnes, *J. Near Infrared Spectrosc.*, **2**, 43, (1994).
- ¹³ H. Martens i T. Naes, *Multivariate Calibration*, Wiley, New York, (1989).
- ¹⁴ P. Geladi i B.R. Kowalski, *Anal. Chim. Acta*, **185**, 1 (1986).
- ¹⁵ S. Wold, *Technometrics*, **20**, 397 (1978).
- ¹⁶ T. Naes, *J. of Chemometrics*, **1**, 121 (1987).
- ¹⁷ International Standards Organisation, Statistical Interpretation of Test Results - Estimation of the Mean - Confidence Interval, **ISO 2602-1980** (E).
- ¹⁸ T. Naes i T. Isaksson, *J. of Chemometrics*, **5**, 49 (1991).

APARTAT IV

ANÀLISI QUALITATIVA

RESOLUCIÓ DE PROBLEMES QUALITATIUS. ESTUDI DE LES DENOMINACIONS D'ORIGEN D'OLIS D'OLIVA CATALANS^A

L'autenticació de productes agraris és un procés que acostuma a ser llarg i costós. Per regla general, i tal i com s'ha vist en l'apartat del fonament teòric, es tracta d'un procés administratiu, en el que s'acostuma a exigir el compliment de certes normatives europees referents a la zona de conreu i processat.

Tot i ser un procés legislat, des del punt de vista analític acostuma a ser de difícil control. No hi ha paràmetres analítics que permetin fer una definició precisa de la categoria del producte i els pocs que s'arriben a tenir en compte no poden ser considerats com a definitius per la facilitat amb la que poden ser alterats. Es per això que es fa necessària una tècnica de 'screening' que permeti de forma ràpida discriminar mostres de procedència desconeguda o amb característiques diferents de les que s'esperen per a la categoria.

En aquest treball es proposa la viabilitat d'un mètode d'autenticació basat en mesures espectrofotomètriques en l'Infrarroig Proper (NIR), per a classificar olis d'oliva de dues denominacions d'origen geogràficament molt properes i amb característiques molt simil-lars. En el cas del treball que es presenta, les denominacions d'origen catalanes seleccionades són les de Siurana i Les Garrigues. A més de les circumstàncies geogràfiques, ambdues denominacions d'origen fan servir

^A Treball publicat sota el títol: *Near Infrared Spectrometry and pattern recognition as screening methods for the authentication of virgin olive oils of very close geographical origins*, a *Journal of Near Infrared Spectroscopy*, **8**, 45-52, (2000)

majoritàriament olives de la varietat Arbequina per a l'elaboració de l'oli, de manera que la diferenciació és complexa.

L'interès en la qualitat dels productes alimentaris per part dels consumidors ha anat creixent de forma important en els darrers anys. Cada cop estan més informats de les seves propietats nutricionals i progressivament demanen més informació dels productes que fan servir. Amb la intenció de complir aquestes demandes, els controls de qualitat alimentaris han esdevingut cada vegada més estrictes.

Com a resposta a aquesta demanda, els organismes internacionals han publicat directives que fixen els paràmetres de qualitat dels productes agroalimentaris. L'objectiu d'aquestes normes és doble: per una banda es busca poder garantir l'absència d'agents adulterants i per l'altra impedir l'accés al mercat de productes que no compleixin certes especificacions. En aquesta direcció, l'Unió Europea ha legislat en la qualitat de determinats productes agrícoles com l'oli d'oliva¹, en funció de diferents factors, incloent el procés de producció.

Basant-se en la normativa general, les autoritats agràries locals poden aplicar legislacions pròpies que, afegides a les europees, permetin afegir especificacions addicionals als productes. En aquest àmbit apareixen les anomenades **Denominacions d'Origen (DO)**², que definen criteris de similitud per a productes de la mateixa categoria fent referència bàsicament a l'àrea de producció, el tipus de producte de partida utilitzat (per exemple, la varietat o l'origen dels fruits utilitzats), les proporcions màximes dels diferents productes de partida i el tipus de processat, entre d'altres. A més a més, és un requisit indispensable que el conreu, el tractament i/o el processat del producte hagi estat fet a l'àrea geogràfica especificada a la DO. Com a conseqüència del fet de complir tots aquests estàndards de qualitat, els productes que responen a aquestes especificacions agafen un valor afegit superior a ulls del consumidor.

Les principals característiques de l'oli d'oliva han estat exposades a l'apartat corresponent del rerafons teòric, però a títol de recordatori es tracta d'un producte natural de qualitat elevada, típic de la zona Mediterrània. Els estudis nutricionals recents han expandit de forma important el seu ús a d'altres països externs a aquesta zona. Com a conseqüència de la pressió social i econòmica ha estat necessari incrementar els esforços de forma considerable, posant especial èmfasi en l'autenticació del producte i en la detecció d'adulteracions^{3,4}.

Donat que l'origen d'un oli d'oliva li dona les característiques organolèptiques i dietètiques, és important que aquest pugui ser identificat d'una manera ràpida i precisa. Amb aquesta intenció alguns autors han estudiat mostres d'olis vegetals de diferent origen i naturalesa, amb l'objectiu d'identificar pautes entre ells que permetin l'aglomeració dels que són simil·lars i la discriminació de la resta. Aquesta aproximació ha estat emprada en treballs que fan servir diferents tècniques analítiques com GC^{5,6}, ¹³C NMR^{7,8} i anàlisi sensorial⁹. La informació proporcionada per aquestes tècniques és complexa. Sovint és difícil d'interpretar i presenta l'inconvenient del temps necessari per a la preparació de les mostres i el consum de reactius que, en major o menor mesura, generen residus difícils de tractar. És per això que tècniques com l'espectrofotometria agafen rellevància en els casos en els que es busquen tècniques de 'screening' al ser més simples i econòmiques.

En aquesta direcció, la conjunció de l'espectrofotometria com a tècnica instrumental i de la quimiometria com a eina de tractament de la informació generada, sembla ser una via apropiada per a la resolució de tots aquells problemes que involucrin processos de classificació i de detecció d'adulteracions. Per a dir-ho d'una altra manera: pot permetre la **qualificació** de les mostres. En aquest sentit s'han desenvolupat tota una sèrie de treballs basats en espectrofotometria FTIR^{10,11}, NIR^{12,13,14}, i FT-Raman¹⁵.

La Quimiometria, com a eina de tractament de la informació, ha estat utilitzada en una gran varietat de camps com la identificació i autenticació de productes agraris, sent l'oli d'oliva un més. Hi han estat aplicades tècniques com l'anàlisi en components principals, PCA¹⁶; l'anàlisi discriminant, DA^{17,18}; el '*Soft Independent Modelling of Class Analogy*' (SIMCA)^{19,20} i els mètodes basats en distàncies²⁰ i les tècniques d'aglomeració.

Una de les tècniques quimiomètriques que ha experimentat una expansió més ràpida en els darrers anys han estat les xarxes neuronals artificials (ANNs)²¹. El seu ús en l'autenticació d'olis d'oliva s'ha comprovat en diferents treballs²², fent servir diferents algorismes d'aprenentatge.

Malgrat estar perfectament descrit el seu àmbit legal, els mètodes analítics que permeten el control de les denominacions d'origen es troba a la seva fase inicial. Es tracta d'un problema d'elevada complexitat², que augmenta quan les mostres procedeixen de zones geogràficament molt properes. Aquests condicionants han estat uns dels responsables de la relativa lentitud de l'evolució dels mètodes analítics en aquest camp.

En aquest apartat es fa un estudi de dues tècniques quimiomètriques com les ANN i la regressió logística (LR) aplicades a dades d'espectrofotometria NIR, com a tècniques de 'screening' per a discriminar entre mostres d'oli d'oliva verge que pertanyen a dues DOs geogràficament molt properes. D'aquesta manera es comparen els avantatges i els desavantatges d'ambdúes tècniques quimiomètriques per a resoldre aquest problema i d'altres de similars dins l'àmbit de les DOs.

Les dues tècniques quimiomètriques proposades es troben relacionades. La regressió logística pot ser considerada un cas particular de xarxa neuronal, d'una única neurona amb x_i variables d'entrada. Els coeficients Beta (β_i) són els pesos que afecten a cadascuna de les variables i β_0 és el biaix de la xarxa. La funció de transferència aplicada en aquest cas és la funció logística. Envers d'altres tècniques de classificació, la regressió logística presenta una major robustesa, especialment quan el número de mostres disponibles és limitat. A més, està preparada per a l'ús de variables dicotòmiques, que poden ser assignades com a variables resposta associades a la categoria.

Metodologia Experimental

Es van obtenir 43 **mostres** d'oli d'oliva verge de la DO 'Siurana' (DOS), provinents de diferents cooperatives de la demarcació, produïdes a les temporades 1996/97 i 1997/98; i 42 mostres d'oli d'oliva verge de la DO 'Les Garrigues' (DOLG), geogràficament veïna de l'anterior, i produïdes al llarg de les mateixes temporades en cooperatives oliveres de la zona. De forma prèvia a l'enregistrament espectral, les mostres van ser conservades en envasos tancats i a les fosques, i van ser filtrades quan la seva torbidesa va ser apreciable.

Les mostres es van distribuir en un conjunt de calibració i en un conjunt de predicció. El primer va estar constituït per 56 mostres (25 de la DOS i 31 de la DOLG). El segon conjunt va estar constituït per un total de 29 mostres, distribuïdes per DO en 18 per a la DOS i 11 per a la DOLG.

És conegut que les xarxes neuronals fan servir dos conjunts de dades pel procés d'aprenentatge. Per aquesta raó, el conjunt de calibració va ser desglosat en 2: un conjunt d'aprenentatge, integrat per 40 mostres (21 DOS i 19 DOLG); i un conjunt de validació que comprenia 16 mostres (4 DOS i 12 DOLG).

Les mesures dels **espectres NIR** de les mostres es van portar a terme en un instrument NIRSystems 6500, fent servir un mòdul de cubeta de líquids com a suport.

El detector de transmissió emprat en aquest cas va ser una unitat del model NR6513-A. El comandament de l'instrument i l'enregistrament dels espectres va ser portat a terme pel software de control i tractament NSAS v. 3.52.

En quant als **softwares** emprats per a aplicar els mètodes quimiomètrics de tractament de les dades, l'anàlisi de components principals, (*Principal Component Analysis, PCA*), es va fer mitjançant *The Unscrambler* v. 6.1. Els càlculs necessaris per a la discriminació d'ambdúes DO es van fer mitjançant dues aplicacions diferents. Mentre la regressió logística, (*Logistic Regression, LR*), es va calcular fent servir el *toolbox* (conjunt de rutines obertes de càlcul) *Stibox* v.1.03²³ per a *MATLAB* v.5.2.1.,; els càlculs de xarxes neuronals es van dur a terme fent servir el paquet estadístic *SPSS* v.7.5.2. i l'extensió de *SPSS* pel seu software, *Neural Connection* v.2.0.

L'enregistrament dels espectres es va fer al llarg de dies diferents, sense un ordre especial, en condicions estàndard de temperatura. Cadascun dels espectres es va enregistrar per duplicat en cubetes de quars de 4mm de camí òptic, fent servir una resolució de 2 nm, i acumulant 32 escombrats al llarg de l'interval de 1100 a 2500 nm.

Preprocessat de les dades

A la figura AIV.1 es mostra l'aspecte d'un espectre NIR d'una mostra d'oli d'oliva verge. Es va fer servir l'interval espectral 1100-2260 perquè, tal i com s'hi pot veure, longituds d'ona més altes presentaven saturació en el detector.

Per a compensar els desplaçaments de la línia de base deguts a la terbolesa residual que presentaven algunes mostres, es va calcular la primera derivada dels espectres fent servir el mètode de Savitzky-Golay amb una finestra de 5 punts i un polinomi de tercer ordre.

Les xarxes neuronals són freqüentment dependents del número de variables inicials. Aquest número, també dicta el número de mostres recomanable per a realitzar l'entrenament de la xarxa. És per això que, freqüentment, es recomana la utilització d'una tècnica de reducció de variables com PCA per a poder fer servir la matriu de *scores* com a conjunt de dades d'entrada²⁴. Al fer servir PCA, la dimensionalitat de les dades es redueix, preservant-ne la informació més significativa. Paral·lelament, aquest pretractament de les dades també pot ser aplicat a la regressió logística, ja que com s'ha esmentat anteriorment, es pot considerar aquesta com un cas particular de les ANN.

En el nostre cas, la dimensionalitat de les dades va ser reduïda, de forma arbitrària, de tal manera que el 95% de la variància total continguda a la matriu de variables predictores del conjunt de calibració restés explicada. La matriu de *scores* resultant va ser considerada a partir d'aleshores el conjunt de dades predictores de partida. Però,

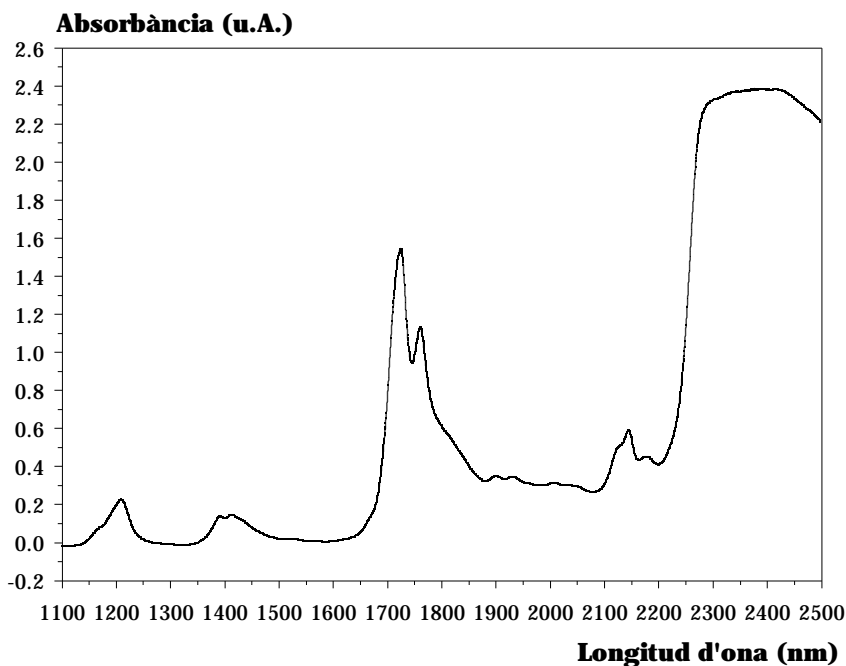


Figura AIV.1. Espectre NIR d'una mostra d'oli d'oliva verge

al fer servir la regressió logística, la reducció de variables mitjançant PCA es va revelar inadequada, ja que el número de mostres disponibles era bastant reduït. Per això va ser necessària una selecció de variables posterior al procés de reducció.

D'aquesta manera, als *scores* corresponents als primers 14 components principals (PC), que explicaven un 95.3% de la variància total, se'ls va aplicar una variant del criteri F. En aquesta variant es comparen les variàncies (inter-classe i intra-classe) de les diferents classes en estudi per a cada variable, amb l'objectiu de conèixer quines d'elles són les que presenten un major poder discriminant. Matemàticament aquest estadístic s'expressa com

$$F = \frac{(1/(g-1)) \sum_{i=1}^g n_i (y_i - \bar{y})^2}{(1/(n-g)) \sum_{i=1}^g \sum_{j=1}^{n_i} (y_{ij} - y_i)^2}$$

On g és el número de classes, n_i el número d'objectes a cada classe i , $n = \sum n_i$, \bar{y}_i és la mitjana del grup, \bar{y} la mitjana total i y_{ij} és el valor de la variable per a l'objecte j de la classe i . Fent servir un nivell de confiança del 95%, només els *scores* dels components principals 1, 3, 6, 8, 9, 11 i 12 varen ser significatius i, com a tals, es van fer servir per a construir el model basat en regressió logística.

Presentació i discussió dels resultats

Xarxes Neuronals Artificials

La matriu de *scores* resultant de l'aplicació de PCA a la matriu de dades d'entrenament original va quedar constituïda per 14 components principals, que explicaven un total del 95.92% de la variància total inicial. Aquesta matriu de *scores* d'entrenament es va fer servir per a construir diferents xarxes MLP (*Multi-Layer Perceptron*) fent servir una única capa oculta. Com a variable de sortida se'n va crear una de dicotòmica, **do**, que agafava el valor de 1 per a les mostres DOS i 0 per a les mostres DOLG. Com a valor frontera de discriminació entre ambdues classes es va agafar el valor central d'aquesta variable dicotòmica (0.5). D'aquesta manera, les mostres van ser assignades a la DOS o a la DOLG, depenent de si el valor calculat per la variable **do** era més gran o més petit que 0.5, respectivament.

Les xarxes assajades es van construir fent servir com a funcions de transferència $\tanh(x)$ i $\text{logsigmoid}(x)$, amb dues distribucions inicials dels pesos diferents (uniforme i gaussiana). Per a fer evolucionar el sistema es va seguir la tècnica del gradient descendent de l'error. Com a criteri de parada de la xarxa es va monitoritzar l'error del conjunt de validació. La xarxa es va aturar en el punt en el que el valor de RMSE per a aquest conjunt es va fer mínim.

A partir dels resultats obtinguts es va poder veure com la funció de transferència que presentava una millor capacitat discriminant era la $\tanh(x)$ fent servir una distribució gaussiana inicial dels pesos. Fixades aquestes condicions, es van construir xarxes de diferents topologies, fent variar el nombre de nodes de la capa oculta. En particular, es van explorar topologies des de la 14,1,1 fins la 14,8,1 (on x_1, x_2, x_3 són el número de variables d'entrada, el nombre de nodes de la capa oculta i el nombre de variables de sortida, respectivament). El valor dels pesos va ser reinicialitzat diferents cops amb l'intenció de comprovar que la xarxa era estable i evitar la convergència cap a un mínim local d'error. De les topologies assajades, es va poder veure com les que eren

construïdes amb una topologia 14,4,1 o superior proporcionaven resultats acceptables i reproduïbles. Així, es va escollir aquesta topologia per la seva simplicitat.

Per a assegurar que la resposta obtinguda estava relacionada directament amb la informació espectral i que reflectia de forma apropiada la seva estructura es va randomitzar un 20% de les variables resposta del conjunt d'entrenament. Així, es va canviar el valor de la variable **do** de 8 mostres d'aquest conjunt. Mantenint aquestes condicions, la xarxa va ser sotmesa de nou a un procés d'entrenament i els resultats van mostrar com aquesta fallava a l'hora d'ajustar la resposta de forma adequada. En aquest cas, el percentatge d'errors en el conjunt d'entrenament va ser del 13.7%, el que equival a dir que 11 mostres varen ser classificades incorrectament. Aquest fet va confirmar la presència d'una estructura real a les dades i la capacitat modelant de la xarxa proposada (14,4,1).

A la taula AIV.1 es mostren les matrius de confusió per a la variable contínua **do** a l'aplicar la xarxa seleccionada (14,4,1). En aquestes matrius s'agrupen en un mateix interval les mostres que presenten valors similars per la variable **do**. Així és més senzill veure que la proporció d'encerts va ser del 100% pels conjunts d'entrenament i de validació. No es va donar cap cas de classificació de la DO incorrecta i els resultats obtinguts varen mostrar una baixa dispersió dels valors de la variable **do**, acumulant-se al voltant de 0 i 1.

Un cop finalitzat aquest procés, es va comprovar la capacitat predictiva de la xarxa obtinguda envers mostres externes^B, incloses en un conjunt de predicció. Els resultats que es van obtenir també apareixen a la taula AIV.1. Tal i com es pot veure, totes les mostres van aparèixer correctament classificades, però tres d'elles amb una dispersió lleugerament superior (dues mostres pertanyents a la DOS i una pertanyent a la DOLG). Malgrat això, els resultats obtinguts per a la variable **do** es van mantenir clarament per sobre i per sota del valor frontera, el que va facilitar la seva assignació sense problemes.

^B No utilitzades en cap pas del procés d'entrenament

DENOMINACIÓ D'ORIGEN													
<i>Conjunt d'entrenament</i>													
	[0,0,0,1]	[0,1,0,2]	[0,2,0,3]	[0,3,0,4]	[0,4,0,5]	[0,5,0,6]	[0,6,0,7]	[0,7,0,8]	[0,8,0,9]	[0,9,1,0]			
0(DOLG)	38	0	0	0	0	0	0	0	0	0	0	0	0
1(DOS)	0	0	0	0	0	0	0	0	0	0	0	0	42
<i>Conjunt de validació</i>													
	[0,0,0,1]	[0,1,0,2]	[0,2,0,3]	[0,3,0,4]	[0,4,0,5]	[0,5,0,6]	[0,6,0,7]	[0,7,0,8]	[0,8,0,9]	[0,9,1,0]			
0(DOLG)	24	0	0	0	0	0	0	0	0	0	0	0	0
1(DOS)	0	0	0	0	0	0	0	0	0	0	0	0	8
<i>Conjunt de predicció</i>													
	[0,0,0,1]	[0,1,0,2]	[0,2,0,3]	[0,3,0,4]	[0,4,0,5]	[0,5,0,6]	[0,6,0,7]	[0,7,0,8]	[0,8,0,9]	[0,9,1,0]			
0(DOLG)	20	0	1	0	1	0	0	0	0	0	0	0	0
1(DOS)	0	0	0	0	0	0	0	1	1	0	1	0	34

Taula AIV.1. ANN-MLP. Matriu de confusió dels processos d'entrenament, validació i predicció. Topologia seleccionada (14,4,1).

Regressió Logística

De forma totalment anàloga al cas anterior, les denominacions es van codificar de la mateixa manera, amb el que el valor de la variable resposta va agafar el valor de 1 per a mostres pertanyents a DOS i 0 per a mostres pertanyents a DOLG. La frontera interclasse es va fixar a un valor de probabilitat de 0.5.

Partint de la matriu de variables seleccionades fent servir el criteri F, en el pas següent es van calcular els coeficients β a partir de les dades del conjunt de calibració aplicant la transformació Logit. Els valors calculats per a aquests coeficients varen donar el polinomi $g(x)$

$$g(x) = 7.35 - 1.91pc_1 + 4.36pc_3 + 3.38pc_6 - 2.64pc_8 + 1.71pc_9 + 0.55pc_{11} - 0.82pc_{12}$$

Amb la intenció de veure com queden afectats els espectres a l'aplicar aquesta transformació basada en els coeficients β , es van comparar les medianes dels espectres de primera derivada per a cadascuna de les classes. Tal i com es veu a la figura AIV.2, les diferències entre ambdós espectres medians eren molt petites. Es trobaven superposats al llarg de pràcticament tot l'interval espectral. Només es veien petites variacions en una zona molt estreta, situada entre els 1850 i 1980 nm. La

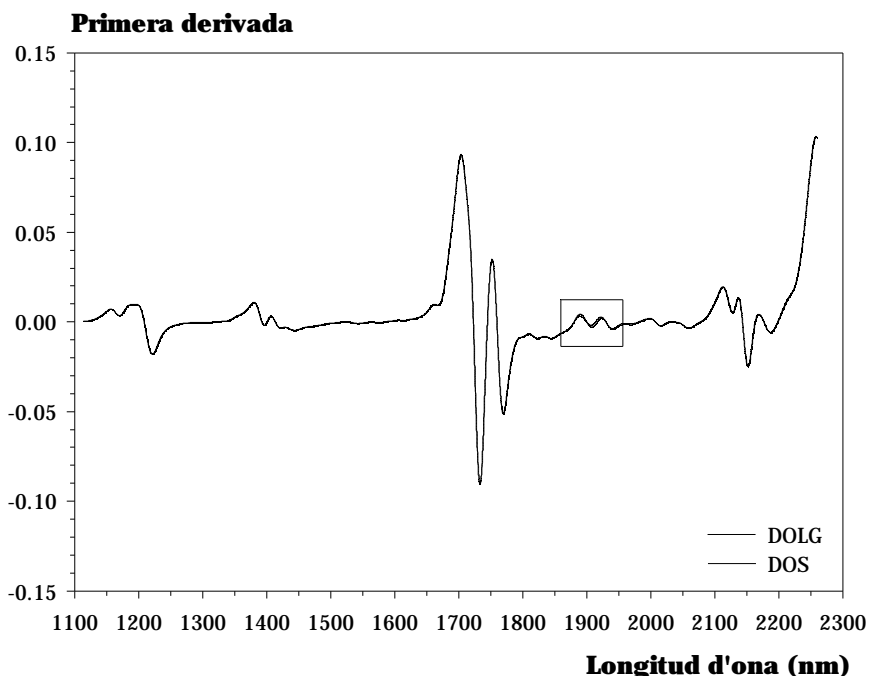
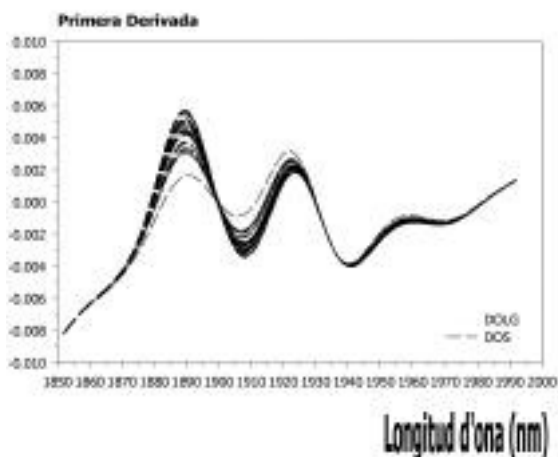
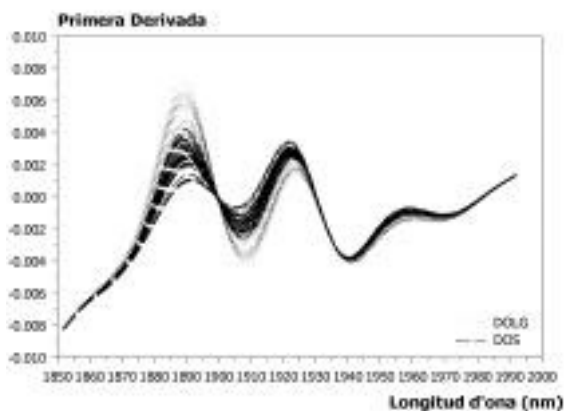


Figura AIV.2. Espectres medians superposats de les dues denominacions d'origen

superposició d'una sèrie d'espectres d'ambdues classes dins d'aquest interval va donar



Figures AIV.3A i AIV.3B. Diferències espectrals a la zona compresa entre 1850 i 1980 nm. Sense efecte de la regressió logística (a dalt) i amb efecte de la regressió logística (a baix)

ser correctament assignades a la seva DO. De la mateixa manera que en el cas anterior, es va randomitzar el valor de la variable resposta amb l'objectiu de verificar si l'equació obtinguda responia a l'estructura present a les dades o bé si els resultats eren fruit de l'atzar. Vuit mostres de calibració van veure alterat el valor de la seva variable resposta i es va recalculer el valor dels coeficients _ fent servir la transformació Logit. Els resultats que es van obtenir mostraven un 11.6% de mostres del conjunt de calibració incorrectament classificades. Un error d'aquest tipus resultava inacceptable, i suggeria la presència d'una certa estructura a les dades i la capacitat de l'equació obtinguda a l'hora de representar-la.

origen a la figura AIV.3(A), en la que es veia com, majoritàriament, els espectres pertanyents les dues DO presentaven diferències en els seus valors d'absorbència. Aquestes diferències es van associar als enllaços carboxílics dels àcids grassos lliures així com a enllaços éster dels triglicèrids.

Per a saber quin era l'efecte dels coeficients de la regressió logística sobre l'espectre, es van aplicar els coeficients _ als scores de cada espectre i es van reconstruir posteriorment a partir dels propis *loadings*, proporcionant els espectres modificats per ambdues denominacions. El resultat d'aquesta transformació es pot veure a la figura AIV.3(B). Després del tractament, les diferències entre ambdós conjunts d'espectres es van reforçar en l'interval 1850-1980 nm de longituds d'ona.

La taula AIV.2 mostra els resultats pel conjunt de calibració. Tal i com es pot veure, totes les mostres van

DENOMINACIÓ D'ORIGEN										
<i>Conjunt de calibració</i>										
	[0,0,0,1]	[0,1,0,2]	[0,2,0,3]	[0,3,0,4]	[0,4,0,5]	[0,5,0,6]	[0,6,0,7]	[0,7,0,8]	[0,8,0,9]	[0,9,1,0]
0(DOLG)	62	0	0	0	0	0	0	0	0	0
1(DOS)	0	0	0	0	0	0	0	0	0	50
<i>Conjunt de predicció</i>										
	[0,0,0,1]	[0,1,0,2]	[0,2,0,3]	[0,3,0,4]	[0,4,0,5]	[0,5,0,6]	[0,6,0,7]	[0,7,0,8]	[0,8,0,9]	[0,9,1,0]
0(DOLG)	21	0	0	1	0	0	0	0	0	0
1(DOS)	3	0	1	0	0	0	0	0	0	32

Taula A1v.2. Regressió logística. Matriu de confusió dels processos de calibració i predicció.

La taula AIV.2 també presenta els resultats de predicció i la probabilitat de cada mostra de pertànyer o no a una DO. Tal i com es pot veure, 4 duplicats de dues mostres de DOS van ser incorrectament classificades com a DOLG. El perquè d'aquest error no pot ser fàcilment explicat, però tenint en compte la simplicitat i rapidesa del mètode de càlcul utilitzat, i la similitud entre ambdues DO, els resultats obtinguts presenten un percentatge d'encert que es pot considerar molt raonable.

Conclusions

A títol de conclusió es pot dir que l'espectrofotometria NIR, conjuntament amb una tècnica adequada de classificació, permet la discriminació d'olis d'oliva verge de dues denominacions d'origen veïnes, sense pretractament de la mostra. La metodologia que es va proposar en aquest treball proporciona una manera ràpida d'obtenir informació útil en el procés d'autenticació d'un producte agroalimentari i té l'avantatge de poder ser utilitzada com a tècnica ràpida de *screening*.

Les xarxes neuronals tenen una major capacitat discriminant si es comparen amb la regressió logística. El seu principal inconvenient prové de la lentitud del procés d'entrenament i de la dependència dels resultats obtinguts de les condicions inicials. Això obliga a una acurada selecció dels paràmetres implicats en el procés d'aprenentatge per a optimitzar el temps de computació i minimitzar el risc que tot el procés convergeixi en un mínim d'error local. Envers la LR, les ANN no necessiten estrictament un pas previ de selecció de variables, perquè els pesos ponderen la informació proporcionada a la xarxa.

La regressió logística presenta l'avantatge de la seva menor complexitat computacional i la seva major rapidesa en la determinació dels paràmetres de l'equació que permet la discriminació. A més, la seva capacitat discriminant és més que acceptable i, per tant, és una bona alternativa a les ANN. Tot i que el càlcul de la LR necessita d'una adequada selecció de les variables, no és un obstacle important perquè, en tot cas, pot fer-se servir un criteri de significació senzill.

-
- ¹ Directiva UE, **2568/91**, No L 248 (1991).
- ² Directiva UE, **2081/92**, No L 208 (1992).
- ³ F. Angerosa, L. Carrera, S. Cumitini, G. Gleixner i F. Renier, *J. Agr. Food Chem.*, **45**, 3044 (1997).
- ⁴ E. Salivaras i A.R. McCurdy, *J. Am. Oil Chem. Soc.*, **69**, 935 (1992).
- ⁵ M. Tsimidou i K. X. Karakostas, *J.Sci. Food Agric.*, **62**, 253 (1993).
- ⁶ G. Alberghina, L. Caruso, S. Fisichella i G. Musumarra, *J.Sci. Food Agric.*, **56**, 445 (1994).
- ⁷ A.D. Shaw, A. di Camillo, G. Vlahov, A. Jones, G. Bianchi, J. Rowland i D.B. Kell, *Analytica Chimica Acta*, **348**, 357 (1997).
- ⁸ R. Zamora, J.L. Navarro i F.J. Hidalgo, *J. Am. Oil Chem. Soc.*, **71**, 361 (1994).
- ⁹ R. Aparicio, M.T. Morales i V. Alonso, *J. Agric. Food Chem.*, **45**, 1076 (1997).
- ¹⁰ D.B. Dahlberg, S.M. Lee, S.J. Wenger i J.A. Vargo, *Appl. Spectrosc.*, **51**, 1118, (1997).
- ¹¹ Y. W. Lai, E. K. Kemsley i R. H. Wilson, *J. Agr. Food Chem.*, **42**, 1154 (1994).
- ¹² I.J. Wesley, R.J. Barnes i A.E.J. Mc Gill, *J. Am. Oil Chem. Soc.*, **72**, 289 (1995).
- ¹³ T.Sato, *J. Am. Oil Chem. Soc.*, **71**, 293 (1994).
- ¹⁴ K. M. Bewig, A. D. Clarke, C. Roberts i N. E. Unklesbay, *J. Am. Oil Chem. Soc.*, **71**, 195 (1994).
- ¹⁵ V. Baeten i M. Meurens, *J. Agr. Food Chem.*, **44**, 2225 (1996).
- ¹⁶ H. Martens i T. Naes, *Multivariate Calibration*. Wiley, New York (1989).
- ¹⁷ L. Kaufman i P.J. Rousseeuw, *Finding groups in data: An introduction to cluster analysis*. John Wiley & Sons, New York (1990).
- ¹⁸ The Unscrambler®. User manual, CAMO ASA (1998).
- ¹⁹ S. Wold, *Pattern Recognition*, **8**, 127 (1976).
- ²⁰ S. Wold i M. Sjöström, *ACS Symposium Series*, **52**, 243 (1977).
- ²¹ J. Zupan i J. Gasteiger, *Neural Networks for Chemists: An Introduction*, VCH Publishers, Berlin, (1993).
- ²² J. Zupan, M. Novic, X. Li i J. Gasteiger, *Anal. Chim. Acta*, **292**, 219 (1994).
- ²³ <http://www.maths.lth.se/matstat/stixbox/> . Darrera consulta: 28.08.01.
- ²⁴ M. Blanco, J. Coello, H. Iturriaga, S. MasPOCH i M. Redón, *Anal. Chem.*, **67**, 4477 (1995).

CONCLUSIONS GENERALS

Establir un recull de conclusions generals d'un treball com aquest és difícil. Les variacions que ha experimentat, des dels objectius inicials fins el resultat final, fa que hagin de ser valorades de forma diferenciada. Així, per una banda es poden enfocar les conclusions des del punt de vista de la viabilitat de l'espectrofotometria com a tècnica instrumental per a l'anàlisi de control de mostres d'oli d'oliva. Segons com, es poden enfocar des del punt de vista de la quimiometria aplicada a dades espectrofotomètriques de mostres d'origen natural. Són plantejaments vàlids i que, en certa manera, es superposen.

L'espectrofotometria com a tècnica d'anàlisi de control d'olis d'oliva

En aquest treball s'ha comprovat com es poden utilitzar les espectroscopies d'infraroig mitjà i proper com a tècniques instrumentals per a realitzar determinacions qualitatives i quantitatives amb mínim tractament de la mostra. Donada la complexitat del senyal que aquestes tècniques proporcionen, originada per la complexitat de la matriu natural, en ambdós casos ha estat necessari l'ús de tècniques quimiomètriques aplicades a les dades instrumentals per a obtenir resultats satisfactoris.

Les determinacions quantitatives s'han basat en PLS com a tècnica de regressió per a relacionar les variables predictores amb la concentració, essent especialment important la selecció adequada dels pretractaments espectrals a aplicar. En la determinació de l'índex d'acidesa per espectrofotometria FTIR-ATR, es va poder comprovar com la utilització de SNV permetia l'obtenció de resultats similars als obtinguts a partir de mesures directes d'absorbància, però fent servir equacions de calibració més senzilles i, alhora, d'àmbit d'aplicació més general. A la determinació dels percentatges de tres àcids grassos insaturats com l'oleic, linoleic i linolènic també es va poder veure la influència dels diferents pretractaments en l'obtenció de resultats. Tots ells van proporcionar resultats força similars. Com en el cas anterior, la diferència més important va estar en la complexitat de les equacions obtingudes.

La classificació de les mostres d'oli d'oliva verge a partir del seu espectre NIR, es va mostrar com un pas inicial per a permetre la qualificació ràpida de mostres en funció del seu origen geogràfic. Va ser possible discriminar entre olis d'oliva de denominacions d'origen molt properes geogràficament, elaborats a partir d'olives de les mateixes varietats i per tant de composició pràcticament igual, en funció de la variabilitat general de l'espectre. Es va poder veure que aquesta proposta, amb un

ventall de mostres adequat, pot permetre obtenir informació ràpida pel procés d'autenticació dels diferents olis d'oliva.

Aspectes Quimiomètrics

En aspectes lligats a l'anàlisi quantitativa, en la determinació de l'índex d'acidesa ha estat necessària una adequada selecció de mostres i el fraccionament de l'interval de concentracions per a obtenir equacions de calibració amb una capacitat predictiva acceptable. Així, la selecció de mostres per a cada conjunt de calibració/predicció ha estat necessària per les diferències existents entre mostres d'origen natural amb un índex d'acidesa molt similar. El fraccionament de l'interval de calibració es mostra com una solució vàlida per a millorar la capacitat predictiva en l'interval de calibració de l'índex d'acidesa entre 0.0° i 2.1°, ja que fa minvar la influència de les mostres d'índex d'acidesa elevat en la determinació dels paràmetres de la regressió.

A la determinació del percentatge d'àcids grassos C18:x (mono-, di- i triinsaturats) s'ha comprovat quin és l'efecte de OSC aplicat a dades de transmitància. Tot i que les equacions de calibració han estat en general més senzilles, no s'ha vist que s'incrementi significativament la capacitat predictiva dels components, tant majoritaris com minoritaris. Els resultats obtinguts semblen assenyalar que OSC elimina els efectes de dispersió amb una influència més suau sobre el senyal de l'analit, però mitjançant un tractament complex que no incrementa la capacitat predictiva. Per les característiques del pretractament, s'ha pogut demostrar que l'efecte de dispersió és ortogonal al senyal.

Quan es va fer l'estudi de l'efecte de les variables latents de la regressió de les variables predictorres (\mathbf{X}) sobre els *scores* ortogonals (\mathbf{t}_c^*) al fer servir OSC en aplicacions quantitatives basades en PLSR, es va poder veure l'existència d'un fenomen de sobreajust lligat a determinades condicions. L'estudi s'ha portat a terme sobre dos conjunts de dades d'origen diferent, amb el que s'ha pogut comprovar que ambdós contenien informació ortogonal que podia ser modelada en el procés de correcció. De l'estudi de les condicions de sobreajust, que afecten tant a les variables latents com als components PLS de l'equació quantitativa final, se'n conclou la necessitat d'una estratègia per a la determinació de les condicions òptimes. Se n'han proposat dues: la validació creuada i l'anomenada de *test set*, proporcionant ambdues uns resultats molt similars. La selecció d'una o una altra ha quedat lligada a la disponibilitat de major o menor número de dades espectrals. De la comprovació de l'efecte dels factors OSC se'n conclou que la utilització d'un únic factor, amb un

número adequat de variables latents, és suficient com per assolir un nivell de correcció sense sobreajustos.

Finalment, també s'ha pogut demostrar que en la discriminació d'olis d'oliva verges de dues denominacions d'origen geogràficament molt properes, de les dues tècniques estudiades, les xarxes neuronals són més complexes a nivell de càlcul, però proporcionen una major capacitat discriminant. En canvi, la regressió logística és un mètode més simple, però que necessita una selecció prèvia de variables per a proporcionar una capacitat predictiva comparable. Tot i així, aquest aspecte no resulta un obstacle especialment important perquè aquesta selecció es pot portar a terme fent servir un criteri de significació senzill.

FE D'ERRADES**APLICACIÓ I DESENVOLUPAMENT DE TÈCNiques QUIMIOMÈTRiques A L'ANÀLISI AGROALIMENTÀRIA**

Sempre són més nombroses del que voldriem, però malauradament n'hi ha. En aquests fulls miro de recollir-ne algunes. Algunes d'elles son petits errors de tipografia, d'altres de compaginació. Totes elles poden haver dificultat el seguiment de text en algun moment, pel que crec que aquest recull us pot donar un cop de mà per aclarir determinats punts. Tot i així, segurament caldria afegir-ne algunes més.

Una salutació ben cordial

Ivan Montoliu i Roura

Pàgina	Hi diu...	Però hi hauria de dir..
42	$(x) = \frac{e^{0 + 1x + 2x_2 + \dots + nx_n}}{1 + e^{0 + 1x + 2x_2 + \dots + nx_n}}$	$(x) = \frac{e^{0 + 1x_1 + 2x_2 + \dots + nx_n}}{1 + e^{0 + 1x_1 + 2x_2 + \dots + nx_n}}$
103	'Albinyana, Albiol, Alcover,...	'Albinyana, L'Albiol , Alcover,...
106	^D Veure l'apartat 1.4.1.5 del rerafons teòric...	^D Veure l'apartat A.4.1.5 del rerafons teòric...
110	$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - x_i)}{n}}$	$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - x_i)^2}{n}}$
112	'Taula AII.1. Variàncies totals explicades...'	'Taula AI.1 . Variàncies totals explicades...'

114	'Taula AI.3. Espectres de correlació de dades d'Absorbància...'	'Figura AI.3. Espectres de correlació de dades d'Absorbància...'
127	'...per a la determinació dels paràmetres <i>i</i> de validació per a determinar...'	'... per a la determinació dels paràmetres i de validació per a determinar...'
141	^B 'Veure l'apartat 1.3.1.5 del rerafons teòric.'	^B 'Veure l'apartat A.4.1.5 del rerafons teòric.'
147	'A la taula AIII.3 apareixen...'	Falta la taula AIII.3. (Veure Annex)
161	'Figures AIV.3a i AIV.3b. Diferències espectrals a la zona compresa entre 1850 i 1980 nm. Sense efecte de la regressió logística (a dalt) i amb efecte de la regressió logística (a baix)'	'Figures AIV.3a i AIV.3b. Diferències espectrals a la zona compresa entre 1850 i 1980 nm. Sense efecte de la regressió logística (a baix) i amb efecte de la regressió logística (a dalt)'
166	'..., s'ha pogut demostrar que l'efecte de dispersió és ortogonal al senyal.'	'..., s'ha pogut demostrar que l'efecte de dispersió és ortogonal a la concentració .

Annex:

Interval	Mode spectral	PC	RMSEC₃	RMSEP₃
1775 -1689 cm ⁻¹	Absorbància	4	0.036	0.049
	Primera derivada	3	0.034	0.058
	Segona derivada	4	0.054	0.054
	SNV	3	0.029	0.040

Taula AIII.3. Valors de RMSEC₃ i RMSEP₃ proporcionats per diferents equacions de calibració en l'interval d'acideses lliures 0.5-2.1%.