



A.D. MDLXII

DOTTORATO DI RICERCA IN SCIENZE CHIMICHE

UNIVERSITA' DEGLI STUDI DI SASSARI

Dipartimento di Chimica e Farmacia

**MESSA A PUNTO E VALIDAZIONE DI METODI
ANALITICI ASPECIFICI ONLINE IMPIEGANTI
STRUMENTAZIONE A TRASFORMATTA DI FOURIER
PER LA DETERMINAZIONE DI MACRO E
MICROANALITI IN LATTE OVINO**

Coordinatore:

Prof. Stefano Enzo

Tesi di Dottorato del

Dr. Marco Caredda

Tutor:

Prof. Gavino Sanna

A mia madre

INDICE

INTRODUZIONE.....	9
1 Il sistema alimentare	13
1.1 La qualità alimentare	13
1.1.1 La sicurezza alimentare	14
1.1.2 Concetti di qualità alimentare.....	15
1.1.3 Proprietà funzionali degli alimenti	16
1.1.4 I prodotti tipici.....	18
1.1.5 I prodotti tipici in Europa ed Italia.....	22
1.2 Il sistema del latte ovino	26
1.2.1 Regolamenti e parametri europei	26
1.2.1.1 Sicurezza alimentare nel latte ovino	27
1.2.2 Qualità chimica e fisica del latte ovino.....	28
1.2.3 La razza ovina Sarda	29
1.2.4 Il sistema di pagamento del latte ovino	32
1.2.5 I controlli di qualità regionali: l'ARA Sardegna.....	36
2 Il latte.....	41
2.1 Introduzione.....	41
2.2 Composizione del latte.....	41
2.3 Composizione chimica del latte	45
2.3.1 Acqua.....	45
2.3.2 Lattosio	45
2.3.3 Proteine	46
2.3.3.1 Caseina.....	49
2.3.3.2 Le micelle caseiniche	50
2.3.3.3 Le proteine del siero	53
2.3.4 Sali minerali	54
2.3.5 Enzimi	56
2.3.6 Vitamine	58
2.4 Lipidi.....	58
2.4.1 I globuli di grasso.....	59
2.4.1.1 La membrana dei globuli di grasso	60
2.4.1.2 Composizione della membrana	61
2.4.1.3 Struttura della membrana	64
2.4.2 Composizione del grasso del latte.....	66
2.4.2.1 Acidi grassi saturi	69
2.4.2.2 Acidi grassi insaturi	70
2.4.2.3 Acidi linoleici coniugati (CLA).....	72
2.4.2.4 Trigliceridi	73
3 La sintesi dei costituenti del latte: la ghiandola mammaria.....	75

4	Cause di variazione della composizione del latte	81
4.1	Fattori genetici	81
4.2	Fattori fisiologici	83
4.3	Fattori alimentari.....	87
4.4	Fattori ambientali.....	89
4.5	Fattori tecnologici.....	90
5	Aspetti nutrizionali dei lipidi	91
5.1	Effetti dei lipidi sulla salute	91
5.1.1	Malattie cardiovascolari	92
5.1.1.1	Gli acidi grassi polinsaturi e gli eicosanoidi.....	92
5.1.1.2	Effetti dei lipidi sul livello ematico di colesterolo	95
5.1.2	Effetti dei lipidi sul cancro	99
5.2	Attività biologica dei CLA.....	100
5.2.1	I CLA e l'aterosclerosi.....	100
5.2.2	I CLA ed il cancro.....	101
5.2.3	Altri effetti dei CLA.....	102
6	Produzione e utilizzo del latte	103
6.1	Produzione del latte	103
6.2	Trattamenti e trasformazioni del latte	105
7	Metodi di caratterizzazione del profilo acido del latte per l'acquisizione dei dati di riferimento	109
7.1	Estrazione della frazione lipidica dalla matrice	109
7.2	Derivatizzazione della frazione lipidica estratta.....	114
7.3	Caratterizzazione analitica dei lipidi derivatizzati	116
8	Analisi Multivariata	119
8.1	Introduzione	119
8.2	Trattamento dei dati	120
8.2.1	Analisi delle componenti principali (PCA).....	123
8.2.1.1	La costruzione matematica della PCA.....	131
8.2.2	Pretrattamento dei dati.....	133
8.3	Calibrazione multivariata e modelli di predizione.....	135
8.3.1	Concetti generali.....	135
8.3.2	Aspetti matematici dei modelli di predizione.....	138
8.3.2.1	La regressione multivariata.....	140
8.3.2.2	Regressione parziale con minimi quadrati: PLS	141
8.3.3	Rischio di overfitting	143
8.3.4	Tecniche di validazione.....	144
8.3.5	Parametri di valutazione dei modelli di regressione	145
8.3.5.1	Valutazioni grafiche dei modelli di regressione	150
8.4	Lo spettro infrarosso come matrice dei predittori.....	155
8.4.1	Pretrattamenti spettrali.....	155

8.4.2	Selezione delle variabili	159
9	Gli algoritmi genetici.....	161
9.1	La teoria evolutiva.....	161
9.2	Teoria evolutiva degli algoritmi genetici.....	163
9.2.1	La codifica delle variabili	164
9.3	Passaggi degli algoritmi genetici	165
9.3.1	Creazione di una popolazione originaria.....	165
9.3.2	Riproduzione	166
9.3.3	Mutazione	168
9.4	Proprietà generali degli algoritmi genetici.....	169
9.5	Applicazioni degli algoritmi genetici	170
9.6	Applicazione degli algoritmi genetici nella selezione di variabili spettrali.....	171
9.6.1	La corsa dell'algoritmo genetico	171
9.6.2	Implementazioni dell'algoritmo per ridurre il rischio di overfitting e l'autocorrelazione tra variabili	174
9.6.3	Il modello finale.....	177
10	La spettroscopia infrarossa nell'analisi alimentare	179
10.1	Introduzione.....	179
10.2	La spettroscopia infrarossa	179
10.2.1	Spettroscopia MIR a Trasformata di Fourier	183
10.3	La spettroscopia infrarossa nell'analisi degli alimenti	185
10.4	La spettroscopia infrarossa nell'analisi del latte.....	187
10.4.1	Spettro MIR del latte	188
10.4.2	Applicazioni della spettroscopia infrarossa nell'analisi del latte	190
11	Parte sperimentale I: acquisizione dei dati di riferimento	195
11.1	Obiettivo della Tesi	195
11.2	Campionamento	195
11.3	Scelta del metodo di caratterizzazione del profilo acidico del latte	197
11.4	Materiali e Metodi	199
11.4.1	Estrazione della frazione lipidica: metodo fisico [Murphy, 1990].....	199
11.4.2	Trasmetilazione basica: metodo FIL-IDF 182:1999/ISO 15884	200
11.4.3	Separazione e quantificazione dei FAME	201
11.4.3.1	Curva di calibrazione.....	203
11.4.4	Acquisizione dello spettro e della concentrazione di grasso nel latte	210
11.4.5	Espressione degli acidi grassi individuali in g/100g di latte	213
11.5	Risultati e discussione	214
12	Parte sperimentale II: validazione del metodo di riferimento per la caratterizzazione del profilo acidico del latte ovino	229
12.1	Precisione.....	229
12.2	Esattezza	232
12.3	Prove di ripetibilità e riproducibilità	234

12.4 Prove di recupero	237
12.5 Variabilità sperimentale del metodo di riferimento	238
13 Parte sperimentale III: analisi delle componenti principali	241
13.1 PCA degli spettri FT-IR in assorbanza	241
13.2 PCA dei valori di riferimento	244
13.2.1 PCA su 242 campioni e 71 variabili	244
13.2.2 PCA su 242 campioni e 19 variabili	248
13.3 PCA degli spettri FT-IR in assorbanza pretrattati con operazioni matematiche	251
13.3.1 PCA degli spettri pretrattati con derivata prima	251
14 Parte Sperimentale IV: elaborazione dei modelli di predizione	257
14.1 Analisi dello spettro.....	258
14.2 Costruzione dei modelli di predizione.....	260
14.2.1 Costruzione dei modelli di predizione con l'utilizzo dello spettro FT-IR intero.....	261
14.2.2 Costruzione dei modelli di predizione con l'utilizzo dello spettro FT-IR ridotto	263
14.2.3 Costruzione dei modelli di predizione con l'utilizzo delle bande spettrali selezionate dall'analisi visiva dello spettro	265
14.2.4 Costruzione dei modelli di predizione con l'utilizzo delle bande spettrali selezionate dagli algoritmi genetici	267
14.2.4.1 Selezione delle bande tramite algoritmi genetici	268
14.2.4.2 Validazione dei modelli di predizione basati sulla selezione delle bande spettrali effettuata dagli algoritmi genetici	285
14.3 Confronto dei modelli di predizione ottenuti tramite selezione delle regioni spettrali ...	287
14.4 Modelli di predizione selezionati	288
15 Conclusioni	301
16 Bibliografia	303
Ringraziamenti	329

Introduzione

L'ampliamento del mercato dei prodotti agroalimentari a livello internazionale ha determinato negli ultimi decenni un'accresciuta possibilità di scelta d'acquisto da parte del consumatore. A tale molteplicità di possibili scelte non ha fatto purtroppo riscontro un pari livello di consapevolezza su potenziali rischi per la salute connessi al consumo di talune produzioni. La perdita del legame tra produttore e consumatore che aveva caratterizzato le produzioni alimentari di qualche decennio fa, e le crescenti preoccupazioni dovute alle sempre più frequente casistica connessa alle contaminazioni alimentari ha portato il consumatore ad esigere una maggiore attenzione verso la sicurezza e la qualità dei prodotti alimentari, che è sfociata in una richiesta crescente di prodotti oggettivamente sani e salutari. Di riflesso è aumentata anche l'attenzione dei produttori agli alimenti di qualità. Tutto questo ha portato alla richiesta da ambo le parti di strumenti di certificazione in campo alimentare, realizzatasi nell'emanazione di disposizioni specifiche per la sicurezza e la tracciabilità, quali:

- i sistemi di autocontrollo HACCP (Hazard Analysis of Critical Control Points) - obbligatori per tutti gli operatori del settore alimentare a partire dal 1993 - atti a valutare i rischi igienico-sanitari che possono influenzare la sicurezza degli alimenti nelle singole fasi del ciclo produttivo;
- il Regolamento CE 178/2002, che stabilisce i principi ed i requisiti generali della legislazione alimentare, comprendendo anche la sicurezza alimentare.

Il mercato globale ha reso necessaria l'introduzione di un sistema di definizione degli standard alimentari, atto a definire le proprietà e le caratteristiche che ogni prodotto deve possedere, permettendo quindi di ricondurre allo stesso nome lo stesso prodotto. A tale scopo, è stato istituito nel 1963 il Codex Alimentarius, che detta inoltre le linee guida per proteggere la salute del consumatore e garantire pratiche eque nel commercio degli alimenti.

L'introduzione dei marchi di origine (DOP, Denominazione d'Origine Protetta; IGP, Identificazione Geografica Protetta; STG, Specialità tradizionale Garantita), ha permesso inoltre di promuovere prodotti di qualità con caratteristiche legate al territorio, garantendo la provenienza delle materie prime e la loro trasformazione.

Oggi giorno l'acquisto dei prodotti alimentari avviene prevalentemente in punti vendita di grandi catene commerciali, e la garanzia della sicurezza e della qualità alimentare sono dati dal marchio commerciale dell'alimento e dalle informazioni in etichetta. La qualità dei prodotti alimentari non può però essere limitata al controllo dei requisiti minimi di qualità e sicurezza. Infatti, da qualche decennio è nata la tendenza a considerare un alimento non solo in base al suo ruolo nutritivo ma anche al suo ruolo "funzionale", ovvero alla presenza in esso di componenti – naturali o aggiunti – che possono apportare effetti positivi sulla salute, grazie all'interazione con una o più funzioni fisiologiche dell'organismo. Il consumatore medio, preoccupato da un lato per lo stile di vita spesso sedentario e dall'altro per la potenziale presenza negli alimenti da produzioni massive di sostanze pericolose per la salute, vede nell'alimento funzionale un prodotto che potrebbe favorire lo stato di salute ed il benessere della persona, agendo come fattore protettivo per l'organismo.

In questo contesto, i prodotti lattiero-caseari costituiscono sia una parte considerevole della nostra alimentazione che il retaggio di una produzione tradizionale di rilevante importanza economica, culturale ed ambientale. Infatti, i prodotti lattiero-caseari tutelati (DOP, IGP e STG) presenti sul mercato sono numerosi e coinvolgono un numero elevato di operatori.

In questo contesto ben si inserisce il sistema di produzione di latte ovino in Sardegna. La sua importanza economica è infatti fondamentale per la regione, in quanto essa costituisce una delle principali attività produttive dell'Isola. La produzione di latte ovino sardo è interamente destinata alla produzione di formaggi, spesso a denominazione di origine protetta quali il Pecorino Romano DOP, il Fiore Sardo DOP ed il Pecorino Sardo DOP.

Mentre in passato in Sardegna il sistema di pagamento del latte ovino era basato esclusivamente sulla quantità di latte prodotto, recentemente è stato introdotto un sistema di pagamento legato al livello igienico-sanitario (conta batterica e numero di cellule somatiche) e a quello di qualità del prodotto (tenore di grasso e proteine). I controlli di qualità regionali vengono effettuati dal laboratorio dell'Associazione Regionale Allevatori della Sardegna (ARA Sardegna, soggetto *super partes* tra gli allevatori ed i caseifici trasformatori). Presso il laboratorio dell'ARA vengono quindi effettuate le analisi per l'oggettivazione dei parametri che determineranno il prezzo del latte. I risultati delle analisi sono disponibili entro le 36 ore dal conferimento del campione grazie all'adozione di tecniche analitiche rapide e non distruttive quali la spettroscopia infrarossa. I parametri che attualmente vengono determinati dal laboratorio dell'ARA sono: Grasso, Proteine, Lattosio, Cellule Somatiche, Carica Batterica, Punto Crioscopico, pH, Urea, Caseina, Ricerca latte vaccino o caprino nel latte ovino e Inibenti. Dal momento che il latte è

l'alimento funzionale per eccellenza, grazie alla presenza di numerosi composti con effetti positivi per la salute quali peptidi bioattivi derivanti dalle caseine, proteine del siero, amminoacidi, acidi grassi polinsaturi e sostanze derivate da essi, acidi grassi omega-3 ed acidi linoleici coniugati, è auspicabile che i controlli analitici rapidi e non distruttivi non si limitino soltanto alla verifica del rispetto delle garanzie minime di qualità e sicurezza, ma si spingano sino alla determinazione di parametri, magari legati ad analiti non maggioritari, che possano valorizzare ulteriormente il prodotto. I nuovi parametri, se inseriti nel sistema di pagamento del latte ovino, potrebbero ulteriormente incentivare l'allevatore alla produzione di un latte di elevata qualità in virtù dell'ottenimento di un maggiore profitto.

Il controllo analitico della qualità tecnologica e funzionale del latte deve però poter essere effettuato, oltretutto in tempi ridotti, anche con costi minimi. Infatti, parte dei costi delle analisi dei controlli di qualità ricadono in qualche modo sugli allevatori, e quindi sui costi di produzione del latte. Le tecniche analitiche online, impieganti spettroscopia infrarossa a Trasformata di Fourier, appaiono anche in questo caso essere le più adatte per effettuare tali valutazioni di qualità in quanto veloci ed economiche. Al momento attuale, tuttavia, non esistono metodi attualmente impiegati su latte ovino che permettano determinazioni di specie presenti in concentrazioni minoritarie nel latte.

Lo scopo generale di questa Tesi di dottorato è stato quello di mettere a punto e validare tecniche analitiche online impieganti strumentazione a Trasformata di Fourier per la determinazione di macro e microanaliti in latte ovino. In particolare, l'attenzione è stata rivolta alla determinazione degli acidi grassi e delle classi di acidi grassi più rappresentativi in termini quantitativi e qualitativi. Sono stati scelti gli acidi grassi saturi con numero pari di atomi di carbonio, a partire dall'acido butirrico fino all'acido stearico (imputati di essere responsabili dell'insorgenza di malattie cardiovascolari), gli acidi grassi insaturi vaccenico, oleico, linoleico, α -linolenico e rumenico (che possiedono effetti positivi sulla salute dell'uomo) e le classi di acidi grassi saturi, insaturi, monoinsaturi, polinsaturi, omega-6 e omega-3.

1 Il sistema alimentare

1.1 La qualità alimentare

Negli ultimi decenni, i prodotti agroalimentari hanno registrato un incremento del loro mercato a livello internazionale. La globalizzazione odierna, favorita dal miglioramento dei trasporti, ha permesso l'interscambio di numerosi prodotti tra Paesi di grandi blocchi geografici ed economici quali l'Europa, la Cina e l'America settentrionale. Un quadro di questo tipo, oltre a garantire un'ampia e diversificata offerta di prodotti agroalimentari nel mercato, ha determinato per il consumatore un'ampia possibilità di scelta.

Oggi il consumatore è sempre più attento alla qualità dei prodotti alimentari in quanto consapevole dei rischi sulla salute derivati da alcuni componenti contenuti in essi, ma quasi mai è in grado di verificarne direttamente le caratteristiche. Stante ciò, la richiesta di prodotti sani e salutaris da parte dei consumatori ha indotto nei produttori la necessità di introdurre sul mercato prodotti di qualità.

È nata quindi l'esigenza di strumenti di certificazione in campo alimentare, sfociata con l'emanazione di disposizioni specifiche per la sicurezza e la tracciabilità, quali i sistemi di autocontrollo HACCP (Hazard Analysis of Critical Control Points) ed il Regolamento (CE) 178/2002. Il compito del sistema HACCP è quello di valutare i rischi igienico-sanitari che possono influenzare la sicurezza degli alimenti nelle singole fasi del ciclo produttivo. L'applicazione del sistema è obbligatoria per tutti gli operatori del settore alimentare a partire dal 1993, anno in cui il sistema è stato introdotto in Europa con la direttiva 43/93/CEE (recepita in Italia con il Decreto Legislativo 155/1997). La normativa è stata sostituita dal Regolamento CE 852/2004, entrato in vigore dal primo gennaio 2006 e recepito in Italia con il Decreto Legislativo 193/2007. Il Regolamento (CE) 178/2002, e successive modifiche, *stabilisce i principi e i requisiti generali della legislazione alimentare, istituisce l'Autorità europea per la sicurezza alimentare e fissa procedure nel campo della sicurezza alimentare.*

Nel settembre 2005, l'Ente di Normazione Internazionale ISO ha pubblicato la norma Standard ISO 22000:2005, applicabile su base volontaria, con lo scopo di conformare gli standard nazionali ed internazionali preesistenti in materia di sicurezza alimentare e HACCP. Lo standard è stato redatto da un gruppo di lavoro composto da esperti provenienti da 23 diverse nazioni e da rappresentanti di organizzazioni internazionali, quali la Global Food Safety Initiative (GFSI), la Codex Alimentarius Commission, l'Associazione internazionale degli hotel e dei ristoranti e la Confederazione delle Industrie

agro-alimentari dell'Unione Europea (CIAA). Lo standard, basato sui principi dell'HACCP definiti dal Codex Alimentarius ed allineato con i precedenti ISO 9000 e ISO 14000, si pone come punto di riferimento per gli operatori per l'applicazione dei regolamenti comunitari in materia di igiene e sicurezza alimentare. Esso può essere adottato sia dagli operatori coinvolti direttamente nel processo di produzione degli alimenti sia dagli operatori indirettamente coinvolti (es. produttori di imballaggi, servizi di pulizia e derattizzazione ecc.). Lo standard viene applicato in ogni singolo stadio della catena di produzione anche se non è volto alla certificazione dell'intera filiera.

La globalizzazione e la libera circolazione dei prodotti ha reso inoltre necessaria la definizione delle proprietà e delle caratteristiche che ogni prodotto deve possedere, e la possibilità di ricondurre allo stesso nome lo stesso prodotto, tutelando così sia il produttore che il consumatore. La FAO (l'Organizzazione delle Nazioni Unite per l'alimentazione e l'agricoltura) ed il WTO (l'Organizzazione mondiale del commercio) hanno istituito nel 1963 il Codex Alimentarius per definire gli standard alimentari. Il Codex Alimentarius è costituito da una Commissione intergovernativa di cui fanno parte 186 membri (185 nazioni più la Comunità Europea) e sviluppa standard alimentari internazionali e linee guida per proteggere la salute del consumatore e garantire pratiche eque nel commercio degli alimenti. Le regole del Codex vengono stabilite e revisionate da esperti di diverse discipline scientifiche dell'alimentazione (autorità di controllo del cibo, esperti scientifici e tecnici alimentari, rappresentanti di associazioni dei consumatori, dei produttori, dell'industria e del commercio) e servono da base per le normative nazionali.

A tutto ciò si aggiunge l'introduzione dei marchi di origine (DOP, Denominazione d'Origine Protetta; IGP, Identificazione Geografica Protetta; STG, Specialità tradizionale Garantita), che garantiscono la provenienza delle materie prime e la loro trasformazione. I marchi di origine, regolamentati dall'Unione Europea, promuovono prodotti di qualità con caratteristiche legate al territorio, tutelando sia i produttori che il consumatore.

1.1.1 La sicurezza alimentare

Col termine "sicurezza alimentare" ci si riferisce alla assenza di possibili impatti negativi causati da un alimento sulla salute dei consumatori. A causa dei grandi cambiamenti che si sono realizzati nel sistema agroalimentare moderno, è progressivamente aumentata la distanza, fisica e culturale, tra il produttore di materie prime agricole ed il consumatore finale, col risultato di una "spersonalizzazione dei rapporti" ed una perdita di quelle

garanzie di sicurezza che, un tempo, il contatto personale tra acquirenti e venditori era in grado di assicurare. L'acquisto dei prodotti alimentari avviene prevalentemente in punti vendita di grandi catene commerciali e gli elementi di garanzia sono dati dal marchio commerciale dell'alimento e dalle informazioni in etichetta. La globalizzazione dei mercati ha portato inoltre ad una diversa percezione dei rischi e ad una maggiore importanza dei controlli sulla sicurezza che vengono effettuati da apparati istituzionali.

1.1.2 Concetti di qualità alimentare

La qualità di un alimento è definita come la *convergenza tra i desideri e le necessità dei consumatori e i requisiti qualitativi intrinseci ed estrinseci dei prodotti alimentari* [Hocquette, 2005]. I requisiti intrinseci si riferiscono al prodotto stesso ed includono:

- aspetti salutistici e di sicurezza (assenza di effetti nocivi per la salute causati dalla presenza di contaminanti chimici o biologici);
- proprietà sensoriali (che includono l'assenza di odori, sapori e colori anomali) e shelf life;
- proprietà chimiche e nutrizionali;
- convenienza ed affidabilità.

I requisiti estrinseci si riferiscono invece a caratteristiche associabili al prodotto:

- sistema di produzione (dall'allevamento ai processi di produzione);
- variabili di mercato (prezzi, marchi di fabbrica, distribuzione, etichettatura, tracciabilità etc...).

Alcuni esperti hanno distinto la qualità in base ai concetti di aspettativa e di esperienza. I consumatori possiedono un'aspettativa della qualità che determina la scelta di un determinato prodotto con l'esclusione di un altro, ma conseguono l'esperienza qualitativa solo quando quel prodotto viene consumato. L'esperienza qualitativa assume una grande importanza in quanto sarà essa che determinerà se un prodotto ha realmente soddisfatto il consumatore e se quindi quel prodotto verrà da lui ricomprato. L'aspettativa di qualità è basata sia su requisiti intrinseci che estrinseci mentre l'esperienza di qualità dipende dall'esperienza pregressa e da parametri di fiducia. Mentre l'esperienza si costruisce con il consumo del cibo (conoscenza del sapore, della texture etc...), i parametri di fiducia non

possono essere valutati dal consumatore e corrispondono a quelle caratteristiche che non possono essere conosciute con certezza neanche dopo il consumo. Ne fanno parte il contenuto di additivi, di conservanti o di sostanze utili alla salute e tutto ciò che dovrebbe rispettare determinate modalità produttive. In questo caso è solo la fiducia nei marchi, nelle informazioni di etichetta o in altri elementi che comunicano nel consumatore una certa “reputazione” del prodotto. Da qui l’importanza dei sistemi di controllo e di garanzia atti a costruire una reputazione e un rapporto di fiducia che risulta centrale per l’apprezzamento e la valorizzazione della qualità di un prodotto [Hocquette, 2005].

Il contesto che si è venuto a formare impone che anche le aziende agricole effettuino una serie di attività e di controlli, che hanno però dei costi elevati. Benché quindi esistano delle esigenze in termini di controlli del rispetto delle garanzie minime di qualità dei prodotti alimentari, c’è il rischio che gli agricoltori siano di fatto impossibilitati a recuperare almeno parte della spesa affrontata.

La qualità dei prodotti alimentari non può però essere limitata al controllo dei requisiti minimi di qualità e sicurezza ma è necessario introdurre nuovi parametri che consentano di valorizzare qualitativamente un prodotto ed ottenere dal mercato un prezzo finale più interessante e remunerativo [ISMEA, 2006].

1.1.3 Proprietà funzionali degli alimenti

Lo scopo principale della dieta è quello di fornire i nutrienti necessari per soddisfare le esigenze dell’organismo. Da qualche decennio è nata la tendenza a considerare un alimento non solo in base al suo ruolo nutritivo ma anche al suo ruolo “funzionale”. Un alimento è definito “funzionale” se in esso sono naturalmente presenti, o aggiunti, componenti che, interagendo con una o più funzioni fisiologiche dell’organismo, portano effetti positivi sulla salute.

Oggi giorno i consumatori mostrano una certa tendenza verso il mantenimento di una dieta equilibrata volta a massimizzare le funzioni fisiologiche dell’individuo, a ridurre il rischio di malattie assicurando così uno stato di benessere. Nell’epoca moderna il consumatore preoccupato per uno stile di vita sempre più sedentario e conscio del largo uso di sostanze pericolose (pesticidi, fertilizzanti, antiparassitari, antibiotici etc) in agricoltura e zootecnia, è sempre più sensibile ai problemi legati all’alimentazione e vede nell’alimento funzionale un alimento che potrebbe favorire lo stato di salute ed il benessere della persona, agendo come fattore protettivo per l’organismo. Il punto di vista comune del consumatore è infatti

il seguente: “i tre più importanti fattori che influenzano la salute sono l’alimentazione, l’esercizio e i fattori genetici” [Hilliarn, 1996].

L’Azione Concertata della Commissione Europea sulla Scienza degli Alimenti Funzionali in Europa ha redatto un documento che definisce il Concetto Scientifico di Alimento Funzionale e cita che “un alimento può essere definito funzionale se è soddisfacentemente dimostrato che apporta benefici ad una o più funzioni dell’organismo, al di là degli adeguati effetti nutrizionali, migliorando lo stato di salute e benessere e/o riducendo il rischio di malattie. Un alimento funzionale deve rimanere un alimento e deve dimostrare i suoi effetti in quantità che possano normalmente essere consumate nella dieta: esso non è una pillola o una capsula ma un normale alimento” [Diplock, 1999].

Secondo la Commissione Europea, un alimento funzionale deve quindi possedere le seguenti caratteristiche:

- essere un alimento;
- essere parte integrante della dieta;
- possedere componenti naturali, non sintetici, anche in concentrazioni superiori alla norma o in alimenti che solitamente non li contengono;
- al di là del proprio valore nutritivo, avere un effetto positivo su una o più attività funzionali;
- migliorare il benessere e la salute dell’uomo e/o ridurre il rischio di malattie in modo tale da migliorare la qualità della vita, incluse le performance fisiche, mentali e comportamentali;
- gli effetti positivi debbono essere riconosciuti su base scientifica.

Va però sottolineato che un alimento funzionale non deve essere necessariamente tale per tutti i membri della popolazione. È compito della ricerca scientifica, grazie al progresso sulla comprensione delle interazioni tra geni e dieta, trovare una corrispondenza tra i singoli bisogni biochimici e l’assunzione di componenti alimentari specifici.

Da un punto di vista pratico, un alimento funzionale può essere:

- un alimento naturale;
- un alimento in cui viene aggiunto un componente;
- un alimento da cui viene rimosso un componente;
- un alimento in cui la natura di uno o più componenti è stata modificata;

- un alimento in cui la biodisponibilità di uno o più componenti è stata modificata;
- una combinazione dei precedenti casi.

Un grande numero di alimenti funzionali in varie forme sono stati introdotti col tempo nel mercato. I principali componenti contenuti in essi sono fibre dietetiche, oligosaccaridi, peptidi e proteine, prebiotici, probiotici, fitochimici, antiossidanti ed acidi grassi polinsaturi. Le sostanze che possono portare alla produzione di alimenti funzionali possono essere classificate nel modo seguente [Roberfroid, 1997]:

- Macronutrienti essenziali con specifiche attività fisiologica, quali gli acidi grassi omega-3;
- Micronutrienti essenziali nel caso in cui l'assunzione conferisca benefici oltre la dose giornaliera raccomandata (RDI);
- Nutrienti non essenziali con specifiche attività fisiologiche quali i fitochimici ed alcuni oligosaccaridi.

1.1.4 I prodotti tipici

La globalizzazione dei mercati, regolati da organizzazioni internazionali, ha inizialmente orientato i consumatori verso prodotti di tipo industriale che garantivano comunque la sicurezza dal punto di vista igienico-sanitario unitamente ad un prezzo sempre più competitivo. La maggiore consapevolezza dell'aspetto nutrizionale dell'alimento e delle malattie legate all'alimentazione ha però portato il consumatore dei Paesi più sviluppati a privilegiare quei prodotti che uniscono gli aspetti di salubrità e di qualità alimentare, nonostante essi presentino normalmente costi superiori rispetto ad altri prodotti della stessa categoria.

La qualità è ben rappresentata da quei prodotti agroalimentari che sono frutto di tecniche di produzione tradizionali che impiegano risorse e materie prime specifiche locali con un'origine geografica ben definita.

Questi prodotti vengono definiti "tipici" e sono il risultato di una sperimentazione alimentare condotta nei secoli. Un prodotto "tipico" viene considerato dal consumatore medio come un prodotto sano e sicuro, e quindi viene visto come sinonimo di qualità garantita.

In generale, con “prodotto tipico” si intende una specialità agroalimentare con una forte caratterizzazione geografica. La definizione di “tipico” include: (i) le caratteristiche sensoriali del prodotto, (ii) l’origine geografica sia delle materie prime che del processo di trasformazione, (iii) la relazione con la tradizione sociale e culturale dell’area di produzione [Scintu, 2007].

Il Regolamento (CE) 2081/92 tenta di unificare differenti concetti di tipicità, definendo i marchi DOP (Denominazione d’Origine Protetta) e IGP (Identificazione Geografica Protetta), oggi normati dal Regolamento (UE) 1151/2012, ponendo così le regole per la protezione dei prodotti agroalimentari. I marchi costituiscono un elemento per la protezione della biodiversità di un territorio (geologica, pedologica, climatica) includendo piante, animali e microbi, pratiche e sistemi di produzione collegati alla storia e alla cultura locale del territorio stesso; allo stesso tempo assicurano un guadagno adeguato per la popolazione rurale proteggendo il prodotto, attraverso una denominazione, contro un sistema di competizione basato sulla globalizzazione e sui costi ridotti.

I marchi DOP e IGP differiscono a seconda della natura e dell’intensità del legame tra il prodotto e l’area geografica alla quale sono legati. Il marchio DOP rappresenta la massima espressione dell’identificazione e verifica delle qualità di un prodotto in rapporto al territorio di produzione, in quanto vincola tutte le fasi produttive ad un territorio specifico. Il marchio IGP possiede una specificità inferiore, in quanto non lega necessariamente tutte le fasi produttive ad un territorio, ma richiede che almeno una di esse sia effettuata nel territorio indicato, mantenendo quindi l’elemento “origine geografica” come punto centrale di riferimento. Un prodotto DOP è quindi associato in modo più profondo all’area geografica d’origine rispetto ad un prodotto IGP.

Un prodotto può essere identificato come DOP solo se è originario della specifica regione di cui porta il nome. Le materie prime devono essere prodotte e lavorate nell’area geografica riportata fino all’ottenimento del prodotto finale, e la qualità o le caratteristiche del prodotto devono essere dovute essenzialmente o esclusivamente alle caratteristiche geografiche dell’area di produzione. Queste ultime includono fattori naturali ed umani quali il clima, la natura del suolo e le conoscenze della popolazione locale. Un prodotto può essere identificato come DOP anche se le materie prime utilizzate provengono da un’area geografica più vasta o differente da quella di lavorazione purché essa sia delimitata e sussistano condizioni particolari per la produzione delle materie prime. L’osservanza delle regole deve essere garantita da un regime di controllo.

Un prodotto può essere identificato come IGP se è prodotto all'interno di un'area geografica. A differenza del marchio DOP, è necessario che almeno uno degli stadi della sua produzione avvenga in suddetta area. Deve esistere un legame tra il prodotto e la regione da cui prende il nome ma questa caratteristica non deve necessariamente essere quella predominante od esclusiva, come nel caso del marchio DOP.

Secondo l'articolo 7 del Regolamento (UE) 1151/2012, per beneficiare di una denominazione di origine protetta (DOP) o di un'indicazione geografica protetta (IGP), un prodotto deve essere conforme a un disciplinare che comprende almeno i seguenti elementi:

- a) il nome da proteggere come denominazione di origine o indicazione geografica, quale utilizzata nel commercio o nel linguaggio comune, e solo nelle lingue attualmente o storicamente utilizzate per descrivere il prodotto specifico nella zona geografica delimitata;
- b) la descrizione del prodotto, comprese (se del caso) le materie prime, nonché le principali caratteristiche fisiche, chimiche, microbiologiche od organolettiche del prodotto;
- c) la definizione della zona geografica delimitata riguardo al legame di cui alla lettera f), punto 1) o punto 2), del presente paragrafo;
- d) gli elementi che dimostrano che il prodotto è originario della zona geografica delimitata;
- e) la descrizione del metodo di ottenimento del prodotto e (se del caso) dei metodi locali e delle informazioni relative al confezionamento qualora esso debba aver luogo nella zona geografica delimitata, al fine di salvaguardare la qualità, garantire l'origine o assicurare il controllo del prodotto stesso;
- f) gli elementi che stabiliscono:
 1. per le DOP le cui fasi produttive si svolgono nella zona geografica di riferimento: il legame fra la qualità o le caratteristiche del prodotto e l'ambiente geografico;
 2. per le IGP la cui produzione si svolge per almeno una delle sue fasi nella zona geografica delimitata: se del caso, il legame fra una data qualità, la reputazione o un'altra caratteristica del prodotto e l'origine geografica;

- g) il nome e l'indirizzo delle autorità o, se disponibili, il nome e l'indirizzo degli organismi che verificano il rispetto delle disposizioni del disciplinare di produzione;
- h) qualsiasi regola specifica per l'etichettatura del prodotto in questione.

Sulla base delle definizioni sopra riportate, è possibile stabilire a quale categoria di designazione geografica (DOP o IGP) appartenga un determinato prodotto e - di conseguenza - per quale tipologia chiederne il riconoscimento. I marchi DOP e IGP garantiscono specifiche proprietà produttive ed implicano che nel prodotto siano presenti particolari caratteristiche. La qualità e la tracciabilità di un prodotto sono quindi garantite dai marchi di designazione geografica.

Il Regolamento (UE) 1151/2012 considera anche la denominazione di Specialità Tradizionale Garantita (STG) che fornisce strumenti di tutela per i prodotti alimentari tradizionali ad alto carattere di "specificità", intesa come elemento o insieme di elementi che, per le loro caratteristiche qualitative e di tradizione, permettono la distinzione del prodotto da altri prodotti analoghi appartenenti alla stessa categoria. La STG si riferisce quindi a prodotti ottenuti secondo un metodo di produzione tradizionale tipico di una particolare area geografica, dove per "tradizionale" si intende l'uso comprovato di un prodotto sul mercato nazionale per un periodo di tempo non inferiore a trenta anni, che permette di tramandare le conoscenze da una generazione all'altra.

L'obiettivo della denominazione STG è la salvaguardia di metodi di produzione e ricette tradizionali che permettono ai produttori di commercializzare i propri prodotti comunicando ai consumatori le proprietà che conferiscono alle loro ricette e valore aggiunto ai loro prodotti tradizionali. La denominazione STG non comprende i prodotti il cui carattere peculiare sia legato all'origine geografica ma deve comprendere gli elementi che ne attestano la "specificità" e la "tradizione".

Secondo l'articolo 18 del Regolamento (UE) 1151/2012, un prodotto può beneficiare della denominazione STG se soddisfa le seguenti condizioni:

- è ottenuto con un metodo di produzione, trasformazione o una composizione che corrispondono a una pratica tradizionale per tale prodotto o alimento;
- è ottenuto da materie prime o ingredienti utilizzati tradizionalmente.

L'articolo 19 dello stesso Regolamento indica il disciplinare per i prodotti STG secondo cui una specialità tradizionale garantita deve indicare (i) il nome di cui è proposta la registrazione, (ii) la descrizione del prodotto, comprese le principali caratteristiche fisiche, chimiche, microbiologiche od organolettiche, a dimostrazione della specificità del prodotto, (iii) la descrizione del metodo di produzione compresi la natura e le caratteristiche delle materie prime o degli ingredienti utilizzati e il metodo di elaborazione del prodotto e (iv) gli elementi fondamentali che attestano il carattere tradizionale del prodotto.

1.1.5 I prodotti tipici in Europa ed Italia

Il numero di prodotti DOP, IGP e STG riconosciuti nell'Unione Europea supera le 1300 unità a cui si aggiungono circa 150 prodotti in attesa di riconoscimento. L'Italia si conferma primo Paese per numero di riconoscimenti conferiti dall'Unione europea (UE) con 256 prodotti di qualità riconosciuti, seguita da Francia (201), Spagna (167), Portogallo (119), Grecia (100) e Germania (94) [UE, 2013].

Diversi sono i settori alimentari considerati, ognuno dei quali prevalente in alcuni paesi piuttosto che in altri: le denominazioni riguardanti prodotti del settore delle carni fresche sono prevalenti in Francia; i prodotti a base di carne (riscaldati, salati, affumicati, ecc.) sono prevalentemente di origine italiana e portoghese; i prodotti della classe dei formaggi sono prevalenti in Francia ed in Italia mentre gli altri prodotti di origine animale (uova, miele, prodotti lattiero-caseari ad eccezione del burro, ecc.) trovano una maggiore produzione in Portogallo; l'Italia risulta essere la maggiore produttrice anche di prodotti contenenti oli e grassi (burro, margarina, olio, ecc.), di prodotti ortofrutticoli e cereali (freschi o trasformati) e di spezie; i prodotti di panetteria, pasticceria, confetteria o biscotteria hanno origine prevalentemente spagnola mentre le bevande (birre e acque minerali) provengono principalmente dalla Germania; il settore di pesci, molluschi, crostacei freschi e prodotti derivati è invece prevalente nel Regno Unito.

I prodotti DOP, IGP e STG si confermano quindi componente significativa della produzione agroalimentare italiana e fattore di competitività delle realtà agricole locali e vanno assumendo nel tempo dimensioni sempre più importanti, pur mantenendo talune caratteristiche tipiche dei prodotti di nicchia. Nel periodo 2004-2013 si registra un consistente aumento del numero delle specialità riconosciute; in particolare, tra il 2004 e il 2012, le DOP, IGP e STG riconosciute dall'Unione Europea passano da 146 a 256 (110

prodotti in più, +75%) [ISTAT, 2012]. Come si vede in Tabella 1.1, dei 256 prodotti italiani riconosciuti dall'Unione Europea, ben 100 sono prodotti di origine ortofrutticola, 45 sono invece le varietà di formaggi tutelate, 43 sono prodotti a base di oli e grassi e 37 sono invece i prodotti a base di carne. Le regioni con più prodotti DOP e IGP sono Emilia-Romagna e Veneto, con 36 e 35 prodotti riconosciuti rispettivamente. Nell'Italia Centrale spiccano invece Toscana con 24 prodotti e Lazio con 23 prodotti, mentre nel sud Italia sono in testa Sicilia e Campania, con 27 e 21 prodotti rispettivamente. In coda alla classifica si trovano Valle d'Aosta e Liguria rispettivamente con 4 e con 3 prodotti tutelati [ISTAT, 2012].

Tabella 1.1 – Divisione per classi di prodotti DOP, IGP e STG in Italia [UE, 2013]

Classe (fonte: Unione Europea)	Prodotti
Carni fresche (e frattaglie)	5
Prodotti a base di carne (riscaldati, salati, affumicati, ecc.)	37
Formaggi	45
Altri prodotti di origine animale	
(uova, miele, prodotti lattiero-caseari ad eccezione del burro, ecc.)	4
Oli e grassi (burro, margarina, olio, ecc.)	43
Ortofrutticoli e cereali, freschi o trasformati	100
Pesci, molluschi, crostacei freschi e prodotti derivati	4
Altri prodotti (spezie, ecc.)	8
Prodotti di panetteria, pasticceria, confetteria o biscotteria	8
Pasta alimentare	1
Oli essenziali	1
Totale	256

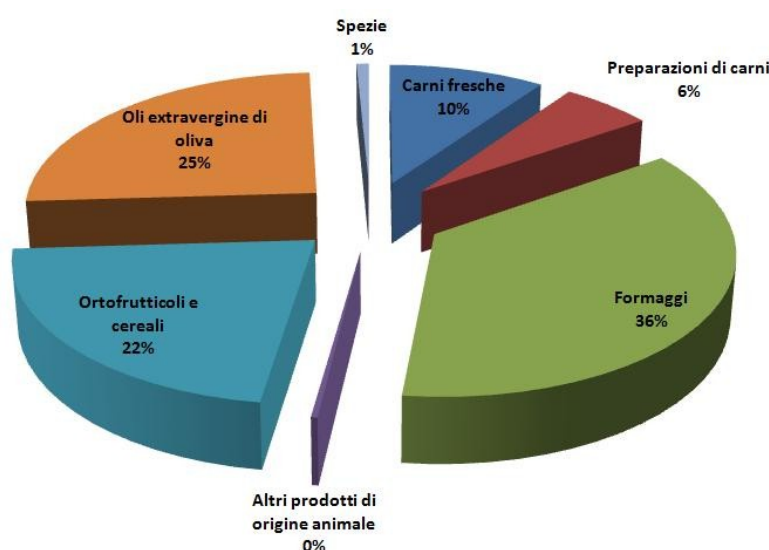


Figura 1.1 – Ripartizione degli operatori dei prodotti agroalimentari italiani DOP, IGP e STG per classe di prodotto [ISTAT, 2012].

Dati ISTAT del 2012 [ISTAT, 2012] riportano che gli operatori dei prodotti di qualità agroalimentari DOP, IGP e STG in Italia hanno raggiunto le 80231 unità e la produzione di prodotti tutelati predispone di un numero di allevamenti pari a 42804 unità e una superficie di terreni utilizzati pari a 159548 ettari.

Nonostante i prodotti tipici nel settore ortofrutticolo siano numericamente superiori rispetto ai prodotti di altri settori agroalimentari, si può vedere in Figura 1.1 che il settore dei formaggi coinvolge il maggiore numero di operatori, corrispondente a circa 29000 unità, ovvero il 36% del totale. Il 38% degli operatori nel settore formaggi (circa 9000 unità) è impegnato nella produzione di formaggi tutelati della Sardegna (Figura 1.2).

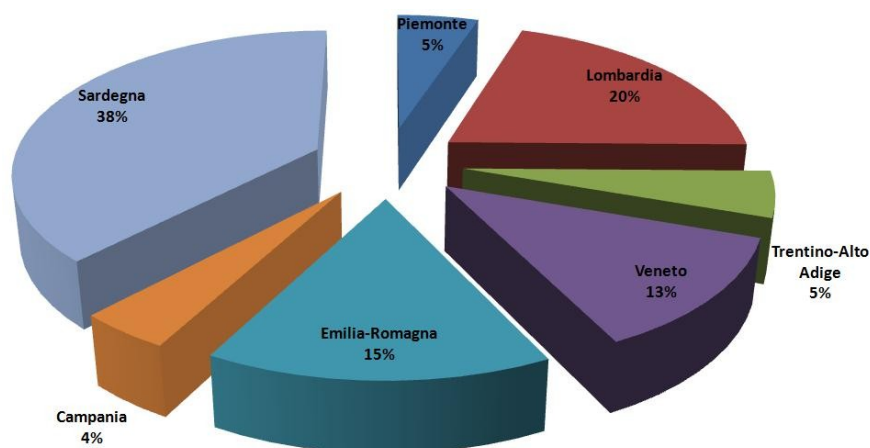


Figura 1.2 – Ripartizione degli operatori dei formaggi DOP, IGP e STG per regioni italiane più attive nel settore [ISTAT, 2012].

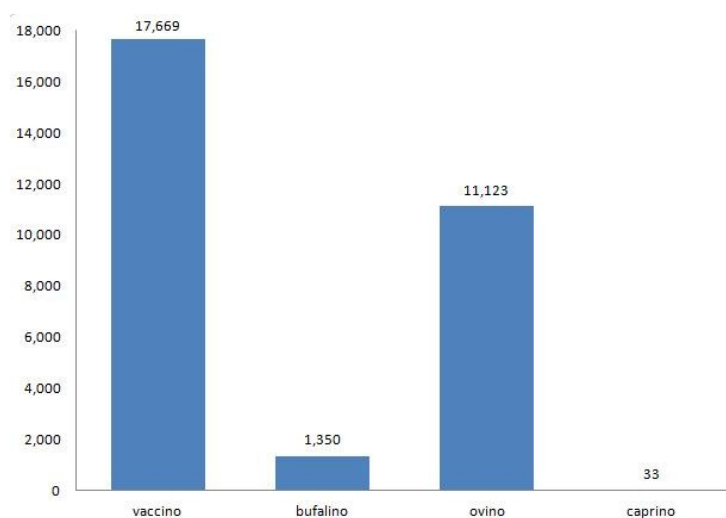


Figura 1.3 – Ripartizione degli allevamenti di specie lattifere coinvolte nella produzione di formaggi italiani tutelati [ISTAT, 2012].

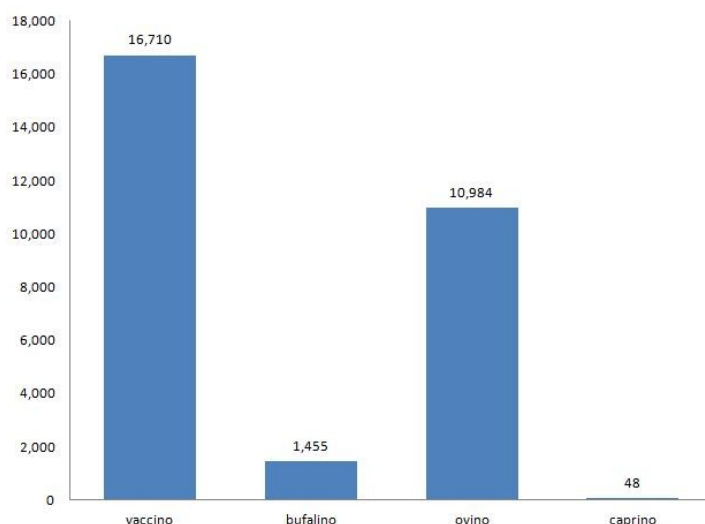


Figura 1.4 – Ripartizione degli operatori coinvolti nella produzione di formaggi italiani tutelati in base al tipo di latte utilizzato [ISTAT, 2012].

Negli istogrammi in Figura 1.3 e Figura 1.4 si può notare che il latte vaccino ed ovino sono i più utilizzati nella produzione dei formaggi italiani tutelati, coinvolgendo un numero consistente di allevamenti e di operatori. Considerato che, in Italia, circa il 94% del latte prodotto è di derivazione bovina mentre solo il 4% è di derivazione ovina, il contributo del latte ovino alla produzione di formaggi tutelati appare essere molto più consistente rispetto a quello del latte bovino.

In Sardegna vengono attualmente prodotte sette denominazioni di origine, sei DOP e una IGP: il Fiore sardo DOP, il Pecorino sardo DOP, il Pecorino romano DOP, l'Olio extravergine di oliva DOP Sardegna, il Carciofo spinoso di Sardegna DOP, lo Zafferano di Sardegna DOP e l'Agnello di Sardegna IGP. Ben tre prodotti DOP sono riconducibili al comparto ovino, a dimostrazione che l'industria del latte e dei suoi derivati è di capitale importanza in Sardegna. La Sardegna ospita infatti circa 10000 allevamenti, che rappresentano circa un terzo del totale nazionale delle strutture per animali. La produzione di latte ovino e dei suoi derivati è una delle principali attività presenti in Sardegna, e presenta un sistema economico solido per l'Isola e per i singoli territori.

1.2 Il sistema del latte ovino

Negli ultimi decenni il sistema di allevamento ovino mondiale è andato sempre più sviluppandosi; la presenza di numerose razze, e quindi la grande diversità tra le popolazioni, ha permesso di ottenere caratteristiche produttive differenti. La variabilità dei sistemi di allevamento e la produzione di latte è però molto ampia e spesso (ad esempio nei paesi meno sviluppati) la resa individuale di latte è molto bassa e la tecnica di allevamento è ancora rudimentale. Inoltre, la consistenza di un gregge è molto differente da zona a zona anche se, nella maggior parte dei casi, l'allevamento è a conduzione familiare. Nei paesi più sviluppati, in cui la qualità del latte prodotto assume maggiore importanza, le dimensioni tipiche di un gregge ovino sono generalmente comprese tra i 100 e i 300 capi.

Le condizioni economiche di produzione, il marketing del prodotto latte nonché l'applicazione di regolamentazioni sanitarie a livello europeo (il Codex Alimentarius e le Direttive dell'Unione Europea) hanno determinato cambiamenti nel sistema di produzione di molti paesi, come la Francia, l'Italia e la Spagna. Il numero medio di capi per gregge è aumentato, le attrezzature sono migliorate, l'alimentazione degli animali è stata razionalizzata e di conseguenza le rese del latte ovino sono aumentate. In principio però, questo nuovo sistema di produzione ha portato ad una diminuzione del contenuto di grasso, proteine e dei solidi totali nel latte prodotto, compensata parzialmente dall'implementazione con programmi di selezione.

Solo alcune razze ovine, tra cui la razza Sarda, sono specializzate nella produzione di latte e sono state sottoposte ad un programma di selezione efficiente. La razza Sarda è preservata grazie anche alla tendenza del mercato, che richiede sempre più prodotti tipici locali.

Negli ultimi anni l'industria casearia del latte ovino ha considerevolmente sviluppato la sua capacità produttiva, incrementando l'industrializzazione e l'automazione dei processi. Oltretutto, i prodotti lattiero caseari ovini sono sempre più esportati a livello internazionale.

1.2.1 Regolamenti e parametri europei

I criteri di qualità igienica e batteriologica del latte ovino sono delineati nelle Direttive Europee 92/46 e 94/71 che regolano i vari aspetti della produzione e della trasformazione del latte di varie specie animali.

1.2.1.1 Sicurezza alimentare nel latte ovino

La conta batterica

La conta batterica si riferisce al numero di microrganismi aerobici che si sviluppano ad una temperatura di 30°C ed è espressa come numero di unità formanti colonia (CFU) per millilitro. Le Direttive Europee 92/46 e 94/71 ne hanno stabilito i limiti massimi ammissibili nel latte ovino (Tabella 1.2) con lo scopo di migliorarne la qualità igienica.

Tabella 1.2 – Limiti di legge per la conta batterica nel latte crudo e trattato termicamente [Pirisi, 2007].

Data di applicazione della norma	Per produzioni da latte trattato termicamente		Per produzioni da latte crudo	
	Gennaio 1995	Dicembre 1999	Gennaio 1995	Dicembre 1999
Conta batterica a 30°C (n/ml)	< 3.000.000	< 1.500.000	< 1.000.000	< 500.000

L'ottenimento di un latte ovino con una ridotta carica batterica è di difficile attuazione in alcuni casi, in quanto il livello di carica batterica è influenzato dal sistema di raccolta e di trasporto del latte (mungitrici, cisterne e veicoli di trasporto), ed in particolari condizioni (pulizia carente, mancanza di refrigerazione e trasporto in veicoli non adatti) può raggiungere livelli elevati.

La modernizzazione degli allevamenti, oltre ad alterare le loro caratteristiche strutturali, richiede un'adeguata formazione da parte dell'allevatore [Pirisi, 2007].

Conta delle cellule somatiche

Le maggiori modificazioni nella composizione del latte ovino sono causate dall'insorgenza delle mastiti. I normali livelli di cellule somatiche nel latte ovino non mastitico sono caratterizzati da una elevata variabilità. Essi sono particolarmente alti nel colostro e nel latte di fine lattazione ma possono essere influenzati da vari fattori quali l'età dell'animale, il suo livello di produzione, il suo stato di salute, etc.

Le cellule somatiche contenute nel latte si possono suddividere in tre categorie: cellule epiteliali, cellule sanguigne e particelle citoplasmatiche. I rapporti tra i tre tipi di cellule

somatiche variano durante il periodo di lattazione e sono influenzati dallo stato di salute dell'animale. Durante un episodio di mastite, le difese immunitarie della mammella vengono attivate, i leucociti polinucleati passano dal sangue alla ghiandola mammaria e il numero di cellule somatiche aumenta. I processi di sintesi e di filtrazione dei componenti del latte vengono alterati causandone sbilanciamenti nella composizione chimica finale del latte. La capacità di filtrazione della ghiandola mammaria viene stimolata, provocando un aumento della frazione proteica non caseinica solubile mentre, soprattutto negli ovini, la sintesi del lattosio viene inibita, riducendone il contenuto nel latte. Anche il pH risulta alterato, e presenta valori superiori nel latte mastitico rispetto al latte normale, riducendo la capacità di coagulazione del caglio e quindi la resa casearia [Pirisi, 2007].

Presenza di sostanze inibitorie

L'utilizzo di sostanze capaci di rallentare o inibire lo sviluppo della carica batterica totale determinano anche l'arresto della crescita dei batteri lattici con la conseguente alterazione dei processi di fermentazione nella produzione dei formaggi, costituendo inoltre un rischio per la salute del consumatore. Tra queste sostanze vi sono residui di antibiotici, farmaci, detergenti e disinfettanti. Il trattamento delle mastiti con antibiotici e sulfamidici è una delle maggiori cause della presenza di sostanze inibitorie nel latte. Il latte e suoi derivati destinati al consumo umano non devono contenere sostanze inibitorie, così come definito nei regolamenti di controllo del latte [Pirisi, 2007].

1.2.2 Qualità chimica e fisica del latte ovino

Il latte è un alimento naturalmente funzionale, il più completo e ricco tra quelli esistenti in natura, destinato ad un organismo in formazione che necessita di nutrizione e di protezione. Tutti i componenti del latte esercitano attività funzionale: acqua, calcio, proteine, grassi, zuccheri, vitamine e sali minerali. Durante i processi digestivi, le caseine si scindono a dare peptidi bioattivi mentre le proteine del siero possiedono tal quali caratteristiche funzionali. Le caseine sono composte da tutti gli amminoacidi essenziali e contengono una percentuale elevata di serina che lega il calcio evitandone la complessazione col fosforo, rendendolo così biodisponibile. I peptidi bioattivi derivati dalle caseine hanno diversi effetti benefici sulla salute dell'uomo, tra cui attività

modulatrice del sistema immunitario e attività anti-proliferatrice delle cellule tumorali [FitzGerald, 2003]. Nella frazione proteica sono inoltre compresi molti enzimi.

La frazione lipidica è costituita essenzialmente da trigliceridi, in cui predominano gli acidi grassi saturi, a lungo tacciati di essere responsabili dell'insorgenza di malattie cardiovascolari. Tale effetto è però attribuibile ai soli acidi a lunga catena in quanto gli acidi grassi a media e corta catena vengono metabolizzati velocemente nell'organismo [Parodi, 2004]. È noto che altre classi di acidi grassi, come gli acidi polinsaturi ed - in particolare - gli acidi omega-3, hanno effetti positivi sulla salute e portano alla formazione di un gruppo di sostanze dette eicosanoidi con effetti antiaterogenici e antinfiammatori [Calder, 2009]. Anche gli acidi linoleici coniugati (CLA) possiedono tali effetti oltre alla dimostrata attività anticancerogena sia a livello preventivo che terapeutico [Ip, 1991] [Liew, 1995] [Belury, 1996] [Thompson, 1997] [Ochoa, 2004]. Gli oligosaccaridi del latte sembra che esercitino invece effetto prebiotico, in quanto, non essendo assorbiti dall'organismo, vengono utilizzati per stimolare la flora intestinale [Boehm, 2003].

Durante la trasformazione in yogurt o in altri prodotti fermentati ad opera di microrganismi probiotici, i componenti del latte subiscono modificazioni a formare molecole con attività benefica per l'uomo, che hanno la capacità di prevenire infezioni intestinali (agendo sul ripristino della flora batterica) e rafforzare il sistema immunitario [Fonden, 2003].

1.2.3 La razza ovina Sarda

La pecora di razza Sarda ha origini autoctone e, grazie alle sue notevoli capacità di ambientamento, si è comunque diffusa in quasi tutta la penisola con maggiori insediamenti nel Lazio e nella Toscana. Nuclei sono altresì presenti in altri paesi del bacino mediterraneo ed in particolare in Tunisia. In Sardegna, la consistenza al 2013 è di oltre 4 milioni e mezzo di capi. Come si vede in Tabella 1.3 la pecora Sarda ha una preminente attitudine alla produzione del latte, con elevate percentuali medie di grasso (6,0%) e di proteine (5,3%).

La qualità del latte ovino sardo è cresciuta notevolmente nel corso del tempo. La carica microbica si è dimezzata dal 1990 ad oggi e - nell'ultimo decennio (Figura 1.5) - si osserva un ulteriore miglioramento, più vistoso negli ultimi 3 anni. Tali risultati sono stati ottenuti grazie agli interventi sul benessere animale e alla sensibilizzazione degli allevatori riguardo l'aspetto igienico-sanitario, cui ha fatto riscontro una remunerazione del latte in base ad aspetti qualitativi.

Tabella 1.3 – Produzione media della razza Sarda al netto del latte poppato dall'agnello (<http://www.ara.sardegna.it/it/specie-allevate-in-sardegna-ovini/index.aspx?m=53&did=413>; ultimo accesso 04/11/2013).

	Allevamento estensivo	Allevamento intensivo
Primipare	60 litri	130 litri in 100 giorni
Secondipare	90 litri	170 litri in 180 giorni
Pluripare	100 litri	180 litri in 180 giorni

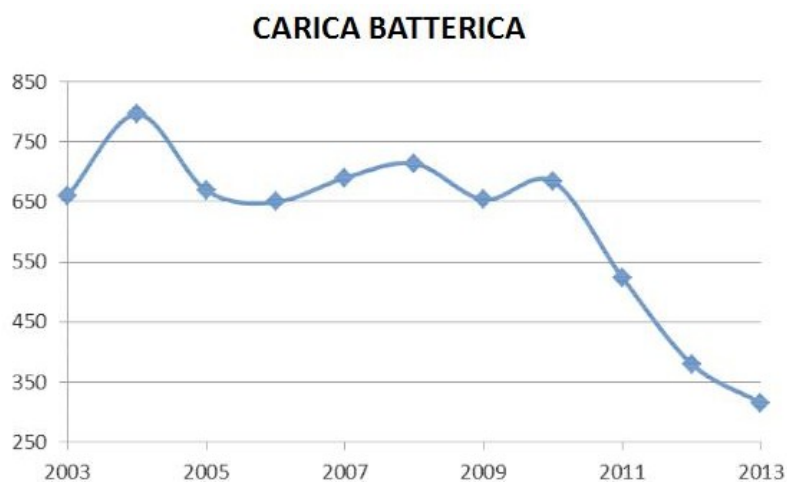


Figura 1.5 - Andamento della carica batterica nel latte ovino di razza Sarda (dati espressi come medie geometriche) [Contu, 2013].

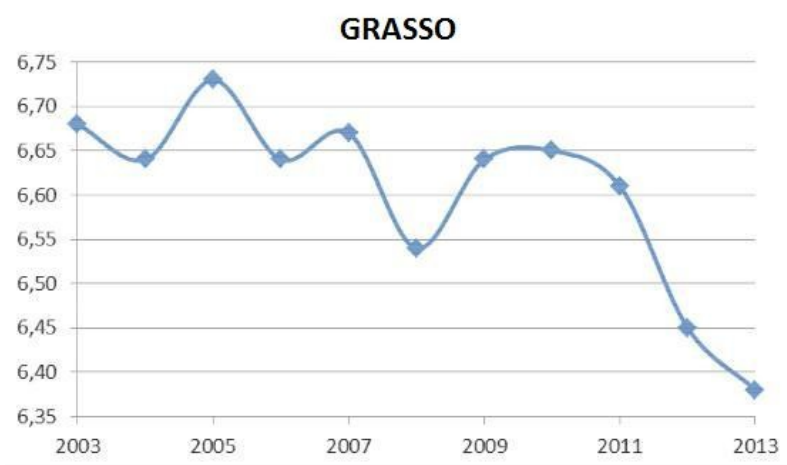


Figura 1.6 – Andamento del contenuto di grasso nel latte ovino di razza Sarda (valori percentuali medi) [Contu, 2013].

L'andamento del grasso (Figura 1.6) subisce oscillazioni annue legate all'andamento climatico ed alla composizione quali-quantitativa delle razioni alimentari. Secondo Contu [Contu, 2013] “un'altra fonte di variazione importante, che potrebbe aver determinato soprattutto la riduzione del 2012-2013 rispetto al 2011, potrebbe consistere nel campionamento che, grazie alla fornitura, da parte della RAS (Regione Autonoma della Sardegna) a tutti i caseifici sardi, di nuovi lattoprelevatori automatici, consente una preparazione di campioni più omogenei e rappresentativi dell'intera massa”.

Nel 2012 il contenuto medio di grasso è del 6,45% mentre il dato provvisorio del 2013 è pari a 6,38%. Anche il contenuto proteico del latte appare in diminuzione negli anni (Figura 1.7). In Figura 1.8 si vede invece che il contenuto percentuale di caseine sul totale delle proteine è in aumento col tempo.

In Figura 1.9 è mostrato l'andamento della conta delle cellule somatiche nel latte ovino di razza Sarda col passare degli anni. Si può vedere una diminuzione progressiva che risulta ancora più marcata nelle ultime due annate, probabilmente dovuta all'individuazione di un valore soglia (1.500.000 cellule per ml) stabilito nel 2006 che ha dato agli allevatori un riferimento sul limite ammissibile (misura 215 - "Pagamenti per il benessere degli animali negli allevamenti ovini-caprini da latte").

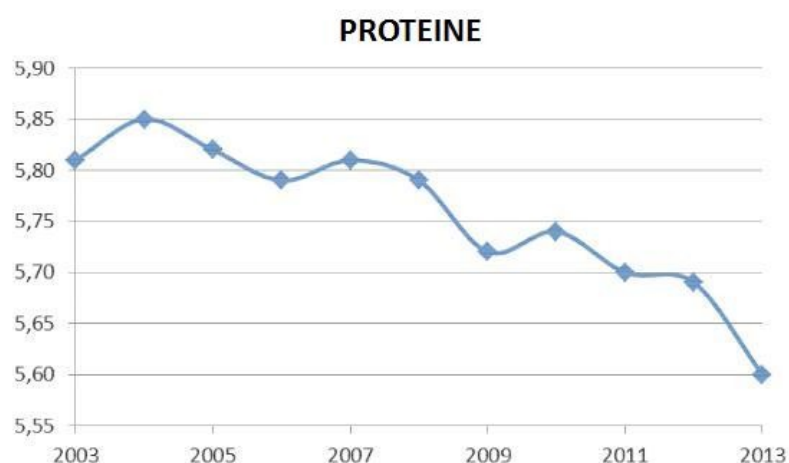


Figura 1.7 – Andamento del contenuto di proteine nel latte ovino di razza Sarda (valori percentuali medi) [Contu, 2013].

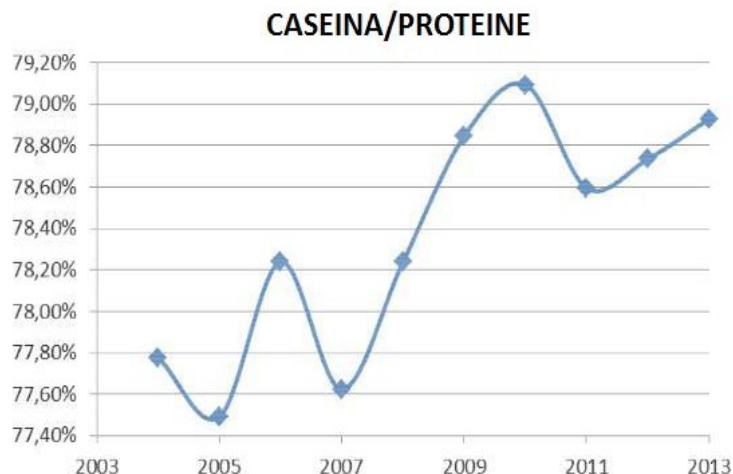


Figura 1.8 - Andamento del contenuto percentuale di caseine sul totale delle proteine nel latte ovino di razza Sarda [Contu, 2013].

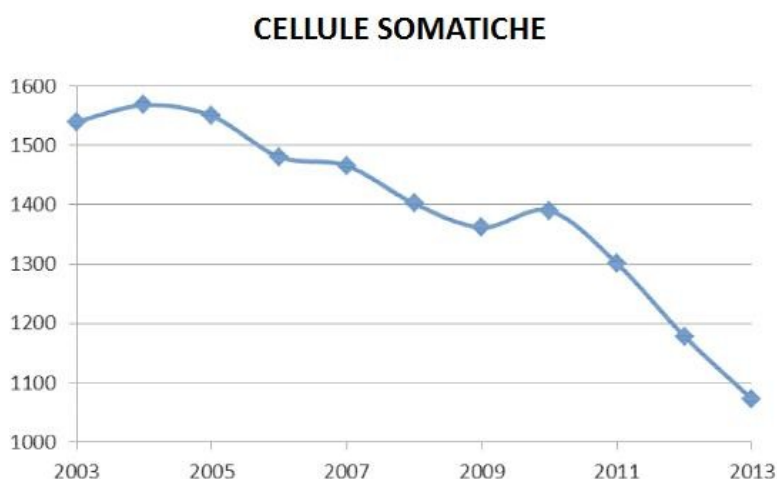


Figura 1.9 – Andamento della conta di cellule somatiche nel latte ovino di razza Sarda (dati espressi come medie geometriche) [Contu, 2013].

1.2.4 Il sistema di pagamento del latte ovino

Nel corso degli ultimi anni la produzione di latte ovino sul territorio nazionale ha subito un calo. Ciò è imputabile alla realtà produttiva sarda, che rappresenta quasi i 2/3 del totale nazionale, che sta attualmente attraversando una crisi di redditività. A partire dal 2004, un surplus produttivo, unito ad una minore richiesta di prodotti caseari sardi all'estero, ha determinato un abbattimento del prezzo unitario del latte pagato dagli industriali, pregiudicando il già debole rapporto costi/ricavi per gli allevatori della regione. Le aziende

con capi ovini si sono ridotte del 4%, di cui ben il 17% è localizzato in Sardegna. Le aziende rimanenti hanno subito un ridimensionamento della consistenza del bestiame e la produzione ne ha subito le conseguenze con progressiva erosione dei margini di redditività del comparto ovino.

Gli allevatori sardi, così come quelli siculi, hanno subito la maggiore penalizzazione economica rispetto ad altri operatori nazionali: nel 2000 il prezzo del latte ovino in Sardegna si attestava mediamente a 0,67 €/l contro i 0,76 €/l in Toscana, con un divario progressivo negli anni ed una differenza massima del 42% nel 2010 (0,62 €/l del latte sardo contro 0,89 €/l del latte toscano) [ARA, 2011]. Nel 2012 la differenza si è lievemente assottigliata e, mentre il latte toscano non ha subito variazioni di prezzo, il latte ovino sardo è stato pagato mediamente 0,68 €/l [ISMEA, 2012]. Si registra inoltre un notevole gap dei risultati tra i bilanci delle aziende ovine sarde e toscane, con un calo del fatturato più consistente per gli allevatori sardi, unito ad una ridotta efficienza del capitale investito ed un maggiore indebitamento. Ciò nonostante, il sistema di produzione di latte ovino e dei suoi derivati rimane una delle principali attività presenti in Sardegna. Inoltre, dato che l'utilizzo principale del latte ovino risiede nella produzione di formaggi, la valutazione della sua qualità è di importanza economica fondamentale.

In passato il sistema di pagamento del latte ovino era basato esclusivamente sulla quantità di latte prodotto, con prezzi dettati dalla contrattazione tra l'allevatore ed il caseificio. Col tempo sono stati introdotti dei sistemi di pagamento in base alla qualità, ovvero in base al contenuto di proteine e grasso, della conta batterica e delle cellule somatiche, ma tuttora non esiste un sistema di pagamento regolato da norme regionali o nazionali, tant'è che, ad oggi, esistono ancora caseifici che effettuano il pagamento esclusivamente in base alla quantità di latte prodotta.

Il pagamento in base alla qualità dovrebbe avere come risultato un vantaggio in termini economici per gli allevatori che potrebbero, in questo modo, ottenere una remunerazione più equa e consistente rispetto al pagamento a volume.

Un sistema di pagamento del latte in base alla qualità dovrebbe richiedere due punti fondamentali:

- stabilire dei reali incentivi, ovvero un beneficio in termini economici quando si superano dei valori soglia prestabiliti per i vari parametri di qualità del latte;

- creare un sistema di assistenza per gli allevatori che fornisca soluzioni e consigli tecnici che siano compatibili con il loro sistema di produzione per migliorare qualitativamente la produzione del latte.

Il metodo di pagamento a qualità attualmente più utilizzato si basa sul sistema bonus/penalità, a seconda del valore assunto nel latte di diversi parametri quali il contenuto totale di grasso e proteine, la conta batterica e la conta delle cellule somatiche. Al prezzo base (contrattato tra l'allevatore ed il caseificio) corrisponde un latte di qualità standard e su di esso vengono applicati bonus o penalità in relazione ad ogni parametro misurato, secondo una griglia che indica diversi importi a seconda della fascia di concentrazione in cui rientrano i diversi parametri del latte analizzato.

In Tabella 1.4 è mostrato un esempio di sistema di pagamento del latte ovino proposto in Sardegna. Come esempio di calcolo, un latte con un contenuto di proteine pari a 5,54% vedrà un decremento di prezzo pari a 0,048 euro in quanto la differenza tra il contenuto di proteine di tale latte ed il requisito minimo (range in cui il prezzo rimane invariato) è pari a $(5,54\% - 5,56\%) = -0,02\%$. Quindi, dato che un decremento proteico dello 0,01% fornisce una diminuzione di prezzo pari a 0,024 euro per il range di concentrazione in cui cade il valore proteico misurato, il decremento di prezzo totale per le proteine sarà pari a 0,048 euro. Se la percentuale di proteine risulta invece essere pari a 4,68, al prezzo base di partenza si decurtano 2,066 centesimi di euro (penalità base per il range considerato) ed ulteriori 0,008 euro dovuti al decremento dello 0,02%.

Il metodo proposto in Sardegna [Pirisi, 2007] è spesso utilizzato dai diversi caseifici presenti nella regione come esempio per il pagamento del latte a qualità ma, non essendo una proposta normata, ogni caseificio effettua modifiche soggettive ai prezzi indicati.

Un secondo metodo di pagamento a qualità è invece basato sulla resa in caseificazione per la produzione di formaggio Pecorino Romano DOP, scelta dettata dal fatto che la maggior parte del latte ovino sardo viene destinato per la produzione di questo formaggio. La resa in caseificazione è espressa come peso di formaggio ottenuto a partire da 100 Kg di latte alle 24 ore dalla fabbricazione e dipende dal tenore in proteina e in grasso del latte. La retta di regressione costruita [Pirisi, 2002] mostra un coefficiente di determinazione R^2 pari a 0,88:

$$\text{resa} = 1,31 \times (\text{tenore in grasso}) + 1,58 \times (\text{tenore in proteina})$$

In base all'equazione ottenuta per il rendimento caseario, è stato proposto un sistema di pagamento del latte ovino che presenta bonus e penalità legate alla resa di caseificazione a partire da una condizione standard (franchigia) che stabilisce il prezzo base. Il tenore di proteina e di grasso è quindi valutato in base al loro rapporto nel latte oltre che al loro aspetto quantitativo. Questo sistema di pagamento è attualmente in uso in alcuni caseifici sardi.

L'organizzazione di un sistema di pagamento del latte in base alla qualità richiede però alcuni prerequisiti: un laboratorio qualificato dove condurre le analisi, la logistica per raccogliere campioni di latte e un'organizzazione interprofessionale che garantisca l'indipendenza delle analisi. Inoltre, dovrebbe essere introdotto un servizio tecnico che assista gli allevatori nel miglioramento della qualità del latte.

Tabella 1.4 – Sistema di pagamento del latte ovino in base alla qualità proposto in Sardegna [Pirisi, 2007].

	Bonus e penalità (stadio iniziale) Euro cents	Variazione	Bonus e penalità (range) Euro
Proteine^a (%)			
< 4,70	-2,066	-0,01%	-0,004
4,71-5,55	Nessuno	-0,01%	-0,024
5,56-5,75	Nessuno	Nessuna	Nessuno
5,76-6,79	Nessuno	+0,01%	+0,021
> 6,80	+2,169	-	-
Grasso (%)			
< 5,40	-1,549	-0,01%	-0,003
5,41-6,20	Nessuno	-0,01%	-0,019
6,21-6,60	Nessuno	Nessuna	Nessuno
6,61-7,59	Nessuno	+0,01%	+0,019
> 7,60	+1,911	-	-
Conta batterica (CFU/ml x 10³)			
< 500	+0,516	-100	+0,103
501-1000	Nessuno	-100	+0,103
1001-1500	Nessuno	Nessuna	Nessuno
1501-3000	Nessuno	+100	-0,052
> 3000	-1,549	-	-
Cellule somatiche (cells/ml x 10³)			
< 500	+0,516	-100	+0,103
501-1000	+0,258	-100	+0,052
1001-1300	Nessuno	-100	+0,086
1301-1700	Nessuno	Nessuna	Nessuno
1701-2500	Nessuno	100	-0,065
2501-3000	-0,516	100	-0,052
> 3000	-1,033	-	-

^a Il bonus per le proteine è cancellato se la conta delle cellule somatiche è superiore a 3000000.

1.2.5 I controlli di qualità regionali: l'ARA Sardegna

I meccanismi di controllo di qualità sono attualmente gestiti da laboratori inter-professionali che raccolgono e analizzano sistematicamente i campioni di latte e sono particolarmente organizzati in alcune zone dell'Italia. In Sardegna è presente l'ARA Sardegna (Associazione Regionale Allevatori della Sardegna) che fornisce servizi zootecnici a circa 5000 aziende e associazioni agricole presenti sull'isola. L'Associazione si avvale della collaborazione di circa 250 professionisti tra veterinari e zootecnici che si occupano dei controlli sanitari dei bovini, degli ovini e dei caprini.

L'organizzazione centrale coordina le attività delle quattro Associazioni Provinciali Allevatori (APA) della Sardegna, che gestiscono i Controlli Funzionali ed i Libri Genealogici delle razze in purezza. Inoltre, su incarico della RAS (Regione Autonoma della Sardegna), in concerto con le APA, gestisce il Programma Regionale di Identificazione Elettronica degli animali iscritti ai Libri Genealogici.

L'ARA svolge la sua attività in tutto il territorio regionale al fine di:

- esercitare funzioni tecniche ed economiche nel settore zootecnico per promuovere ed attuare tutte le iniziative di portata regionale volte all'incremento ed al miglioramento delle produzioni animali ed alla valorizzazione zootecnica;
- promuovere ed incoraggiare studi e ricerche diretti a risolvere particolari problemi tecnici, di assistenza tecnica ed economici in collaborazione e d'intesa con Organismi pubblici, Istituti di sperimentazione e ricerca, Università, ecc.;
- promuovere, d'intesa ed in collaborazione con le Autorità, Istituti sanitari competenti, azioni di profilassi e di lotta contro le malattie infettive e diffuse del bestiame;
- promuovere ed attuare ogni altra iniziativa ritenuta utile al potenziamento della zootecnia regionale;
- tenere i registri anagrafici delle razze bovine Sarde, Sardo-Bruna e Sardo-Modicana.

L'ARA Sardegna ha a disposizione un Laboratorio per l'analisi del latte, con sede a Nuraxinieddu (OR), che nel 1997 ha ottenuto l'accreditamento SINAL in conformità alle prescrizioni della norma EN 45001 e, successivamente della ISO IEC 17025:2005, e ai criteri applicabili delle norme della serie ISO 9000:2000.

Il Laboratorio è parte integrante del sistema di assistenza tecnica e, con il Programma “Qualità Latte”, istituito alla fine degli anni '80 del secolo scorso, rappresenta uno strumento indispensabile per l'attività dell'équipe zootecnico-veterinario. Il laboratorio svolge, per conto delle Associazioni Provinciali, le analisi su campioni individuali di latte, relative ai controlli funzionali, previste dai Libri Genealogici, rappresentando un punto di riferimento essenziale per i caseifici, gli enti universitari e di ricerca al fine di consentire loro verifiche mirate o particolari programmi sperimentali. Quasi 5000 aziende del settore ovi-caprino (ed un numero di capi superiore al milione di unità) vengono assistite nel Programma “Qualità Latte” le cui attività, volte al miglioramento della qualità del latte, comprendono [\(<http://www.ara.sardegna.it/it/attivita/assistenza-tecnica-p-o-a-t/index.aspx?m=53&did=236>; ultimo accesso 04/11/2013\)](http://www.ara.sardegna.it/it/attivita/assistenza-tecnica-p-o-a-t/index.aspx?m=53&did=236):

- assistenza per il controllo dei parametri quantitativi, merceologici, microbiologici ed igienico-sanitari del latte;
- consulenza sul corretto funzionamento degli impianti di refrigerazione;
- consulenza con i tecnici delle Associazioni Provinciali Allevatori (APA) sul funzionamento degli impianti di mungitura;
- programma profilassi mastiti con individuazione degli agenti patogeni, relativo antibiogramma e conseguente terapia;
- collaborazione con i caseifici nella predisposizione e simulazione dei modelli di griglie di pagamento del latte in base alla qualità.

Nel dettaglio, il Programma si è sviluppato seguendo direttrici precise ([\(<http://www.ara.sardegna.it/index.aspx?m=81&f=3&idf=287>; ultimo accesso 04/11/2013\)](http://www.ara.sardegna.it/index.aspx?m=81&f=3&idf=287)):

- attivazione e standardizzazione delle modalità operative relative al sistema di accurato prelievo di campioni del latte prodotto nelle aziende aderenti al Piano;
- attuazione di un sistema di trasporto e di conservazione dei campioni prelevati garantendo il rispetto la catena del freddo ed evitando contaminazioni esterne;
- rendere operativo un sistema di analisi rapide (entro 36 ore dal prelievo) dei campioni, presso il laboratorio ARA, garantendo l'immediato trasferimento dei dati analitici agli attori di filiera (allevatori, l'industria di trasformazione, il sistema sanitario di sorveglianza e l'assistenza tecnica). Inoltre, i dati forniti debbono avere il requisito di essere immediatamente interpretabili dall'utente e/o dal tecnico tramite formule grafiche e cromatiche di sicuro impatto.

- consolidamento di un protocollo di intervento in azienda degli zootecnici e dei veterinari ARA, in collaborazione e integrazione con i tecnici dell'Agenzia Laore Sardegna, per l'assistenza agli impianti di mungitura e ai refrigeratori, onde attuare gli interventi ritenuti opportuni (ad es. riformulazione delle razioni, consulenza sugli impianti di mungitura, ecc.) e per l'eventuale prelievo di campioni (ad es. foraggi, mangimi, ecc.);
- attivazione di un trasferimento rapido di informazioni al sistema sanitario (Istituto Zooprofilattico Sperimentale) soprattutto per i problemi legati alla diagnosi e profilassi delle mastiti subcliniche;
- standardizzazione di un sistema di valutazione continua dell'efficacia delle azioni ed interventi allo scopo di adottare protocolli dinamici sempre più rispondenti alle esigenze operative.

Il Programma ha attualmente raggiunto molti degli scopi prefissati ed è tuttora in via di sviluppo.

In merito ai controlli di qualità, l'ARA funge quindi da soggetto *super partes* tra gli allevatori ed i caseifici, effettuando i controlli di qualità sul latte, i cui risultati determineranno il pagamento dello stesso in base ai sistemi descritti nel paragrafo 1.2.4 applicati dai singoli caseifici. L'ARA stabilisce le date in cui effettuare controlli a campione in diversi allevamenti da cui vengono prelevati campioni di latte che verranno analizzati nel Laboratorio per l'Analisi del Latte di Nuraxinieddu.

In Tabella 1.5 sono elencate le attuali determinazioni igienico-sanitarie e quantitative effettuate dal laboratorio del latte dell'ARA sul latte ovino crudo. I parametri individuati per i controlli sull'aspetto igienico-sanitario riguardano la conta delle cellule somatiche, della carica batterica totale, del pH e delle sostanze inibitorie. Per quanto riguarda i componenti del latte, le determinazioni effettuate riguardano il contenuto totale di grasso, di lattosio, di proteine e di caseine. Il valore dell'urea consente invece di valutare l'adeguatezza dell'apporto proteico della razione alimentare degli animali ed il corretto rapporto tra energia e proteine della razione in rapporto alla quantità di latte prodotto.

Alcune determinazioni vengono effettuate tramite tecniche che utilizzano la spettroscopia infrarossa, metodo che assicura rapidità di analisi e costi ridotti, per garantirne i risultati entro le 36 ore indicate nel Programma; il Laboratorio del latte ARA utilizza uno spettrometro Milkoscan FT6000 della FOSS, che registra uno spettro FT (in Trasformata di

Fourier) nel medio-infrarosso (tra i 926 ed i 5012 cm^{-1}). La metodica utilizzata per la determinazione di grasso, proteine e lattosio totali è la FIL-IDF 141C:2000.

Tabella 1.5 – Determinazioni dei parametri del latte ovino crudo effettuate dal Laboratorio ARA.

Descrizione	Unità di misura	Metodo di prova	Accreditamento ACCREDIA	Denominazione	Utilizzato dal
Grasso	g/100ml	Infrarosso	Prova accreditata	FIL-IDF 141C:2000	15.01.03
Proteine	g/100ml	Infrarosso	Prova accreditata	FIL-IDF 141C:2000	15.01.03
Lattosio	g/100ml	Infrarosso	Prova accreditata	FIL-IDF 141C:2000	15.01.03
Cellule Somatiche	CELL/ μ l	Citofluorimetria di flusso	Prova accreditata	FIL-IDF 148-2:2006 ISO13366-2	22.06.10
Carica Batterica	UFC/ μ l	Citofluorimetria di flusso	Prova accreditata	P03 rev. 13:2013	18.06.13
Punto Crioscopico	$^{\circ}$ H	Crioscopio a termistori	Prova non accreditata	P04 rev 7:2008	29.02.08
pH	pH	pHmetro	Prova accreditata	P05 rev 12:2011	7.12.11
Urea	mg/dl	Enzimatico - pHmetria differenziale	Prova accreditata	FIL-IDF 195:2004 ISO 14637	8.10.04
Urea	mg/dl	Infrarosso	Prova non accreditata	Manuali Foss Italia	5.04.04
Caseina	g/100ml	Infrarosso	Prova non accreditata	Manuali Foss Italia	1.03.08
Ricerca latte vaccino o caprino nel latte ovino	positivo o negativo	Immunodiffusione radiale	Prova accreditata	P07 rev 12:2013	26.06.13
Inibenti	positivo, negativo o tracce	Microbiologico	Prova accreditata	P08 rev 13:2013	4.09.13

I risultati delle analisi vengono attualmente comunicati entro le 24 ore, e sulla base del dato ricevuto i caseifici, che nel tempo hanno ottenuto uno storico della qualità del latte del singolo allevatore, applicano il sistema di pagamento in base alla qualità. I costi delle analisi sono a carico del caseificio richiedente, che però detrae a tal fine parte dal compenso dovuto all'allevatore. Quest'ultimo ha quindi in carico un ulteriore onere, che va a gravare sui costi di produzione del latte. È quindi necessario operare affinché i costi delle analisi siano minimi. Le tecniche analitiche online impieganti spettroscopia infrarossa a Trasformata di Fourier sono quelle che appaiono più adatte per i controlli di qualità del latte in quanto veloci ed economiche.

L'introduzione di altri parametri di qualità, quali il contenuto di acidi grassi nel latte, permetterebbe di incentivare ancora di più la produzione di un latte di qualità, andando ad aumentare il possibile guadagno per l'allevatore. Il pagamento in base al contenuto di acidi grassi potrebbe essere effettuato tramite una griglia, così come già avviene per altri parametri, oppure tramite remunerazione forfettaria. Inoltre, in tal modo sarebbe possibile oggettivare prodotti di elevato livello qualitativo, adatti quindi a lavorazioni particolari e non a caseificazioni di massa (ad esempio, latte con alto contenuto di acido rumenico o acido α -linolenico potrebbe esser orientato verso la produzione di formaggi DOP con caratteristiche funzionali).

L'obiettivo principale del presente lavoro di tesi è quello di proporre la spettroscopia medio-infrarossa a Trasformata di Fourier come veloce metodo predittivo di analisi anche per la determinazione di analiti minoritari del latte ovino quali le diverse categorie di acidi grassi (saturi, insaturi e polinsaturi). Ove l'obiettivo risulti conseguibile, la costruzione di modelli di calibrazione atti alla predizione del profilo acidico del latte ovino permetterà di poter estendere, con costi ragionevoli per le aziende, il sistema di remunerazione a qualità anche ai costituenti non maggioritari del latte ovino, con auspicabile vantaggio economico per gli operatori del settore.

2 Il latte

2.1 Introduzione

Il latte è il liquido secreto dalle ghiandole mammarie delle femmine dei mammiferi (più di 4000 specie) dopo la nascita del piccolo, ed è spesso l'unica risorsa di cibo per i giovani mammiferi durante il periodo più critico della loro esistenza. Il ruolo del latte è di nutrire, tramite l'apporto di proteine, vitamine, lipidi, zuccheri e sali minerali, e di provvedere allo sviluppo del sistema immunitario e delle funzioni di tipo fisiologico grazie alla presenza di proteine e peptidi che includono immunoglobuline, enzimi e agenti antibatterici. Per l'uomo la funzione nutritiva del latte continua fino all'età adulta, essendone un'importante componente del regime alimentare [Fox, 1998].

La prima attività produttiva dell'uomo, la pastorizia, è nata proprio per poter avere a disposizione sufficiente quantità di latte per l'alimentazione quotidiana. Le specie animali individuate come idonee alla sua produzione furono numerose, permettendo quindi l'allevamento di specie animali lattifere in diverse aree geografiche con diverse condizioni pedo-climatiche che, indirettamente, hanno determinato una diversa scelta delle condizioni di conservazione del latte. Oggigiorno l'interesse economico ruota attorno a poche specie animali lattifere, in prevalenza vacca, bufala, pecora e capra, dal cui latte si ottiene una vasta gamma di derivati alimentari, ognuno con le proprie caratteristiche sensoriali, nutrizionali e di conservabilità [Corradini, 1995].

Il latte è legalmente definito come *prodotto ottenuto dalla mungitura regolare, ininterrotta e completa di animali in buono stato di salute e nutrizione* [R.D. 9/5/29 n. 994 e successive modifiche]. La definizione, valida per tutte le specie animali lattifere, riprende le disposizioni stabilite dal Primo Congresso Internazionale per la Repressione delle Frodi Alimentari (Ginevra, 1908). Il latte, inoltre, non deve essere maleodorante, non deve presentare colorazioni anomale e non deve contenere specie patogene. Non è quindi adatto al consumo umano il latte proveniente da animali malnutriti o malati ed il latte contenente colostro, liquido che si differenzia dal latte soprattutto per le sue parti proteiche e saline, ottenuto dalla mungitura effettuata nei sette giorni successivi al parto [Alais, 2010].

2.2 Composizione del latte

Il latte è un liquido uniforme, bianco e opaco; l'opacità ne suggerisce una disomogeneità dei costituenti. I diversi elementi strutturali che lo compongono stabiliscono rapporti di

emulsione, dispersione o soluzione colloidale, soluzione vera con la fase disperdente, l'acqua. L'acqua è infatti il composto maggioritario mentre gli altri costituenti principali sono grasso, lattosio e proteine (principalmente caseine e proteine del siero). In rapporto all'acqua, i globuli di grasso si trovano in stato di emulsione, le micelle caseiniche in dispersione pseudo colloidale, le proteine del siero in soluzione e le particelle lipoproteiche in dispersione colloidale [Corradini, 1995]. Il latte contiene anche piccole quantità di minerali e specifiche proteine del sangue. I costituenti in eterofase possono essere individuati tramite l'uso di un microscopio: il microscopio ottico ci permette di notare goccioline sferiche che si muovono nel liquido torbido definito come plasma, ma è il microscopio elettronico che ci permette di identificare aggregati proteici sferici (micelle e sub micelle caseiniche) e i globuli di grasso. La separazione della frazione lipidica e proteica ci permette di distinguere il siero, un liquido opalescente che contiene ancora una frazione corpuscolare, formata dall'associazione delle proteine globulari (dette proteine del siero) con particelle lipoproteiche. Nel siero, stavolta in soluzione vera, sono presenti gli altri componenti del latte: lattosio ed altri carboidrati, sali minerali, acidi organici ed inorganici, composti azotati ed elementi in tracce. I principali costituenti del latte possono essere separati ed isolati con relativa facilità; di conseguenza essi sono stati ampiamente studiati e le loro principali caratteristiche sono ben note.

Come in precedenza accennato, il latte viene prodotto da migliaia di mammiferi, terrestri e marini, che possono vivere in zone climatiche estremamente diverse. È quindi inevitabile che ci siano delle differenze nella composizione del loro latte. Il confronto di latte da mammiferi diversi deve tenere conto di molti fattori, quali la fisiologia dell'animale (ruminante o non ruminante), la frequenza dell'allattamento, i fattori di sviluppo e crescita e quelli ambientali. Tra le 4000 diverse specie di mammiferi, solo il latte di 180 di esse è stato analizzato e di queste solamente 50 hanno risultati che possono essere considerati statisticamente attendibili [Fox, 1998]. L'ultimo grande inventario è stato effettuato da Jenness e Sloan (1970) in cui è possibile notare che il latte di tutte le specie studiate contiene proteine e sali, mentre la concentrazione di importanti componenti spesso maggioritari, quali ad esempio i lipidi ed il lattosio, può subire importanti fluttuazioni. Ad esempio, il latte dei pinnipedi non contiene lattosio ma contiene elevate concentrazioni di lipidi, cui è affidato il ruolo di fonte primaria di energia. In linea generale si può affermare che il latte di una specie animale è tanto più ricco in sostanze azotate e sali quanto meno è completo lo sviluppo nell'utero del nascituro. Il tenore in lattosio segue l'andamento inverso ma è legato allo sviluppo della massa cerebrale (il latte umano è infatti ricco di

lattosio). Le sostanze grasse sono invece in percentuale maggiore nel latte dei mammiferi delle regioni fredde e degli oceani [Alais, 2010]. Oltre quello umano il latte più studiato appartiene alle specie bovina, ovina, caprina e bufalina ed ha costituenti in quantità percentuale relativamente simile se paragonato a quello di altre specie. La composizione del latte varia quindi in base alla specie animale ma, all'interno della stessa specie, la sua variabilità è dovuta anche ad altri fattori quali il periodo dello stato di lattazione, l'alimentazione dell'animale ed il suo stato di salute.

Tabella 2.1 - Composizione (%) del latte di alcune specie [Fox, 1998].

Specie	Solidi totali	Grasso	Proteine	Lattosio
Donna	12,2	3,8	1,0	7,0
Vacca	12,7	3,7	3,4	4,8
Capra	12,3	4,5	2,9	4,1
Pecora	19,3	7,4	4,5	4,8
Cavallo	11,7	1,9	2,5	6,2
Coniglio	32,8	18,3	11,9	2,1
Bisonte	14,6	3,5	4,5	5,1
Orso polare	47,6	33,1	10,9	0,3
Foca grigia	67,7	53,1	11,2	0,7

In Tabella 2.1 è illustrata la composizione percentuale dei composti maggioritari del latte di alcune specie. Si nota che il latte di pecora è più ricco nei costituenti principali rispetto al latte di vacca e di capra, e ciò ne detta la sua importanza nutrizionale; esso è infatti una risorsa importante di lipidi e proteine. Il latte di pecora possiede inoltre elevate caratteristiche tecnologiche atte alla trasformazione casearia. In Tabella 2.2 è invece riportata la composizione media del latte di vacca.

Il latte è un prodotto alimentare facilmente alterabile; già a temperatura ambiente, se non omogeneizzato, si separa in tre parti: la crema, formata dallo strato di globuli di grasso che si raggruppano per effetto della gravità; la cagliata, formata dalla caseina coagulata a causa di attività microbiche, ed il siero, contenente composti solubili, che si separa in un secondo momento dalla cagliata. Nel latte non alterato, i costituenti presentano delle interdipendenze più o meno strette che vanno a perdersi quando essi vengono isolati; ne sono un esempio la caseina ed il fosfato di calcio o l'acqua che si lega alle proteine. Il latte è quindi un'eterofase metastabile che può essere compromessa da diverse azioni quali il riscaldamento, il raffreddamento e l'attività microbica.

Il riscaldamento riduce la ionizzazione del calcio in quanto determina l'aumento della stabilità del fosfocaseinato di calcio; il raffreddamento al contrario induce la solubilizzazione del fosfato di calcio micellare e l'aumento degli ioni calcio in soluzione, nonché la solubilizzazione della β -caseina. L'aumento della carica microbica favorisce invece la degradazione del lattosio con produzione di acido lattico determinando la flocculazione da parte di alcune proteine. L'applicazione di un trattamento al latte porta sempre ad un cambiamento dello stato di uno o più suoi componenti, ed esso può avere ripercussioni secondarie anche inattese sia per il latte che per i suoi derivati.

Tabella 2.2 - Latte di vacca: composizione media [Alais, 2010].

Composizione	grammi / litro	Stato fisico dei composti
Acqua	905	Acqua libera (solvente) + acqua legata (3,7%)
Glucidi: lattosio	49	Soluzione
Lipidi:	35	
Sostanze grasse propriamente dette	34	Emulsione di globuli di grasso (3-5 μ m)
Lecitina (fosfolipidi)	0,5	
Parte insaponificabile (steroli, caroteni, tocoferoli)	0,5	
Protidi:	34	Sospensione micellare di fosfocaseinato di calcio (0,08-0,12 μ m)
Caseina	27	
Proteine solubili (globulina, albumina)	5,5	Soluzione colloidale
Sostanze azotate non proteiche	1,5	Soluzione
Sali:	9	Soluzione o stato colloidale (P e Ca)
dell'acido citrico	2	(Sali di K, Ca, Na, Mg, ecc)
dell'acido fosforico (come P₂O₅)	2,6	
dell'acido cloridrico (come NaCl)	1,7	
Costituenti diversi (vitamine, enzimi, gas disciolti)	tracce	
Estratto secco (totale)	127	
Residuo magro	92	

2.3 Composizione chimica del latte

2.3.1 Acqua

L'acqua è il componente maggioritario del latte ed è presente in diverse percentuali a seconda della specie animale variando notevolmente tra il 32% per il latte della foca grigia e l'89% per quello del cavallo. Nel latte vaccino il contenuto di acqua è in media l'87% mentre nel latte ovino è dell'81%.

2.3.2 Lattosio

Il lattosio, disaccaride costituito da glucosio e galattosio (Figura 2.1), è il principale carboidrato del latte. È presente in due forme anomeriche, α e β , che sono in equilibrio tra di loro con una percentuale rispettivamente del 38% e del 62%. Altri zuccheri, come il glucosio o il fruttosio, sono contenuti solo in tracce.

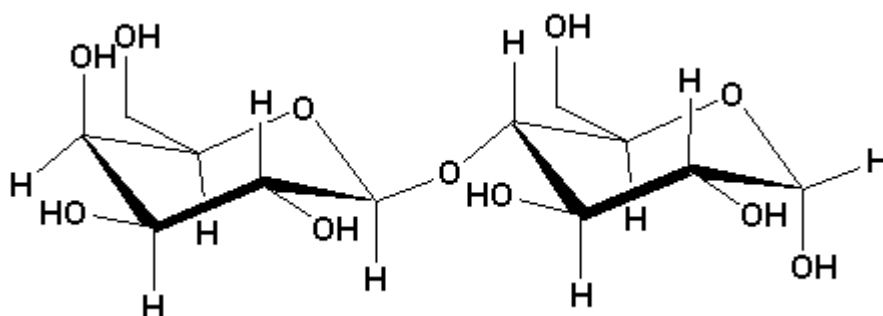


Figura 2.1 - Lattosio

La concentrazione del lattosio nel latte varia a seconda della specie animale (Tabella 2.3) anche se la variabilità è molto più bassa rispetto a quella di altri costituenti, quali ad esempio il grasso. Per quel che riguarda le specie ruminanti lattifere principali, il lattosio si trova in concentrazione pressoché simile, leggermente minore solo nel latte di capra. Per il latte di pecora la sua concentrazione è compresa tra 4,5 e 5,0 g/100 ml ed è molto simile alla concentrazione presente nel latte di vacca. Il lattosio è anche fonte di galattosio, necessario per la costituzione dei tessuti nervosi, e costituente di molte glicoproteine e molti glicolipidi, importanti costituenti delle membrane cellulari. Il lattosio, assieme agli ioni sodio, potassio e cloruro, gioca un ruolo importante nel mantenimento della pressione

osmotica del sistema mammario, per cui ogni suo aumento o decremento di concentrazione deve essere compensato da un aumento o decremento della concentrazione dei sali. Questa relazione spiega perché il tenore in lattosio sia inversamente proporzionale al contenuto in ceneri del latte.

Nel latte avvengono moltissimi processi fermentativi, essenziali nella produzione di molti derivati del latte, la maggior parte dei quali ha origine proprio dal lattosio. Di particolare importanza è la fermentazione lattica, nella quale il lattosio viene trasformato in acido lattico da batteri lattici, punto di partenza per la coagulazione acida tramite la quale viene prodotto lo yogurt [Corradini, 1995] [Fox, 1998].

Tabella 2.3 - Concentrazione del lattosio in diverse specie animali (g/100ml).

Specie	Lattosio	Specie	Lattosio	Specie	Lattosio
Foca	0,0	Cane	3,1	Gatto	4,8
Orso	0,4	Capra	4,1	Maiale	5,5
Delfino	0,6	Vacca	4,8	Cavallo	6,2
Coniglio	2,1	Pecora	4,8	Donna	7,0
Ratto	2,6	Bufalo	4,8	Asino	7,4

2.3.3 Proteine

La funzione delle proteine del latte è quella di fornire al nascituro sia gli amminoacidi essenziali richiesti per lo sviluppo dei tessuti a base proteica (es: muscoli), sia le proteine biologicamente attive come le immunoglobuline, gli ormoni proteici e le proteine metallo- e vitamine-leganti. I piccoli delle diverse specie hanno alla nascita diversi stadi di maturità, cui fanno riscontro diverse richieste nutritive e fisiologiche. Il contenuto proteico del latte è direttamente correlato alla velocità di crescita del piccolo e riflette le esigenze proteiche per la crescita, come si evince dalla Figura 2.2 in cui è mostrata la relazione tra la percentuale di calorie ottenute dalle proteine ed il tempo necessario affinché il neonato raggiunga il doppio del suo peso alla nascita [Bernhart, 1961].

In Tabella 2.4 è mostrata la concentrazione percentuale delle proteine nel latte di alcune specie animali. Oltre alla loro principale funzione di fonte alimentare, le proteine hanno una grande importanza nelle proprietà nutrizionali e strutturali degli alimenti a lunga conservazione ottenuti tramite modificazioni tecnologiche del latte. Un esempio è dato dalle caseine che, grazie alla loro alta stabilità al calore, permettono la produzione di diversi prodotti sterilizzati e disidratati. L'alto contenuto di proteine è quindi diventato

indice di qualità, determinando, insieme al grasso, il valore economico di un latte ricco di questi componenti quale quello ovino.

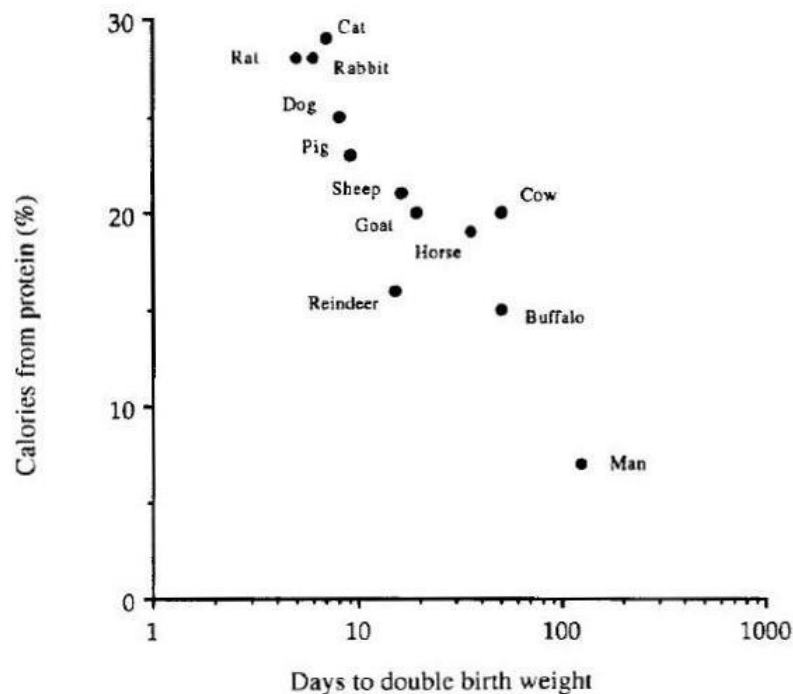


Figura 2.2 - Relazione tra le calorie ottenute dalle proteine ed tempo di crescita del neonato.

Le proteine del latte si dividono in due grandi famiglie a seconda delle loro proprietà chimico-fisiche; acidificando il latte delle specie mammifere ruminanti fino a pH 4,6, circa l'80% delle proteine precipita mentre il restante 20% rimane in soluzione nel siero. Il primo gruppo è formato da caseine mentre il secondo da proteine non caseiniche o più comunemente definite come proteine del siero. Entrambi i gruppi sono eterogenei, formati quindi da diverse proteine costituite da numerosi singoli amminoacidi, e presentano proprietà molto diverse che vengono sfruttate in campo industriale e tecnologico. Le principali differenze sono di seguito riassunte:

- La proprietà delle caseine di precipitare a pH 4,6 (loro punto isoelettrico) viene sfruttata nella produzione industriale dei formaggi. Le proteine del siero sono invece solubili in queste condizioni.
- La chimosina ed altre proteasi (enzimi presenti nei diversi tipi di caglio) determinano cambiamenti specifici nelle caseine che hanno come risultato finale la loro coagulazione in presenza di ioni calcio. Il processo di coagulazione è alla base

della tecnologia di produzione delle diverse varietà di formaggio. Le proteine del siero non subiscono invece alcun cambiamento per azione di tali enzimi.

- Le caseine sono molto stabili al calore; mantenendone il suo pH normale (6,7), il latte può essere portato ad una temperatura di 100°C per un periodo di 24 ore e, ancora, ad una temperatura di 140°C per 20 minuti senza che le caseine coagulino. Il raggiungimento di elevate temperature comporta comunque diverse trasformazioni nel latte, come il cambiamento del bilancio dei sali e la formazione di acidi a partire dal lattosio che, a causa della diminuzione del pH, può indurre la precipitazione delle caseine. Le proteine del siero sono invece termolabili e vengono completamente denaturate riscaldando il latte a 90°C per soli 10 minuti.
- Contrariamente alle proteine del siero, le caseine sono fosfoproteine. Esse contengono in media lo 0,85% di tale elemento. I gruppi fosfato permettono alle caseine di legare il calcio rendendole un costituente ad alto valore nutritivo, soprattutto per i neonati. I gruppi fosfato sono inoltre i principali responsabili dell'alta stabilità termica delle caseine e intervengono nel meccanismo della coagulazione presamica in presenza di calcio. Il fosfato è esterificato alle caseine tramite il gruppo ossidrilico delle serine (fosfato organico) oppure è associato alle caseine nella forma colloidale di fosfato di calcio (fosfato inorganico).
- Le caseine sono per contro povere di zolfo (0,8%) mentre le proteine del siero ne contengono una quantità maggiore (1,7%); la differenza è dovuta alla natura degli amminoacidi contenenti zolfo presenti nelle diverse classi di proteine. Lo zolfo delle caseine è presente principalmente nella metionina, con basse concentrazioni di cisteina e cistina, mentre le proteine del siero ne contengono quantità significative. Questi amminoacidi sono in parte responsabili della denaturazione delle proteine del siero a seguito di trattamenti termici.
- Le caseine sono sintetizzate esclusivamente nella ghiandola mammaria mentre parte delle proteine del siero provengono direttamente dal sangue.
- Le proteine del siero si trovano disperse in soluzione ed hanno semplici strutture quaternarie. Le caseine hanno invece strutture quaternarie più complicate e circa il 95% di esse è presente nel latte sotto forma di aggregati colloidali, le micelle caseiniche, di massa compresa tra 10^6 e 10^9 Dalton. [Fox, 1998]

Tabella 2.4 - Composizione percentuale in proteine del latte di alcune specie animali.

Specie	Caseine	Proteine del siero	Totale
Donna	0,4	0,6	1,0
Vacca	2,8	0,6	3,4
Capra	2,5	0,4	2,9
Pecora	4,6	0,9	5,5
Asino	1,0	1,0	2,0
Cavallo	1,3	1,2	2,5
Maiale	2,8	2,0	4,8
Bisonte	3,7	0,8	4,5

2.3.3.1 Caseina

La caseina costituisce la parte proteica del latte che, a 20°C, precipita per effetto dell'acidificazione a pH 4,6. Nelle caseine è localizzato circa il 78% dell'azoto totale del latte. La caseina è formata da quattro costituenti principali (le caseine α_{s1} , α_{s2} , β e k) e da altri componenti da essi derivati (le caseine γ e la caseina λ). Le caseine α_{s1} , α_{s2} , β e k costituiscono rispettivamente circa il 38%, il 10%, il 36% ed il 13% delle caseine totali mentre la somma delle tre caseine γ raggiunge solo il 2% e la caseina λ non arriva all'1%. Le caseine differiscono tra loro per la loro struttura primaria, data da una diversa sequenza amminoacidica, e per il grado di fosforilazione e di glicosilazione, mentre presentano tutte una scarsa organizzazione in strutture secondarie e terziarie: la loro struttura aperta le rende suscettibili all'azione delle proteasi assai più di quanto avvenga per le proteine del siero. Le caseine hanno la capacità di associarsi in aggregati (micelle e submicelle caseiniche) che ne aumentano la stabilità nel latte e ne determinano il loro comportamento nella caseificazione.

L'associazione in micelle è spinta dalla presenza di ioni calcio, che determinano una riduzione della carica negativa delle caseine permettendone l'avvicinamento. Se però la concentrazione di ioni calcio è molto elevata, le micelle che vengono a formarsi saranno di dimensioni talmente grandi da non riuscire più a rimanere in soluzione colloidale. Le caseine α_{s1} , β e k (mostrate in Figura 2.3) hanno carattere anfifilico in quanto nella loro catena amminoacidica sono presenti sia zone idrofile che zone idrofobe. Le zone idrofobe permettono le associazioni tra le diverse caseine anche in assenza di ioni calcio mediante interazioni apolari, mentre le zone idrofile, in particolare un tratto comune ai tre tipi di caseine, caratterizzato da una concentrazione di cariche negative e definito come peptide

acidico, rendono possibile la formazione di uno strato di solvatazione nel loro intorno [Corradini, 1995] [Fox, 1998].

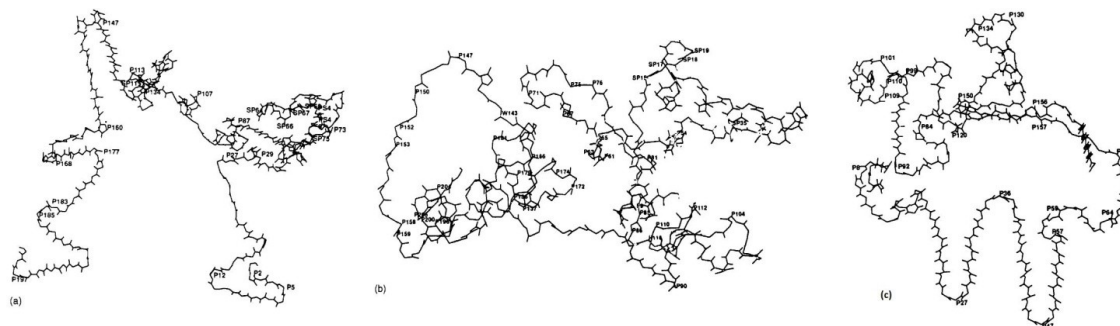


Figura 2.3 - Struttura terziaria delle caseine α_{s1} (a), β (b) e κ (c).

2.3.3.2 Le micelle caseiniche

Il 95% delle caseine sono presenti nel latte sotto forma di micelle sferiche. La sostanza secca ad esse riferite è formata al 94% da proteine e al 6% da specie chimiche a basso peso molecolare (principalmente fosfato di calcio colloidale). Le micelle hanno un diametro compreso tra 50 e 500 nm. L'osservazione al microscopio elettronico rivela una numerosa quantità di submicelle di diametro compreso tra 12 e 20 nm, che rappresentano però una frazione molto piccola del peso totale delle micelle. La concentrazione micellare è compresa tra 10^{14} e 10^{16} micelle per millilitro di latte.

La struttura delle micelle caseiniche è nota già da diversi decenni; nel 1967 Morr propose una struttura in cui le submicelle sono unite tramite il fosfato di calcio colloidale, conferendo alle micelle una struttura porosa (Figura 2.4).

Rimuovendo il fosfato di calcio tramite acidificazione si può notare che le micelle si disgregano. Lo stesso risultato si può ottenere per trattamento con urea o rendendo l'ambiente basico ($\text{pH} > 9$). Il modello submicellare è stato negli anni affinato [Schmidt, 1982] [Wakstra, 1984] [Ono, 1989]; è stata inoltre riportata la variabilità del contenuto di κ -caseina nelle diverse submicelle. In particolare è stato appurato che le submicelle povere di κ -caseina si posizionano nel cuore della micella mentre quelle con alta concentrazione di κ -caseina si trovano nella parte superficiale della stessa. La regione idrofila C-terminale delle κ -caseine è rivolta verso l'esterno della micella caseinica formando uno strato di

spessore 5-10 nm che conferisce stabilità alla micella stessa. Se questo strato venisse rimosso, la stabilità colloidale delle micelle verrebbe a mancare con conseguente coagulazione o precipitazione.

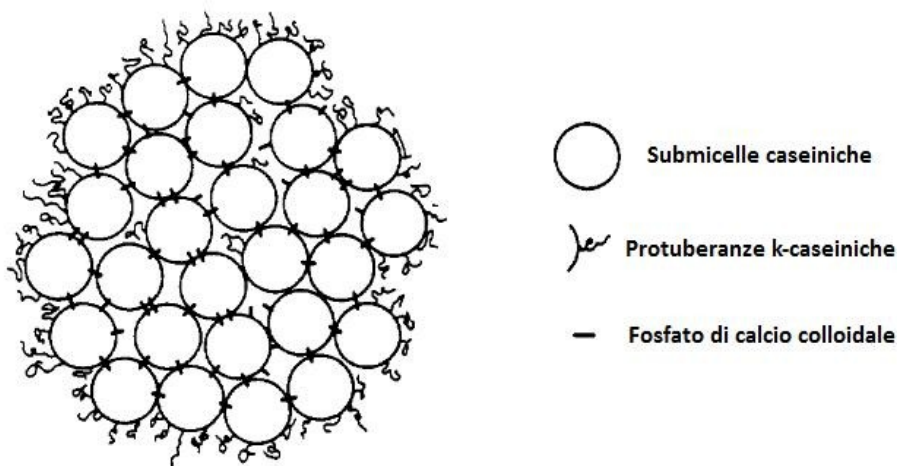


Figura 2.4 - Rappresentazione delle micelle caseiniche.

Il modello submicellare è stato accettato in larga parte della comunità scientifica ma Holt, nel 1994, propose un modello simile al modello di Morr in cui però le micelle hanno una struttura gelatinosa formata da una “ragnatela” di caseine flessibili tenute insieme da microgranuli di fosfato di calcio colloidale (Figura 2.5). A causa di interazioni non covalenti, ogni colloide di fosfato di calcio è circondato da circa 50 catene di caseine fosforilate (caseine α_s e β), formando un nanocluster. I diversi nanocluster si aggregano tramite interazioni non covalenti sia tra di essi che con altre caseine, portando alla formazione delle micelle. Le k-caseine, non possedendo gruppi fosforici, non partecipano alla formazione dei nanocluster ma si associano in maniera non covalente ad altre proteine, formando lo strato di superficie della micella. Il modello finale non risulta quindi essere formato da un insieme di subunità identiche ma da un’unica struttura proteica [Dalglish, 2012]. Tra i due modelli si ha comunque concordanza sul fatto che il fosfato di calcio colloidale funga da collante per le caseine e sulla presenza delle k-caseine sulla superficie della micella con le loro regioni idrofiliche rivolte verso l’esterno. Holt introdusse anche il concetto di funzione biologica delle caseine, grazie alla loro capacità di trasportare il calcio nel latte mantenendolo in forma stabile. Infatti, senza l’effetto stabilizzante delle caseine, il

fosfato di calcio precipiterebbe nelle cellule mammarie determinando la morte delle ghiandole mammarie e dell'animale stesso [Fox, 1998].

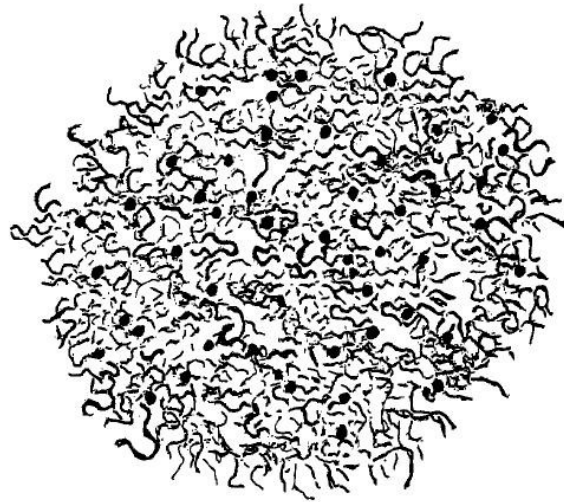


Figura 2.5 - Modello delle micelle caseiniche [Holt, 1994].

Studi più recenti hanno permesso di postulare una nuova ipotesi di modello micellare come di una struttura di tipo spugnoso formata dai nanocluster proposti da Holt in cui sono presenti regioni acquose stabilizzate dall'interazione della caseina β con porzioni idrofobiche dei nanocluster [McMahon, 2008] [Bouchoux, 2010] [Dalglish, 2011]. In Figura 2.6 è raffigurato il modello caseinico: le caseine α_s e β (colorate in arancione) sono unite ai nanocluster di fosfato di calcio colloidale (sfere grigie); parte delle caseine β (colorata in blu) è legata idrofobicamente ad altre caseine mentre le para-k-caseine (colorate in verde) e i macro peptidi di caseina (colorati in nero) si trovano nella parte più esterna della micella. All'interno della struttura si possono vedere le regioni di acqua. La presenza delle regioni di acqua spiega la porosità delle micelle, dimostrata dal fatto che le caseine β possono essere rimosse dalle micelle tramite raffreddamento [Creamer, 1997] e dalla capacità di penetrazione di grandi molecole all'interno della micella [Colsenet, 2005].

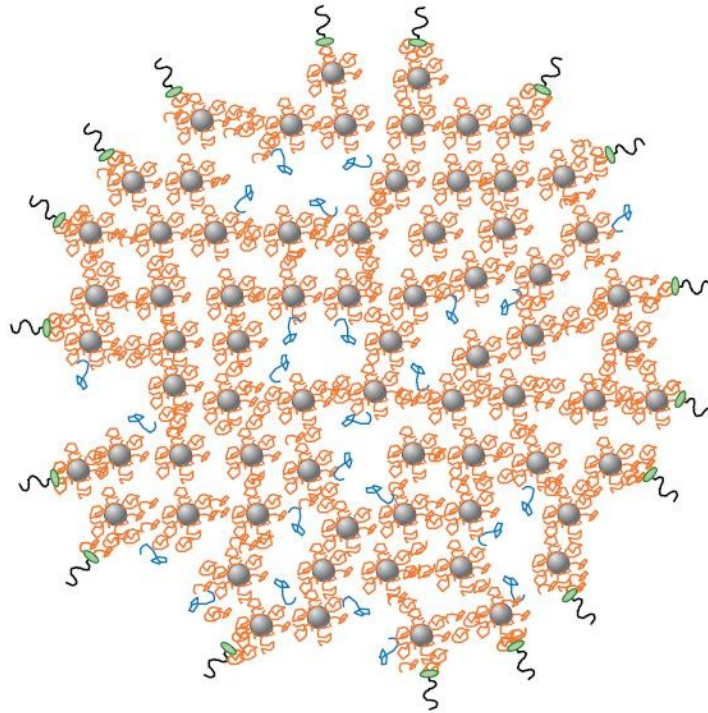


Figura 2.6 – Schema del modello caseinico con le regioni di acqua nella struttura [Dalglish, 2012].

2.3.3.3 Le proteine del siero

Le proteine che, contrariamente alle caseine, non precipitano a pH 4,6 ma rimangono stabili in soluzione sono definite proteine del siero. Sono molto solubili in soluzione acquosa diluita (è possibile ivi disciogliere sino al 5% di proteine) anche a valori di pH prossimi al punto isoelettrico (pH 4,5-5,5) e costituiscono circa il 20% delle proteine totali del latte. Esse sono principalmente rappresentate dalla β -lattoglobulina, dall' α -lattalbumina, dalla sieralbumina, dalle immunoglobuline, dai componenti della frazione dei proteoso peptoni e da altre componenti presenti in concentrazione molto più bassa come la lattoferrina, la transferrina e la ceruloplasmina. Oltre alle proteine, nel siero vi è anche una frazione azotata non proteica (NPN), formata in particolare da urea. Nella composizione amminoacidica delle proteine del siero vi sono diversi residui cisteinici, molti dei quali formano legami a ponte disolfuro intramolecolari. La presenza dei residui cisteinici determina la scarsa stabilità al calore delle proteine del siero. I ponti disolfuro si rompono liberando acido solfidrico, attivando i gruppi tiolici e formando legami disolfuri intermolecolari originando così grosse macromolecole che precipitano dalla soluzione [Corradini, 1995].

2.3.4 Sali minerali

I sali presenti nel latte sono principalmente fosfati, citrati, cloruri, solfati, carbonati e bicarbonati di sodio, potassio, calcio e magnesio. Altri elementi, come rame, ferro, silicio, zinco e iodio sono presenti in tracce. Le proteine stesse possono essere incluse tra i sali, avendo dei gruppi carichi positivamente e negativamente che possono formare sali con i contro-ioni, ma di norma non vengono trattate come tali. Il contenuto in ceneri del latte, che rappresenta l'insieme delle sostanze minerali in esso contenute, è pressoché costante, con un quantitativo che si aggira intorno al 0,7-0,8% nel latte di vacca e 1,0-1,1% nel latte di pecora. Il contenuto dei singoli ioni può invece variare considerevolmente; in Tabella 2.5 è mostrato il contenuto medio dei singoli ioni del latte vaccino, il loro range tipico ed il range dei valori estremi registrati. I valori estremi riportati sono da considerarsi anomali in quanto relativi a colostro, a latte di tarda lattazione o a latte di animali con infezioni mastitiche.

Tabella 2.5 - Concentrazioni dei diversi costituenti salini del latte (mg/l) [Fox, 2008].

Costituente	Contenuto medio	Range tipico	Valori estremi
Sodio	500	350-600	110-1150
Potassio	1450	1350-1550	1150-2000
Calcio	1200	1000-1400	650-2650
Magnesio	130	100-150	20-230
Fosforo totale^a	950	750-1110	470-1440
Fosforo inorganico^b	750		
Cloruri	1000	800-1400	540-1420
Fosfati	100		
Carbonati (CO₂)	200		
Citrati	1750		

^aIl fosforo totale include il fosfato inorganico colloidale, il fosfato delle caseine, il fosfato inorganico solubile e i fosfolipidi.

^bIl fosforo inorganico comprende il fosfato inorganico colloidale ed il fosfato inorganico solubile.

Anche la concentrazione dei sali nel latte differisce a seconda della specie animale. Come si può vedere in Tabella 2.6 il latte di capra e di vacca hanno composizione simile e sono i più ricchi di potassio mentre il latte di pecora ha maggiori quantità di calcio e fosforo. In generale i tenori in fosforo e calcio del latte delle diverse specie animali sono direttamente correlati alla velocità di crescita del piccolo.

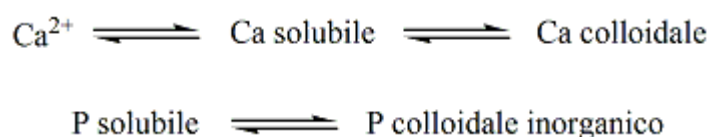
Alcuni sali sono sufficientemente solubili da essere presenti solo in soluzione; altri, in particolare il fosfato di calcio, hanno concentrazioni troppo elevate per poter rimanere esclusivamente in soluzione, per cui una parte di essi viene a trovarsi in forma colloidale,

associata alle caseine. Nelle micelle caseiniche, il fosfato di calcio colloidale ha infatti il compito di tenere unite le submicelle legandosi ai residui serinici delle caseine (Figura 2.7).

Tabella 2.6 - Contenuto medio di elementi e composti nel latte di diversi mammiferi (g/Kg) [Alais, 2010].

Costituente	Capra	Pecora	Donna	Vacca
Potassio	1,90	1,30	0,60	1,45
Sodio	0,4	0,45	0,17	0,50
Calcio	1,30	2,10	0,30	1,25
Magnesio	0,13	0,16	0,04	0,12
Fosforo	1,00	1,50	0,16	1,00
Cloro	1,50	0,10	0,45	1,10
Zolfo	0,20	-	0,15	0,35
Acido citrico	1,5	-	0,8	1,8
Ceneri	8,0	11,0	2,5	8,0

Calcio e fosforo si trovano, nel latte, in forme solubili ed insolubili in equilibrio tra di loro come mostrato di seguito:



Gli equilibri riportati variano a seconda della temperatura e del pH, con conseguente variazione della stabilità delle micelle caseiniche in cui le proteine sono unite dai gruppi di fosfato di calcio colloidale. Anche la presenza dell'anione citrato altera l'equilibrio del calcio in quanto "sequestra" lo ione rendendolo meno disponibile per la formazione di fosfato colloidale determinando una diminuzione della capacità aggregante delle caseine [Alais 2010] [Corradini, 1995] [Fox, 1998]. Il contenuto totale di sali è in equilibrio col contenuto di lattosio nel latte in quanto entrambi partecipano al mantenimento della pressione osmotica del sistema mammario.

Secondo la classificazione della International Union of Biochemistry and Molecular Biology gli enzimi vengono suddivisi in sei classi a seconda della loro funzione catalitica [Webb, 1992]:

- EC 1: Ossidoreduttasi (catalizzano le reazioni di ossidoriduzione);
- EC 2: Transferasi (catalizzano il trasferimento di un gruppo funzionale);
- EC 3: Idrolasi (catalizzano l'idrolisi dei legami chimici);
- EC 4: Liasi (catalizzano la rottura di legami chimici non causata da idrolisi);
- EC 5: Isomerasi (catalizzano le reazioni di isomerizzazione);
- EC 6: Ligasi (catalizzano la formazione di un legame covalente tra due molecole).

Gli enzimi presenti nel latte appartengono soprattutto alle categorie EC 1, EC 2 ed EC 3 e in quantità minore alle categorie EC 4 ed EC 5. Non sono invece presenti enzimi della classe EC 6. I più importanti enzimi sono le lipasi, le proteasi e le fosfatasi, appartenenti alla classe EC 3, che catalizzano rispettivamente le reazioni di idrolisi dei trigliceridi, delle caseine e dei loro gruppi fosforici. I principali enzimi appartenenti alla classe EC 1 sono la lattoperossidasi, che porta alla formazione di sostanze antibatteriche; la xantino-reduttasi, che favorisce la formazione di acido urico; e la catalasi, che catalizza le reazioni di ossidazione degli acidi grassi insaturi. I principali enzimi della classe EC 2 sono invece la ribonucleasi, che catalizza la trans-fosforilazione degli acidi ribonucleici, di fatto idrolizzandoli; e la lattosio-sintetasi, che catalizza la biosintesi del lattosio a partire dai suoi monomeri.

La presenza e l'attività degli enzimi è influenzata da fattori fisiologici, genetici e zootecnici, quali il periodo della lattazione, la stagione, la razza, i fenomeni mastitici e l'alimentazione. Il latte comunque non contiene substrati per molti degli enzimi presenti e non offre un ambiente adatto alla loro attività, inibita soprattutto dal pH quasi neutro e dalla bassa temperatura a cui il latte viene di fatto tenuto durante la conservazione e prima della trasformazione tecnologica.

Gli enzimi svolgono una attività importante durante i processi tecnologici di trasformazione del latte, ma alcuni di essi possono comunque avere effetti negativi sulla qualità del latte e dei suoi derivati. Alcuni enzimi possono causare modificazioni indesiderabili quali l'irrancidimento della frazione lipidica del latte e dei suoi derivati, alcune proteasi e lipasi termoresistenti, prodotte da batteri psicrotrofi, possono indurre fenomeni di gelificazione nel latte UHT accorciandone il periodo di conservabilità. Per

questo motivo in molti processi tecnologici viene previsto il trattamento termico, il cui fine è quello di sanitizzare il latte, riducendone la carica microbica ed enzimatica [Alais, 2010] [Corradini, 1995] [Fox, 1998].

2.3.6 Vitamine

Le vitamine sono molecole di varia struttura necessarie per il funzionamento dei processi vitali dell'organismo che è però incapace di sintetizzarle. L'apporto vitaminico deve pertanto necessariamente provenire dall'alimentazione. Il latte è uno degli alimenti più ricchi di vitamine in quanto primo alimento della vita dei mammiferi; esso contiene vitamine sia liposolubili che idrosolubili.

Le vitamine liposolubili, in particolare quelle A, D ed E, sono associate al grasso e, dopo la scrematura, si ritrovano quasi completamente nella crema e nel burro. La loro concentrazione nel latte è assai variabile, in quanto influenzata da fattori esogeni quali il tipo di alimentazione dell'animale, il periodo di lattazione ed il clima.

Le vitamine idrosolubili, in particolare quelle B₁, B₂, C etc..., si trovano invece nella fase acquosa, cioè permangono prevalentemente nel latte scremato e nel siero. A differenza delle vitamine liposolubili, la concentrazione delle vitamine idrosolubili nel latte varia in misura esigua, in quanto essa non è influenzata da fattori esogeni [Alais, 2010].

2.4 Lipidi

I lipidi hanno la funzione primaria di provvedere all'apporto energetico del neonato dei mammiferi. Il loro tenore nel latte varia tra il 2% ed il 50% a seconda della specie lattifera esaminata, riflettendo in ciò le esigenze energetiche del nascituro. I lipidi fungono anche da risorsa di acidi grassi essenziali, non sintetizzabili dall'organismo, come l'acido linoleico, e di vitamine liposolubili (A, D, E, K).

Come mostrato in Tabella 2.7, il contenuto di grasso del latte varia considerevolmente a seconda della specie animale, a causa dei già citati fattori genetici, fisiologici e ambientali. All'interno di una stessa specie, sia il contenuto di grasso che la sua composizione acidica variano anche al variare della razza, del tipo di alimentazione e dello stato di lattazione.

Il latte bovino presenta una concentrazione del grasso variabile tra 3 e 6%, ma il range tipico è compreso tra 3,5 e 4,7%. Tra le specie ruminanti più comuni, il latte di pecora è quello che ne presenta il contenuto maggiore; esso infatti raggiunge concentrazioni anche

doppie rispetto a quelle del latte di vacca. I cambiamenti della composizione acidica possono portare a modifiche delle proprietà del grasso [Fox, 1998] [MacGibbon, 2006].

Tabella 2.7 - Contenuto in grasso del latte di diverse specie (g/l) [Christie, 1995].

Specie	Contenuto di grasso	Specie	Contenuto di grasso
Vacca	33-47	Maiale	68
Bufala	47	Coniglio	183
Pecora	40-99	Scimmia	10-51
Capra	41-45	Bue muschiato	109
Elefante	85-190	Delfino	62-330
Donna	38	Foca	502-532
Cavallo	19	Orso (quattro specie)	108-331

2.4.1 I globuli di grasso

La sostanza grassa si trova nel latte sotto forma di globuli sferici costituiti dall'associazione di gliceridi e da materiale interfacciale dello spessore di 10 nm, chiamato membrana, che determina le tipiche proprietà di stabilità dei globuli di grasso in emulsione. L'assenza di questo tipo di interazione determinerebbe l'insolubilità dei lipidi nel latte.

L'esistenza dei globuli venne riportata già nel 1674 da Van Leeuwenhoek a seguito dell'analisi del latte al microscopio [Huppertz, 2006]. Le proprietà fisiche e colloidali dei globuli contribuiscono, in maniera parziale o esclusiva a seconda dei casi, ad alcune delle proprietà del latte, come il colore, e di fenomeni fisici quali l'affioramento della crema. Le emulsioni sono infatti fisicamente instabili, portando ad un'alterazione della distribuzione spaziale o dell'organizzazione strutturale dei globuli.

I meccanismi che agiscono sui globuli possono essere distinti in *separazione gravitazionale* e *aggregazione dei globuli*. La separazione gravitazionale è data dal movimento delle goccioline di grasso causato dalla diversa densità rispetto al liquido circostante. Se la densità del globulo è minore, si ha affioramento verso la superficie (formazione della crema), se invece la densità del globulo è maggiore si ha sedimentazione (movimento verso il fondo). L'aggregazione avviene per definizione quando i globuli rimangono uniti per un periodo di tempo maggiore del periodo di tempo in cui rimarrebbero uniti in assenza di interazioni colloidali. I meccanismi responsabili dell'aggregazione sono la flocculazione e la coalescenza. La flocculazione è definita come l'aggregazione dei globuli a dare flocculi tridimensionali in cui i singoli globuli rimangono

entità individuali legati tra loro da interazioni deboli e reversibili. La coalescenza è il processo in cui, a causa della rottura delle membrana esterna, due o più globuli di grasso si uniscono a dare un unico globulo sferico di dimensioni maggiori [Huppertz, 2006]. I globuli hanno la tendenza ad aggregarsi in grappoli e sono in genere classificati in base alle loro dimensioni:

- Piccoli (diametro minore di 3 μm);
- Medi (diametro compreso tra 4 e 9 μm);
- Grossi (diametro maggiore di 10 μm , fino ad un massimo di 20 μm);

Il loro diametro varia a seconda della specie animale; per le specie bovine esso tende a diminuire all'aumentare del periodo di lattazione e nel corso di una specifica mungitura le prime frazioni presentano globuli con diametro minore rispetto a quelli raccolti nelle ultime frazioni. I globuli di grasso del latte di vacca hanno diametro medio compreso tra 3 e 5 μm , maggiore di quello dei globuli del latte di pecora, preponderantemente presenti con diametro inferiore a 3,5 μm [Park, 2007]. Questa differenza rende il latte di pecora più digeribile di quello di vacca.

Nel globulo di grasso possiamo individuare tre zone: la parte esterna, chiamata membrana, composta da diverse sostanze tra cui proteine, fosfolipidi e colesterolo che determina la stabilità del globulo sotto forma di emulsione; la parte intermedia, sottostante la membrana, formata da lipidi ad alto punto di fusione e la parte interna, che presenta lipidi a basso punto di fusione, generalmente presenti allo stato liquido a temperatura ambiente [Alais, 2010] [Corradini, 1995] [Fox, 1998].

2.4.1.1 La membrana dei globuli di grasso

La membrana costituisce il 2-6% della massa del globulo di grasso. Il suo studio è di difficile realizzazione a causa della sua alta fragilità. I metodi per l'isolamento della membrana sono diversi: in generale il primo passaggio consiste nella separazione della crema, contenente i lipidi, per centrifugazione del latte (tale trattamento può però causare danni alla membrana) oppure per effetto gravità. La crema viene lavata da 3 a 6 volte con acqua oppure diluita e successivamente separata per gravità. In questo modo i sali solubili ed altre piccole molecole andranno in soluzione nel siero. I globuli di dimensioni minori potranno venir persi durante le fasi di lavaggio e può comunque avvenire una perdita dello strato più esterno e più debolmente legato della membrana su cui può inoltre avvenire

l'adsorbimento di alcune sostanze solubili del siero. La crema lavata viene poi destabilizzata da processi quali il congelamento o la zangolatura. In seguito il grasso viene sciolto e separato dalla membrana per centrifugazione; in questa fase può avvenire una contaminazione della membrana da parte dei costituenti interni del globulo di grasso. Alternativamente, per separare la membrana dal grasso, si può trattare con un tensioattivo non caustico come il desossicolato di sodio. La separazione della membrana può essere effettuata anche direttamente dalla crema per ultracentrifugazione a 25°C e a 200.000 x g. [Alais, 2010] [Fox, 1998].

2.4.1.2 Composizione della membrana

La composizione chimica generale della membrana del globulo di grasso è ben nota anche se essa può variare leggermente a seconda del trattamento subito dal latte, probabilmente a causa del fatto che alcuni costituenti sono in equilibrio tra membrana e siero. La sua composizione è molto più simile a quella delle membrane cellulari che a quella del grasso o del siero del latte. Le sostanze principali che compongono la membrana sono:

- Fosfolipidi e colesterolo;
- Cerebrosidi;
- Proteine e glicoproteine;
- Enzimi, principalmente la fosfatasi alcalina e la xantina-ossidasi (riduttasi aldeidica);
- Sostanze varie quali riboflavina (vitamina B₂), acido ribonucleico, etc...;
- Acqua.

Proteine e lipidi formano più del 90% della sostanza secca della membrana. Le proteine di membrana costituiscono circa l'1% delle proteine totali del latte ma la loro composizione differisce alquanto dalle caseine e dalle proteine che si trovano nel siero del latte. La maggior parte sono glicoproteine contenenti una percentuale considerevole di carboidrati (esosì, 2,8 ÷ 4,15%; esosammine, 2,5 ÷ 4,2% e acido sialico, 1,3 ÷ 1,8%). A seconda del tipo di metodo di estrazione e del danno subito, la membrana può contenere anche caseine e proteine del siero.

Tabella 2.8 - Composizione della membrana dei globuli di grasso del latte bovino [Keenan, 1983].

Costituente	Quantità
Proteine	25-60% della sostanza secca
Lipidi totali	0,5-1,2 mg su mg di proteine
Fosfolipidi	0,13-0,34 mg su mg di proteine
- Fosfatidilcolina	34% dei fosfolipidi totali
- Fosfatidil-etanolamina	28% dei fosfolipidi totali
- Sfingomieline	22% dei fosfolipidi totali
- Fosfatidil-inositolo	10% dei fosfolipidi totali
- Fosfatidilserina	6% dei fosfolipidi totali
Lipidi neutri	56-80% dei lipidi totali
- Idrocarburi	1,2% del totale dei lipidi neutri
- Steroli	0,2-5,2% del totale dei lipidi neutri
- Esteri sterolici	0,1-0,8% del totale dei lipidi neutri
- Gliceridi	53-74% del totale dei lipidi neutri
- Acidi grassi liberi	0,6-6,3% del totale dei lipidi neutri
Cerebrosidi	3,5 nmoli su mg di proteine
Gangliosidi	6-7,4 nmoli di acido sialico su mg di proteine
Acidi sialici totali	63 nmoli su mg di proteine
Esosi	0,6 μ moli su mg di proteine
Esosoammine	0,3 μ moli su mg di proteine
Acidi uronici	99 ng su mg di proteine
RNA	20 μ g su mg di proteine

Tramite la tecnica SDS-PAGE (*Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis*, ossia elettroforesi su gel di poliacrilammide in presenza di sodio dodecilsolfato) è stato possibile analizzare gli estratti proteici della membrana, risolvendo le proteine in una serie di 60 bande discrete, con una massa molecolare variabile tra 11 e 250 KDa [Keenan, 1995]. Le principali proteine sono la xantina-ossidasi, la butirrofilina e la glicoproteina B, rispettivamente di massa molecolare pari a 155, 67 e 48 KDa. La maggior parte delle altre proteine si trova a concentrazioni molto basse. La xantina-ossidasi, enzima che favorisce l'ossidazione dei lipidi in un processo radicalico, costituisce circa il 20% delle proteine della membrana ed una parte di essa viene persa nei processi di raffreddamento durante l'estrazione. Ne esistono almeno quattro diverse varianti che hanno punto isoelettrico compreso tra 7,0 e 7,5. La butirrofilina, la principale proteina di membrana, è una glicoproteina poco solubile presente in almeno quattro varianti dotate di punto isoelettrico compreso tra 5,2 e 5,3. È costituita da 526 amminoacidi ed ha una massa molecolare (esclusi i carboidrati) di 56 KDa. Ha un ruolo attivo nell'avvolgimento dei globuli di grasso, riuscendo a creare legami forti con i fosfolipidi.

Tabella 2.9 – Composizione acidica della membrana e del cuore del globulo di grasso nel latte ovino [Martini, 2013].

Acido grasso	g/100g dei lipidi totali di membrana	g/100g dei lipidi totali del cuore del globulo
C4:0	0,01	1,58
C6:0	< 0,01	1,62
C8:0	0,03	1,87
C10:0	1,77	6,52
C12:0	2,50	4,14
C14:0	10,49	11,81
C16:0	34,53	28,42
C18:0	19,81	11,82
C18:1-<i>cis</i>-9	19,62	21,98
C18:2-<i>cis</i>-9, <i>cis</i>-12	0,56	0,97
C18:3-<i>cis</i>-9, <i>cis</i>-12, <i>cis</i>-15	0,17	0,07
C18:2-<i>cis</i>-9, <i>trans</i>-11	0,59	0,26
Saturi	72,89	70,85
Monoinsaturi	24,21	27,20
Polinsaturi	2,89	1,95
Omega-3	0,32	0,12
Omega-6	2,06	1,64
Insaturi/Saturi	0,37	0,44

I lipidi della membrana dei globuli di grasso costituiscono circa lo 0,5-1,0% dei lipidi totali del latte. La loro composizione è data da fosfolipidi e lipidi neutri presenti in rapporto 2:1, e da altri lipidi presenti in minore concentrazione. I fosfolipidi sono principalmente fosfatidilcoline, fosfatidiletanolammine e sfingomieline in rapporto 2:2:1. I fosfolipidi della membrana hanno un alto contenuto di acidi grassi insaturi (maggiore del 30%) e perciò la membrana è molto suscettibile all'ossidazione. I principali acidi grassi dei fosfolipidi di membrana e la loro percentuale relativa nel latte bovino sono i C14:0 (5%), C16:0 (25%), C18:0 (14%), C18:1 (25%), C18:2 (9%), C22:0 (3%), C24:0 (3%). La membrana contiene anche numerosi glicolipidi e cerebrosidi; questi ultimi, grazie agli acidi grassi a lunga catena di cui sono composti, determinano l'alta stabilità della membrana. I lipidi neutri della membrana sono invece formati da trigliceridi (83-88%), digliceridi (5-14%) e acidi grassi liberi (1-5%). Gli acidi grassi che li formano sono principalmente a catena lunga, come il palmitico (C16:0), lo stearico (C18:0), il miristico (C14:0), l'oleico (C18:1) ed il laurico (C12:0), e quindi ad alto punto di fusione. In Tabella 2.8 è mostrata la composizione più dettagliata della membrana dei globuli di grasso contenuti nel latte bovino. Per quanto riguarda il latte ovino, gli acidi grassi a corta e media catena sono presenti in maggiore percentuale nel cuore del globulo mentre l'acido palmitico (C16:0) e l'acido stearico (C18:0) presentano percentuali maggiori nella membrana. La membrana

presenta inoltre maggiori percentuali di acidi polinsaturi, di acidi linoleici coniugati e di acidi omega-3 (Tabella 2.9) [Recio, 2009] [Martini, 2013].

La membrana contiene anche diversi enzimi ed alcuni metallocomposti (prioritari quelli di Cu e Fe) mentre numerosi altri metalli (Co, Ca, Na, K, Mg, Mo, Zn) sono presenti in tracce [Fox, 1998].

2.4.1.3 Struttura della membrana

Nonostante sia stata studiata per oltre 40 anni, la struttura della membrana non è stata ancora ben definita nel dettaglio ed è l'aspetto meno conosciuto dei globuli di grasso. Gli studi degli ultimi anni convergono verso un sistema formato da tre strati, di spessore totale compreso tra 10 e 50 nm. Lo strato più interno è costituito da proteine e lipidi polari che hanno origine dal reticolo endoplasmatico delle cellule epiteliali mammarie, mentre i due strati più esterni sono costituiti da lipidi polari, aventi invece origine da precise regioni della membrana plasmatica apicale delle stesse cellule che circondano i globuli di grasso quando essi vengono secreti [Heid, 2005]. In Figura 2.8 è mostrato lo schema del sistema a tre strati della membrana. La membrana costituisce quindi lo strato interfacciale tra fase acquosa (siero) e fase grassa (globulo di grasso).

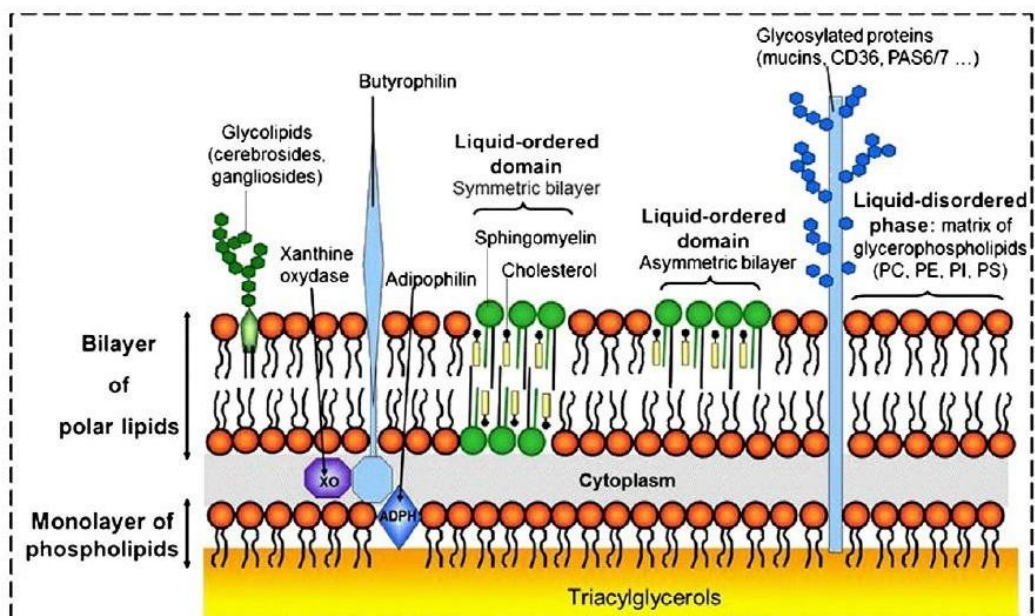


Figura 2.8 - Membrana dei globuli di grasso del latte: struttura a tre strati [Lopez, 2011].

La superficie esterna è perciò composta prevalentemente da composti idrofilici mentre quella interna da composti idrofobici. La sua importanza risiede nella capacità di dare stabilità nel latte al globulo di grasso, in quanto la sua superficie esterna è la prima ad essere coinvolta in reazioni e interazioni riguardanti i globuli. Ha quindi la funzione di proteggere i trigliceridi contenuti nel cuore dei globuli da alterazioni di tipo enzimatico o chimico che ne determinano i fenomeni di irrancidimento.

La membrana può però subire alterazione a causa di fattori di varia natura che possono avvenire sia prima della secrezione del latte (alimentazione, malattie e stato di lattazione dell'animale) sia dopo la mungitura (trattamenti di conservazione e di trasformazione tecnologica del latte).

Sulla superficie della membrana dei globuli avvengono interazioni con altri componenti del latte, quali le caseine ed altre proteine del siero come quelle che avvengono, ad esempio, durante il trattamento di omogeneizzazione del latte, che causa alterazioni della membrana stessa. Possono avvenire anche solo semplici fenomeni di aggregazione tra i globuli, quali quelli che determinano il naturale affioramento della crema, che avvengono ad esempio quando il latte, dopo la mungitura, viene stoccato per lunghi periodi. La presenza dei fenomeni di aggregazione dei globuli di grasso determina un aumento della velocità di separazione dei lipidi.

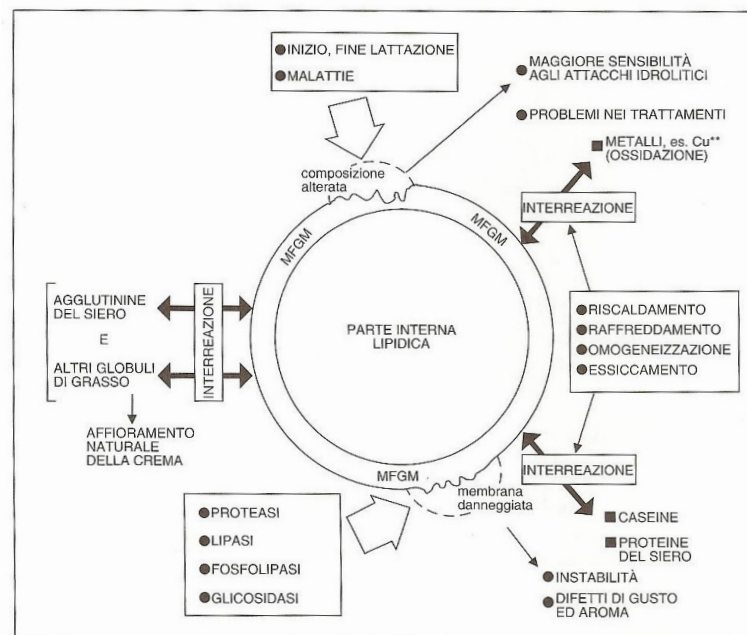


Figura 2.9 - Azioni e interazioni che avvengono sulla superficie dei globuli di grasso del latte [McPherson, 1983].

In Figura 2.9 sono mostrate, oltre alle interazioni tra globuli diversi, le azioni e le interazioni causate da enzimi, fattori fisiologici o legati al benessere animale o trattamenti tecnologici del latte, che avvengono sulla superficie dei globuli di grasso del latte e le loro conseguenze [Corradini, 1995].

2.4.2 Composizione del grasso del latte

I lipidi del latte sono composti nella loro quasi totalità da trigliceridi. La parte rimanente è divisa tra digliceridi, monogliceridi, acidi grassi liberi, fosfolipidi e steroli con presenza in tracce di vitamine liposolubili e carotenoidi. In Tabella 2.10 è mostrata la composizione percentuale del latte bovino.

Tabella 2.10 - Composizione generale dei lipidi nel latte bovino [Walstra, 1984].

Classe lipidica	Quantità (% peso su peso)
Trigliceridi	98,30
Digliceridi	0,30
Monogliceridi	0,03
Acidi grassi liberi	0,10
Fosfolipidi	0,80
Steroli	0,30
Carotenoidi	tracce
Vitamine liposolubili	tracce

Essendo i trigliceridi i componenti maggioritari del grasso, essi hanno un effetto importante sulle proprietà del grasso del latte come l'idrofobicità, la densità ed il punto di fusione. I trigliceridi sono una classe di composti assai diversificata, causa l'elevato numero di acidi grassi di diverso peso molecolare e grado di insaturazione. Grazie all'utilizzo di tecniche cromatografiche e spettroscopiche, sono stati identificati circa 400 acidi grassi presenti nel latte, tuttavia solo 40 di essi sono presenti in concentrazioni superiori allo 0,01% e solo 13 superano l'1% del totale degli acidi grassi. In Tabella 2.11 sono indicati i principali acidi grassi del latte, il loro nome IUPAC ed il loro nome comune. I digliceridi e i monogliceridi sono probabilmente il prodotto di una sintesi incompleta dei trigliceridi, oppure derivano da una parziale idrolisi di questi ultimi, a seguito di danneggiamenti della membrana del globulo di grasso durante la mungitura e la

conservazione del latte, che spiega in parte la presenza di acidi grassi liberi nel latte [Fox, 1998].

Tabella 2.11 - Acidi grassi principali del latte.

Acido grasso	Nome IUPAC	Nome comune
C4:0	Acido butanoico	Acido butirrico
C6:0	Acido esanoico	Acido caproico
C8:0	Acido ottanoico	Acido caprilico
C10:0	Acido decanoico	Acido caprinico
C12:0	Acido dodecanoico	Acido laurico
C14:0	Acido tetradecanoico	Acido miristico
C16:0	Acido esadecanoico	Acido palmitico
C18:0	Acido ottadecanoico	Acido stearico
C18:1-<i>trans</i>-11	Acido <i>trans</i> -11-ottadecenoico	Acido vaccenico
C18:1-<i>cis</i>-9	Acido <i>cis</i> -9-ottadecenoico	Acido oleico
C18:2-<i>cis</i>-9, <i>cis</i>-12	Acido <i>cis</i> -9- <i>cis</i> -12-ottadecadienoico	Acido linoleico
C18:3-<i>cis</i>-9, <i>cis</i>-12, <i>cis</i>-15	Acido <i>cis</i> -9- <i>cis</i> -12- <i>cis</i> -15-ottadecatrienoico	Acido α -linolenico
C18:2-<i>cis</i>-9, <i>trans</i>-11	Acido <i>cis</i> -9- <i>trans</i> -11-ottadecadienoico	Acido rumenico

Gli acidi grassi sono raggruppabili in 3 categorie a seconda della lunghezza della loro catena acilica:

- Acidi grassi a catena corta, con numero di atomi di carbonio minore di 10;
- Acidi grassi a catena media, con numero di atomi di carbonio compreso tra 11 e 17;
- Acidi grassi a catena lunga, con numero di atomi di carbonio pari (o superiore) a 18.

La catena alifatica degli acidi grassi può presentare uno o più doppi legami. A seconda della assenza o presenza di doppi legami, gli acidi grassi possono essere classificati in:

- Acidi grassi saturi (Saturated Fatty Acids, SFA), quando la catena alifatica non contiene doppi legami;
- Acidi grassi insaturi (Unsaturated Fatty Acids, UFA), quando la catena alifatica contiene almeno un doppio legame che può essere presente in configurazione *cis* o *trans*. In presenza di un solo doppio legame, gli acidi grassi vengono definiti monoinsaturi (Mono Unsaturated Fatty Acids, MUFA); in presenza di più doppi legami, gli acidi grassi vengono definiti polinsaturi (Poly Unsaturated Fatty Acids, PUFA).

La presenza di doppi legami conferisce all'acido grasso una struttura non lineare. Mentre gli acidi con un doppio legame *trans* possiedono una struttura che non risulta essere molto dissimile da quella lineare, gli acidi grassi con un doppio legame *cis* hanno una geometria ripiegata che risulta essere molto più accentuata nel caso di presenza di più doppi legami *cis* (Figura 2.10). Questa differenza ha effetto sul punto di fusione dei trigliceridi. I trigliceridi formati da acidi grassi contenenti doppi legami *cis* presentano una densità di impaccamento più bassa ed un corrispondente punto di fusione più basso.

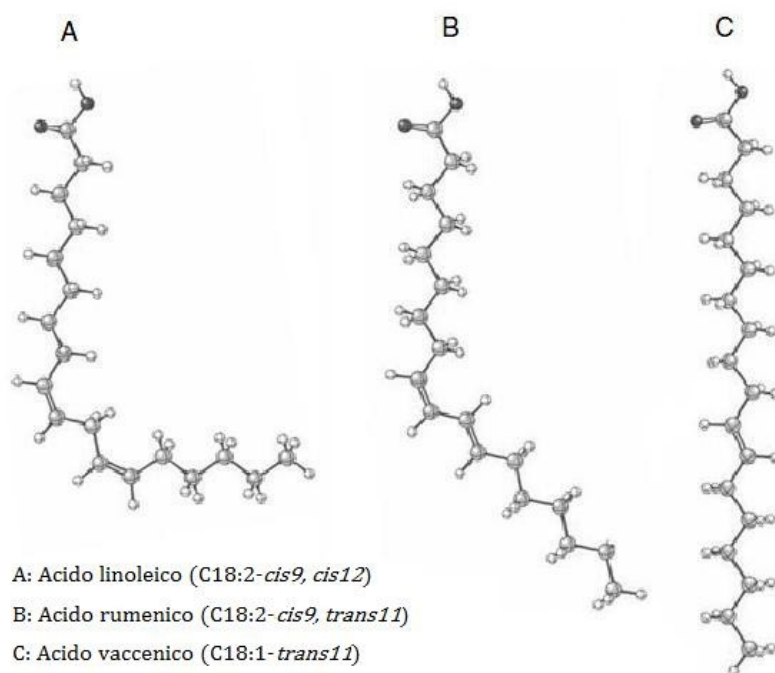


Figura 2.10 – Geometria degli acidi grassi con diversi tipi di doppi legami [Bauman, 2006].

La classe dei fosfolipidi rappresenta meno dell'1% dei lipidi totali ma essi giocano un ruolo importante nella struttura della membrana dei globuli di grasso; il latticello e il latte scremato ne sono ricchi indicando che questi composti presentano grandi quantità di materiale di membrana.

Il colesterolo è il maggiore componente della categoria degli steroli; rappresenta più del 95% degli steroli totali e circa lo 0,3% dei lipidi totali del latte, un valore basso se confrontato con quello di altri alimenti. Il 90% del colesterolo è presente nella sua forma libera mentre la parte rimanente è esterificato con un acido grasso.

I carotenoidi, rappresentati maggiormente dal β -carotene e presenti nel latte bovino in tracce, contribuiscono all'attività della vitamina A nel latte fornendogli una colorazione giallognola. I carotenoidi provengono dall'alimentazione dell'animale ma non tutti i ruminanti trasmettono queste sostanze nel loro latte. I carotenoidi sono presenti nel latte bovino ma assenti in quello ovino, che quindi non presenta colorazione giallognola.

Altre sostanze in tracce sono date dalle vitamine liposolubili, come la vitamina A e la vitamina E, la cui presenza nel latte è influenzata dai vari fattori genetici e fisiologici dell'animale, ed è superiore nel latte di inizio lattazione rispetto al latte dei periodi più avanzati [Fox, 1998] [MacGibbon, 2006].

2.4.2.1 Acidi grassi saturi

Gli acidi grassi saturi maggioritari nel latte sono costituiti da catene idrocarburiche lineari contenenti un numero pari di atomi di carbonio variabile tra 4 e 26. Il loro contenuto varia tra il 55 ed il 75% del totale degli acidi grassi sia nel latte bovino che in quello ovino. Da un punto di vista quantitativo, l'acido grasso saturo più importante è l'acido palmitico C16:0 che, da solo, costituisce il 25-30% del totale mentre gli acidi miristico C14:0 e stearico C18:0 contribuiscono al totale con circa il 10% [MacGibbon, 2006]. L'acido butirrico C4:0 è contenuto in percentuali rilevanti nel latte dei ruminanti in quanto ottenuto, all'interno della ghiandola mammaria, dalla riduzione del β -idrossibutirrato proveniente dal rumine, ivi sintetizzato a partire da carboidrati per azione di microorganismi [Fox, 1998].

Contribuiscono al totale degli acidi grassi saturi, entrambi con una percentuale vicina al 2%, anche gli acidi grassi saturi con numero dispari di atomi di carbonio (compreso tra 3 e 23) e gli acidi grassi saturi ramificati, costituiti da ramificazioni monometiliche, etiliche o dimetiliche. I più abbondanti sono gli acidi grassi con ramificazioni monometiliche, rappresentati dai nove seguenti acidi in cui il metile si trova in posizione iso o anteiso: iso-C13:0, anteiso-C13:0, iso-C14:0, iso-C15:0, anteiso-C15:0, iso-C16:0, iso-C17:0, anteiso-C17:0 e iso-C18:0.

In Tabella 2.12 sono mostrati i valori medi percentuali dei principali acidi grassi saturi presenti nel latte ovino e bovino [Addis, 2005] [Creamer, 1996]. Si può notare che gli acidi grassi a catena medio-corta sono maggiormente rappresentati nel latte ovino. Questi acidi sono associati alle proprietà organolettiche dei prodotti caseari da essi derivati in quanto posseggono aromi forti e caratteristici; in particolare l'acido butirrico (C4:0) è direttamente

responsabile della comparsa del sapore piccante nei formaggi [MacGibbon, 2006]. Anche gli acidi ramificati a catena più corta contribuiscono alla caratterizzazione dei prodotti caseari; la loro importanza è infatti dovuta alla loro volatilità più che alla loro quantità [Park, 2007].

I trigliceridi costituiti da acidi grassi a catena corta o media presentano punti di fusione più bassi rispetto ai trigliceridi costituiti da acidi grassi a catena lunga, e ciò rende il grasso del latte capace di rimanere liquido a temperature fisiologiche [MacGibbon, 2006].

Tabella 2.12 - Valore medio degli acidi grassi saturi nel latte ovino e bovino.

Acido grasso	Pecora g/(100g FAME)*	Vacca g/(100g FAME)*
C4:0	4,5	3,9
C6:0	2,4	2,5
C8:0	1,7	1,5
C10:0	8,4	3,2
C12:0	4,0	3,6
C14:0	10,1	11,1
C15:0	1,4	1,2
C16:0	22,8	27,9
C17:0	0,8	0,6
C18:0	10,8	12,2

* FAME = totale degli esteri metilici degli acidi grassi

2.4.2.2 Acidi grassi insaturi

Il latte dei ruminanti contiene sia acidi monoinsaturi (MUFA) che acidi polinsaturi (PUFA), questi ultimi fino ad un massimo di sei doppi legami presenti in configurazione *cis* o *trans*. Il contenuto di acidi grassi insaturi varia tra il 20 ed il 45% del totale degli acidi grassi contenuti nel latte sia ovino che bovino; il costituente contenuto in quantità maggiore è l'acido oleico, C18:1-*cis*-9, che - da solo - costituisce il 10-20% del totale degli acidi grassi ed è l'unico acido monoinsaturo *cis* presente con concentrazioni medie superiori del 1%.

I restanti acidi monoinsaturi in configurazione *cis*, una decina nel totale, molti dei quali isomeri *cis* dell'acido oleico, contribuiscono nell'insieme ad una percentuale variabile tra l'1 ed il 10% del totale. Il più abbondante è l'acido palmitoleico, C16:1-*cis*-9, che da solo può raggiungere l'1%.

Gli acidi polinsaturi in configurazione *cis* sono presenti a basse concentrazioni a causa delle reazioni di bioidrogenazione che avvengono nel rumine. Sono rappresentati nella loro quasi totalità dall'acido linoleico, C18:2-*cis*-9, *cis*-12, e dall'acido α -linolenico, C18:3-*cis*-9, *cis*-12, *cis*-15, ognuno dei quali presenti in concentrazioni che raggiungono il 4-5% del totale degli acidi grassi.

Gli acidi grassi insaturi *trans* sono presenti nel latte grazie all'incompleta bioidrogenazione dei lipidi alimentari nel rumine e sono costituiti in maggioranza da isomeri *trans* del C18:1. Il piú abbondante di questi è l'acido vaccenico, C18:1-*trans*-11, che costituisce da solo il 2 – 4 % del totale degli acidi grassi.

Una percentuale bassa del totale degli acidi grassi, dal 2 al 5%, è data dalla somma di acidi polinsaturi contenenti doppi legami sia in configurazione *trans* che in configurazione *cis*.

Vi è poi un'ultima classe formata dagli isomeri posizionali e geometrici dell'acido linoleico, comunemente chiamati CLA (Conjugated Linoleic Acid), che contengono un sistema di doppi legami coniugati e il cui principale isomero è l'acido rumenico C18:2-*cis*-9, *trans*-11, presente in concentrazioni variabili tra lo 0,5 ed il 4% del totale degli acidi grassi del latte [MacGibbon, 2006].

Tabella 2.13 - Valore medio degli acidi grassi insaturi principali nel latte ovino e bovino.

Acido grasso	Pecora g/(100g FAME)*	Vacca g/(100g FAME)*
C14:1 <i>cis</i>-9	0,2	0,8
C16:1 <i>cis</i>-9	0,7	1,5
C18:1 <i>cis</i>-9	23,7	17,2
C18:1 <i>trans</i>-11	2,8	3,9
C18:2 <i>cis</i>-9, <i>cis</i>-12	2,2	1,4
C18:2 <i>cis</i>-9, <i>trans</i>-11	1,7	1,1
C18:3 <i>cis</i>-9, <i>cis</i>-12, <i>cis</i>-15	2,0	1,0

* FAME = totale degli esteri metilici degli acidi grassi

In Tabella 2.13 sono mostrati i valori medi percentuali dei principali acidi grassi insaturi presenti nel latte ovino e bovino [Addis, 2005] [Creamer, 1996]. Si può notare che il latte di pecora contiene maggiori quantità di acidi insaturi, in particolare di acido rumenico, i cui precursori sono l'acido linoleico e l'acido α -linolenico.

Tra gli acidi grassi polinsaturi (PUFA) si trovano acidi grassi che possono essere raggruppati in due categorie a seconda della posizione dell'ultimo doppio legame rispetto alla coda della catena acilica:

- Acidi grassi ω -3: presentano l'ultimo doppio legame sul terzo carbonio a partire dalla fine della catena alifatica. Ne fanno parte l'acido α -linolenico C18:3 ω -3, l'acido eicosapentaenoico (EPA) C20:5 ω -3, l'acido docosapentaenoico (DPA) C22:5 ω -3 e l'acido docosaesaenoico (DHA) C22:6 ω -3.
- Acidi grassi ω -6: presentano l'ultimo doppio legame sul sesto carbonio a partire dalla fine della catena alifatica. Ne fanno parte l'acido linoleico (LA) C18:2 ω -6, l'acido γ -linolenico (GLA) C18:3 ω -6 e l'acido arachidonico (AA) C20:4 ω -6.

2.4.2.3 Acidi linoleici coniugati (CLA)

Gli acidi linoleici coniugati sono dati da una miscela di isomeri posizionali e geometrici dell'acido linoleico C18:2-*cis*-9, *cis*-12. La loro importanza risiede nel fatto che possiedono numerose proprietà benefiche per la salute dell'uomo, quali antitumorali, antiobesità, antiaterogeniche ed antidiabetiche.

Come si può intuire dal nome, la differenza strutturale principale dei CLA rispetto all'acido linoleico è data dalla coniugazione dei due doppi legami presenti nella catena carboniosa. I doppi legami possono trovarsi, nei diversi isomeri, in diverse posizioni, variando tra le posizioni 7-9 e 12-14, formando così sei diversi isomeri posizionali. Ogni isomero posizionale possiede quattro possibili disposizioni geometriche dei doppi legami: *cis-cis*, *trans-cis*, *cis-trans*, *trans-trans*, generando così quattro isomeri geometrici. Nel totale, la classe dei CLA è formata da 24 possibili isomeri ma solo una parte è presente nel latte dei ruminanti.

Gli isomeri del CLA, prodotti nel rumine, sono intermedi transienti della bioidrogenazione dell'acido linoleico (C18:2-*cis*-9, *cis*-12) introdotto con la dieta. L'isomero più abbondante (dal 75 al 90% del totale) è il C18:2-*cis*-9, *trans*-11 (chiamato anche acido rumenico, nome che deriva proprio dal fatto che i CLA vengono sintetizzati nel rumine) prodotto durante la bioidrogenazione ruminale del C18:2-*cis*-9, *cis*-12 ad acido stearico (C18:0). L'intermedio di reazione più abbondante di questo processo di bioidrogenazione è l'acido vaccenico (C18:1-*trans*-11) che a sua volta viene convertito, in parte, in acido rumenico dall'enzima Δ 9-desaturasi nella ghiandola mammaria [Parodi, 1977] [Bauman, 2006]. L'acido vaccenico può derivare inoltre anche dalla bioidrogenazione ruminale del C18:3-*cis*-9, *cis*-12, *cis*-15).

Tra il latte delle diverse specie ruminanti, come visto anche in Tabella 2.13, il grasso del latte ovino contiene i più alti livelli di CLA. In Tabella 2.14 sono mostrati i range degli isomeri del CLA nel latte ovino [Luna, 2005].

Tabella 2.14 - Composizione dei CLA nel latte ovino.

Isomero	Range (% sul totale dei CLA)
<i>trans</i> -12 <i>trans</i> -14	1,31-3,47
<i>trans</i> -11 <i>trans</i> -13	1,21-5,08
<i>trans</i> -10 <i>trans</i> -12	1,17-1,77
<i>trans</i> -9 <i>trans</i> -11	1,13-1,99
<i>trans</i> -8 <i>trans</i> -10	1,05-1,37
<i>trans</i> -7 <i>trans</i> -9	0,48-0,61
12-14 (<i>cis-trans</i> + <i>trans-cis</i>)	0,52-1,83
11-13 (<i>cis-trans</i> + <i>trans-cis</i>)	0,76-4,23
10-12 (<i>cis-trans</i> + <i>trans-cis</i>)	0,28-0,41
9-11 (<i>cis-trans</i> + <i>trans-cis</i>)	76,5-82,4
8-10 (<i>cis-trans</i> + <i>trans-cis</i>)	0,11-0,71
7-9 (<i>cis-trans</i> + <i>trans-cis</i>)	3,31-9,69

2.4.2.4 Trigliceridi

I trigliceridi costituiscono il 98% dei lipidi contenuti nel latte dei ruminanti. Essendo formati da diversi tipi di acidi grassi con diversa lunghezza della catena alifatica, i trigliceridi variano in maniera consistente per quanto riguarda il grado di insaturazione ed il peso molecolare. Come si vede dalla Tabella 2.15, in cui sono mostrati i dati relativi ai trigliceridi contenuti nel latte ovino e nel latte bovino, il numero di atomi di carbonio contenuto nei trigliceridi varia tra 26 e 56. La composizione è abbastanza variabile ma circa il 35% in peso è dato dai trigliceridi con numero di atomi di carbonio compreso tra 36 e 40. Il latte ovino contiene maggiori percentuali di trigliceridi contenenti un basso numero di atomi di carbonio (da 26 a 36) ma minori percentuali di quelli con alto numero di atomi di carbonio (da 46 a 54), riflettendo le differenze dei singoli acidi grassi nel latte delle due specie.

I trigliceridi vengono sintetizzati nella ghiandola mammaria dei ruminanti tramite meccanismi enzimatici che mostrano una certa selettività nell'esterificare le diverse posizioni *sn*- della catena del glicerolo con i singoli acidi grassi. In Figura 2.11 è visualizzata la proiezione di Fischer di un trigliceride in cui sono mostrate le tre diverse posizioni *sn*-. Il C4:0, il C6:0, il C8:0 ed il C10 occupano preferenzialmente la posizione

sn-3; il C12:0 ed il C14:0 occupano di preferenza la posizione *sn*-2; il C16:0 occupa in maggioranza le posizioni *sn*-1 ed *sn*-2 mentre il C18:0 occupa preferenzialmente la posizione *sn*-1. Il C18:1 si trova in maggiori percentuali nelle posizioni *sn*-1 ed *sn*-3 mentre l'acido C18:2 si trova di preferenza nella posizione *sn*-2 [Jensen, 1991] [Fox, 2008].

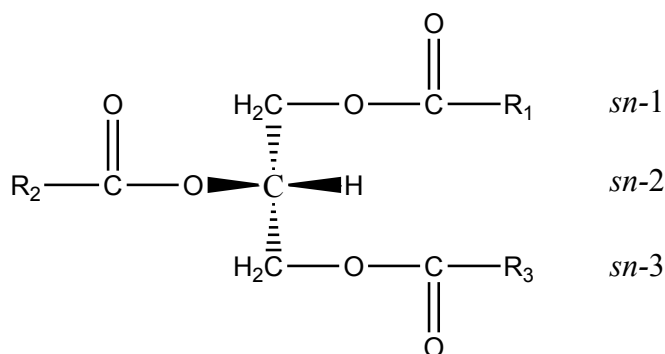


Figura 2.11 - Proiezione di Fischer di un trigliceride.

Tabella 2.15 - Composizione in trigliceridi del latte ovino e bovino [Recio, 2009].

Trigliceride (numero totale di atomi di carbonio delle catene aciliche)	Pecora (% in peso)	Vacca (% in peso)
C26	0,60-1,00	0,20-0,30
C28	1,30-2,50	0,40-0,80
C30	2,00-3,40	0,80-1,90
C32	3,10-5,00	1,80-3,20
C34	5,70-7,00	4,40-6,90
C36	9,20-10,30	9,10-12,40
C38	12,50-14,00	11,80-14,60
C40	11,00-12,50	9,50-12,10
C42	8,00-9,70	6,20-7,90
C44	7,30-8,60	5,40-7,80
C46	6,50-6,90	5,60-8,30
C48	6,10-7,80	6,90-10,70
C50	4,70-9,10	9,70-12,80
C52	5,00-9,40	7,20-12,60
C54	3,10-4,30	2,70-7,80
C56	< 0,50	0,40-0,60

3 La sintesi dei costituenti del latte: la ghiandola mammaria

Il latte si forma nelle cellule dell'epitelio ghiandolare della mammella, riempiendo gli alveoli, o acini, che ivi si raggruppano numerosi. La sintesi della maggior parte dei costituenti organici del latte avviene nella ghiandola mammaria; il sangue trasporta le sostanze base (glucosio, acetati, β -idrossibutirrato, amminoacidi, acidi grassi etc...) che, raggiunte le cellule epiteliali, vengono trasformate, tramite processi enzimatici, nei grassi, nelle proteine e nello zucchero (il lattosio) caratteristici del latte. Altri costituenti minori del latte provengono direttamente dal sangue, senza subire alcun tipo di trasformazione chimica. Le ghiandole mammarie hanno la stessa struttura di base per tutti i mammiferi, sono localizzate esternamente al corpo dell'animale e regolano la successione delle varie fasi della sintesi e della secrezione del latte. Il latte si accumula nelle cellule secretorie che formano uno strato all'interno degli alveoli della ghiandola mammaria, intorno al lumen. Le sostanze base, trasportate dal sangue, entrano nelle cellule secretorie attraverso la membrana basale dove avviene la trasformazione nei vari costituenti del latte che sono secreti nel lumen attraverso la membrana apicale. I vari costituenti si formano in diverse regioni della cellula: i globuli di grasso vengono a formarsi nella parte basale della cellula, circondandosi da una sottile membrana e vanno a muoversi verso la parte apicale. Quando il globulo di grasso emerge nel lumen, viene involuppato da uno strato di membrana apicale la cui rigenerazione è legata alla sintesi delle catene polipeptidiche. Essa avviene nel reticolo endoplasmatico su cui si trovano i ribosomi. Le proteine formate entrano poi nell'apparato del Golgi dove si uniscono a gruppi prostetici e possono dar luogo ad associazioni con altri costituenti non lipidici del latte. Le vescicole secretorie muovono dall'apparato del Golgi alla regione apicale della cellula dove si ha il passaggio del loro contenuto nel lumen, mentre la loro membrana si fonde con la membrana apicale (Figura 3.1) [Corradini, 1995].

Molteplici sono le reazioni chimiche che portano alla formazione dei costituenti del latte. Il lattosio viene formato principalmente a partire dal glucosio che viene isomerizzato in galattosio nel tessuto mammario attraverso il passaggio di due forme intermedie, il glucosio fosfato e l'UDP-glucosio. L'enzima lattosio-sintetasi condensa i monomeri glucosio e galattosio per formare il lattosio. Nei ruminanti, una piccola percentuale di lattosio (circa il 10%) è sintetizzato a partire da acidi grassi volatili quali l'acido propionico ed acetico.

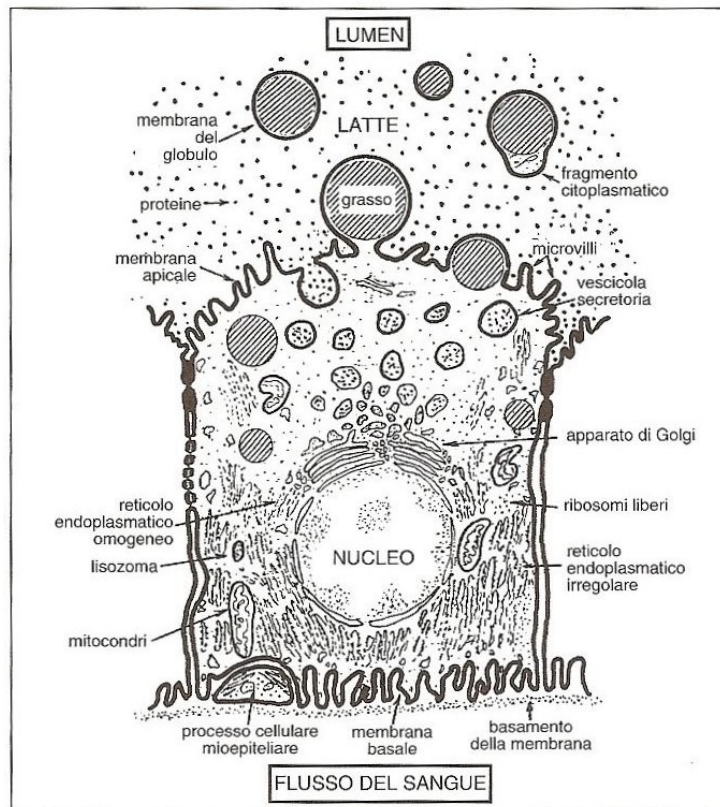


Figura 3.1 – Rappresentazione di una cellula dell'epitelio secretorio di una ghiandola mammaria [Corradini, 1995].

Le proteine vengono sintetizzate nella ghiandola mammaria a partire dagli amminoacidi liberi. La formazione delle catene polipeptidiche avviene probabilmente per attivazione, transfer per mezzo di un RNA e successiva condensazione dei singoli amminoacidi, processi che avvengono anche in altri tessuti dell'organismo. Gli amminoacidi liberi provengono principalmente dal sangue ma una parte di essi, quelli non essenziali, viene sintetizzata nella ghiandola mammaria a partire da glucosio od acetato e da una sorgente di azoto non ancora identificata.

Il fosforo inorganico proveniente dal sangue viene incorporato nella catena polipeptidica delle caseine formate grazie all'intervento di una fosfo-proteina kinasi. Nella caseina il fosforo si trova sotto forma di monoestere ortofosforico. In Figura 3.2 sono schematizzate le vie generali della lattogenesi mentre in Tabella 3.1 sono mostrate le composizioni medie del latte e del sangue [Alais, 2010].

Gli acidi grassi presenti nel latte provengono da due diversi processi: possono venir sintetizzati *de novo* all'interno della ghiandola mammaria oppure provenire direttamente dall'alimentazione, passando per il plasma sanguigno. Una delle differenze principali tra i

due processi è il tipo di acidi grassi coinvolto: gli acidi a catena corta e media, dal C4:0 al C14:0 e parte del C16:0 sono sintetizzati *de novo* mentre gli acidi a lunga catena, tra cui i C18 e la restante parte del C16:0, provengono dal sangue. Il 55% in peso degli acidi grassi totali del latte proviene dall'alimentazione mentre la sintesi *de novo* contribuisce con il restante 45% [Moore, 1979].

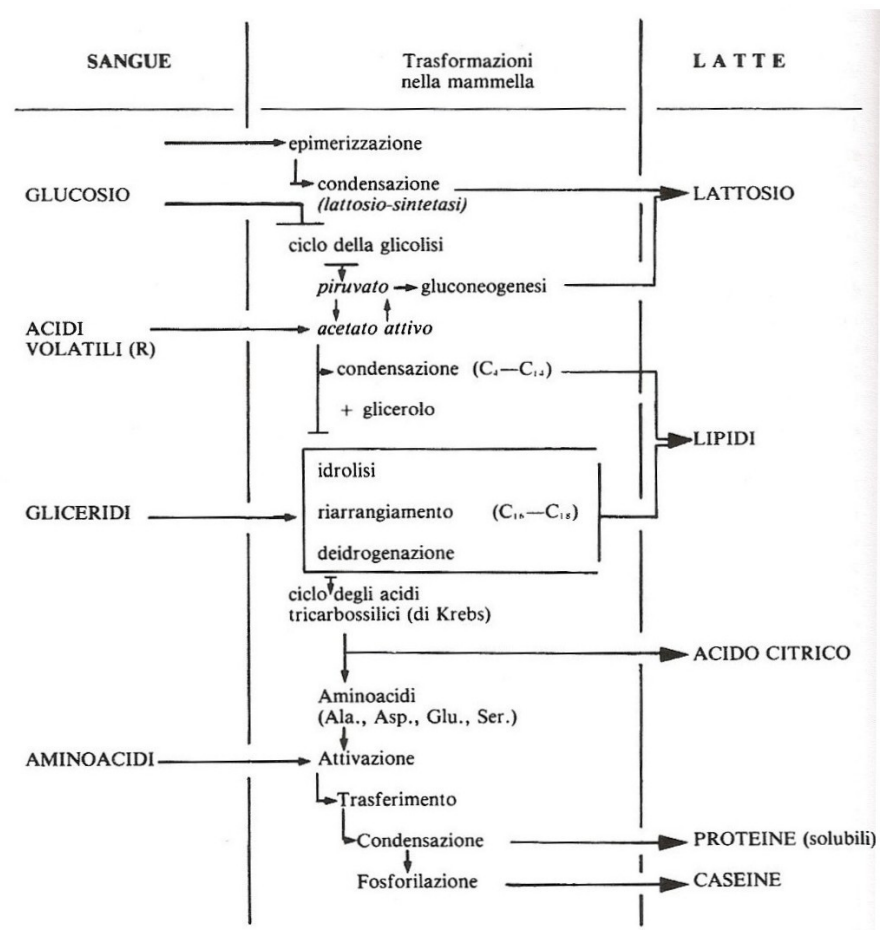


Figura 3.2 – Vie generali della lattogenesi

La sintesi *de novo* avviene a partire da acetato e β -idrossibutirrato provenienti dalla fermentazione microbica della cellulosa del rumine. All'interno della ghiandola mammaria, l'acetato viene convertito in Acetil-CoA che, per azione dell'enzima Acetil-CoA Carbossilasi (ACC), forma il malonil-CoA, essenziale nel processo di elongazione della catena alchilica che avviene gradualmente tramite la condensazione di gruppi di due atomi di carbonio per volta, catalizzata dall'enzima Fatty Acid Synthase (FAS). Vengono quindi a formarsi gli acidi grassi a catena corta e media con numero pari di atomi di

carbonio fino al C16:0. Parte del C14:0 e del C16:0 viene trasformata, nella ghiandola mammaria, in C14:1 e C16:1 per azione dell'enzima Δ^9 -desaturasi.

Gli acidi grassi con numero di atomi di carbonio dispari, così come gli acidi grassi ramificati, vengono sintetizzati a partire dal propionato anziché dall'acetato e β -idrossibutirrato.

Gli acidi grassi a catena lunga del latte provengono invece dalla dieta dell'animale; i lipidi che vengono ingeriti dall'animale sono principalmente glicolipidi, fosfolipidi e trigliceridi formati in maggioranza da acido linoleico C18:2-*cis*-9, *cis*-12, acido α -linolenico C18:3-*cis*-9, *cis*-12, *cis*-15 e acido γ -linolenico C18:3-*cis*-6, *cis*-9, *cis*-12. Nel rumine i lipidi vengono idrolizzati e successivamente bioidrogenati da microorganismi [Jenkins, 1993].

Tabella 3.1 - Composizioni medie del latte e del plasma sanguigno della vacca [Alais, 2010].

	grammi / litro	
	Latte	Plasma sanguigno
Grasso	35	3
Lattosio	49	0
Caseina	27	0
α-lattalbumina + β-lattoglobulina	4	0
Albumina + globulina	1,5	75
Acido citrico	2	0
Cloruri	1,6	6
(in P₂O₅) fosfati	2,5	0,3

La sequenza di bioidrogenazione per l'acido linoleico inizia con un passaggio di isomerizzazione che produce acido rumenico C18:2-*cis*-9, *trans*-11 (CLA) il cui legame *cis* viene idrogenato a formare l'acido vaccenico C18:1- *trans*-11 dal quale si forma, per ulteriore riduzione, l'acido stearico C18:0. Le bioidrogenazioni dei due acidi linolenici seguono un percorso simile, come mostrato in Figura 3.3.

Ogni reazione coinvolta nella biosintesi degli acidi grassi non va a completamento e la miscela degli acidi grassi formati viene esterificata a trigliceridi che, passando dal rumine al sangue, arrivano alla ghiandola mammaria [MacGibbon, 2006]. La maggior parte degli acidi grassi provenienti dal rumine passa nel tessuto mammario e quindi nel latte senza subire variazioni; l'acido stearico viene invece convertito in parte in acido oleico mentre l'acido vaccenico è in parte riconvertito in acido rumenico (Figura 3.4).

Le reazioni sono entrambe catalizzate dall'enzima Δ^9 -desaturasi. Come si può vedere in Figura 3.5 vi è una forte correlazione tra la quantità di acido vaccenico e di acido rumenico nel latte ed il rapporto tra le loro concentrazioni è di circa 3:1 [Bauman, 2011].

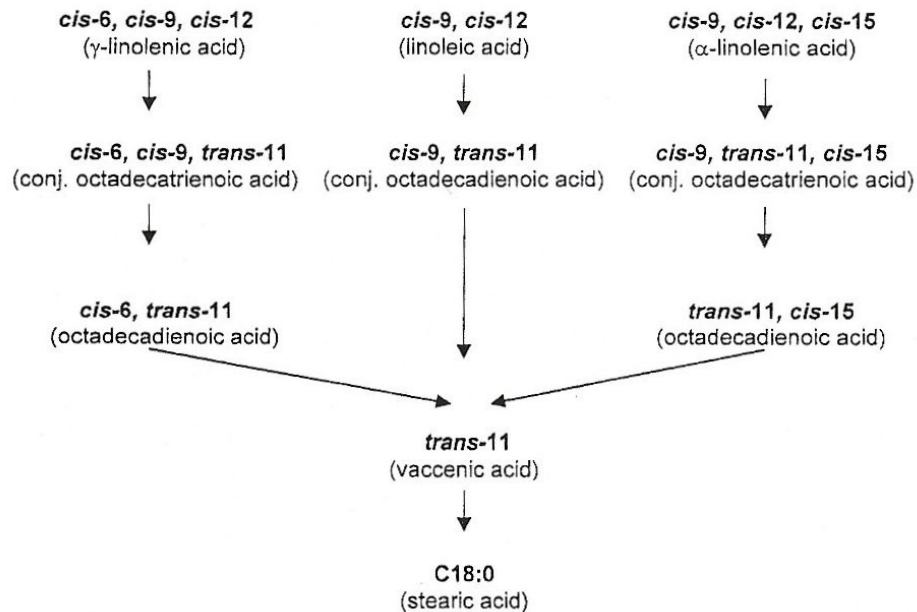


Figura 3.3 - Schema della bioidrogenazione degli acidi grassi insaturi C18:2 e C18:3 [Griinari, 1999].

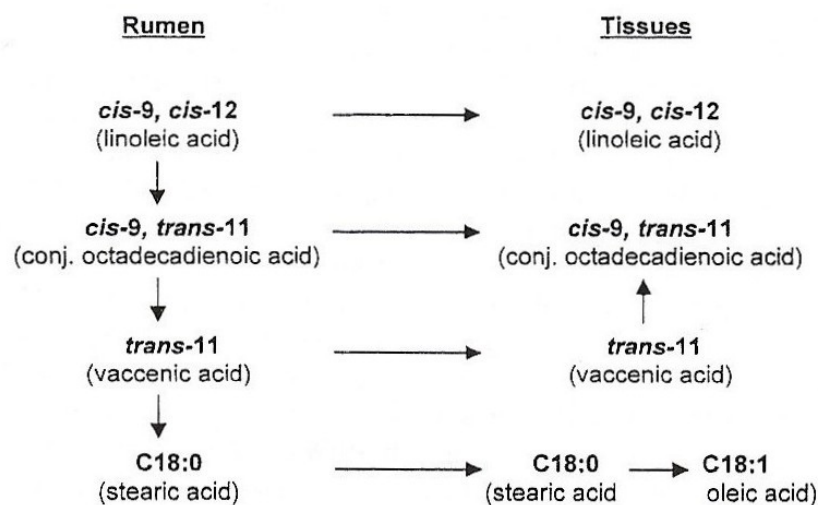


Figura 3.4 - Schema della biosintesi degli acidi linoleici coniugati [Griinari, 1999].

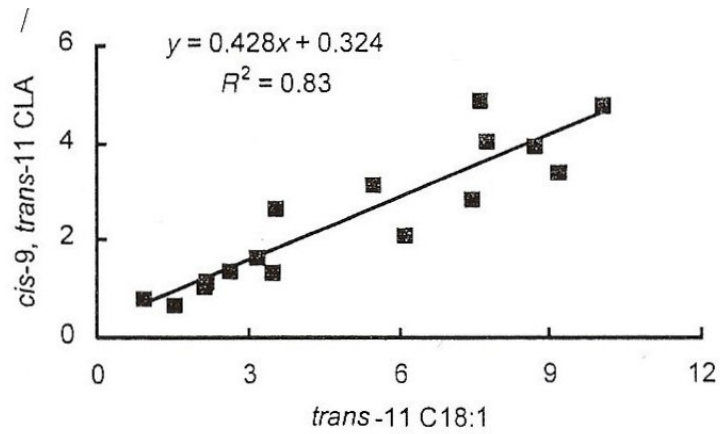


Figura 3.5 - Relazione tra acido vaccenico e rumenico nel grasso del latte (g/100g FAME) [Grinari, 1999].

4 Cause di variazione della composizione del latte

Come già più volte sottolineato in precedenza, la quantità di latte prodotta da un mammifero, la sua composizione, e di conseguenza le sue proprietà chimico-fisiche, possono variare considerevolmente in funzione di molteplici fattori, quali:

- Fattori genetici (razza, individuo; effetto della selezione).
- Fattori fisiologici (età, periodo di lattazione, stato di salute).
- Fattori alimentari (quantità, livello energetico e composizione degli alimenti).
- Fattori ambientali (condizioni igieniche, stagione, clima, altitudine).
- Fattori tecnologici (mungitura, trattamenti di conservazione).

La presenza di così numerosi fattori rende difficoltosa la possibilità di distinguere l'influenza specifica di ciascuno di essi. Per questo motivo vengono studiati maggiormente i fattori che causano le variazioni più significative sulle proprietà del latte e sugli equilibri dei suoi elementi strutturali. Un aspetto da considerare della produzione del latte è dato dalle fluttuazioni giornaliere che avvengono nel singolo individuo anche quando tutti i fattori di variazione vengono mantenuti costanti; gli scarti giornalieri della quantità di latte prodotto dall'individuo sono in media del 5-6%; maggiori invece sono gli scarti del tenore di grasso che in media si aggirano tra 7-8%, raggiungendo in certi giorni ed in alcuni individui anche il 20%. Proteine e lattosio hanno invece una variabilità media del 2,5%. È probabile che le fluttuazioni siano dovute alle condizioni climatiche o alle modalità di mungitura [Corradini, 1995] [Alais, 2010].

4.1 Fattori genetici

La produzione e la composizione del latte di animali della stessa razza, messi nelle medesime condizioni geoclimatiche e di alimentazione, presentano differenze significative e riproducibili. È quindi importante conoscere il tasso di variabilità dei componenti del latte in particolare per quel che riguarda gli studi relativi alla selezione e al miglioramento della produttività e della qualità. Per quel che riguarda i ruminanti, di seguito possiamo vedere i valori medi del coefficiente di variazione individuale:

- Produzione del latte: 12-20%
- Grasso: 8%
- Proteine: 6%

- Lattosio: 4%
- Residuo magro e sali: 3%

Dai dati riportati pare evidente che la variabilità maggiore la si riscontra per la quantità di latte prodotto, mentre tra i costituenti maggioritari quello soggetto alla maggiore variabilità individuale è il grasso [Alais, 2010] [Corradini, 1995].

Differenze importanti si hanno anche nella produzione e nella composizione del latte proveniente da razze diverse di una stessa specie animale. In Tabella 4.1 sono mostrati i valori medi di alcuni costituenti maggioritari e del profilo acidico del latte ovino di pecore di tre diverse razze, tra cui la razza Sarda, allevate nello stesso gregge. Si può osservare che le pecore della razza Sarda sono più produttive rispetto alle pecore delle altre due razze ma il latte è caratterizzato da un minore tenore di grasso e proteine. Esiste infatti una relazione inversa tra quantità di latte prodotto e quantità dei costituenti; si può affermare che in generale la produzione delle razze il cui latte è ricco in costituenti è minore di quella delle razze il cui latte è più povero di costituenti. Il tenore dei singoli costituenti è invece correlato positivamente e la variabilità è più marcata per il grasso, così come visto nei singoli individui. Tra individui della stessa razza vi è una correlazione positiva tra grasso e proteine con coefficiente di correlazione r pari a 0,99 [Signorelli, 2008].

Relativamente agli acidi grassi, in generale il loro contenuto non è invece correlato con la quantità di latte prodotto; solo la somma degli acidi polinsaturi sembra aumentare all'aumentare del volume del latte prodotto. Vi sono invece correlazioni positive e negative tra la concentrazione degli acidi grassi e la percentuale di grasso nel latte; in particolare, gli acidi grassi a corta e media catena, così come il CLA, sono presenti in minore quantità quando aumenta il tenore di grasso mentre l'acido stearico è correlato positivamente con il grasso totale [Signorelli, 2008].

Oltre che dalla specie, il profilo acidico del latte è influenzato chiaramente anche dalla razza dell'animale produttore: si può osservare che il latte della razza Sarda ha quantità minori di acidi grassi saturi e maggiori di acidi grassi monoinsaturi e polinsaturi, con concentrazioni di C18:2, C18:3 e CLA *cis*-9, *trans*-11 superiori anche del 30% rispetto a quelle presenti nel latte di altre due razze quali l'Altamura e la Gentile di Puglia [Signorelli, 2008].

Tabella 4.1 - Composizione del latte di diverse razze ovine [Signorelli, 2008].

	razza Altamura	razza Gentile di Puglia	razza Sarda
Produzione latte (cl)	320,08	195,00	519,54
Grasso %	9,53	9,65	7,46
Proteine %	6,73	6,93	5,45
	mg / g grasso	mg / g grasso	mg / g grasso
C4:0	4,58	4,18	5,13
C6:0	2,87	2,83	2,75
C8:0	2,39	2,52	2,20
C10:0	6,32	6,94	5,56
C12:0	3,96	4,07	3,51
C14:0	10,48	10,03	9,25
C16:0	25,26	23,51	23,24
C16:1	1,27	1,28	1,11
C18:0	9,27	9,91	9,43
C18:1	23,66	24,84	26,00
C18:2	3,14	2,94	4,18
C18:3	0,60	0,55	0,83
CLA c9,t11	1,56	1,54	2,00
Saturi	66,70	65,62	62,85
Polinsaturi	5,31	5,03	7,00
Monoinsaturi	25,74	26,99	27,80

4.2 Fattori fisiologici

Le variazioni più significative della produzione di latte, nonché della sua composizione, sono dovute al benessere animale. La produzione giornaliera di latte e la sua composizione variano in maniera non uniforme durante la lattazione. Per una vacca giovane, subito dopo e per tutto il primo mese dopo il parto si osserva un aumento progressivo della quantità di latte prodotta fino a raggiungere un massimo nel corso del secondo mese. Da questo periodo in poi si ha un calo graduale della produzione che si protrae fino all'ottavo mese, a partire dal quale inizia il periodo di fine lattazione in cui si ha il calo vero e proprio fino all'azzeramento della produzione del latte che normalmente avviene al decimo mese. Una vacca in età adulta presenta invece durante il primo mese una produzione di latte più elevata rispetto a quella di una giovane vacca, fino al raggiungimento del massimo nel secondo mese, con però una successiva diminuzione molto più accentuata, con un calo medio del 10% per ogni mese. Un andamento del genere si ha anche nel caso di animali non alimentati in maniera corretta o con sistemi di mungitura non adeguati.

La composizione dei componenti maggioritari del latte varia durante il periodo di lattazione come mostrato in Figura 4.1. Il grasso e le proteine tendono a diminuire

rispettivamente durante le prime 15 e 10 settimane per poi aumentare in misura sempre più marcata fino al periodo di fine lattazione. Il lattosio, per contro, subisce un lieve aumento durante le prime 10 settimane circa per poi calare una volta superato questo periodo e fino al periodo di fine lattazione.

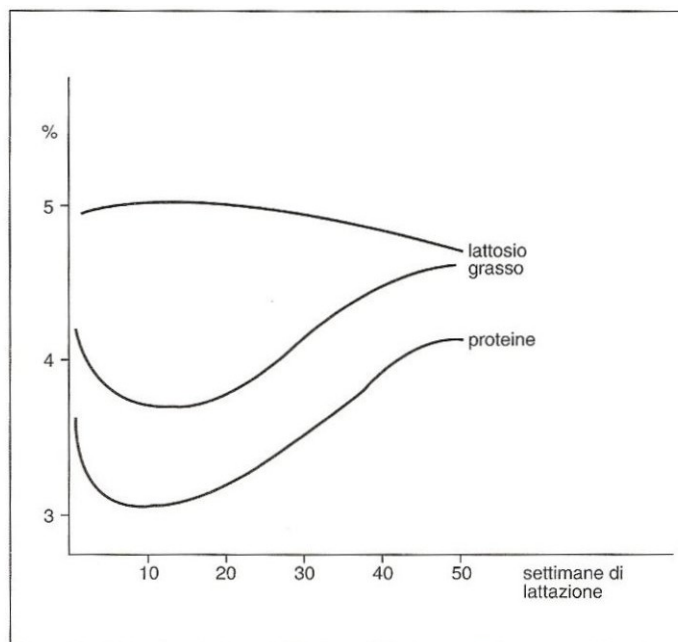


Figura 4.1 - Andamento dei costituenti maggioritari del latte durante il periodo di lattazione.

Le variazioni della composizione sono quindi più evidenti all'inizio e alla fine del periodo di lattazione. Nei sette giorni dopo il parto l'animale secerne un liquido, definito colostro, che si differenzia notevolmente dal latte per composizione essendo molto più ricco di grasso e proteine e più povero d'acqua, lattosio e sali (potassio). Il colostro è particolarmente ricco di immunoglobuline dato che il piccolo ha bisogno di formare anticorpi nei suoi primi giorni di vita. La composizione del colostro muta molto velocemente e nel giro di 5-6 giorni si discosta poco da quella del latte normale. Il colostro è meno stabile del latte ai trattamenti termici e non è adatto alla caseificazione dato che il coagulo ottenuto ha caratteristiche reologiche (compattezza, morbidezza, gommosità, elasticità) non favorevoli. Il latte di fine lattazione ha un contenuto di materia secca più elevato, avendo maggiori concentrazioni di grasso, proteine e ceneri, e presenta differenze anche nella concentrazione dei componenti salini; in particolare è più ricco di calcio e sodio e più povero di potassio e fosforo. Per questo motivo, e soprattutto per il variare degli equilibri calcio/fosforo, il latte di fine lattazione è meno stabile ai trattamenti termici.

In Figura 4.2 si può vedere l'andamento del calcio e del fosforo durante il periodo di lattazione.

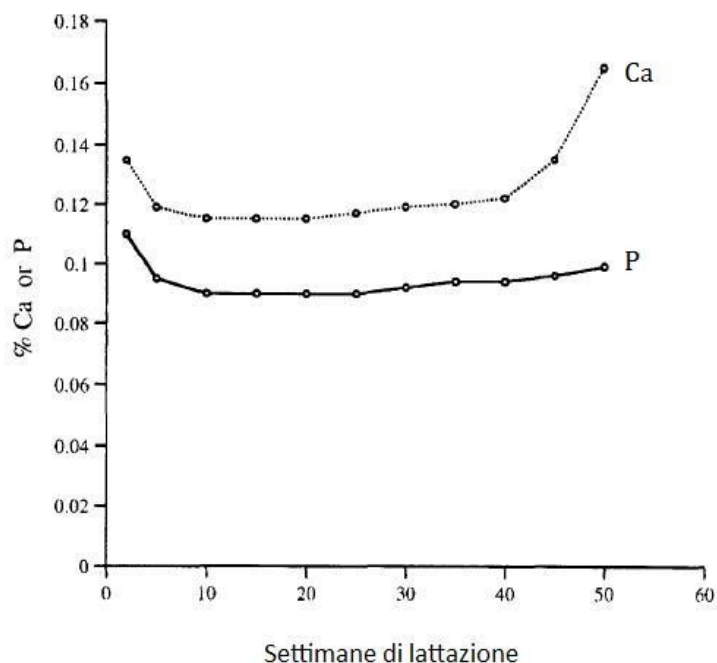


Figura 4.2 - Andamento della concentrazione di calcio e fosforo durante la lattazione [Fox, 1998].

Una diminuzione della produzione del latte e l'alterazione della sua composizione avvengono anche a causa di malattie dell'animale, in particolare a causa dei fenomeni mastitici, ovvero delle infezioni a carico della mammella. Il latte prodotto dalla mammella dell'animale malato è ricco di cellule somatiche (globuli bianchi o leucociti), la cui quantità è proporzionale alla gravità dell'infezione, e di immunoglobuline, conseguentemente alle azioni di difesa dell'organismo dell'animale. La ghiandola mammaria infetta ha ridotta capacità di sintesi, perciò il latte secreto ha minori contenuti di grasso, lattosio e caseina, nonché di fosforo e calcio, in quanto componenti delle micelle caseiniche.

Il minore contenuto di lattosio comporta un aumento nel latte di sostanze provenienti dal sangue come la sieralbumina, l'immunoglobulina, la lattoferrina, i sali di sodio e potassio ed alcuni enzimi che determinano una diminuzione del pH del latte che si porta a 6,6 contrariamente ai valori normali 6,7 - 6,8. Le micelle caseiniche del latte mastitico hanno un diametro minore del normale, e ciò porta a difficoltà nella coagulazione. La cagliata

formata da latte normale miscelato a latte mastitico presenta delle zone con diverse proprietà reologiche, presupponendo che le micelle caseiniche del latte mastitico coagulino con diverse modalità e tempi di reazione. La presenza di enzimi di origine cellulare possono inoltre determinare la comparsa di difetti di gusto nei formaggi prodotti [Corradini, 1995].

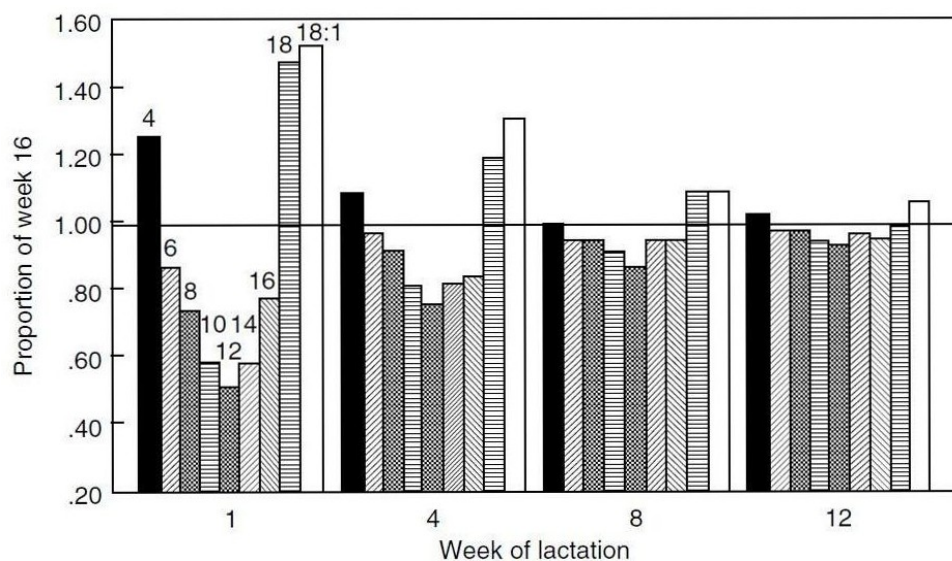


Figura 4.3 - Proporzioni degli acidi grassi alle settimane di lattazione 1,4,8,12 relativamente ai valori della settimana 16 [Palmquist, 1993].

Anche il profilo acido del latte, come si vede dalla Figura 4.3, varia in funzione dello stato fisiologico dell'animale. La concentrazione del C4:0 è alta ad inizio lattazione e decresce fino all'ottava settimana, mantenendo da lì in poi una concentrazione pressoché costante fino alla fine della lattazione. La concentrazione degli acidi grassi sintetizzati *de novo* nella ghiandola mammaria (dal C6:0 al C16:0) è inizialmente bassa ed aumenta nel primo periodo della lattazione; per contro, la concentrazione degli acidi con 18 atomi di carbonio diminuisce. Questo è dovuto al fatto che, ad inizio lattazione, il bilancio energetico dell'animale è negativo, causando la mobilitazione del grasso del tessuto adiposo, costituito da acidi a lunga catena, ed il suo passaggio nel latte. Questo processo inibisce la sintesi *de novo* degli acidi a catena media e corta. L'inibizione diminuisce sempre più con il ristabilimento dell'animale, con conseguente aumento della concentrazione degli acidi a corta e media catena e la diminuzione della concentrazione degli acidi a lunga catena [Palmquist, 1993].

4.3 Fattori alimentari

L'alimentazione è un fattore importante non solo perché legata alla quantità di cibo ingerito, ma perché relazionata al territorio in cui gli animali vengono allevati. I due sistemi principali di allevamento dei ruminanti sono il pascolo (sistema estensivo) e l'allevamento in stalla (sistema intensivo); il primo viene generalmente considerato come il sistema migliore in quanto di solito più vario e ricco rispetto al secondo, ma vi sono zone in cui il pascolo povero deve essere comunque compensato da un'alimentazione supplementare.

L'alimentazione dell'animale ha una grande influenza sulla quantità e sulla qualità del latte prodotto, tanto da giocare un ruolo importante nella produttività di un allevamento. Se la dieta di un animale, solitamente alimentato secondo il suo fabbisogno, viene ridotta bruscamente e temporaneamente, si avrà, nel giro di soli due giorni, un calo della produzione del latte ed un aumento dell'estratto secco; il grasso avrà una maggiore quantità di acidi insaturi a causa dell'aumento dell'acido oleico. Quando la riduzione dell'apporto alimentare viene prolungata per lunghi periodi di tempo, il latte prodotto dall'animale sarà non solo minore in quantità, ma anche più povero in tutti i suoi costituenti principali, mentre l'animale subirà un dimagrimento causato dall'utilizzo delle sue riserve corporee per la secrezione del latte. Un'alimentazione che invece copre ogni fabbisogno dell'animale ha come conseguenza un aumento della produzione del latte a fronte di una sostanziale invarianza della sua composizione. In ultimo, se si varia l'apporto alimentare da un livello inferiore ad uno superiore del fabbisogno dell'animale, si avrà un aumento continuo del titolo proteico e dell'estratto secco mentre grasso e lattosio non variano significativamente.

Un'alimentazione povera di grassi ha come effetto una riduzione della produzione del latte e della quantità di grasso in esso contenuto. Gli acidi grassi che subiscono una maggiore variazione sono quelli a 18 atomi di carbonio: la loro concentrazione viene ridotta notevolmente, a tutto vantaggio del C16:0, che può raggiungere concentrazioni anche pari al 50% con conseguente depauperamento del valore nutrizionale del latte. Esso infatti diminuisce all'aumentare del rapporto C16:0/C18:0, a causa dell'elevato potere aterogenico del C16:0. Un'alimentazione a base di concentrati poveri in acido palmitico, come i semi di colza e di girasole, ha come effetto l'aumento dei livelli di acido stearico a discapito della quantità di acido palmitico. Un'alimentazione che invece è ricca di acidi grassi a con 18 atomi di carbonio fa aumentare la resa di tutti C18 nel latte. La percentuale di C18 assorbito dalla dieta e trasferito al latte non supera comunque mai il 60% [Banks,

1976]. L'alimentazione a base di oli e semi protetti, ricchi in acido oleico (soia e colza), causa un aumento della quantità di acido oleico e diminuzione dello stearico con conseguente aumento del valore nutrizionale del latte e della fluidità del grasso del latte, utile per fini tecnologici.

L'alimentazione ha una grande influenza sulla sintesi *de novo* degli acidi grassi, con effetto sulla proporzione di acetato, β -idrossibutirrato e propionato prodotti nel rumine. Questi composti sono sintetizzati, attraverso una serie di reazioni enzimatiche catalizzate dagli enzimi Acetil-CoA Carbossilasi (ACC) e Fatty Acid Synthase (FAS), a partire da carboidrati introdotti con la dieta e sono i precursori degli acidi grassi a catena media e corta sintetizzati *de novo* nella ghiandola mammaria. Per ottenere una produzione equilibrata di questi acidi, la razione alimentare somministrata agli animali deve provvedere un rapporto foraggio/concentrato non inferiore a 60:40 con conseguente rapporto dei precursori nel rumine pari a 65:20:10. La sintesi *de novo* viene inoltre inibita dalla presenza di acidi grassi a lunga catena in quanto ostacolano l'attività dell'Acetil-CoA Carbossilasi e l'inibizione è tanto maggiore quanto più è alto il numero di atomi di carbonio della catena ed il suo grado di insaturazione. I maggiori inibitori della sintesi *de novo* nella ghiandola mammaria sono alcuni isomeri del C18:1 *trans* e del CLA, in particolare il C18:1 *trans*-10 ed il CLA *trans*-10, *cis*-12 [Chilliard, 2002].

Gli acidi grassi a lunga catena che non vengono sintetizzati nella ghiandola mammaria ma che provengono direttamente dall'alimentazione presentano concentrazioni nel latte che dipendono direttamente dal loro contenuto nella razione alimentare, dalle entità della bioidrogenazione che subiscono nel rumine e del trasferimento dal rumine alla ghiandola mammaria. Fanno parte di questa categoria gli acidi polinsaturi come l'acido linoleico e l'acido linolenico. I metodi per aumentare il livello di questi acidi grassi nel latte si basano sull'introduzione nella dieta dell'animale di alimenti che ne sono ricchi. L'aumento dell'acido linoleico può essere ottenuto con l'utilizzo di mangimi concentrati additivati a base di olio di girasole, di soia, di colza o di cotone, sotto forma di capsule protette o a base di semi di girasole mentre quello dell'acido linolenico attraverso l'utilizzo di mangimi additivati a base di semi di lino o di olio di lino. Un'alternativa poco utilizzata è quella di agire su fattori che diminuiscono la bioidrogenazione a livello ruminale come la velocità di transito e la protezione. La fonte principale di acido linolenico è comunque l'erba del pascolo e il foraggio fresco mentre le operazioni di sfalcio e di affienatura del foraggio ne abbassano la concentrazione. Infatti il latte degli animali alimentati con erba fresca è più ricco di acido linolenico rispetto a quello di animali nutriti con fieno e concentrato. La

percentuale di grasso - e di conseguenza la concentrazione di acido linolenico nell'alimento - varia a seconda della specie foraggiera, della varietà, della fase fenologica e della stagione.

Gli acidi linoleico e linolenico sono i precursori per la sintesi di acido vaccenico e rumenico (CLA) importanti dal punto di vista nutraceutico. Un'alimentazione ottimale per incrementare il contenuto di acido vaccenico e rumenico nel latte è quella che prevede l'utilizzo di pascoli di buona qualità ed una ridotta integrazione con mangimi concentrati [White, 2001]. In particolare per quanto riguarda le pecore di razza Sarda, le maggiori quantità di acido rumenico sono state ottenute nel latte prodotto da pecore alimentate in pascoli costituiti da specie appartenenti alle famiglie delle leguminose o delle compositae, quest'ultime sia in purezza sia in consociazione binaria, rispetto al pascolo su graminacee [Cabiddu, 2003] [Addis, 2005]. Il contenuto di CLA nel latte può essere incrementato anche attraverso una dieta integrata con mangimi integrati con semi o oli di girasole e di lino [Zhang, 2006] [Mele, 2007] [Nudda, 2006] [Addis, 2009], ma alcuni studi hanno evidenziato che tali alimentazioni, oltre ad incrementare il contenuto di CLA, favoriscono anche la formazione di alcuni acidi grassi *trans*, dannosi per la salute del consumatore [Dhiman, 2000] [Bauman, 2003].

4.4 Fattori ambientali

La stagionalità determina una modificazione della curva di lattazione. Ad esempio per la specie bovina, in estate, il latte è più povero in grasso, proteine, ed in residuo magro mentre in autunno presenta le concentrazioni maggiori. La quantità di latte prodotto ha invece l'andamento inverso, diminuendo dall'estate all'autunno. Secondo alcuni studi effettuati in locali condizionati, l'effetto stagionale è determinato probabilmente dalla temperatura registrata nei periodi citati ed è indipendente dalla regione geografica e dall'alimentazione dell'animale. A temperature maggiori di 27°C si ha una diminuzione rapida della quantità di latte prodotto ed un aumento del grasso (causato soprattutto dalla perdita di appetito dell'animale) mentre al di sotto di 5°C si ha lo stesso effetto ma in maniera più graduale. Le temperature ideali sono comprese tra 5 e 27°C. Probabilmente però, è più corretto ricercare la causa delle variazioni del latte nella diversificazione del tipo di alimentazione a seguito del cambio di stagione [Alais, 2010] [Corradini, 1995].

4.5 Fattori tecnologici

I fattori zootecnici che influenzano la quantità e la composizione del latte sono dati dalle operazioni di mungitura e dai trattamenti di conservazione a cui viene sottoposto il latte prima del suo utilizzo in fase industriale. La mungitura influisce soprattutto sul tenore in grasso che, come si vede per i bovini in Tabella 4.2, aumenta dalla prima alla quarta frazione della stessa. Una mungitura incompleta può quindi portare all'ottenimento di un latte più povero di grasso. La produzione del latte può essere aumentata, entro certi limiti, aumentando il numero di mungiture effettuate giornalmente. Il latte così prodotto ha infatti minore tenore in grasso e in residuo secco magro, è più difficilmente coagulabile e quindi meno utilizzabile nei processi caseari. Durante il trasporto dal luogo della mungitura fino al caseificio o alla centrale di raccolta, il latte viene refrigerato. L'abbassamento della temperatura provoca variazioni negli equilibri salini e nelle associazioni sia delle proteine che dei globuli di grasso, variazioni che sono tanto maggiori quanto minori sono le temperature raggiunte e quanto più lunghi sono i tempi di raffreddamento. Si ha così un aumento della frazione solubile delle caseine con conseguente peggioramento della capacità di coagulazione presamica del latte ed un più difficile affioramento della crema [Corradini, 1995].

Tabella 4.2 - Variazioni del contenuto percentuale di grasso durante la mungitura [Webb, 1974].

	Vacca n.1	Vacca n.2	Vacca n.3
Prima frazione	0,57	2,09	1,73
Seconda frazione	1,82	2,66	2,65
Terza frazione	4,15	3,66	3,82
Quarta frazione	5,56	6,42	4,80

5 Aspetti nutrizionali dei lipidi

Il grasso è una fonte concentrata di energia che soddisfa al 30-40% il fabbisogno energetico fornito con l'alimentazione. Il grasso impartisce gustosità agli alimenti, funge da trasporto per le vitamine liposolubili A, D, E e K e fornisce gli acidi grassi essenziali per l'organismo. I lipidi, presenti sotto forma di trigliceridi, hanno nell'organismo funzione di riserva energetica e di tessuti (tessuto adiposo) cui è in parte demandata la termostatazione dell'organismo. Sono inoltre una componente vitale delle membrane cellulari e partecipano alle segnalazioni a cascata inter- ed intra- cellulari.

La digestione dei lipidi ha inizio nello stomaco, ove si ha il distacco dei lipidi dal cibo (nel quale sono generalmente associati alle proteine) e dove fenomeni di agitazione formano emulsioni olio in acqua stabilizzate dai fosfolipidi. L'emulsione di grasso entra nella parte superiore dell'intestino dove, per azione delle lipasi pancreatiche, i trigliceridi vengono idrolizzati con conseguente rilascio dei singoli acidi grassi [Gurr, 1999]. Gli acidi grassi a catena corta e media, assorbiti nell'intestino, raggiungono il fegato dove vengono rapidamente metabolizzati attraverso una β -ossidazione mentre gli acidi grassi a catena lunga vengono riconvertiti in trigliceridi e trasportati principalmente nei tessuti adiposo e muscolare. Nel tessuto adiposo, i trigliceridi vengono immagazzinati negli adipociti per fungere da riserva energetica mentre nel tessuto muscolare i trigliceridi vengono idrolizzati dall'enzima lipoproteina lipasi che libera i singoli acidi grassi utilizzati poi come *carburante* dalle cellule muscolari. Quando l'organismo ha bisogno di attingere alle riserve energetiche, i trigliceridi del tessuto adiposo vengono degradati ad Acetil-CoA attraverso una β -ossidazione ed in seguito convertiti in ATP [Parodi, 2004].

5.1 Effetti dei lipidi sulla salute

Negli ultimi decenni, i lipidi sono stati considerati quasi esclusivamente per il loro effetto negativo sulla salute del consumatore; è stato stimato che nei paesi più sviluppati circa il 35% degli adulti è sovrappeso ed il 30% è obeso. Al di là dell'aspetto estetico, l'eccesso di grasso può causare anomalie nei processi metabolici dell'organismo con conseguente sviluppo di disturbi e malattie che costituiscono una seria minaccia per la salute dell'uomo. Il peso in eccesso è responsabile del 30-40% delle malattie coronariche, quali l'aterosclerosi, legate anche ad altri fattori di rischio tra cui l'ipertensione, il diabete di tipo 2, la dislipidemia e l'insorgenza di tumori [Willett, 2002] [Parodi, 2006].

5.1.1 Malattie cardiovascolari

L'aterosclerosi è un processo infiammatorio che coinvolge diversi tipi di cellule ed ha come effetto l'alterazione della struttura e delle funzioni vascolari, lo sviluppo e la progressione di malattie cardiovascolari. L'aterosclerosi è caratterizzata da cambiamenti degenerativi, deposizione del colesterolo, proliferazione di cellule muscolari lisce, circolazione di cellule pro-infiammatorie e fibrosi. L'effetto risultante è dato da una condizione in cui le principali arterie coronarie non sono più capaci di far affluire al cuore sufficiente sangue ed ossigeno. La causa principale del flusso ridotto è data dall'accumulo di placche di grasso all'interno dell'arteria. A seconda della velocità di restringimento delle arterie possono verificarsi quattro diverse sindromi, di pericolosità crescente: l'angina pectoris, l'angina pectoris instabile, l'infarto miocardico e la morte cardiaca improvvisa.

Le cause dell'aumento dell'incidenza dell'obesità sono da ricercarsi nel disequilibrio tra il fabbisogno energetico dell'individuo e l'apporto energetico tramite l'alimentazione, ricca di cibi energetici appetibili ed economici, unito ad uno stile di vita sempre più sedentario; i grassi sono tacciati come i maggiori responsabili in quanto il loro valore energetico (9,0 kcal/g) è superiore a quello di proteine e carboidrati (4,0 kcal/g). Il valore energetico degli acidi grassi dipende però dal rapporto tra il contenuto di atomi di carbonio e quello di atomi di ossigeno. Per esempio l'acido stearico C18:0 ha un alto rapporto carbonio/ossigeno ed un valore energetico pari a 9,48 kcal/g mentre l'acido butirrico C4:0 ha un basso rapporto carbonio/ossigeno ed un valore energetico pari a 5,92 kcal/g. Inoltre, dato che circa un terzo degli acidi grassi contenuti nel latte sono costituiti da meno di 14 atomi di carbonio e vengono quindi metabolizzati rapidamente nel fegato, il grasso del latte dovrebbe contribuire al sovrappeso dell'organismo in quantità minore rispetto ad una quantità equivalente di altri grassi alimentari [Parodi, 2004].

5.1.1.1 Gli acidi grassi polinsaturi e gli eicosanoidi

Oltre al loro valore energetico ed al loro ruolo importante nella struttura delle membrane cellulari, gli acidi grassi polinsaturi agiscono come precursori di acidi grassi ossigenati, gli eicosanoidi, in grado di esercitare forti attività biologiche anche a basse concentrazioni. I tessuti dei mammiferi sono in grado di convertire gli acidi grassi saturi in acidi grassi monoinsaturi *cis* tramite l'azione dell'enzima $\Delta 9$ -desaturasi; sono possibili ulteriori desaturazioni per formare altri doppi legami nella catena, ma gli organismi animali non sono capaci di inserire doppi legami nelle posizioni 12 e 15. Due degli acidi grassi più

importanti contenuti nel latte presentano doppi legami nelle posizioni indicate: l'acido linoleico (LA) C18:2-*cis*-9, *cis*-12 e l'acido α -linolenico (ALA) C18:3-*cis*-9, *cis*-12, *cis*-15. I due acidi grassi, essenziali per l'uomo, vengono sintetizzati esclusivamente negli organismi vegetali e devono essere introdotti necessariamente con l'alimentazione. Tramite una serie di reazioni di elongazione e desaturazione che vengono a formarsi nei tessuti degli organismi animali, a partire dagli acidi linoleico e α -linolenico, si formano altri acidi grassi delle classi ω -6 e ω -3 rispettivamente, caratterizzati da una catena acilica piú lunga ed un grado di insaturazione maggiore rispetto ai loro precursori (Figura 5.1). Gli acidi ω -6 ed ω -3 possiedono funzioni biologiche distinte spesso in contrasto tra di loro, tali che le due classi non possono essere interscambiabili. Entrambe agiscono come componenti strutturali dei fosfolipidi del doppio strato lipidico delle membrane cellulari, aumentandone l'impermeabilit  verso la maggior parte delle molecole idrosolubili. La lunghezza ed il grado di saturazione degli acidi grassi determinano il grado di impaccamento dei fosfolipidi, influenzando la fluidit  della membrana, la trasduzione del segnale e l'espressione dei recettori cellulari.

Come si vede nella Figura 5.1, l'ultimo acido sintetizzato a partire dall'acido linoleico   l'acido arachidonico (AA) C20:4-*cis*-5, *cis*-8, *cis*-11, *cis*-14 mentre l'ultimo acido sintetizzato a partire dall'acido α -linolenico   l'acido eicosapentaenoico (EPA) C20:5-*cis*-5, *cis*-8, *cis*-11, *cis*-14, *cis*-17. Entrambi sono acidi eicosanoici, contenenti cio  20 atomi di carbonio. In realt , in molti tessuti, le reazioni che hanno come precursore l'acido α -linolenico non si arrestano con la formazione dell'EPA ma proseguono fino alla formazione dell'acido docosaesaenoico (DHA) C22:6-*cis*-4, *cis*-7, *cis*-10, *cis*-13, *cis*-16, *cis*-19. Vi sono due ipotesi riguardo le reazioni di formazione del DHA a partire dall'EPA. La prima prevede una reazione di elongazione della catena ed una reazione di desaturazione con conseguente formazione di un doppio legame in posizione 4 catalizzata dall'enzima Δ 4-desaturasi. La seconda ipotesi prevede un percorso piú lungo in cui si ha l'elongazione della catena dell'EPA a formare prima l'acido docosapentaenoico C22:5-*cis*-7, *cis*-10, *cis*-13, *cis*-16, *cis*-19, poi l'acido tetracosapentaenoico C24:5-*cis*-9, *cis*-12, *cis*-15, *cis*-18, *cis*-21 su cui agisce l'enzima Δ 6-desaturasi. Si forma cos  l'acido tetracosaeasenoico C24:6-*cis*-6, *cis*-9, *cis*-12, *cis*-15, *cis*-18, *cis*-21, in cui   presente un nuovo doppio legame in posizione 6, che subisce un processo di riduzione della catena tramite β -ossidazione. (Figura 5.2).

Gli acidi arachidonico (AA) e eicosapentaenoico (EPA) sono i precursori degli eicosanoidi, una famiglia di composti simili ad ormoni che includono prostaglandine, trombossani,

lipossine, leucotrieni ed altri composti ossidati. Gli eicosanoidi sono mediatori e regolatori dei processi infiammatori ed influiscono principalmente sulle funzioni renali, polmonari e vascolari. Si distinguono dagli ormoni in quanto la loro azione avviene localmente, nello stesso sito in cui vengono sintetizzati e spesso presentano funzioni opposte, con un effetto risultante dato dal bilanciamento delle singole attività [Lewis, 1990] [Tilley, 2001].

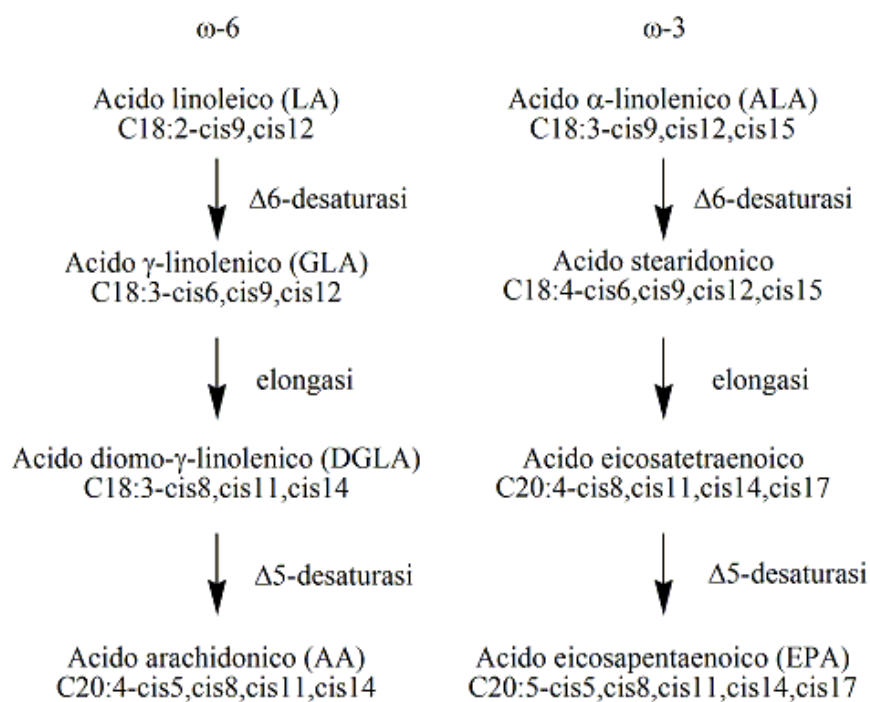


Figura 5.1 - Reazioni metaboliche delle classi ω -6 e ω -3.

Il livello degli eicosanoidi è in genere controllato da meccanismi omeostatici, ma può essere significativamente alterato in presenza di fenomeni infiammatori [Mori, 2013]. In generale, gli eicosanoidi derivati dall'acido arachidonico hanno una potente attività pro-trombotica e pro-infiammatoria, ma alcuni di essi, in particolare le prostaglandine, hanno un duplice ruolo, agendo sia a favore che contro i processi infiammatori. Altri eicosanoidi, come alcune lipossine, sono invece indispensabili per la risoluzione dello stato infiammatorio [Serhan, 2003] [Gewirtz, 2003]. Gli eicosanoidi derivati dall'acido eicosapentaenoico (EPA) e dall'acido docosaesaenoico (DHA) sono invece noti come potenti agenti antinfiammatori, avendo oltretutto un effetto inibitore sulla produzione degli eicosanoidi provenienti dall'acido arachidonico [Calder, 2009].

Nei primi studi, si riteneva che gli eicosanoidi derivati dall'EPA e dal DHA avessero un'attività molto più blanda rispetto a quelli formati dall'AA, ma studi più recenti hanno

dimostrato che l'effetto è presente solo per alcuni tipi di eicosanoidi, che costituiscono una minima parte del totale [Goldman, 1983] [Lee, 1984] [Dooper, 2002] [Miles, 2002].

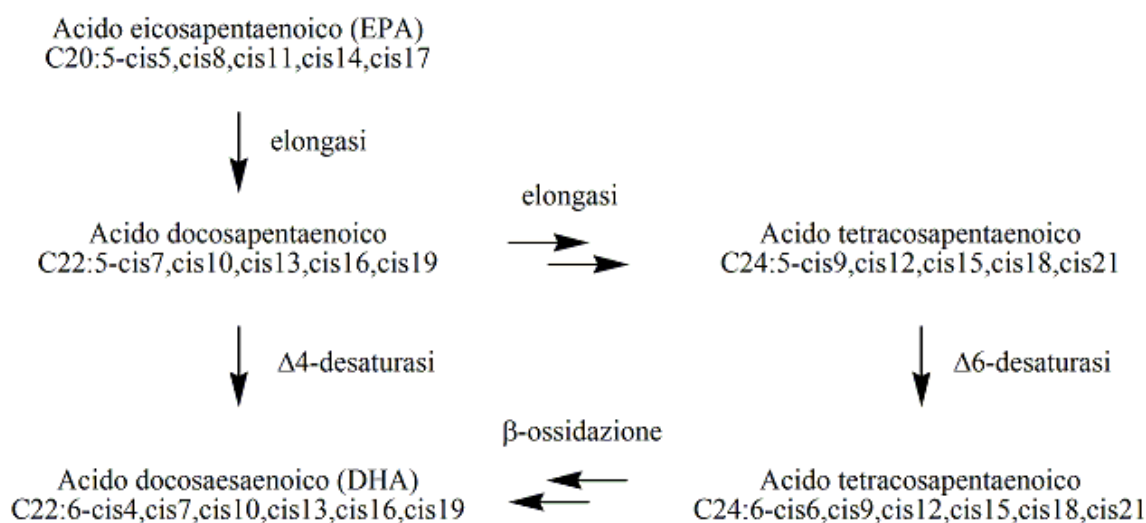


Figura 5.2 - Processi di formazione del DHA.

5.1.1.2 Effetti dei lipidi sul livello ematico di colesterolo

Gli studi riguardanti l'influenza degli acidi grassi sulle malattie coronariche hanno origine già dalla metà del XX secolo, periodo in cui è stata introdotta per la prima volta la relazione tra il grasso ingerito con l'alimentazione, i livelli di colesterolo nel sangue ed il rischio di contrarre malattie coronariche, quest'ultima tra le principali cause di mortalità dell'epoca moderna. È stato evidenziato che le diverse classi lipidiche hanno un diverso effetto sui livelli del colesterolo nel sangue ed in particolare che gli alimenti ricchi di acidi grassi saturi influenzano positivamente il livello di colesterolo più di quanto facciano gli acidi grassi insaturi [Ahrens, 1957].

La teoria del legame tra grasso alimentare, acidi grassi saturi, colesterolo e malattie coronariche, definita "ipotesi lipidica" [Keys, 1957], non è stata universalmente accettata tanto da generare la cosiddetta "controversia del colesterolo" [Steinberg, 2006]. Col passare degli anni si sono susseguiti numerosi studi che hanno confermato l'ipotesi lipidica [Anderson, 1987] [Stamler, 1986] [Pekkanen, 1990] [Verschuren, 1995] ma che tenevano conto solo del contenuto totale di colesterolo presente nel sangue. Il colesterolo però, così come altri lipidi, è poco solubile in soluzione acquosa ed è trasportato nel sangue da proteine chiamate apoproteine. Le lipoproteine risultanti formano una miscela

estremamente eterogenea con composti di struttura chimica, proprietà chimiche e funzioni metaboliche differenti.

Le lipoproteine vengono divise in cinque classi in base alla loro densità, determinata tramite ultracentrifugazione, di seguito riportate ordinate dalla meno densa alla più densa:

- Chilomicroni, o CM;
- lipoproteine a densità molto bassa, o VLDL (*very low density lipoproteins*);
- lipoproteine a densità intermedia, o IDL (*intermediate density lipoproteins*);
- lipoproteine a bassa densità, o LDL (*low density lipoproteins*);
- lipoproteine ad alta densità o, HDL (*high density lipoproteins*).

All'aumentare della densità delle particelle aumenta sempre più il contenuto proteico e diminuisce quello lipidico. Studi sul trasporto del colesterolo da parte dei diversi tipi di lipoproteine hanno permesso di rilevare che le lipoproteine LDL sono le maggiori trasportatrici di colesterolo mentre le lipoproteine HDL sono le minori, per cui le prime sono associate positivamente al rischio di contrarre malattie coronariche e le seconde lo sono negativamente. Risulta quindi importante il rapporto (colesterolo LDL)/(colesterolo HDL) che fornisce una stima del rischio di contrarre malattie coronariche che cresce all'aumentare del valore del suddetto rapporto [Miller, 1975] [Gordon, 1989] [Stampfer, 1991] [Gurr, 2002] [Parodi, 2009]. Le lipoproteine VLDL, ed in quantità minore i chilomicroni, sono invece le maggiori trasportatrici di trigliceridi mentre le lipoproteine IDL trasportano sia colesterolo che trigliceridi. In Tabella 5.1 è mostrato l'effetto generale delle diverse classi di lipidi sui livelli dei diversi tipi di colesterolo nel sangue.

Si può vedere che gli acidi grassi saturi determinano l'aumento del colesterolo totale e del colesterolo LDL ma non hanno effetto sul livello del colesterolo HDL. Il rapporto (colesterolo LDL)/(colesterolo HDL) subisce quindi un aumento. I livelli del colesterolo nel sangue non sono però relazionati nello stesso modo con i singoli acidi grassi saturi e dipendono da fattori quali la struttura chimica dei trigliceridi e l'attività recettoriale delle lipoproteine LDL [Hayes, 1992] [Hegsted, 1993]. L'acido palmitico C16:0, maggiore acido grasso saturo contenuto nel grasso alimentare, insieme agli acidi laurico C12:0 e miristico C14:0, esercita la maggiore influenza sui livelli di colesterolo mentre l'acido stearico C18:0 e gli acidi a media catena, ovvero il caprilico C8:0 ed il caprinico C10:0, esercitano la minore influenza [Mensink, 1993] [Mensink, 2003]. Studi più approfonditi sui diversi tipi di lipidi nel sangue indicano che l'acido palmitico è l'acido che

maggiormente influisce positivamente sulle malattie coronariche in quanto i suoi livelli nel grasso alimentare sono correlati all'aumento del colesterolo LDL e del colesterolo totale mentre non hanno effetto sui livelli del colesterolo HDL e sui trigliceridi nel sangue [Denke, 1992].

Tabella 5.1 - Effetto delle diverse classi lipidiche sui livelli del colesterolo e dei trigliceridi del sangue (adattata da Lovegrove (2003)).

Classe	Colesterolo totale	Colesterolo LDL	Colesterolo HDL	Trigliceridi
Saturi	Aumento	Aumento	Nessun effetto	Nessun effetto
Monoinsaturi	Diminuzione	Diminuzione	Nessun effetto	Nessun effetto
Polinsaturi ω-6	Diminuzione	Diminuzione	Diminuzione	Nessun effetto
Polinsaturi ω-3	Variabile	Variabile	Variabile	Diminuzione
Acidi grassi <i>trans</i>	Aumento	Aumento	Diminuzione	n.d.
CLA	Diminuzione	Diminuzione	n.d.	Diminuzione

n.d.: non disponibile

Gli acidi grassi monoinsaturi, costituiti prevalentemente dall'acido oleico C18:1-*cis*-9, determinano una diminuzione del contenuto di colesterolo LDL e totale, mentre non hanno effetto significativo sul contenuto del colesterolo HDL [Kurushima, 1995] [Lopez-Huertas, 2010].

Nonostante i numerosi studi sull'attività degli acidi grassi ω -6 e ω -3, non è ancora ben chiaro il loro effetto diretto sulla salute dell'organismo umano. Un'alimentazione povera di grassi ω -6, principalmente di acido linoleico, sembra essere associata ad un aumento del rischio di malattie cardiovascolari. Un aumento del loro contenuto, a partire da un alimentazione povera fino ad un livello moderato di grassi ω -6, in concomitanza con una diminuzione del contenuto di acidi grassi saturi, può invece avere effetti benefici sul metabolismo delle lipoproteine, può abbassare la pressione sanguigna e ridurre il rischio di malattie cardiovascolari. L'effetto degli acidi grassi ω -6 è però strettamente legato al livello di antiossidanti presenti nell'alimentazione; dato che i grassi sono suscettibili all'ossidazione, un moderato contenuto di acidi grassi ω -6, in presenza di un basso contenuto di antiossidanti, ha come effetto un aumento del rischio di contrarre malattie cardiovascolari.

Essendo l'acido linoleico il maggiore componente degli acidi grassi ω -6, il suo effetto sul livello del colesterolo nel sangue è assimilabile all'effetto prodotto dall'intera categoria. Un aumento del contenuto di acido linoleico ha come risultato la diminuzione del livello di

colesterolo LDL nel sangue. L'effetto è però lineare solamente nell'intervallo di apporto energetico compreso tra il 2 ed il 5%. La concentrazione di colesterolo HDL viene invece diminuita solo da alti livelli di acido linoleico, mentre livelli moderati non hanno alcun effetto [Harris, 2009] [Mori, 2013].

Gli studi sugli acidi grassi ω -3, ed in particolare sull'acido α -linolenico, sono quantitativamente inferiori rispetto a quelli sugli acidi grassi ω -6, a causa della loro minore presenza nei grassi sia vegetali che animali. Il loro effetto è ancora fonte di dibattito anche se sono numerosi i casi in cui il loro apporto è relazionato ad aspetti benefici per la salute dell'uomo. In generale, sembra che gli acidi grassi ω -3 abbassino il livello di lipidi nel sangue e riducano la pressione sanguigna con conseguente diminuzione del rischio di contrarre malattie cardiovascolari. La loro presenza nell'alimentazione ha effetto sui trigliceridi del sangue che subiscono una netta diminuzione [Bang, 1971] [Cunnane, 1992] [Lovegrove, 2004] [Motoyama, 2009]. L'effetto sul colesterolo di tipo LDL e HDL è invece poco chiaro ed in letteratura vi è un'alternanza tra l'aumento e la diminuzione dei loro livelli nel sangue ed è spesso riportata anche l'assenza di effetti significativi [Gurr, 1999]. L'effetto complessivo degli acidi polinsaturi ω -6 e ω -3 sulla salute dell'uomo è relazionato al rapporto tra le loro concentrazioni, di valore positivo in quanto gli alimenti presentano maggiori quantità di LA rispetto a ALA. Più il rapporto è elevato, più aumenta il rischio di contrarre malattie cardiovascolari mentre, al decrescere del rapporto, il rischio subisce un decremento [Weber, 1989] [Simopoulos, 2003]. La presenza di acido linoleico nella dieta ha influenza sulla formazione degli acidi ω -3 in quanto inibisce la formazione dell'EPA; quindi al crescere del contenuto di LA, il rapporto ω -6/ ω -3 aumenta non solo per l'aumento degli ω -6, ma anche per la diminuzione degli ω -3 [Liou, 2007]. Tenendo anche conto del tipo di eicosanoidi formati dalle due classi, un corretto regime alimentare deve prevedere l'apporto di entrambe le famiglie di PUFA ω -6 e ω -3, e quindi di un corretto equilibrio di fattori pro-infiammatori e anti-infiammatori. La loro concentrazione varia però notevolmente a seconda dei livelli di LA e ALA nella dieta che differisce a seconda della zona geografica. È stato stimato, per esempio, che un'alimentazione tipica del nord Europa presenta un rapporto ω -6/ ω -3 più elevato rispetto a quello presente nell'alimentazione di tipo mediterraneo [Ambring, 2006]. I valori tipici del rapporto ω -6/ ω -3 variano tra 1 e 10, ma è stato stimato che un alimento con un rapporto ω -6/ ω -3 < 4 ha un elevato valore nutrizionale [Department Of Health, 1994]. Tale rapporto non presenta però gli stessi effetti benefici sulle diverse popolazioni in quanto è stato evidenziato che la capacità di formazione degli acidi AA, EPA e DHA nell'organismo

umano differisce a seconda dell'etnia a causa della diversa attività enzimatica delle desaturasi [Patel, 2010].

Gli acidi grassi *trans* hanno come effetto l'aumento del colesterolo LDL e del colesterolo totale in misura simile a quella ottenuta dagli acidi grassi saturi. Contrariamente a questi ultimi, gli acidi grassi *trans* hanno però influenza anche sul colesterolo HDL, abbassandone i livelli nel sangue. Ne segue che il rapporto (colesterolo LDL)/(colesterolo HDL) ha valori circa doppi rispetto a quelli ottenuti considerando gli acidi grassi saturi, con conseguente aumento del rischio di contrarre malattie coronariche [Ascherio, 1999] [Simopoulos, 2008]. Inoltre, gli acidi grassi *trans* interferiscono con le reazioni di elongazione e desaturazione degli acidi grassi ω -6 e ω -3, con conseguente riduzione delle quantità di acidi AA, EPA e DHA [Simopoulos, 1995].

Gli isomeri del CLA hanno come effetto la diminuzione del contenuto di trigliceridi, di colesterolo totale e di colesterolo LDL mentre non si hanno evidenze sull'effetto sul colesterolo HDL. I CLA ed in particolare l'isomero C18:2 *cis*-9, *trans*-11 sono quindi associati alla diminuzione del rischio di formazione di malattie coronariche [Bauman, 2006].

L'organizzazione mondiale della salute raccomanda, per prevenire disturbi cronici, un apporto di grasso totale che non superi il 30% dell'energia totale, con un massimo del 10% costituito da acidi grassi saturi, con una quantità compresa tra il 10 ed il 15% per gli acidi grassi monoinsaturi e tra il 6 ed il 10% per gli acidi grassi polinsaturi [World Health Organization, 2003].

5.1.2 Effetti dei lipidi sul cancro

Le neoplasie principali legate al consumo di lipidi sono quelle del colon, della prostata e del seno. Diversi studi hanno messo in evidenza la correlazione tra lo sviluppo dei tumori del colon e della prostata con la quantità di grasso ingerita, con l'apporto energetico totale e con lo stile di vita sedentario tipico dei paesi occidentali. Allo stesso tempo non si è però trovata evidenza dell'influenza delle diverse categorie di acidi grassi nel rischio dello sviluppo dei tumori del colon e della prostata indipendentemente dal bilancio energetico totale [Howe, 1997] [Giovannucci, 1997] [Willett, 2001] [Kolonel, 2001] [Kushi, 2002]. Gli studi riguardanti il tumore al seno mostrano lo stesso tipo di correlazione, identificando però gli acidi grassi ω -6 come i responsabili dell'insorgenza del tumore. Indipendentemente dall'apporto energetico totale, si è visto infatti che l'aumento dei livelli

di acido linoleico entro il range 0,5-4,4% favorisce la comparsa del tumore al seno mentre gli acidi grassi saturi, i monoinsaturi e gli acidi grassi *trans* hanno influenza minima o nulla [Parodi, 2006]. Al contrario, gli acidi linoleici coniugati e gli acidi grassi appartenenti alla classe ω -3 inibiscono la comparsa del tumore; i primi hanno il maggiore effetto inibitore, considerato anche che il loro livello nel grasso del latte si aggira intorno all'1% [Ip, 1990] [Ip, 1997].

5.2 Attività biologica dei CLA

5.2.1 I CLA e l'aterosclerosi

Nonostante molti isomeri del CLA, tra cui l'acido rumenico (C18:2 *cis*-9, *trans*-11), contengano doppi legami *trans*, diversi studi hanno evidenziato l'effetto positivo di questa classe di acidi grassi sui livelli di colesterolo nel sangue e la diminuzione dello sviluppo dell'aterosclerosi negli animali da laboratorio, dimostrando che i CLA sembrano essere in grado di agire sulla produzione nel fegato delle lipoproteine aterogeniche VLDL e LDL; il colesterolo LDL ed i trigliceridi subiscono un decremento dei loro livelli nel sangue sia di conigli che di criceti alimentati con una dieta integrata con CLA, con riduzione della formazione di lesioni aterosclerotiche [Lee, 1994] [Kritchevsky, 1999] [Kritchevsky, 2003]. Gli animali alimentati con una dieta ricca di grasso (20%) e di colesterolo (0,2%) integrata con acido vaccenico e rumenico presentano livelli di colesterolo VLDL e LDL molto più bassi di quelli presenti negli animali a cui non è stata data l'integrazione [Lock, 2005]. Alcuni studi evidenziano effetti variabili o non significativi sui cambiamenti dei livelli di colesterolo nel sangue e sul metabolismo delle lipoproteine in animali da laboratorio la cui dieta è integrata con CLA. Questi studi sono comunque accomunati dall'aver inserito nella dieta degli animali una miscela sintetica di isomeri del CLA, che sembra non presenti gli stessi effetti biologici della classe degli isomeri presenti naturalmente negli alimenti di origine animale/ruminale [Bauman, 2005].

L'inibizione dell'aterosclerosi dovuta ai CLA non è da attribuire esclusivamente alla diminuzione dei livelli di colesterolo aterogenico nel sangue. I CLA sono anche in grado di ridurre l'effetto infiammatorio che contribuisce alla formazione dell'aterosclerosi, inibendo l'azione di alcuni eicosanoidi pro-infiammatori [Urquhart, 2002] [Yu, 2002] [Toomey, 2003].

Gli studi sull'effetto anti-aterogenico dei CLA sull'uomo sono numericamente meno consistenti rispetto a quelli effettuati sugli animali da laboratorio. Il metabolismo dei trigliceridi e del colesterolo VLDL viene accelerato negli individui sani la cui alimentazione è integrata con una miscela 50:50 di acido rumenico e dell'isomero *trans*-10, *cis*-12. La riduzione del colesterolo VLDL nel sangue avviene anche nel caso il rapporto dei costituenti della miscela sia 80:20. La separazione dei due isomeri, introdotti puri nella dieta, ha permesso di evidenziare una maggiore attività anti-aterogenica dell'acido rumenico rispetto a quella dell'isomero *trans*-10, *cis*-12, con riduzioni dei trigliceridi, del colesterolo totale, del colesterolo LDL e del rapporto (colesterolo LDL)/(colesterolo HDL) [Noone, 2002] [Tricon, 2004].

5.2.2 I CLA ed il cancro

Sin da quando è stato evidenziato per la prima volta l'effetto antimutagenico dei CLA [Pariza, 1985], l'interesse su questa classe di acidi grassi è cresciuta in modo esponenziale. Gli studi sull'effetto antitumorale dei CLA sono oggi numerosi ed hanno permesso di evidenziare la loro efficacia su neoplasie di diverso tipo.

L'effetto anticancerogeno dei CLA è particolarmente marcato nel caso del tumore al seno; l'incidenza del tumore diminuisce notevolmente negli animali da laboratorio all'aumentare della concentrazione dei CLA, raggiungendo un decremento del 60% nel caso di un alimentazione integrata con 1,5% di CLA [Ip, 1991]. Oltretutto, l'integrazione dell'alimentazione con CLA durante il periodo peripuberale ha permesso la protezione contro il tumore mammario anche quando la sostanza cancerogena è stata somministrata all'animale in età adulta [Thompson, 1997]. L'inibizione tumorale dei CLA è in contrasto con l'effetto dell'acido linoleico che favorisce invece lo sviluppo del tumore. La maggior parte degli studi hanno utilizzato miscele di due o più isomeri del CLA, prodotte sinteticamente da oli vegetali, contenenti principalmente uguali concentrazioni di acido rumenico e dell'isomero *trans*-10, *cis*-12. In seguito è stato dimostrato che i due isomeri mostrano lo stesso potere inibitore verso la formazione del tumore mammario [Ip, 2003]. Tra i diversi isomeri, l'acido rumenico è comunque il più studiato essendo presente negli alimenti in quantità maggiore.

Numerosi altri studi sugli animali da laboratorio hanno evidenziato l'efficacia dei CLA nell'inibizione di altri tipi di tumori maligni: si è visto che i CLA interferiscono col metabolismo di agenti cancerogeni promotori del tumore al colon [Liew, 1995], sono

efficaci nella riduzione della proliferazione delle cellule tumorali della prostata [Ochoa, 2004] e sono inibitori dei tumori della pelle nei topi da laboratorio [Belury, 1996].

Gli studi che riportano l'influenza dell'apporto di CLA con l'alimentazione e l'insorgenza di tumori negli esseri umani sono numericamente limitati, sia per la difficoltà del tipo di studio, sia per il dilatato arco temporale in cui un tumore può formarsi. Il contenuto di CLA nel tessuto adiposo del seno di donne malate di tumore risulta essere inferiore a quello presente in soggetti non malati [Lavillonniere, 1999]. Esiste inoltre una relazione inversa tra il consumo di latte e l'insorgenza di tumore al seno che suggerisce un effetto anticancerogeno del CLA presente nel latte [Knekt, 1996]. Il CLA inibisce inoltre la crescita di un tipo di cellula di adenocarcinoma polmonare [Schønberg, 1995]. Studi in vitro su cellule tumorali umane hanno evidenziato l'efficacia del CLA nell'ostacolare la proliferazione delle cellule tumorali del colon, del seno e della pelle [Shultz, 1992]. Molti studi non hanno però portato a risultati concreti, cosa non sorprendente in quanto la stima del contenuto del CLA negli alimenti è resa difficoltosa dall'ampia variabilità dei costituenti del grasso nel latte e nei prodotti caseari, così come dalla diversa concentrazione dei CLA nei diversi tipi di latte che spesso viene riportata con valori poco accurati [Bauman, 2006].

5.2.3 Altri effetti dei CLA

I CLA sono inoltre responsabili dell'effetto di riduzione della massa corporea degli animali da laboratorio e degli esseri umani, con diminuzione della massa grassa a favore di quella magra [Roche, 2001] [Larsen, 2003] [Martin, 2002] [Gaulhier, 2005]. L'effetto è però causato principalmente dall'isomero *trans*-10, *cis*-12 mentre l'acido rumenico sembra essere neutrale [Martin, 2002] [Terpstra, 2004].

L'isomero CLA *trans*-10, *cis*-12 pare avere influenza sul metabolismo del glucosio anche se il suo ruolo non è ancora chiaro. È stato documentato sia il suo ruolo anti-diabetico, innalzando la sensibilità dell'insulina e diminuendo i livelli di glucosio [Khanal, 2004] sia il suo ruolo pro-diabetico [Moloney, 2004] [Wang, 2004]. L'acido rumenico sembra invece non avere un ruolo attivo nel metabolismo del glucosio [Martin, 2002].

6 Produzione e utilizzo del latte

6.1 Produzione del latte

Il latte, insieme ai prodotti da esso derivati, è entrato a far parte della dieta quotidiana degli esseri umani, rivestendone un ruolo essenziale come alimento nutritivo con proprietà organolettiche caratteristiche. Il latte viene consumato probabilmente in tutte le aree geografiche del mondo e, in relazione ad una richiesta sempre maggiore, la sua produzione è in continuo aumento. La produzione mondiale del latte da tutte le specie supera le diverse centinaia di milioni di tonnellate l'anno, raggiungendo, nel 2011, 749 milioni di tonnellate (Figura 6.1). Di queste, l'84% è latte vaccino, il 13% è latte di bufala mentre il latte di capra e di pecora si aggirano intorno al 2% ognuno; il restante 1% è dato da latte di altre specie, prevalentemente quello di cammella.

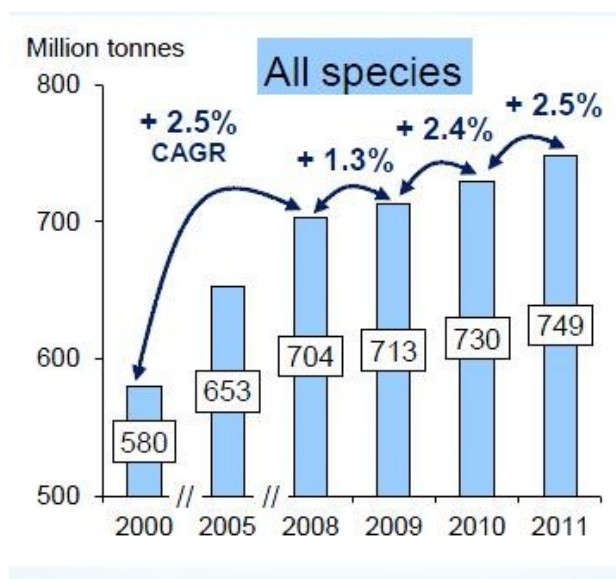


Figura 6.1 – Crescita annua della produzione mondiale del latte [FIL-IDF, 2012].

La crescita della produzione ha riguardato il latte di tutte le diverse specie. Nel decennio tra il 2000 ed il 2011 il latte bovino ha registrato un aumento del 28%, passando da 489 a 629 milioni di tonnellate; il latte bufalino ha avuto la percentuale di incremento più elevata, 44%, passando da 67 a 97 milioni di tonnellate; il latte caprino è passato da 13 a 18 milioni di tonnellate con un incremento del 38%; anche la produzione del latte ovino è aumentata, passando da 8,1 a 10,1 milioni di tonnellate, con un incremento del 25%. Un aumento così consistente è dovuto anche alla selezione delle razze “lattifere”, ossia di

specie improntate verso l'aumento della produzione di latte. La produzione di un individuo "selezionato" può raggiungere quantità 20 volte superiori rispetto a quelle di un individuo "aspecifico".

Le aree geografiche in cui si ha la maggiore produzione del latte sono quelle più densamente popolate, come l'Asia e l'Europa che complessivamente contribuiscono al 67% del totale mentre la restante parte è suddivisa tra gli altri continenti, con un contributo maggiore dato dal Nord e dal Sud America. La produzione del latte delle diverse specie animali non è egualmente suddivisa tra le diverse zone. Il latte bovino, pur essendo prodotto in tutte le aree geografiche, è prevalente in Europa, in Asia (in particolare in India) e nelle due Americhe. Il latte ovino è caratteristico delle zone più asciutte del pianeta ed è maggiormente prodotto nei paesi mediterranei dell'Europa e dell'Africa ed in numerosi paesi asiatici, soprattutto Cina. Il latte caprino è prodotto in prevalenza nel subcontinente indiano dell'Asia ed in numerosi paesi europei, specialmente Grecia, Francia e Spagna. Il latte di bufala è prodotto quasi esclusivamente nelle regioni indiana e cinese dell'Asia ed in Egitto mentre il latte di cammella è prodotto solo nelle regioni arabe dell'Asia e nelle regioni mediterranea, sahariana ed orientale dell'Africa.

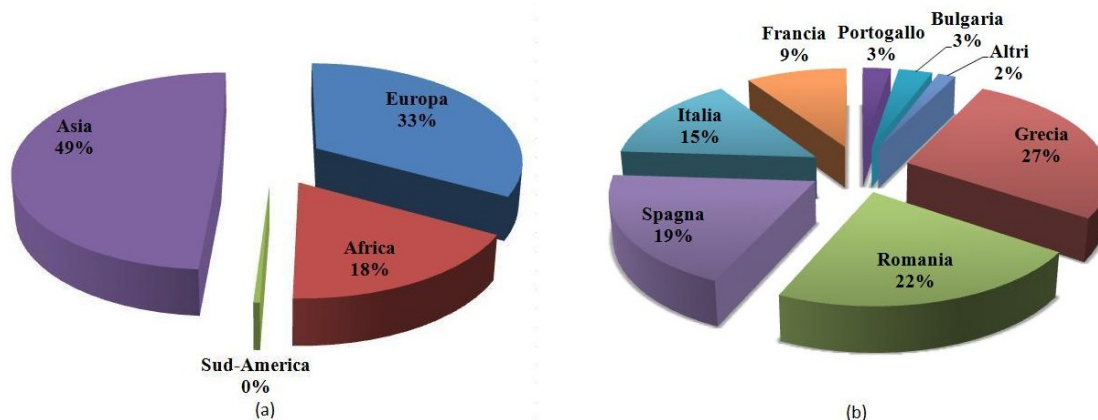


Figura 6.2 - Ripartizione della produzione mondiale (a) ed europea (b) di latte ovino nel 2011.

In Italia, nel 2011, sono state prodotte più di 11 milioni di tonnellate di latte di cui il 94% solo di latte bovino (circa 10 milioni di tonnellate), il 4% di latte ovino (420.000 tonnellate) e la restante parte divisa tra latte caprino e bufalino (200.000 tonnellate). L'Italia è tra le maggiori nazioni produttrici di latte ovino, contribuendo con il 15% al totale del latte ovino prodotto in Europa ed con il 5% al totale del latte ovino prodotto nel mondo (Figura 6.2) [FAOSTAT, 2011].

Quasi il 75% del latte ovino italiano viene prodotto nel Mezzogiorno e nelle isole, il 25% nella parte centrale della penisola e meno dell'1% nel nord Italia. La Sardegna è la prima regione italiana per la produzione del latte ovino coprendo, con le sue 300.000 tonnellate, circa il 67% della produzione italiana del 2011, seguita da Toscana (14%) e Lazio (10%) [ISTAT, 2011]. La produzione del latte ovino in Sardegna copre quindi più del 3% della produzione mondiale, una percentuale di assoluto rilievo specie in relazione alla ridotta estensione della regione.

6.2 Trattamenti e trasformazioni del latte

Il latte crudo è un alimento facilmente alterabile a causa della proliferazione di batteri e microrganismi in esso presenti. Il suo tempo di conservazione può essere comunque protratto tramite l'impiego di trattamenti termici o mediante la riduzione del suo contenuto di acqua.

I trattamenti termici, differenti per tempo e temperatura, variano a seconda del tipo di microrganismo che si vuole debellare (Tabella 6.1). La termizzazione riduce la flora anticasearia, evitando così i difetti tipici dei formaggi ottenuti in loro presenza; la pastorizzazione agisce sui germi patogeni, incrementando così la conservabilità del latte a bassa temperatura; la sterilizzazione viene invece effettuata per abbattere la flora microbica totale, permettendo una conservazione più lunga del latte anche a temperatura ambiente.

La riduzione del contenuto di acqua del latte viene effettuata per ridurre l'ambiente di proliferazione dei batteri; i trattamenti effettuati sul latte sono la concentrazione, la condensazione e l'essiccamento. La concentrazione, effettuata tramite parziale evaporazione dell'acqua sottovuoto a temperature di 40-50°C, riduce il contenuto di acqua del 30%. La condensazione, effettuata tramite concentrazione ed aggiunta di saccarosio, permette una riduzione del contenuto di acqua fino al 75%. L'essiccazione riduce quasi totalmente il contenuto di acqua, ottenendo latte in polvere con un residuo di acqua di circa il 2-7%. Il latte liquido viene rigenerato semplicemente con l'aggiunta di acqua.

La produzione del latte non si limita a soddisfare la richiesta del mercato per il suo consumo diretto; parte di esso viene convertita in numerosi derivati che possiedono proprie caratteristiche organolettiche e che presentano tempi di conservazione maggiori rispetto a quelli del latte di partenza. Il latte prodotto dai ruminanti è particolarmente adatto alla trasformazione in quanto presenta maggiori concentrazioni di caseina rispetto al latte dei

mammiferi non ruminanti. Il latte cosiddetto “caseinoso” coagula facilmente per aggiunta del caglio o per acidificazione fino a pH 4,6.

Tabella 6.1 – Trattamenti termici del latte [Raynal-Ljutovac, 2007].

Trattamento	Obiettivo	Tempo/Temperatura
Termizzazione	Distruzione degli <i>psicrotrofi</i>	15-20 s / 63-65°C
Pastorizzazione	Distruzione del <i>Mycobacterium tuberculosis</i>	15-30 s / 75-85°C
Pastorizzazione di alta qualità	Distruzione del <i>Mycobacterium tuberculosis</i>	15-30 s / 72-75°C
Sterilizzazione	Distruzione di enzimi e germi patogeni	20 min / 100- 120°C
Sterilizzazione UHT	Distruzione di enzimi e germi patogeni	2-3 s / 135-150°C

I trattamenti tecnologici a cui il latte può essere sottoposto rendono possibile la formazione di una serie di derivati che differiscono a seconda del tipo di processo utilizzato:

- la separazione per centrifugazione o per affioramento permette la separazione della crema (o panna), che è costituita dal 40% di materia grassa e da circa il 54% di acqua. Tramite successiva zangolatura della crema, con cui si ottiene la separazione della parte acquosa, viene prodotto il burro. Il burro deve avere per legge un tenore minimo di materia grassa ad almeno l'82% ed un contenuto d'acqua del 16%; in commercio è tuttavia presente anche il cosiddetto *burro leggero* che presenta percentuale di grasso del 60%, ed il *burro leggero a basso tenore di grasso*, con un percentuale massima di lipidi pari al 40%;
- la coagulazione enzimatica (o presamica) delle caseine, ottenuta per azione degli enzimi coagulanti del caglio, forma la cagliata dalla cui lavorazione si ottengono diverse varietà di formaggi. In seguito alla separazione della cagliata, che trattiene anche buona parte della materia grassa residua, dal siero è possibile ottenere, per termocoagulazione delle siero proteine, un altro interessante prodotto derivato: la ricotta;
- la coagulazione acida delle caseine, favorita dalle alte temperature, permette di ottenere una cagliata più friabile di quella ottenuta per coagulazione enzimatica, in quanto il calcio fosfato delle caseine viene solubilizzato nel siero con conseguente disgregazione delle micelle in sub-micelle. Da questa tipologia di cagliata si ottiene una vasta gamma di formaggi freschi e molli;

- la fermentazione lattica che, per azione di batteri quali lo streptococcus thermophilus ed il lactobacillus bulgaricus, forma acido lattico a partire dal lattosio con conseguente precipitazione delle caseine e formazione del coagulo. Vengono così formati i prodotti da latte fermentato, come gli yogurt, il cui contenuto di fermenti lattici vivi deve essere elevato fino alla scadenza;
- il raffreddamento eterogeneo del latte dal quale è possibile ottenere prodotti come i gelati.

L'utilizzo principale del latte differisce a seconda della specie. Il latte bovino è quello maggiormente consumato per uso alimentare come bevanda; nonostante ciò, circa due terzi del latte bovino prodotto viene trasformato tecnologicamente in derivati. Il latte di bufala, capra e pecora viene invece quasi esclusivamente impiegato nella produzione dei derivati del latte. Il latte di queste specie contiene una quantità maggiore di grasso e proteine ed è pertanto più adatto alla caseificazione. Generalmente il tempo di coagulazione delle caseine è minore rispetto a quello del latte bovino e la cagliata ottenuta è meno friabile di quella prodotta dal latte bovino.

In Italia, i derivati prodotti dal latte bovino sono prevalentemente formaggi a pasta fresca o di tipo grana; il latte di bufala viene quasi esclusivamente adoperato per la produzione della mozzarella di bufala; il latte caprino viene utilizzato per produrre i formaggi caprini ed altri formaggi freschi mentre il latte ovino è quasi interamente utilizzato per la produzione di Pecorino Romano e Fiore Sardo, due formaggi a pasta dura, e di Pecorino Sardo, presente nelle due varianti a pasta molle e semi-dura [Corradini, 1995] [Fox, 1998] [Pirisi, 2007] [Alais, 2010] [Pirisi, 2011].

7 Metodi di caratterizzazione del profilo acido del latte per l'acquisizione dei dati di riferimento

Lo sviluppo di un metodo alternativo per l'analisi degli acidi grassi, basato sulle proprietà di assorbimento nel medio infrarosso dei loro gruppi funzionali, necessita del cosiddetto "valore vero", ottenibile con un metodo di riferimento validato per la caratterizzazione del profilo acido del latte.

Il metodo di riferimento prevede tre passaggi principali: l'estrazione della frazione lipidica dalla matrice; la derivatizzazione della frazione lipidica estratta e la caratterizzazione analitica dei lipidi derivatizzati.

7.1 Estrazione della frazione lipidica dalla matrice

La caratterizzazione della frazione lipidica del latte comporta diversi passaggi chiave: l'estrazione della frazione grassa totale, la separazione e quantificazione delle varie classi lipidiche, la loro derivatizzazione, l'identificazione e la quantificazione degli acidi grassi che le compongono.

E' evidente che dall'esito del passaggio iniziale di estrazione dipenderà quindi la qualità dei risultati ottenuti in tutte le fasi successive. Il metodo analitico utilizzato deve quindi essere in grado di estrarre i lipidi in modo quantitativo, qualora ne sia richiesta la determinazione totale; o quanto meno rappresentativo senza alterarne i reciproci rapporti ponderali, qualora ne sia richiesta la caratterizzazione in termini di costituenti.

Il metodo analitico di estrazione dovrebbe anche possedere i seguenti requisiti:

- capacità di inibire l'azione di enzimi come lipasi e fosfolipasi, in modo da evitare fenomeni di degradazione lipolitica della materia grassa durante l'analisi;
- capacità di limitare il più possibile i fenomeni autossidativi e di isomerizzazione dei doppi legami durante l'estrazione (indotti generalmente dall'impiego di elevate temperature e da solventi non idonei);
- capacità di estrarre in modo aspecifico tutti i costituenti della frazione lipidica;
- capacità di separare i lipidi dai componenti a ridotta polarità (interferenti apolari) che potrebbero introdurre errori nelle successive determinazioni.

I metodi di estrazione della materia grassa si dividono in due categorie:

1. metodi ufficiali di riferimento
2. metodi non ufficiali o sperimentali

I metodi ufficiali sono metodi validati che derivano da una estesa sperimentazione volta a definirne le caratteristiche di accuratezza e precisione. Tali metodi sono stati messi a punto e validati, nella maggior parte dei casi, nell'ambito di gruppi di lavoro che fanno capo ad organizzazioni internazionali che riuniscono esperti del settore scientifico. Nel caso del latte e dei prodotti lattiero caseari la FIL-IDF (Fédération International de Laiterie – International Dairy Federation) e la ISO (International Standard Organization) sono le due organizzazioni internazionali che, con attività congiunta, provvedono allo sviluppo di metodi di analisi e di campionamento o norme di prodotto su latte e derivati. Tali norme hanno carattere volontario, ma vengono frequentemente recepite a livello nazionale o internazionale come norme ufficiali. Inoltre sia a livello nazionale che europeo, gruppi di esperti nominati dai Ministeri provvedono alla messa a punto di specifici metodi analitici. La scelta, da parte dell'analista, di applicare un metodo sperimentale o ufficiale può dipendere da molti fattori, ma deve necessariamente dare priorità a quest'ultimo, nel caso si tratti di un controllo analitico di tipo ufficiale [Antongiovanni, 2007].

I metodi ufficiali

Nell'ambito dei metodi ufficiali, spesso sono disponibili due norme, di cui una definita "metodo di riferimento" e l'altra "metodo di routine". La prima è la metodica che deve essere applicata, necessariamente, in caso di controversie legali; la seconda è solitamente un metodo più rapido, applicabile in caso di maggiore numerosità dei campioni. Nel caso del latte è disponibile anche una norma che riporta le linee guida per l'utilizzo degli strumenti che operano nel medio infrarosso, per la valutazione dei principali costituenti del latte, tra cui anche il grasso.

- Determinazione quantitativa del contenuto in grasso – IDF 1D:1996/ISO1211:1999 "Milk – Determination of fat content- Gravimetric method (Reference method)".

Questa procedura consente la determinazione precisa del contenuto in materia grassa. Il tenore in materia grassa è determinato gravimetricamente mediante

estrazione della materia grassa da una soluzione ammoniaco-alcolica del latte con etere dietilico ed etere di petrolio, evaporazione dei solventi e pesata del residuo, secondo il principio del metodo di Röse-Gottlieb. Il metodo prevede l'idrolisi basica del latte che favorisce l'eliminazione dell'interazione dei lipidi con le proteine. I lipidi vengono successivamente estratti con etere etilico o etere di petrolio o una miscela dei due solventi. A causa delle elevate temperature raggiunte durante l'aggiunta dell'ammoniaca, il processo può causare fenomeni di idrolisi, di isomerizzazione o di ossidazione degli acidi grassi coniugati e dei fosfolipidi, alterandone quindi i mutui rapporti di concentrazione.

- Determinazione quantitativa del contenuto in grasso secondo il metodo Gerber – ISO 2446: 1976 “Determination of fat content (Routine method)”.

Questa norma riporta le indicazioni per la determinazione del contenuto in grasso del latte, con il metodo rapido che utilizza i butirrometri Gerber. Il principio del metodo si basa sulla separazione del grasso mediante centrifugazione del latte, dopo aver dissolto le proteine con acido solforico e favorito la separazione delle fasi con alcol isoamilico. La reazione avviene all'interno di un'apparecchiatura graduata (butirrometro) e fornisce direttamente la percentuale in materia grassa mediante lettura, su idonea scala, del livello di grasso limpido separato. Benché il metodo sia stato testato in differenti analisi interlaboratoriali, non sono ancora disponibili valori definitivi di precisione e di accuratezza, soprattutto in confronto al metodo gravimetrico di riferimento, a causa dei controversi risultati ottenuti con i diversi sistemi di misura dell'aliquota del campione da sottoporre ad analisi. E' possibile infatti effettuare un prelievo in volume con pipette da 11 o da 10,77 ml, esprimendo quindi il risultato in p/v o pesando direttamente $11,13 \pm 0,03\text{g}$, esprimendo quindi il risultato in p/p. Poiché tale metodo è utilizzato in maniera routinaria, effettuare una pesata di tale accuratezza comporta necessariamente un notevole consumo di tempo e contrasta quindi con le peculiarità di questa procedura.

Nella pratica il metodo Gerber è comunemente utilizzato effettuando il prelievo con la pipetta da 11 ml. Bisogna comunque tener presente che, sia per le caratteristiche del metodo, sia per questa problematica di prelievo, i risultati possono leggermente differire dal metodo di riferimento sia in termini di accuratezza che in termini di precisione.

- Determinazione quantitativa del contenuto in grasso con apparecchiatura automatica – IDF 141C:2000/ISO 21543:1999 “Whole milk – Determination of milkfat, protein & lactose content – Guide for the operation of mid-infra-red instruments”.

Si tratta di una procedura automatica, non distruttiva, che utilizza la correlazione esistente tra l'assorbimento a specifiche lunghezze d'onda e il contenuto in macrocostituenti del latte, tra cui il grasso, per determinarne la concentrazione. La norma ufficiale non riporta i dettagli del procedimento analitico, in quanto può variare a seconda del tipo di strumento, ma indica delle linee guida applicabili a tutti i tipi di strumentazione. I dati di riferimento utilizzati per la costruzione dei modelli di predizione sono acquisiti tramite metodica IDF 1D:1996/ISO1211:1999. L'analisi, replicata in doppio, permette l'ottenimento del valore predetto nell'arco di pochi secondi senza alcun tipo di pretrattamento del campione. La metodica è validata su latte bovino ed applicabile al latte di altre specie animali.

- Metodo di estrazione del grasso e delle sostanze liposolubili – IDF172:2001/ISO 14156:2001 – “Milk and Milk products – Extraction methods for lipids and liposoluble compounds”.

Questa norma descrive le procedure idonee ad ottenere la separazione o l'estrazione di una quota rappresentativa della frazione lipidica del latte, nel caso in cui si debba procedere ad altre determinazioni dei singoli costituenti della materia grassa. L'estrazione effettuata con questo metodo prevede l'utilizzo di ammoniaca, etanolo, etere etilico e pentano. Il procedimento è lungo ed è specificamente applicabile nel caso si debba effettuare la valutazione della composizione in acidi grassi e trigliceridi, mentre non consente un recupero rappresentativo dei costituenti lipidici più polari, come ad esempio i fosfolipidi. Non consente inoltre l'estrazione degli acidi grassi liberi. Le procedure descritte non sono idonee ad ottenere un valore quantitativo del contenuto in grasso.

I metodi non ufficiali o sperimentali

- Metodo Röse-Gottlieb.

Il tenore in materia grassa è determinato gravimetricamente mediante estrazione della materia grassa da una soluzione ammoniaco-alcolica del latte con etere

dietilico ed etere di petrolio, evaporazione dei solventi e pesata del residuo, secondo il principio del metodo di Röse-Gottlieb.

- Metodo rapido per l'estrazione dei trigliceridi.

Il tenore in materia grassa è determinato gravimetricamente mediante estrazione della materia grassa da una soluzione ammoniaco-alcolica del latte con esano, evaporazione dei solventi e pesata del residuo, secondo Buccioni et al. [Buccioni, 2004].

- Metodo di Hara and Radin [Hara, 1978].

Il metodo di estrazione Hara Radin nasce originariamente per l'estrazione dei lipidi dai tessuti animali. Il metodo prevede l'utilizzo di una miscela di solventi esano/isopropanolo; l'alcol causa la dissociazione delle interazioni dei lipidi mentre il secondo solvente agisce da mezzo di estrazione. La miscela proposta ha una tossicità inferiore della miscela cloroformio/metanolo del classico metodo di estrazione Folch [Folch, 1957]. L'utilizzo del cloroformio può infatti provocare l'insorgenza di tumori, mentre il metanolo può indurre danni al sistema visivo. L'esano, pur essendo una neurotossina se utilizzato ad elevate concentrazioni, comunque nel normale utilizzo in laboratorio è considerato relativamente non tossico [Hara, 1978].

Il metodo è stato applicato con successo anche all'estrazione dei lipidi dal latte [Jiang, 1996] e dalla crema del latte [Chouinard, 1999].

- Metodo fisico [Murphy, 1990].

Il metodo non prevede l'utilizzo di solventi chimici per estrarre la materia grassa dal latte ma utilizza metodi fisici (shock termico e centrifugazione) per separare i lipidi tramite rottura fisica della membrana dei globuli di grasso. Il metodo prevede una preliminare separazione della materia grassa del latte tramite centrifugazione (30 min, 3000 rpm, 4°C). Sulla crema separata viene effettuato uno shock termico (60°C per 15 min) che induce la rottura della membrana dei globuli di grasso. La successiva centrifugazione (30 min, 3000 rpm, 37°C) consente il recupero della fase grassa. Il metodo, più veloce dei precedenti, non è adatto alla quantificazione del grasso totale presente nel latte ma ne estrae una quantità rappresentativa.

7.2 Derivatizzazione della frazione lipidica estratta

Le principali tecniche analitiche strumentali utilizzate per l'identificazione e la quantificazione degli acidi grassi individuali del latte sono tecniche cromatografiche quali la gascromatografia (GC) e la cromatografia liquida ad elevate prestazioni (HPLC, High Performance Liquid Chromatography). La gascromatografia è la tecnica più comune e conveniente in quanto, grazie all'utilizzo di un detector a ionizzazione di fiamma (FID), permette di determinare un numero elevato di acidi grassi in un'unica corsa analitica.

Perché la composizione lipidica in acidi grassi possa essere determinata per via gascromatografica, è necessario convertire gli acidi grassi, legati ai componenti della materia grassa (trigliceridi, di e monogliceridi, fosfolipidi, glicolipidi etc.) in derivati non polari a basso peso molecolare, quali i loro esteri metilici.

Successivamente al passaggio di estrazione, la materia grassa deve quindi essere sottoposta ad idrolisi ed esterificazione (metilazione), il cui risultato è l'ottenimento di una miscela di esteri metilici degli acidi grassi, originariamente contenuti nel latte, facilmente separabili, identificabili e quantificabili tramite gascromatografia.

La trans-metilazione degli acidi grassi da gliceridi ad esteri metilici viene effettuata in presenza di un catalizzatore, che può essere acido o basico.

La trans-metilazione acida permette di formare esteri metilici a partire sia da acidi grassi legati al glicerolo che da acidi grassi liberi (Figura 7.1).

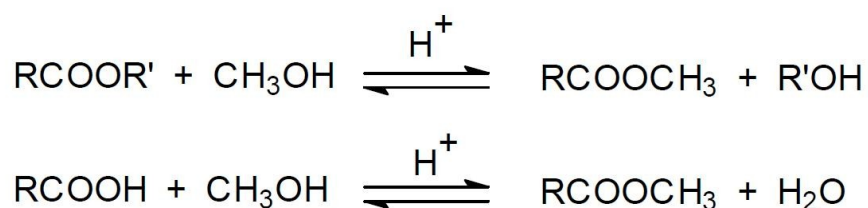


Figura 7.1 – Reazioni di trans-metilazione acida

La reazione avviene ad alte temperature (80-100°C) e tempi di circa un'ora in presenza di metanolo anidro e di un catalizzatore acido. I catalizzatori acidi maggiormente utilizzati sono HCl, H₂SO₄, BF₃ e CH₃COCl.

L'utilizzo di HCl e H₂SO₄ a temperature elevate e per tempi prolungati può causare reazioni di isomerizzazione degli acidi coniugati con conseguente formazione di isomeri *trans-trans* [Yurawecz, 1999]. HCl e H₂SO₄ possono inoltre indurre la formazione di metossi e idrossi derivati e causare la migrazione dei doppi legami, modificando quindi il

profilo acidico originario del campione. Le elevate temperature di reazione possono inoltre causare l'evaporazione dei metil esteri più volatili, ovvero quelli a corta catena come il butirrato ed il caproato.

L'utilizzo di acetil cloruro (CH_3COCl) consente di ridurre sia i tempi di reazione che la temperatura permettendo di completare la reazione di trans-metilazione in 1 ora a 60°C . Il metil acetato che viene a formarsi non interferisce con la reazione di trans-metilazione [Chin, 1992]. La temperatura è però ancora troppo elevata per evitare l'evaporazione dei metilesteri. Il BF_3 consente invece di lavorare anche a temperatura ambiente ma ha vita breve e l'utilizzo del reattivo vecchio, o anche troppo concentrato, provoca la formazione di artefatti e la perdita di un quantitativo apprezzabile di acidi polinsaturi [Christie, 1990].

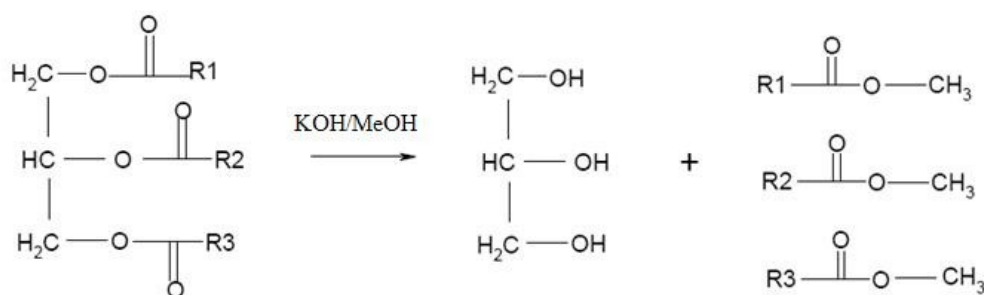


Figura 7.2 – Trans-metilazione basica con KOH in metanolo.

La trans-metilazione basica, descritta come metodo ufficiale nella norma: Metodo di preparazione degli esteri metilici del grasso del latte IDF 182:2002/ISO 15884:2002 – “Milkfat – Preparation of fatty acid methylesters”, consente la formazione degli esteri metilici a partire da acidi grassi esterificati col glicerolo ma non da acidi grassi liberi. L'ambiente di reazione deve essere quindi anidro per evitare l'idrolisi dei trigliceridi e seguente formazione di acidi grassi liberi. La reazione avviene in eccesso di metanolo anidro, in presenza di un catalizzatore basico, a temperatura ambiente e in tempi molto ridotti rispetto a quelli necessari per la trans-esterificazione in ambiente acido (Figura 7.2). In queste condizioni blande non esistono reazioni collaterali che possano interferire con l'analisi; inoltre, la volatilità dei metilesteri è notevolmente ridotta. I catalizzatori utilizzati (NaOCH_3 , KOH e NaOH) hanno un'ottima attività catalitica anche a temperatura ambiente [Christie, 1993]. Lo svantaggio della trans-metilazione basica risiede nel fatto che non è possibile esterificare la totalità degli acidi grassi di alcuni tipi di prodotti (i formaggi ed il

latte che ha subito fenomeni di lipolisi) in quanto contenenti un quantitativo elevato di acidi grassi liberi [De la Fuente, 2006].

7.3 Caratterizzazione analitica dei lipidi derivatizzati

I metilesteri, ottenuti in seguito a trans metilazione acida o basica della materia grassa, vengono separati, identificati e quantificati tramite gascromatografia.

La gascromatografia è una tecnica che permette la separazione, l'identificazione e la quantificazione di più composti di una miscela, in base alla loro diversa ripartizione tra una fase stazionaria ed una mobile; la fase stazionaria, solida o liquida, è contenuta in una colonna mentre la fase mobile, gassosa, fluisce nella colonna. La separazione dei componenti della miscela è determinata dalla natura chimico-fisica della fase stazionaria in cui i diversi composti vengono trattenuti per un periodo di tempo differente a seconda del loro grado di affinità ad essa. I diversi composti emergono dalla colonna, trasportati dalla fase mobile, in tempi differenti e vengono rilevati da un detector, solitamente un rivelatore a ionizzatore di fiamma (FID), che converte la concentrazione del componente nella fase mobile in un segnale elettrico che viene amplificato e trasmesso ad un sistema di elaborazione che registra il gascromatogramma. Il tempo impiegato dal composto per percorrere la fase stazionaria e raggiungere il FID è chiamato tempo di ritenzione (RT). Ogni composto della miscela ha il suo tipico tempo di ritenzione, e questo concorre al suo riconoscimento.

Lo strumento utilizzato, il gascromatografo, è formato da cinque componenti principali:

- la sorgente del gas di trasporto (He, N₂, H₂, Ar), di solito costituita da normali bombole di acciaio;
- il sistema di iniezione e la zona di vaporizzazione della miscela in esame, in cui la miscela introdotta viene istantaneamente vaporizzata e catturata dal gas di trasporto;
- la colonna cromatografica (di tipo capillare o impaccato), contenuta in una camera termostata a circolazione di aria calda (forno). L'eluizione dei composti può avvenire in isoterma (temperatura costante) o in programmata di temperatura;
- il FID, rivelatore universale e distruttivo, produce una fiamma (alimentata da aria e H₂) che brucia i singoli componenti separati dalla colonna trasformandoli in stato di ioni in fase gassosa. Gli ioni producono una corrente, proporzionale al numero di

atomi che l'ha generata, che viene registrata, amplificata e trasmessa al sistema di acquisizione. La sensibilità del FID è molto elevata: il limite di rivelabilità, in condizioni ottimali, può arrivare a circa 10^{-9} - 10^{-12} g con dinamiche di risposta lineare di 10^6 - 10^8 ;

- il sistema di acquisizione ed elaborazione dei dati, solitamente un Personal Computer con un software di gestione dello strumento e di registrazione ed elaborazione dei cromatogrammi.

Le prestazioni di un sistema gascromatografico sono dettate in maggior parte dall'efficienza della colonna che dipende da fattori quali:

- natura e velocità del gas di trasporto;
- lunghezza della colonna: un aumento della lunghezza della colonna consente di ottenere separazioni più efficaci ma prolunga i tempi di analisi;
- natura della fase stazionaria: scelta in base al tipo di composti da separare;
- temperatura della colonna;

L'ottimizzazione dei parametri sperimentali permette di ottenere un'elevata risoluzione dei picchi cromatografici che – in termini ottimali – si caratterizza con il loro isolamento, rispetto ai picchi contigui, a livello della linea di base.

La tecnica gascromatografica è oggi la più utilizzata per la caratterizzazione del profilo acidico del grasso degli alimenti. Gli acidi grassi individuali vengono identificati con ragionevole certezza per confronto dei tempi di ritenzione degli analiti con quelli dei composti autentici, e per additivazione di questi ultimi alla miscela analitica in esame. Le colonne che permettono la migliore separazione degli acidi grassi sono colonne capillari polari con una fase stazionaria liquida depositata su un supporto solido costituita da polimeri silossanici modificati o poliglicoli esterificati con succinati o adipati. In genere, per l'analisi degli acidi grassi, si utilizzano colonne molto polari della lunghezza di 100 metri, costituite da cianopropil-polisilossano. Questo tipo di colonne permette una separazione ottimale in cui l'eluizione dei diversi tipi di acidi grassi avviene in base alle seguenti priorità:

1. lunghezza della catena alifatica: l'eluizione procede a partire dall'acido a catena più corta;
2. grado di insaturazione: i tempi di eluizione aumentano all'aumentare del grado di insaturazione;

3. posizione del doppio legame: maggiore è la lunghezza del tratto di catena che separa il doppio legame dal metile terminale, minori sono i tempi di eluizione;
4. configurazione *cis-trans*; gli isomeri *cis* hanno tempi di ritenzione maggiori rispetto agli isomeri *trans*;
5. posizione della ramificazione, se presente: gli acidi grassi con ramificazioni *iso* vengono eluiti prima degli acidi grassi con ramificazioni *anteiso*.

Gli acidi linoleici coniugati (CLA) eluiscono invece dopo il metil-estere dell'acido α -linolenico. L'eluizione avviene secondo la seguente regola: a parità di isomero posizionale, gli isomeri *cis-trans* sono i primi ad essere eluiti seguiti dagli isomeri *trans-cis*. Seguono gli isomeri *cis-cis* e per ultimi i *trans-trans* (Figura 7.3). A causa della maggiore concentrazione dell'isomero *cis-9, trans-11*, la separazione tra i primi due isomeri non è in genere ottimale. È inoltre possibile che ci sia coeluizione tra isomeri del CLA e altri acidi grassi, come nel caso dell'isomero *trans-10, cis-12* che viene eluito insieme al C21:0 [Christie, 1990] [Cozzi, 2001b] [De la Fuente, 2006].

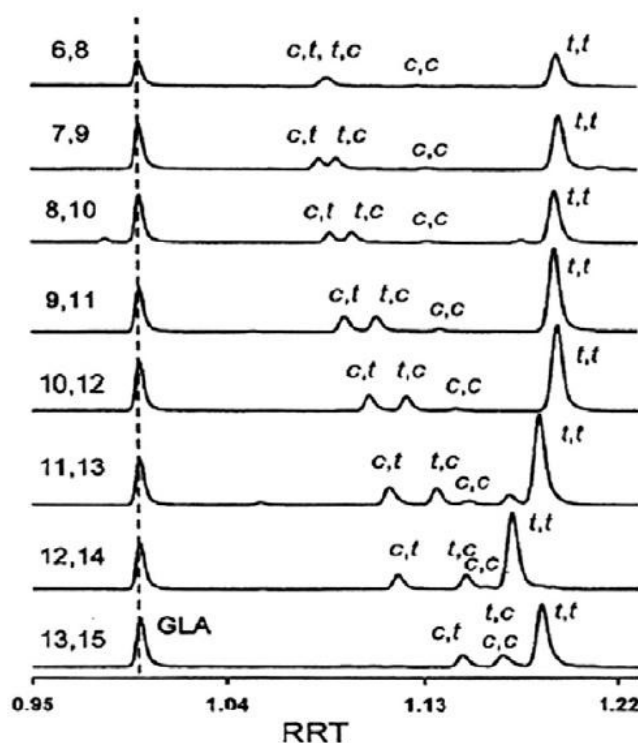


Figura 7.3 – Ordine di eluizione degli isomeri del CLA in relazione all'eluizione dell'acido γ -linolenico (GLA). RRT = Tempo di ritenzione relativo.

8 Analisi Multivariata

8.1 Introduzione

Negli ambienti di lavoro ed universitari ci si ritrova spesso a dover affrontare problemi complessi che necessitano di risposte immediate. In molti casi il processo o il sistema in esame non può essere descritto da una teoria ben definita e la risoluzione del problema è spesso difficile da trovare. Esempi sono dati dai problemi di natura ambientale, farmacologica e geologica, e da problemi di ottimizzazione o monitoraggio dei processi industriali. In questo tipo di sistemi, le variabili implicate nei processi sono numerose e spesso non è possibile controllarle tutte con la precisione desiderata. Nella maggior parte dei casi, le variabili non hanno tutte la stessa importanza nel processo studiato e la risoluzione del problema può essere mascherata dalla presenza di una o più di esse che, invece che fornire informazione, portano con sé rumore sperimentale, confondendo i veri effetti delle altre variabili. Inoltre, un effetto può essere causato dal ruolo combinato di due o più variabili e non essere presente se si considera solo una di esse. È quindi necessario che venga studiato il grado di correlazione e la dipendenza tra le variabili che descrivono il sistema in esame. I metodi chemiometrici permettono di lavorare sui dati relativi al sistema separando il contenuto di informazione utile per risolvere il problema da tutto ciò che invece è irrilevante (informazione non legata al problema da risolvere), ridondante (informazione dovuta alla correlazione tra più variabili) e dal rumore sperimentale (Figura 8.1).

L'obiettivo principale dei metodi chemiometrici è quindi quello di estrarre, dai dati in possesso, l'informazione utile per risolvere il nostro problema. Con l'analisi chemiometrica è possibile esplorare i dati visualizzandone la struttura e l'andamento evidenziando così relazioni, correlazioni e possibili problemi di incongruità. È possibile costruire modelli qualitativi e quantitativi evidenziando graficamente la presenza di raggruppamenti di dati relativi a situazioni tra loro differenti.

La chemiometria utilizza metodi matematici e statistici per l'analisi multivariata di dati. Al suo interno sono presenti metodi di modellamento, di classificazione e di regressione, l'analisi di similarità, l'analisi delle componenti principali e i metodi di intelligenza artificiale oltre che i metodi di disegno sperimentale e di ottimizzazione [Todeschini, 1998].

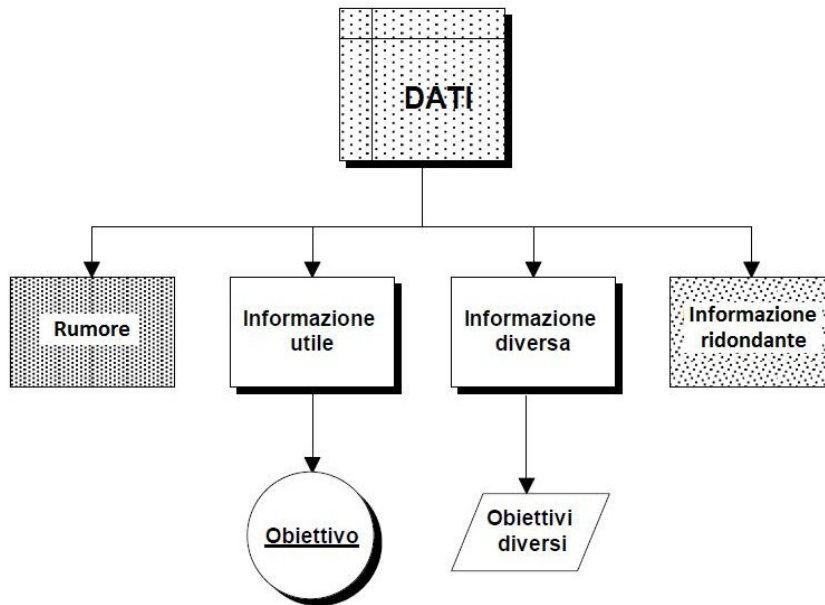


Figura 8.1 – Schema generico dell'informazione contenuta in un sistema di dati [Todeschini, 1998].

8.2 Trattamento dei dati

Quando si studia un sistema, si analizzano uno o più oggetti sui quali vengono determinati i valori di uno o più parametri. Ogni oggetto (campione) è quindi descritto da un dato numero di variabili. La chemiometria lavora su matrici di dati perciò i set di dati, per essere utilizzabili, devono avere la struttura riportata in Tabella 8.1.

Tabella 8.1 – Struttura di un set di dati

	Variabile 1	Variabile 2	Variabile 3	Variabile 4	...	Variabile p
Oggetto 1						
Oggetto 2						
Oggetto 3						
...						
Oggetto n						

Un set di dati può essere visualizzato graficamente ed il cervello umano decodifica un grafico in maniera più intuitiva rispetto ad una tabella di dati. Questa affermazione è dimostrabile anche per set di dati in cui ogni campione è descritto da un'unica variabile.

Guardando i dati in Tabella 8.2 ed osservando il grafico in Figura 8.2, ci si rende subito conto dell'importanza di un grafico. La distinzione dei campioni in due gruppi appare subito evidente nel grafico, mentre l'osservazione dei dati in tabella richiede uno sforzo maggiore per evidenziare i due gruppi. Oltretutto, dal grafico è possibile notare con facilità che il gruppo che presenta i valori più alti risulta essere più ristretto rispetto all'altro gruppo.

Tabella 8.2 – Dieci campioni descritti da una sola variabile [Leardi, 2013].

Campione	Variabile
1	25,3
2	22,1
3	25,5
4	25,6
5	19,4
6	25,7
7	20,2
8	21,3
9	25,9
10	21,8

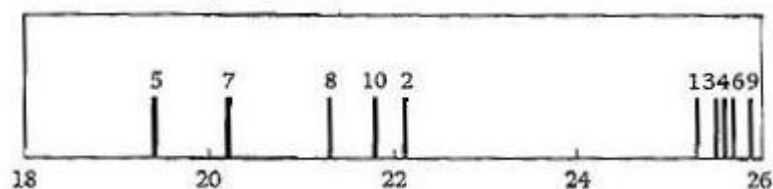


Figura 8.2 – Rappresentazione grafica dei dati della Tabella 8.2 [Leardi, 2013].

Prendendo in considerazione un set di dati leggermente più complesso, in cui ogni osservazione è descritta da due variabili (Tabella 8.3) ed il rispettivo grafico (Figura 8.3), si può dimostrare, oltre alla maggiore comprensione di un grafico rispetto ad una tabella da parte del cervello umano, che quando si considera solo una variabile del sistema, si può giungere a conclusioni errate. Il grafico delle due variabili mostrato in Figura 8.3 mette infatti in evidenza due gruppi di campioni che non verrebbero invece individuati considerando una variabile alla volta dato che nessuna delle due variabili è capace di

discriminare da sola i due gruppi. I campioni che presentano i valori delle due variabili più bassi sono inoltre più raggruppati rispetto agli altri campioni.

Con un set di dati in cui ogni campione è descritto da tre variabili, sarebbe sempre possibile visualizzare i campioni e l'intera informazione in un grafico tridimensionale, ma quando le variabili che descrivono il sistema sono più di tre bisogna ricorrere a delle tecniche che permettono di visualizzare in due o tre dimensioni la maggior parte dell'informazione di un sistema dimensionalmente numeroso.

La tecnica utilizzabile per questo scopo è l'analisi delle componenti principali (PCA), uno dei metodi di analisi multivariata più semplici e più utilizzati. La PCA è utilizzata soprattutto nell'analisi esplorativa dei dati per ottenere un quadro generale del sistema [Leardi, 2013].

Tabella 8.3 – Venti campioni descritti da due variabili [Leardi, 2013].

Campione	Variabile 1	Variabile 2
1	21,2	32,5
2	16,2	21,0
3	13,1	21,7
4	11,6	21,3
5	20,8	29,9
6	10,4	20,6
7	19,5	26,8
8	9,8	25,2
9	15,2	31,2
10	12,0	26,0
11	17,6	28,5
12	24,0	30,0
13	17,8	33,1
14	15,0	24,0
15	11,0	24,2
16	24,8	25,3
17	12,8	23,3
18	26,5	30,6
19	22,9	27,5
20	9,7	22,8

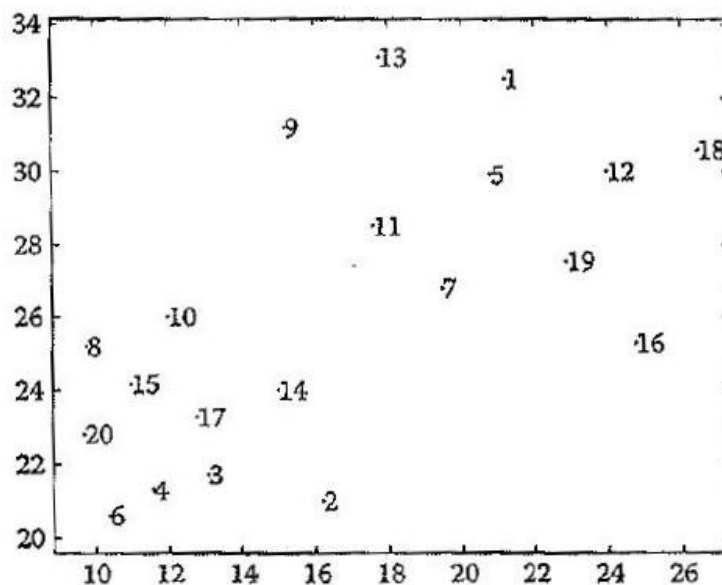


Figura 8.3 – Rappresentazione grafica dei dati della Tabella 8.3 (ascissa = variabile 1; ordinata = variabile 2) [Leardi, 2013]

8.2.1 Analisi delle componenti principali (PCA)

Spesso si ha a che fare con set di dati in cui i campioni sono descritti da un consistente numero di variabili, come per esempio gli acidi grassi nell'olio di oliva, o uno spettro di assorbimento, come per esempio gli spettri FT-IR di matrici alimentari, in cui le variabili che descrivono il campione possono superare le mille unità. In casi come questi sarebbe impossibile ottenere informazione valida senza l'aiuto di tecniche multivariate.

Da un punto di vista geometrico, si può considerare che i campioni giacciono in uno spazio di p dimensioni, ognuna di esse associata ad una variabile misurata. Un campione avrà coordinate che corrispondono ai valori delle variabili che lo descrivono. Sarà impossibile visualizzare tutta l'informazione in un unico grafico, ma si potrebbe visualizzare parte dell'informazione totale in grafici che considerano due o tre dimensioni. In questo modo però, l'informazione visualizzata sarà totalmente differente da grafico a grafico dato che ognuno è dato da combinazioni di diverse variabili. Inoltre, se alcune variabili sono altamente correlate, l'informazione in essa contenuta sarà pressoché la stessa; se sono perfettamente correlate, l'informazione sarà perfettamente identica ed una di esse potrà essere eliminata senza rischio di perdere informazione. La dimensionalità dello spazio verrà quindi ridotta di una variabile.

Non tutte le dimensioni hanno la stessa importanza e, a causa della correlazione presente tra di esse, la dimensionalità "reale" del set di dati sarà minore del numero totale di

variabili. Per questo motivo è necessario l'utilizzo di una tecnica capace di concentrare in poche "nuove" variabili, e quindi in poche dimensioni, la maggior parte dell'informazione del sistema. L'analisi delle componenti principali (PCA) è capace di raggiungere un simile scopo, riducendo la dimensionalità del set di dati ed estraendo la maggior parte dell'informazione rilevante, lasciando alle ultime dimensioni l'informazione non strutturata, ovvero il rumore.

La riduzione avviene tramite la costruzione di nuove variabili, chiamate componenti principali, non correlate tra di loro, ottenute per combinazione lineare delle variabili originali. Le componenti principali sono ordinate in modo tale che la maggior parte della variabilità del sistema sia contenuta nelle prime componenti. Le componenti principali rappresentano quindi gli assi cartesiani del nuovo sistema di riferimento in cui gli oggetti vengono proiettati ortogonalmente a partire dalla loro posizione nello spazio originale [Leardi, 2013].

La costruzione delle componenti principali avviene tramite un processo di rotazione dei dati, che permette di ottenere nuovi assi tali che il primo sia orientato nella direzione di massima varianza dei dati, il secondo sia ortogonale al primo e sia orientato nella direzione della successiva massima varianza dei dati e così di seguito per tutti i nuovi assi. In questo modo, ogni successiva componente principale descriverà una varianza inferiore rispetto a quella descritta dalla componente precedente.

La condizione di ortogonalità delle componenti principali impone che non esistano correlazioni tra di esse. Ciò è di grande importanza in quanto in questo modo ogni componente principale descrive differenti "aspetti" dei dati originali.

In Figura 8.4 è mostrata la costruzione delle componenti principali a partire da un sistema di riferimento costituito da tre variabili in cui vengono raffigurati un dato numero di campioni descritti dalle tre variabili. La prima componente principale descriverà la direzione di massima varianza dei dati (linea rossa), la seconda componente principale descriverà la direzione della seconda massima varianza dei dati, perpendicolare alla prima massima varianza (linea viola) e la terza componente principale descriverà la direzione della terza massima varianza dei dati perpendicolare alla prima ed alla seconda massima varianza (linea verde).

Il totale della varianza del sistema sarà dato dalla somma delle varianze descritte da ogni componente, ma la maggior parte della varianza totale sarà compresa nelle prime componenti. Le ultime componenti descriveranno una minima parte della varianza totale che può essere attribuibile a variabilità spuria e rumore sperimentale.

La PCA può essere effettuata su set di dati su cui è stata eseguita un'operazione di autoscaling (Paragrafo 8.2.2) che consente di osservare il sistema dal suo punto centrale e di considerare tutte le variabili come aventi la stessa importanza. In questo caso, l'origine del nuovo sistema coincide con il baricentro dei dati.

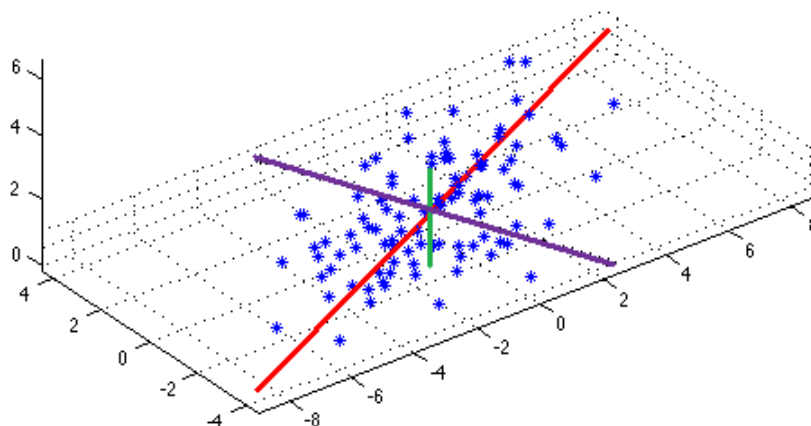


Figura 8.4 – Costruzione delle componenti principali a partire da un set di tre variabili (linea rossa = prima componente principale; linea viola = seconda componente principale; linea verde = terza componente principale).

Nell'esempio in Figura 8.4, tutta l'informazione è contenuta nelle prime due componenti principali mentre la terza componente descrive il rumore, ovvero la variabilità non legata ad informazione.

Lo scopo della PCA è quindi quello di massimizzare la varianza e di minimizzare la covarianza del sistema. I nuovi assi, ovvero le componenti principali, sono dati da una combinazione lineare delle variabili originali e sono di solito in numero inferiore rispetto alle variabili originali. Se le variabili originali sono correlate, sarà necessaria una sola componente principale per descrivere la loro variabilità, in quanto hanno un peso simile nella costruzione della componente. Se due variabili sono inversamente correlate, avranno peso simile ma segno opposto.

Come primo esempio applicativo della PCA, si può considerare un set di dati costituito da 130 campioni di acqua potabile ognuno descritto da 10 variabili [Leardi, 2013]: torbidità (1), solidi totali disciolti (2), temperatura (3), pH (4), conducibilità (5), cloruri (6), durezza totale (7), nitrati (8), ferro (9) e cloro residuo (10). I 130 campioni sono stati raccolti, tra luglio 2001 e febbraio 2003, in sette differenti siti di campionamento. I campioni sono quindi suddivisibili in sette gruppi composti rispettivamente da 21, 20, 18, 10, 21, 20 e 20

campioni. Per questo set di dati, lo scopo della PCA è trovare differenze e similarità tra i siti di campionamento e verificare se esiste un trend stagionale.

In Figura 8.5 è mostrato il grafico degli scores ottenuto dalla PCA del set di dati considerato. Il grafico degli scores rappresenta i campioni del set di dati nel nuovo sistema di assi cartesiani, codificati secondo la loro origine. In ascissa è infatti mostrata la prima componente principale (che descrive la direzione di maggiore varianza del sistema) mentre in ordinata è mostrata la seconda componente principale (che descrive la seconda direzione di massima varianza del sistema). Nel grafico si può notare una certa distinzione tra i gruppi, in particolare si vede che la variabilità all'interno di ogni gruppo è descritta dalla seconda componente (varia lungo la seconda componente), mentre la distinzione tra i diversi gruppi viene effettuata lungo la prima componente.

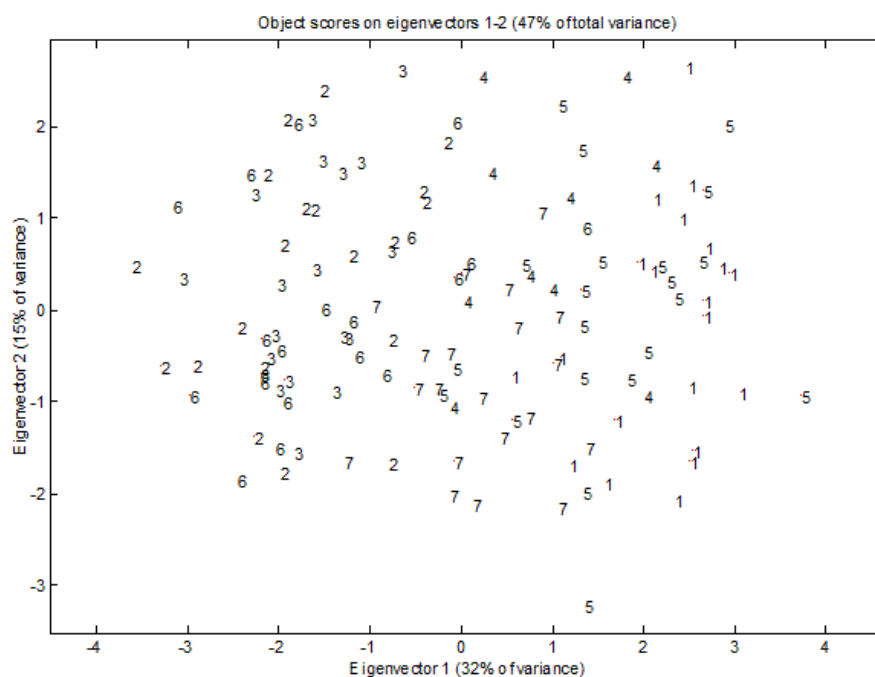


Figura 8.5 – Grafico degli scores del set di dati delle acque potabili (ascissa = prima componente principale; ordinata = seconda componente principale) [Leardi, 2013].

In Figura 8.6 è mostrato il grafico dei loadings che raffigura le dieci variabili originali in riferimento all'origine del nuovo sistema di assi cartesiani. Il grafico dei loadings mostra le relazioni tra le variabili originali ed il loro ruolo nelle diverse componenti, ovvero quanto contribuiscono alla costruzione di ogni componente principale. La prima componente

principale è composta prevalentemente dalle variabili 2 e 5 e, in maniera meno importante, dalle variabili 6, 7, 8, e 10, mentre la seconda componente principale è composta prevalentemente dalle variabili 1, 3, 4 e 9. Immaginando di sovrapporre i due grafici, si può dire che i campioni che si ritrovano nella direzione di una delle variabili originali, possiedono elevati valori di quella variabile. I campioni dei gruppi 1 e 5 hanno elevati valori delle variabili 2 (solidi totali disciolti) e 5 (conducibilità) mentre i campioni dei gruppi 2, 3 e 6 presentano i valori più bassi delle stesse variabili, in quanto posizionati nella direzione opposta della prima componente principale.

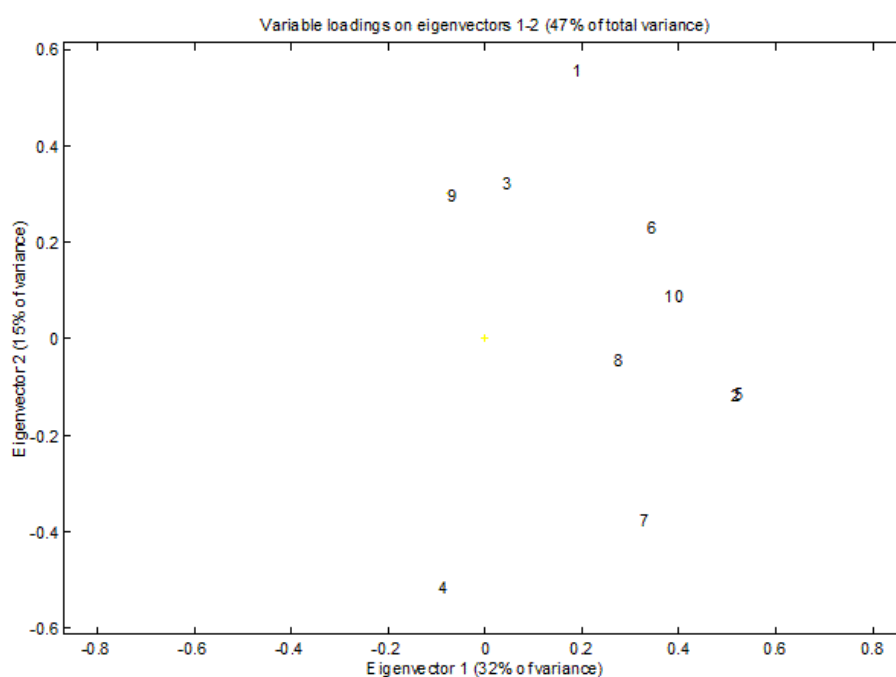


Figura 8.6 – Grafico dei loadings ottenuto dal set di dati delle acque potabili [Leardi, 2013].

Il grafico dei loadings consente di visualizzare la correlazione tra le variabili originali. Le variabili 2 (solidi totali disciolti) e 5 (conducibilità) sono altamente correlate, le variabili 6 (cloruri) e 10 (cloro residuo) sono correlate mentre le variabili 1 (torbidità), 3 (temperatura) e 9 (ferro) sono inversamente correlate alla variabile 4 (pH) in quanto situate in direzioni opposte rispetto al centro. La seconda componente descrive il contrasto tra la variabile 1 e la variabile 4. Osservando il grafico degli scores (Figura 8.5), si può vedere che i campioni dei gruppi 2, 3 e 6 sono sovrapposti, perciò dai tre siti di raccolta si ottiene la stessa acqua. La stessa conclusione si può ottenere per i gruppi 1 e 5 e per i gruppi 4 e 7.

Dato che la seconda componente principale descrive la variabilità all'interno dei gruppi e che le variabili originali che vengono descritte dalla seconda componente sono le variabili 1, 3, 4, 7 e 9, si può dire che la variabilità tra i campioni di acqua di uno stesso sito è data dalle variabili torbidità (1), temperatura (3), pH (4), durezza totale (7) e ferro (9).

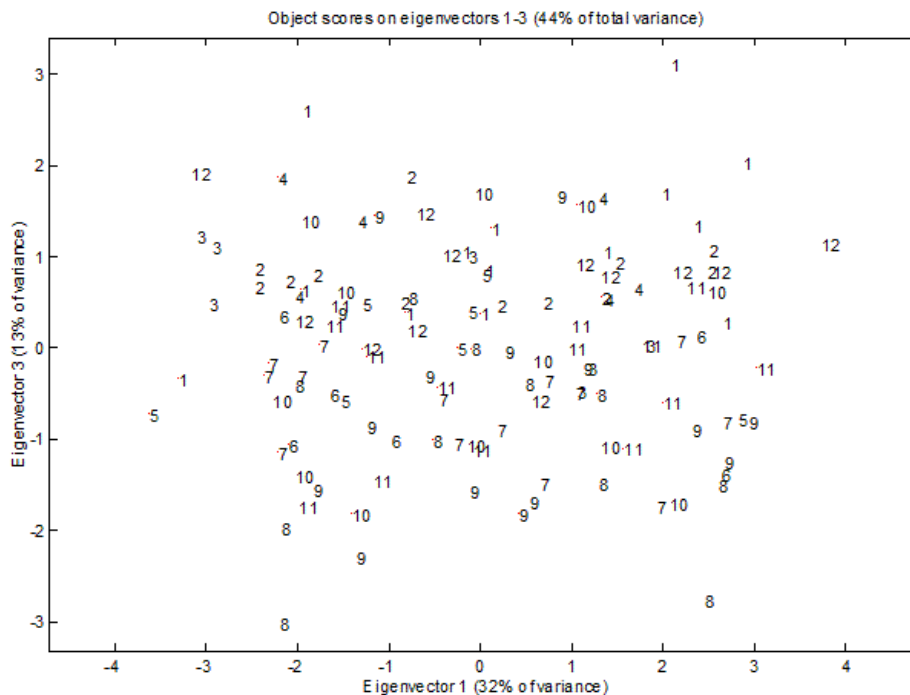


Figura 8.7 – Grafico degli scores del set di dati delle acque potabili (Prima e terza componente)

La Figura 8.7 mostra il grafico degli scores in cui vengono visualizzate la prima e la terza componente principale. I campioni sono codificati per mese di raccolta, da 1 a 12 corrispondenti ai mesi da gennaio a dicembre. Si può subito notare che la terza componente spiega la variabilità legata al periodo in quanto vi è una separazione tra i campioni invernali (codificati con 12, 1, 2, 3) dai campioni estivi (codificati con 7, 8, 9). I restanti campioni (primaverili ed autunnali) sono situati tra i gruppi estivi ed invernali. Tutte le considerazioni effettuate finora sarebbero impossibili da evidenziare osservando la tabella di dati del set di campioni [Leardi, 2013].

Un secondo set di dati, definito “Wine” (<http://ftp.ics.uci.edu/pub/machine-learning-databases/wine/> ultimo accesso 09/11/2013), contiene 178 campioni di vino descritti da 13 variabili: alcool (1), acido malico (2), ceneri (3), alcalinità delle ceneri (4), magnesio (5), fenoli totali (6), flavonoidi (7), fenoli non flavonoidi (8), proantocianidine (9), intensità del colore (10), tonalità del colore (11), OD280/OD315 di vini diluiti (12), prolina (13). Tutti i

campioni di vino provengono dalla stessa regione (Piemonte) ma derivano da tre diverse cultivar.

L'analisi delle componenti principali del set di dati "Wine" ha portato al grafico degli scores in Figura 8.8 in cui viene evidenziata la variabile categoria (che descrive a quale categoria appartiene ogni campione) tramite colorazione. In Figura 8.9 è mostrato il grafico dei loadings. Si può vedere che la prima componente principale descrive la variabilità legata alle variabili fenoli totali (6), flavonoidi (7) e proantocianidine (9) (variabili direttamente correlate tra di loro) e alle variabili fenoli non flavonoidi (8) e alcalinità delle ceneri (4) (variabili direttamente correlate tra di loro). I due gruppi di variabili sono inversamente correlati. La prima componente descrive anche parte della variabilità della variabile acido malico (2), prolina (13), OD280/OD315 di vini diluiti (12) e tonalità del colore (11). La seconda componente descrive la variabilità delle variabili intensità del colore (10), ceneri (3), alcool (1) e magnesio (5). Confrontando grafico degli scores e grafico dei loadings si può dire che il primo gruppo (campioni in rosso, numerati da 1 a 59), presentano i valori più elevati di prolina (13), alcool (1), magnesio (5), fenoli totali (6) e flavonoidi (7) mentre i valori più bassi di queste variabili sono contenuti nei campioni del terzo gruppo (campioni in verde, numerati da 131 a 178). Il terzo gruppo presenta i valori più elevati di acido malico (2), fenoli non flavonoidi (8) e di intensità del colore (10). Il secondo gruppo (campioni in marrone, numerati da 60 a 130) presenta i valori più elevati di tonalità del colore (11) ed i valori più bassi di intensità del colore (10), alcool (1) e prolina (13).

Con la PCA è quindi possibile ottenere informazione semplicemente osservando dei grafici, senza aver postulato a priori delle ipotesi.

L'analisi delle componenti principali ci consente quindi di ottenere visualizzazioni grafiche molto efficaci, quali il grafico degli scores ed il grafico dei loadings, permettendo di individuare la posizione degli oggetti rispetto alle variabilità del sistema intero, evidenziando possibili raggruppamenti di campioni; di analizzare il comportamento dei campioni nelle diverse componenti cercandone similarità; di mostrare le relazioni e le correlazioni tra le diverse variabili originali, il loro ruolo nelle diverse componenti e quindi nei campioni del sistema.

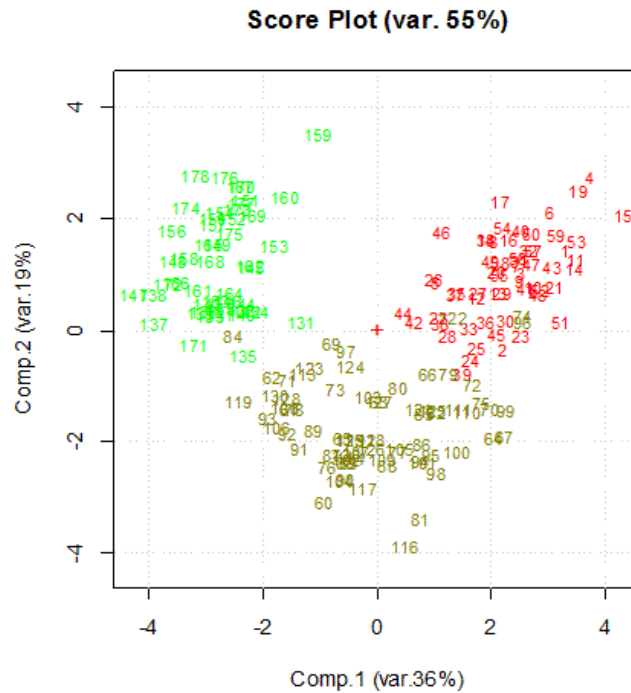


Figura 8.8 – Grafico degli scores del set di dati “Wine” con l’evidenza dei gruppi

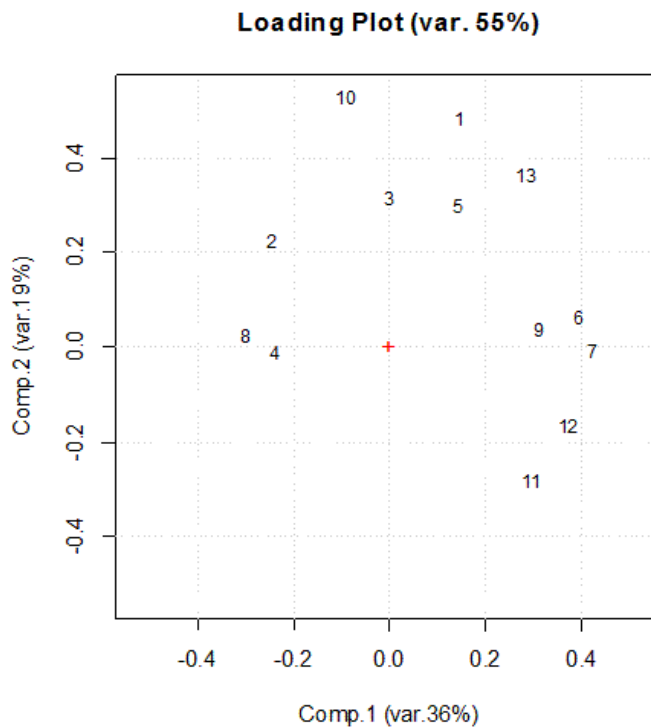


Figura 8.9 – Grafico dei loadings del set di dati “Wine”.

8.2.1.1 La costruzione matematica della PCA

La matrice dei dati utilizzabile in chemiometria ha la seguente forma:

$$X = \begin{bmatrix} X_{11} & X_{12} & \dots & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ X_{n1} & \dots & \dots & \dots & X_{np} \end{bmatrix}$$

La matrice X è costituita da n osservazioni (righe) ognuna descritta da p variabili (colonne). Le osservazioni sono gli oggetti (o campioni) che si hanno a disposizione per studiare il sistema mentre le variabili sono le grandezze (fisiche, chimiche etc...) che descrivono le osservazioni e quindi il nostro sistema. Ogni valore x_{ij} , con i compreso tra 1 e n , e j compreso tra 1 e p , corrisponde al dato della i -esima riga e j -esima colonna.

Tramite la PCA, da un numero p di variabili, si ottengono altrettante combinazioni lineari delle variabili p , che possiedono due importanti caratteristiche: non essere correlate ed essere ordinate in base alla varianza spiegata (ovvero all'informazione che contengono). Nel sistema originale, ogni dimensione (ovvero ogni variabile originale) porta con sé, teoricamente, una frazione dell'informazione totale pari a $1/p$. Nel nuovo sistema, l'informazione è concentrata nella prima direzione, e decresce progressivamente in modo che l'ultima direzione non contenga informazione ma rumore. La dimensione globale del sistema è sempre quella del sistema originale ma, visto che le ultime dimensioni spiegano solo una piccola parte di informazione, esse possono essere scartate. Si possono così prendere in considerazione solo le prime dimensioni (componenti significative), il cui esame permette di ottenere una buona rappresentazione dell'informazione del set di dati studiato.

Indicando con $\text{var}(z_j)$ la varianza espressa dalla componente principale z_j , si può scrivere:

$$\text{var}(z_1) > \text{var}(z_2) > \dots > \text{var}(z_p)$$

e che:

$$\sum \text{var}(x_j) = \sum \text{var}(z_j)$$

ovvero, la varianza totale del sistema nei due sistemi di assi è la stessa.

La prima componente principale esprime la massima varianza del sistema ed è data dalla combinazione lineare:

$$z_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p$$

La seconda componente principale esprime la seconda massima varianza del sistema ed è data dalla combinazione lineare:

$$z_2 = a_{21}x_1 + a_{22}x_2 + \dots + a_{2p}x_p$$

e così via per le componenti successive.

Da un punto di vista matematico, la PCA trova gli autovalori della matrice varianza-covarianza; essi corrispondono alla varianza spiegata dalla corrispondente componente principale.

Dato che l'informazione rilevante è concentrata nelle prime componenti principali, solo poche di esse vengono mantenute per successivi utilizzi. Il numero c di componenti principali mantenute sono quelle ritenute significative, mentre le componenti rimanenti sono quelle maggiormente associate al rumore.

Per trovare la posizione di un singolo oggetto nel nuovo sistema di assi, è sufficiente sostituire il termine x_j nelle equazioni di z_1 , z_2 , e successive, con il valore della variabile originale corrispondente per l'oggetto considerato. Il valore z_1 ottenuto definisce la coordinata dell'oggetto sulla prima componente principale. Il valore z_2 ottenuto definisce la coordinata dell'oggetto sulla seconda componente principale, e così via. Le coordinate ottenute sono definite "scores", mentre i coefficienti che moltiplicano ogni variabile originale x_j sono definiti "loadings".

Prendendo in considerazione i loadings delle variabili originali sulle differenti componenti principali, si può vedere quanto ogni singola variabile originale influisca sulla costruzione di ogni componente. Un elevato valore assoluto di loading di una variabile originale significa che essa ha un'influenza elevata sulla costruzione della componente principale in esame. Un basso valore assoluto significa che la variabile originale ha poca influenza sulla costruzione della componente principale in esame. Se il valore del loading ha segno positivo, l'oggetto con un alto valore della variabile corrispondente possiede un valore di score elevato e di segno positivo sulla componente in esame. Se il valore del loading ha segno negativo, l'oggetto con un alto valore della variabile corrispondente possiede un

valore di score elevato e di segno negativo sulla componente in esame. Come già menzionato, dopo aver effettuato una PCA, l'informazione risulta concentrata nelle prime componenti principali. Come visto negli esempi mostrati nel paragrafo 8.2.1, nel grafico degli scores è possibile rilevare similarità tra gli oggetti, mentre tramite confronto con il grafico dei loadings, è possibile interpretare le differenze tra gli oggetti (o gruppi di oggetti).

Matematicamente, si può dire che la matrice X originale (con numero di righe pari al numero di oggetti e numero di colonne pari al numero di variabili originali) è stata decomposta in una matrice S degli scores (con numero di righe pari al numero di oggetti e numero di colonne pari al numero c di componenti principali utilizzate) ed una matrice L dei loadings (con numero di righe pari al numero c di componenti principali utilizzate e numero di colonne pari al numero di variabili originali). Di solito, $c < p$, per cui è presente anche una matrice, la matrice E dei residui, che contiene la differenza tra i dati originali ed i dati ricostruiti dal modello della PCA (più piccoli sono i valori della matrice E , maggiore è la varianza spiegata dal modello) [Leardi, 2013]. La relazione descritta è la seguente:

$$X_{n,p} = S_{n,c} \times L_{c,p} + E_{n,p}$$

8.2.2 Pretrattamento dei dati

Il pretrattamento dei dati delle variabili permette di modificarli per raggiungere due scopi principali: (1) eliminare - o attenuare - caratteristiche indesiderate nella distribuzione dei dati, che può per esempio essere non normale o presentare asimmetrie accentuate, oppure (2) rendere confrontabili le variabili in modo da far sì che esse possiedano lo stesso livello di importanza nell'elaborazione dei dati. Il pretrattamento dei dati influenza quindi i risultati delle elaborazioni e deve essere scelto con estrema cura e solo in caso di reale utilità.

I pretrattamenti utilizzati per raggiungere il primo scopo sono dati da funzioni matematiche - quali le logaritmiche - che permettono di ridurre l'asimmetria delle distribuzioni.

La trasformazione più utilizzata è quella logaritmica:

$$x'_{ij} = \log(x_{ij}) \quad \text{oppure} \quad x'_{ij} = \log(1 + x_{ij})$$

I pretrattamenti utilizzati per raggiungere il secondo scopo sono dati dalle operazioni di scalatura dei dati che comprendono le operazioni di centraggio e quelle di pesata. Le

operazioni di centraggio permettono di avere una differente visualizzazione dei dati che possono essere osservati rispetto ad un punto preciso interno al data-set. Le operazioni di pesata vengono utilizzate quando le variabili sono descritte da quantità molto diverse tra di loro o quando sono espresse con differenti unità di misura. In questi casi, le varianze delle diverse variabili possono differire molto tra di loro e il risultato di un'elaborazione e la sua interpretazione saranno condizionati da quelle variabili con elevati valori numerici ed alta varianza, apparentemente più importanti, che mascherano il contributo delle variabili minori. Le operazioni di pesata permettono di rendere ogni variabile ugualmente importante, dando loro lo stesso peso *a priori*.

L'operazione di scalatura viene effettuata contemporaneamente su tutte le variabili, se pur indipendentemente l'una dall'altra [Todeschini, 1998] [Hair Jr, 2009].

Il più comune tipo di operazione di centraggio è il centraggio rispetto al valor medio:

$$x'_{ij} = x_{ij} - \bar{x}_j$$

i dati vengono centrati rispetto al valore medio di ciascuna variabile. È quindi un centraggio di colonna. In questo modo, il valore medio di ogni variabile ha valore zero e corrisponde al punto centrale del set di dati (Figura 8.10).

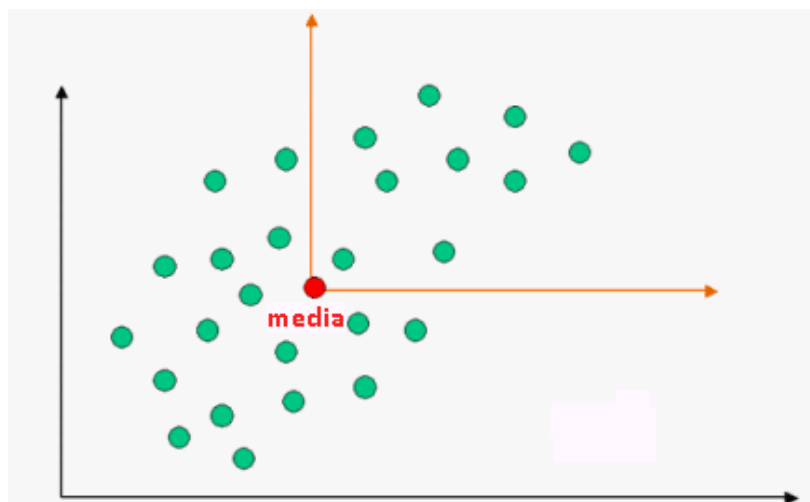


Figura 8.10 – Operazione di centraggio dei dati rispetto al valore medio

I più comuni tipi di operazioni di pesata (o *weighting*) sono i seguenti:

- Scalatura a varianza unitaria: i dati di ciascuna variabile vengono divisi per la deviazione standard (s_j) dei dati di quella variabile. In questo modo, ogni variabile ha deviazione standard, e quindi varianza, unitaria. Ogni variabile avrà così la stessa influenza nell'elaborazione dei dati e nella costruzione delle nuove coordinate del sistema nell'analisi delle componenti principali.

$$x'_{ij} = \frac{x_{ij}}{s_j}$$

- Autoscalatura (autoscaling): è il tipo di scalatura più utilizzato in chemiometria. I dati vengono centrati rispetto al valore medio e scalati a varianza unitaria. In questo modo, il valore medio sarà centrato sullo zero e la varianza dei dati sarà uguale ad 1.

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$$

8.3 Calibrazione multivariata e modelli di predizione

8.3.1 Concetti generali

L'analisi multivariata è utile non solo per osservare un set di dati ma anche per ottenere modelli empirici da cui ricavare un'informazione. I modelli possono essere di tipo qualitativo o quantitativo e racchiudono tra di essi metodi di classificazione e di regressione. Mentre i metodi di classificazione permettono di assegnare ad un campione incognito una classe di appartenenza, i metodi di regressione permettono di ottenere dati quantitativi di parametri in cui esiste una dipendenza con altre variabili.

Se per esempio si ha un set di dati formato da campioni di vino di cui sono stati registrati gli spettri FT-IR e di cui si conoscono dei parametri chimici quali il contenuto alcolico, il pH, e l'acidità totale, sarà possibile utilizzare questi dati per costruire un modello di predizione dei parametri chimici in base allo spettro dei campioni di vino. In pratica si va a cercare una relazione tra ogni parametro chimico e lo spettro. Una volta stabilita la relazione e una volta validata, sarà possibile predire il contenuto delle variabili chimiche di

campioni di cui è noto solo lo spettro. L'utilizzo di un modello di predizione permette un'analisi rapida ed economica, contrariamente all'analisi chimica.

Una calibrazione multivariata trova la relazione esistente tra una o più variabili risposta y e il gruppo di variabili dei predittori x . La risposta y è dipendente quindi dalle variabili x per cui le variabili y sono definite variabili dipendenti mentre le variabili x sono definite variabili indipendenti. Lo scopo finale della calibrazione multivariata non è solo quella di descrivere la relazione tra le variabili x e y del set di campioni, ma di trovare un'applicazione pratica per campioni di cui, in un tempo successivo, son state misurate la variabili x .

Il modello che viene costruito è un polinomio lineare ($y = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p + E$) dove b_0 è l'intercetta, b_p ($p = 1, \dots, P$) sono i coefficienti di regressione e E è il residuo. Il metodo di calcolo "tradizionale" dei coefficienti di regressione si basa sul metodo dei minimi quadrati che lo rendono però inapplicabile per svariati set di dati in quanto con esso non si possono trattare set di dati in cui il numero di variabili supera il numero di campioni ed è sensibile alla collinearità delle variabili. Non è quindi possibile applicare il metodo dei minimi quadrati per una regressione in cui si prendono in considerazione gli spettri dei campioni in quanto le variabili sono numerose e molto correlate tra di loro.

Queste limitazioni sono superate quando si utilizza la costruzione delle componenti principali. Nella PCA, infatti, il numero di componenti principali è inferiore al numero delle variabili iniziali che descrivono il sistema ed inoltre le componenti principali sono ortogonali tra di loro, per cui descrivono variabilità non correlate, eliminando di fatto la correlazione tra le dimensioni utilizzate. È quindi possibile applicare il metodo dei minimi quadrati ai valori degli scores originati dalla PCA. La tecnica risultante è definita regressione sulle componenti principali (PCR). I componenti della PCR sono ottenuti prendendo in considerazione solo le variabili dipendenti x , senza cercare relazioni con la variabile risposta y , per cui non è detto che le prime componenti principali costruite, che esprimono la massima varianza, siano quelle che contengono l'informazione più utile per predire il valore della variabile risposta y . La tecnica di regressione che permette di superare questo ostacolo è la regressione sui minimi quadrati parziali (PLS). La PLS massimizza la covarianza tra le variabili dipendenti e la variabile indipendente. In questo modo, le prime componenti (dette anche "variabili latenti") costruite saranno quelle che meglio predicono la variabile risposta y . Inoltre, una versione più avanzata di PLS (PLS2) consente di modellare più risposte contemporaneamente.

Il punto critico di PCR e PLS è l'identificazione del numero di variabili latenti significative. Sebbene l'informazione sia contenuta nelle prime variabili latenti e il rumore nelle ultime, non è sempre facile determinare quando finisce l'informazione ed inizia il rumore. La selezione di un basso numero di variabili latenti potrebbe causare perdita di informazione (underfitting), mentre la selezione di troppe variabili latenti potrebbe causare l'inglobamento nel modello del rumore (overfitting).

Prima di effettuare una calibrazione è molto importante individuare eventuali outliers. Esistono tre tipi di outliers: outliers nello spazio x (campioni per cui le variabili x sono molto differenti da quelle del resto dei campioni; tali outliers possono essere rilevati osservando la PCA delle variabili x), outliers nello spazio y (campioni per cui le variabili y sono molto differenti da quelle del resto dei campioni), e campioni per cui la calibrazione del modello non è valida.

La bontà di una calibrazione può essere riassunta in due parametri: la percentuale di varianza spiegata dal modello e l'errore quadratico medio in calibrazione (Root Mean Square Error in Calibration, RMSEC). La prima dà un'idea di quanta variabilità del set di dati è catturata nel modello; il secondo dà informazioni sulla grandezza dell'errore.

Per poter essere utilizzato per la predizione di campioni incogniti, un modello deve essere validato.

La calibrazione di un modello viene effettuata su un set di campioni definito training set, ma la validazione deve essere effettuata su un set di campioni non utilizzato nella calibrazione, definito come test set. La composizione del test set è di grande importanza in quanto fattori esterni contribuiscono alla variabilità del sistema. Per esempio, è noto che, per prodotti naturali, la variabilità aumenta considerando diversi periodi e origini geografiche.

La validazione può essere effettuata su un test set che contiene campioni della stessa popolazione dei campioni presenti nel training set ma, per avere la certezza che il modello sia veramente affidabile nella predizione, il test set dovrebbe contenere campioni di periodi e origini differenti dai campioni utilizzati per la calibrazione. Nel primo caso, la validazione potrebbe presentare risultati più ottimistici rispetto alla validazione del secondo caso, e quindi la reale capacità predittiva del modello potrebbe risultare sopravvalutata. In particolare, i valori del bias nella predizione saranno inferiori.

La valutazione della predizione sul set esterno è effettuata confrontando i valori predetti dal modello con i valori "veri", ottenuti con un metodo di riferimento. Si ottiene così l'errore quadratico medio in predizione (Root Mean Square Error in Prediction, RMSEP).

Una volta che il modello ha superato la fase di validazione, potrà essere utilizzato per predire eventi incogniti di sistemi di cui si conoscono le variabili indipendenti. È necessario che il gruppo di campioni che formano il training set, descriva una variabilità che sia la più ampia possibile in modo da poter avere un modello più universale possibile. Il test set deve invece contenere campioni che descrivono una variabilità compresa nel range di variabilità dei campioni utilizzati per costruire il modello, altrimenti la capacità predittiva su di loro sarà scarsa.

In Figura 8.11 è mostrato lo schema generale della costruzione di un modello di predizione: dall'utilizzo del training set si costruisce un modello parziale utilizzato per predire la variabile risposta degli oggetti del test set. Nessuna parte dell'informazione contenuta nei dati del test set deve essere utilizzata per la costruzione del modello parziale. La validazione sul set esterno consente di valutare la capacità predittiva del modello, ottenendo tutti i dati necessari per descrivere il modello finale.

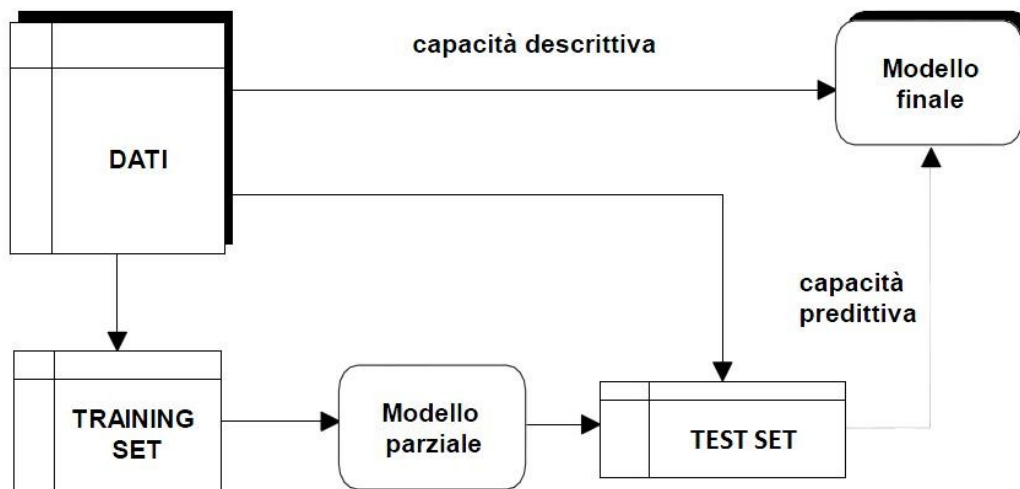


Figura 8.11 – Schema di costruzione di un modello di predizione [Todeschini, 1998].

8.3.2 Aspetti matematici dei modelli di predizione

Un modello è dato da un'equazione matematica che definisce una relazione causale tra diverse variabili indipendenti (predittori) e una, o più, variabile risposta.

Le variabili indipendenti sono rappresentate da una matrice denominata X mentre le variabili dipendenti sono rappresentate da una matrice denominata Y. Le due matrici sono utilizzate nel caso in cui il valore di una serie di variabili, le indipendenti, determinino il valore di un'altra serie di variabili, le dipendenti. La matrice X è definita “blocco X dei predittori” mentre la matrice Y “blocco Y delle risposte”. Il numero delle osservazioni (n) delle due matrici è ovviamente identico. Se i campioni provengono da diverse popolazioni o appartengono a categorie differenti è possibile inserire questa informazione nel set di dati tramite la variabile *categoria*, rappresentata da un vettore che presenta numeri interi che designano le classi. In presenza di più *categorie*, il vettore diventa una matrice usualmente denominata matrice C (Figura 8.12).

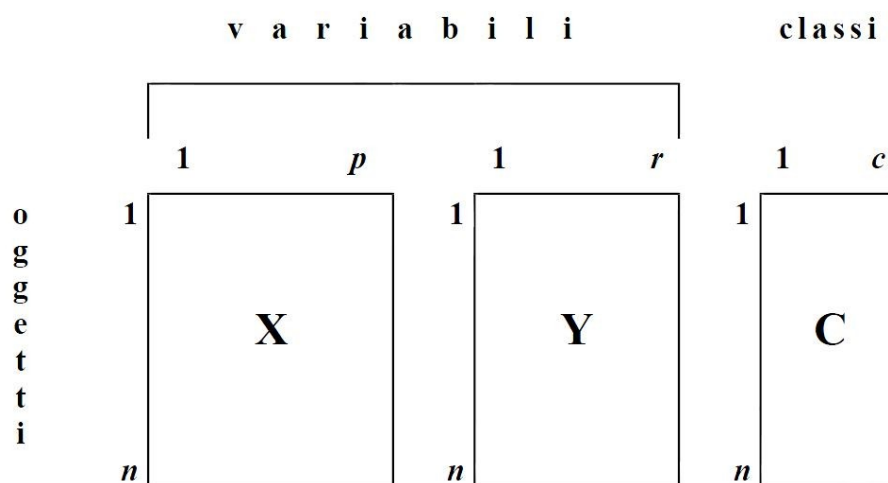


Figura 8.12 – Schema di un set di dati multivariato.

In un sistema di questo tipo, l'elaborazione delle matrici permette di ottenere un modello di predizione qualitativa, attraverso il quale i campioni incogniti possono essere assegnati ad una classe di appartenenza, o quantitativa, attraverso il quale si ottiene la predizione del valore delle variabili del gruppo Y in campioni incogniti.

Una volta definito il sistema e la variabile risposta che si vuole ottenere, si costruisce il modello a partire dal set di dati del sistema (training set), passando da una prima forma generale ad una forma più specifica, in cui si sono eliminate le fonti di incertezza, ovvero le cause che apportano rumore nei dati e le fonti di errore. Questa fase è denominata fase di calibrazione. Il modello viene modificato per migliorare le sue capacità predittive e

renderlo stabile a piccole alterazioni di dati e delle condizioni iniziali. Il modello verrà poi testato, in una fase detta di validazione, misurando la sua capacità predittiva rispetto ad un nuovo sistema di dati (test set), descritto dalle stesse variabili utilizzate nella costruzione del modello stesso, di cui si conosce a priori la variabile risposta.

8.3.2.1 La regressione multivariata

I metodi di regressione multivariata sono capaci di trovare la migliore relazione tra un insieme di variabili che descrive gli oggetti studiati (matrice X) e un insieme di risposte misurate per gli stessi oggetti (matrice Y). La forma della relazione descrive la modalità con cui la descrizione del sistema si raccorda con la misura sperimentale (il cosiddetto fitting) mentre il modello che si va ad ottenere, una volta validato, sarà utilizzabile per predire le future risposte di oggetti di cui sono note solo le variabili che li descrivono, ma non le misure sperimentali. I metodi di regressione sono quindi metodi matematici che forniscono informazioni sulle relazioni quantitative tra un numero di descrittori indipendenti (i predittori) e la variabile risposta. Ogni variabile indipendente viene pesata tramite un coefficiente di regressione per assicurare la massima predizione. Il set delle variabili indipendenti pesate forma una combinazione lineare delle variabili indipendenti, detto modello di regressione, che meglio predice la variabile dipendente.

Il modello di regressione che viene costruito è un'equazione lineare della seguente forma:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p + E = b_0 + \sum_{j=1}^p b_jx_j + E$$

dove y è il valore sperimentale della variabile risposta, x descrive le p variabili indipendenti e b i loro coefficienti di regressione, mentre E è il vettore del rumore del sistema.

La costruzione del modello viene effettuata tramite il calcolo matriciale, per cui un modello può essere indicato con la seguente notazione:

$$Y_{n,m} = X_{n,p}B_{p,m} + E_{n,m}$$

con Y matrice risposta (formata da n oggetti e m variabili risposta), X matrice dei predittori (formata da n oggetti e p numero di variabili), B matrice dei coefficienti di regressione (di grandezza $p \times m$), ed E matrice dei residui (di stessa grandezza di Y), che esprime il rumore del modello.

8.3.2.2 Regressione parziale con minimi quadrati: PLS

La principale tecnica di regressione utilizzata in chemiometria è la regressione dei minimi quadrati parziali (Partial Least Square Regression, PLS), che combina i metodi di regressione con la costruzione delle componenti principali (paragrafo 8.2.1.1). La PLS è una tecnica *biased* in quanto tiene conto della presenza degli errori sistematici ed è quindi utile anche nel caso in cui il rapporto tra gli oggetti e le variabili sia molto basso o le variabili molto correlate tra loro, oppure in presenza di più risposte da predire. Un esempio è dato dalle variabili spettrali, di solito molto più numerose degli oggetti e molto correlate tra loro.

La regressione PLS mette in relazione due matrici di dati X e Y attraverso un modello multivariato utilizzando una regressione parziale con minimi quadrati. Il principio della PLS si basa sulla costruzione di nuove direzioni, dette variabili latenti, a partire dalle matrici X e Y, costruite però in modo tale da massimizzare anche la covarianza tra i predittori e le variabili risposta. Le variabili latenti differiscono quindi dalle componenti principali in quanto la loro costruzione è influenzata dalla variabile risposta Y. Le prime variabili latenti costruite, che massimizzano la varianza di ogni matrice, corrispondono alle variabili che meglio predicono la variabile risposta Y, in quanto esprimono la migliore correlazione tra le due matrici.

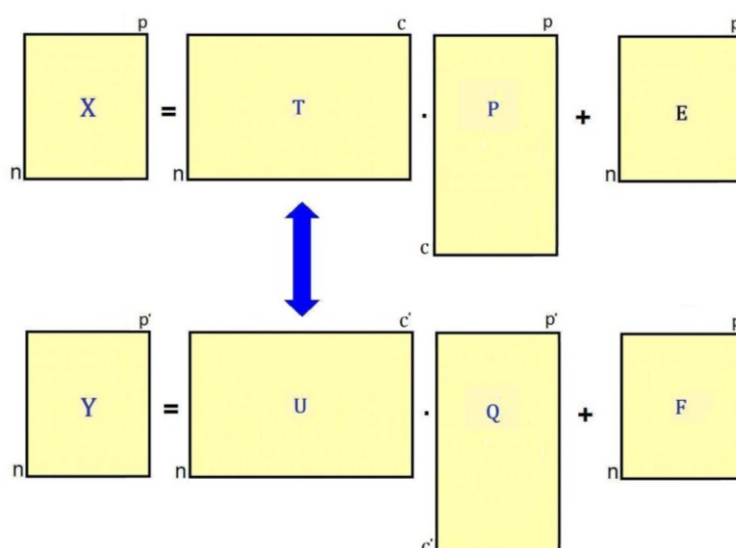


Figura 8.13 - Schema del calcolo dell'algoritmo PLS

Nella pratica, l'algoritmo PLS decompone entrambe le matrici X e Y. Dalla matrice X si ottiene la matrice T degli scores, la matrice P dei loadings e la matrice E dei residui che descrive il rumore mentre dalla matrice Y si ottiene la matrice U degli scores, la matrice Q dei loadings e la matrice F dei residui:

$$X = T \times P + E$$

$$Y = U \times Q + F$$

Le due matrici degli scores vengono calcolate in modo tale da massimizzare sia la varianza all'interno delle matrici X e Y sia la covarianza tra le matrici T e U e minimizzare i residui nelle matrici E ed F (Figura 8.13).

Il calcolo delle variabili latenti non viene quindi effettuato separatamente per le due matrici, ma simultaneamente. L'algoritmo PLS cerca quindi di trovare la direzione multidimensionale dello spazio X che spiega la direzione di varianza multidimensionale massima dello spazio Y. Gli scores della matrice T vengono poi utilizzati per predire gli scores della matrice U (regressione di U su T) che a loro volta vengono utilizzati per predire la variabile risposta Y (Figura 8.14).

Quindi, se:

$$U = T \times B$$

l'equazione finale del modello sarà:

$$Y = T \times B \times Q + F$$

dove B è la matrice dei coefficienti di regressione che mettono in relazione le matrici degli scores T e U. Questo processo non viene in realtà visualizzato e quello che l'algoritmo produce è il valore della variabile risposta Y a partire dalla matrice X dei predittori, motivo per cui le componenti principali vengono definite variabili latenti.

La PLS utilizza quindi il set di dati delle risposte in modo attivo durante l'analisi statistica, permettendo di bilanciare meglio l'informazione contenuta nelle matrici X e Y, riducendo l'effetto di grandi variazioni dei predittori, irrilevanti ai fini della descrizione del fenomeno e della predizione di oggetti incogniti. La PLS è il metodo di regressione multivariata meno restrittivo che ne permette l'utilizzo anche in casi speciali quali la presenza di un numero ridotto di osservazioni rispetto al numero di predittori. Il numero di variabili latenti ottimali viene poi scelto tramite uno dei metodi di validazione indicati nel paragrafo 8.3.4.

Nel caso in cui la risposta da predire sia solo una, e quindi che Y non sia una matrice ma un vettore, l'algoritmo PLS non calcola le variabili latenti per Y ma costruisce la matrice T degli scores, a partire dalla matrice X , sempre massimizzando la covarianza con il vettore risposta Y [Sadocchi, 1980] [Todeschini, 1998] [Nicolai, 2007] [Hair Jr, 2009].

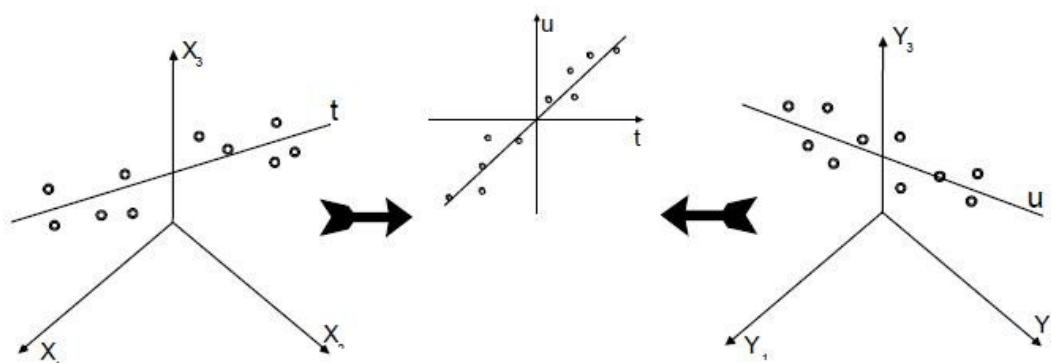


Figura 8.14 – Schema grafico della costruzione PLS [CAMO, 2006b].

8.3.3 Rischio di overfitting

La complessità di un modello è descritta dalla sua varianza totale perciò un aumento di complessità (dovuto ad un aumento del numero di variabili) determina l'aumento della varianza del sistema. La qualità descrittiva di un modello subisce un aumento ma ciò non è detto che avvenga anche per la capacità predittiva del modello. Il numero delle componenti principali considerate significative aumenta ma, come mostrato nell'esempio in Figura 8.15, un non controllato aumento della complessità del sistema fa decrescere la capacità predittiva del modello (overfitting): il modello parziale si adatta in pratica ai dati osservati usando un numero eccessivo di parametri.

Nel caso in Figura 8.15, il massimo potere predittivo è raggiunto con cinque componenti nonostante la varianza espressa in calibrazione aumenti all'aumentare del numero di componenti. La struttura del modello deve perciò essere controllata tramite validazione in modo da valutare la presenza di overfitting, dovuta a correlazione, rumore, specificità del campione, caratteristiche del metodo utilizzato o ingiustificata complessità del modello.

I metodi di validazione utilizzati differiscono principalmente dal modo con cui vengono ripartiti gli oggetti del training set e del test set. In alcuni metodi, detti di cross-validazione, l'intero data set è suddiviso in diversi gruppi, utilizzati a turno nella costruzione e nella

validazione del modello finale. In questo modo tutti i dati vengono utilizzati in entrambe le fasi di sviluppo del modello, procedimento utile in caso il numero di campioni presenti nel set di dati non sia consistente. L'utilizzo dell'intero data set permette di ridurre il bias.

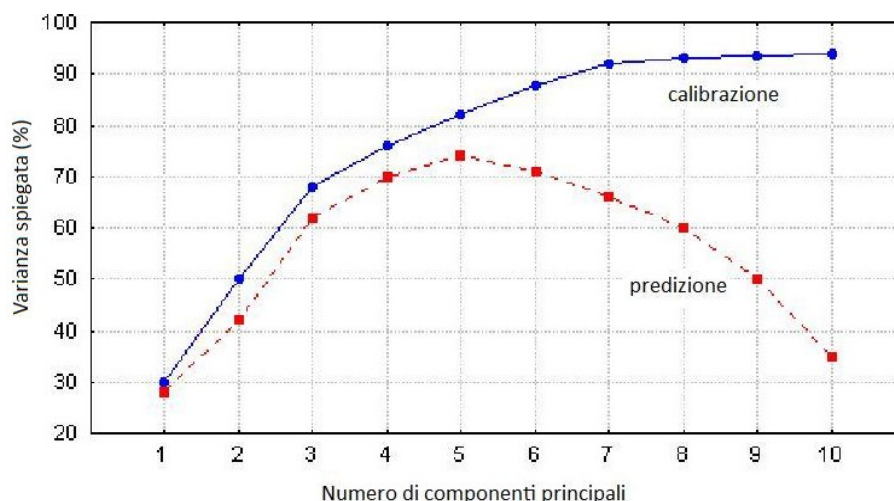


Figura 8.15 – Andamento della varianza spiegata all'aumentare della complessità del sistema [Todeschini, 1998].

8.3.4 Tecniche di validazione

Le tecniche di validazione più utilizzate sono le seguenti:

- External Evaluation Set: si basa sulla divisione dei campioni tra training set e test (evaluation) set. Il modello viene costruito sul training set e validato sul test set. La capacità predittiva è affidabile in quanto i campioni del set esterno non vengono mai utilizzati in nessuno degli step che servono per la costruzione del modello. La selezione dei campioni può essere effettuata in modo random oppure tramite una selezione manuale.
- Leave More Out: è il metodo di cross-validazione più utilizzato. Consiste nel dividere il set di dati in G gruppi, definiti gruppi di cancellazione, e costruire un numero G di modelli ognuno dei quali è costruito su $G-1$ gruppi e validato sul gruppo escluso. In pratica si elimina a turno un gruppo G_i , si costruisce il modello sui restanti gruppi e lo si valida sul gruppo escluso. Nel caso si abbiano tanti gruppi quanti sono i campioni ogni gruppo sarà costituito da un solo campione (Leave One

Out, da non usare se non in casi particolari perché, a causa della ridotta perturbazione che si ottiene rimuovendo un solo campione, porta a una sovrastima della capacità predittive del modello). L'errore in cross-validazione sarà dato dalla somma degli errori su tutti i campioni (ogni campione viene predetto una volta). I risultati del metodo cambiano però a seconda di quali campioni vengono a trovarsi nello stesso gruppo e la costruzione di un modello identico a partire dallo stesso set di dati porterà a capacità predittive leggermente differenti dal punto di vista numerico (anche se non significativamente). Il metodo Leave More Out generalmente risulta essere più ottimista rispetto ad un modello validato su un set totalmente esterno.

8.3.5 Parametri di valutazione dei modelli di regressione

Ogni modello di regressione può essere valutato in base ad una serie di parametri, calcolati nella calibrazione del modello e nella sua validazione, che forniscono un'informazione completa sull'attendibilità del modello stesso. L'informazione più utile proviene dal valore dei residui, utilizzati in diverse operazioni.

Si definisce residuo la differenza tra il valore sperimentale della variabile risposta (y_i) e quella calcolata dal modello (\hat{y}_i):

$$r = y_i - \hat{y}_i$$

Un modello di regressione è tanto migliore quanto più piccola è la somma dei quadrati dei residui RSS (Residual Sum of Squares):

$$\text{RSS} = \sum_i r^2 = \sum_i (y_i - \hat{y}_i)^2$$

Si definisce TSS (Total Sum of Squares), chiamata anche devianza, la somma dei quadrati delle differenze tra il valore i della variabile sperimentale (y_i) e la media dei valori sperimentali (\bar{y}):

$$\text{TSS} = \sum_i (y_i - \bar{y})^2$$

La correlazione tra la risposta sperimentale ed i predittori è dettata dal valore R , coefficiente di correlazione multipla, definito come:

$$R = \sqrt{1 - \frac{\text{RSS}}{\text{TSS}}} = \sqrt{1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}}$$

Il quadrato del coefficiente di correlazione multipla R^2 è definito coefficiente di determinazione e, quando viene moltiplicato per cento, rappresenta la varianza percentuale spiegata dal modello, ovvero la porzione di varianza di y che il modello è in grado di spiegare tramite predizione. R^2 è dato da:

Il quadrato del coefficiente di correlazione multipla R^2 è definito coefficiente di determinazione e, quando viene moltiplicato per cento, rappresenta la varianza percentuale spiegata dal modello, ovvero la porzione di varianza di y che il modello è in grado di spiegare tramite predizione. R^2 è dato da:

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

Viene inoltre definito R^2 adjusted, il valore R^2 che tiene conto dei gradi di libertà con cui è stato calcolato:

$$R^2_{\text{adj}} = 1 - \frac{\text{RSS}_{(p_i)}/(n - p_i)}{\text{TSS}/(n - 1)} = 1 - (1 - R^2) \times \left(\frac{n - 1}{n - p_i} \right)$$

dove p_i è il numero di variabili latenti utilizzate e $\text{RSS}_{(p_i)}$ è il valore RSS calcolato per un numero p_i di variabili latenti.

Il valore R^2_{adj} presenta un massimo quando il numero di p_i usati per il modello è ottimale e ridiscende quando vengono aggiunte variabili non significative. R^2 e R^2_{adj} misurano il fitting, ovvero la capacità del modello di raccordarsi contemporaneamente con tutti gli n oggetti utilizzati per costruire il modello.

Il valore RSS è utilizzato per calcolare il valore RMSEC (Root Mean Square Error of Calibration), l'errore quadratico medio in calibrazione:

$$\text{RMSEC} = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}}$$

con N numero di campioni utilizzati per costruire il modello. Minore è l'errore del modello, più piccola è la deviazione tra i valori sperimentali e quelli calcolati dal modello e quindi maggiore è la capacità di fitting dei dati.

Il valore RMSEC include anche il bias [Davies, 2006] e può essere scritto anche come:

$$\text{RMSEC} = \sqrt{\text{SEC}^2 + \text{bias}^2}$$

dove SEC (Standard Error of Calibration) è:

$$\text{SEC} = \sqrt{\frac{\sum_i (y_i - \hat{y}_i - \text{bias})^2}{N - 1}}$$

e dove il bias è dato da:

$$\text{bias} = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)}{N}$$

Spesso il valore del bias è di grandezza trascurabile e SEC e RMSEC differiscono in maniera non significativa.

I parametri che misurano la capacità predittiva del modello sono invece calcolati nella fase di validazione del metodo.

Così come sono stati definiti i residui del modello (in calibrazione), è possibile definire i residui in predizione (validazione) come la differenza tra y_i valore sperimentale della risposta, e $\hat{y}_{i(\text{val})}$ valore predetto nella validazione del modello per l' i -esimo campione non considerato nella costruzione del modello stesso:

$$r_{\text{val}} = y_i - \hat{y}_{i(\text{val})}$$

Normalmente i residui calcolati per la validazione sono più grandi di quelli calcolati per la calibrazione.

La somma dei quadrati dei residui in validazione è definita PRESS (Predictive Error Sum of Squares):

$$\text{PRESS} = \sum_i (y_i - \hat{y}_{i(\text{val})})^2$$

L'errore in predizione del modello è calcolato in base al tipo di validazione effettuata. Nel caso della cross-validazione bisogna considerare i gruppi di cancellazione utilizzati. L'errore, definito come RMSECV (Root Mean Square Error of Cross-Validation), è dato da:

$$\text{RMSECV} = \sqrt{\frac{\sum_{j=1}^Z \sum_{i=1}^N (y_{i,j} - \hat{y}_{i,j(\text{val})})^2}{N_T}}$$

con Z numero di gruppi di cancellazione utilizzati in cross validazione, N numero di campioni presenti in ogni di gruppo di cancellazione e N_T il numero di campioni totali. Allo stesso modo dell'RMSEC, l'RMSECV è comprensivo del bias ed è quindi dato da:

$$\text{RMSECV} = \sqrt{\text{SECV}^2 + \text{bias}^2}$$

dove SECV (Standard Error of Cross-Validation) è:

$$\text{SECV} = \sqrt{\frac{\sum_{j=1}^Z \sum_{i=1}^N (y_{i,j} - \hat{y}_{i,j(\text{val})} - \text{bias})^2}{N - 1}}$$

e dove il bias è dato da:

$$\text{bias} = \frac{\sum_{j=1}^Z \sum_{i=1}^N (y_{i,j} - \hat{y}_{i,j(\text{val})})}{N}$$

La varianza spiegata del modello in cross-validazione è data da Q^2 , definito come:

$$Q^2 = 1 - \frac{\text{PRESS}}{\text{TSS}} = 1 - \frac{\sum_i (y_i - \hat{y}_{i(\text{val})})^2}{\sum_i (y_i - \bar{y})^2}$$

Il parametro Q^2 è quindi valutato rispetto al potere predittivo del modello validato con una cross-validazione.

Nel caso in cui la validazione del metodo venga effettuata su di un test esterno, l'errore in predizione è dato dal valore RMSEP (Root Mean Square Error of Prediction):

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_{i(\text{val})})^2}{N}}$$

dove N è il numero di campioni presenti nel set esterno usato per la validazione. Minore è l'errore in predizione, migliore è la capacità predittiva del modello.

Anche l'RMSEP è comprensivo del bias:

$$\text{RMSEP} = \sqrt{\text{SEP}^2 + \text{bias}^2}$$

dove SEP (Standard Error of Prediction) è:

$$\text{SEP} = \sqrt{\frac{\sum_i (y_i - \hat{y}_{i(\text{val})} - \text{bias})^2}{N - 1}}$$

e dove il bias è dato da:

$$\text{bias} = \frac{\sum_{i=1}^N (y_i - \hat{y}_{i(\text{val})})}{N}$$

La varianza spiegata del modello validato su set esterno è data da R^2_v , definito come:

$$R^2_v = 1 - \frac{\text{PRESS}}{\text{TSS}} = 1 - \frac{\sum_i (y_i - \hat{y}_{i(\text{val})})^2}{\sum_i (y_i - \bar{y})^2}$$

in cui vengono sostituiti i valori sperimentali e predetti del set esterno.

La valutazione del modello costruito può essere effettuata confrontando l'errore in predizione del modello con la variabilità sperimentale massima del metodo analitico di riferimento s_{ref}^2 . La variabilità sperimentale si calcola in base ai valori di riferimento ottenuti da diversi campioni, replicati almeno due volte, analizzati in modo tale che le repliche dello stesso campione siano effettuate in tempi diversi e, se possibile, da operatori diversi, per massimizzare la deviazione tra le repliche. La variabilità sperimentale corrisponde alla deviazione standard raggruppata:

$$s_{\text{ref}}^2 = \sqrt{\frac{\sum_i (s_i^2 \times \text{gdl}_i)}{\sum_i \text{gdl}_i}}$$

dove i gradi di libertà del campione i sono dati dal numero di repliche effettuate sul campione i sottratto di un'unità. Nel caso ideale i due valori coincidono mentre, nei casi reali, l'errore in predizione del modello costruito è sempre numericamente maggiore rispetto alla variabilità del metodo analitico di riferimento [Todeschini, 1998] [CAMO, 2006a] [CAMO, 2006b] [Soyeurt, 2011] [Ferrand, 2011].

8.3.5.1 Valutazioni grafiche dei modelli di regressione

È possibile valutare anche graficamente la bontà di un modello. Il grafico “Predetto contro misurato” mostra la correlazione tra la variabile predetta dal modello e la variabile sperimentale (Figura 8.16).

La correlazione r tra le due variabili è calcolata tramite la formula di Pearson:

$$r = \frac{\text{COV}(X_1, X_2)}{S_{X_1} \times S_{X_2}}$$

dove $\text{COV}(X_1, X_2)$ indica la covarianza tra le due variabili X_1 e X_2 (nel caso del modello rappresentate dalle variabili Y predetta e Y sperimentale) e S_{X_1} e S_{X_2} sono le deviazioni standard delle due variabili. La covarianza è data da:

$$\text{COV}(X_1, X_2) = \frac{\sum_i (X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2)}{n - 1}$$

dove i valori indicati col simbolo \bar{X} sono i valori medi delle due variabili.

La retta costruita è della forma:

$$y = ax + b$$

dove a, il coefficiente angolare della retta, o slope, è dato da:

$$a = \frac{N \sum yx - \sum y \sum x}{N \sum x^2 - (\sum x)^2}$$

e dove b, l'intercetta o offset, è dato da:

$$b = \frac{1}{N} \left(\sum y - a \sum x \right)$$

Il valore di correlazione è però influenzato dal range di concentrazione studiato [Davies, 2006], assumendo valori maggiori all'aumentare di esso. Ha quindi solo un valore indicativo per cui l'efficacia del modello non può essere valutata in base ad esso.

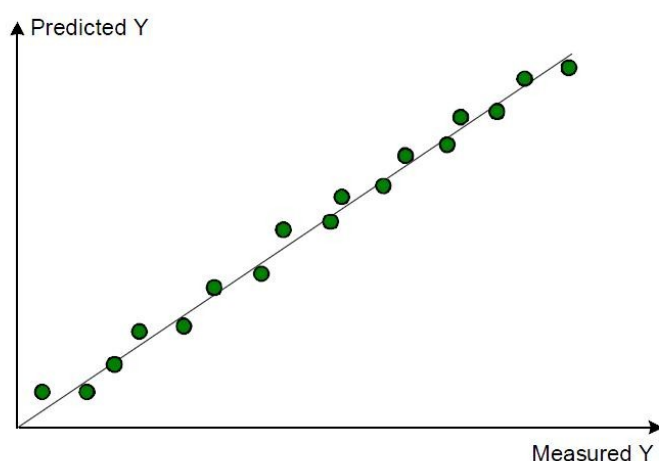


Figura 8.16 – Grafico “Predetto contro Misurato”.

Dal grafico “Predetto contro misurato” è possibile individuare graficamente i campioni che possiedono un residuo molto alto. Se un campione ha un residuo molto grande, e quindi chiaramente distante dalle altre osservazioni disponibili, è probabile che sia un outlier, ovvero un campione anomalo e aberrante (Figura 8.17). Un oggetto con un residuo molto alto può influire in modo maggiore sulla pendenza della retta rispetto ad un oggetto con un

residuo più piccolo. Questo effetto, chiamato leverage, definisce quindi quanto un dato oggetto ha influenza sulla costruzione del modello. Il leverage di un dato oggetto aumenta all'aumentare della sua distanza dal centroide dello spazio del modello. Nel centroide il leverage è nullo, e tale è la sua influenza sulla costruzione del modello, mentre è massimo nei punti più distanti dal centro del modello. Gli oggetti con leverage molto alto hanno una notevole influenza sul modello.

In Figura 8.18, si può vedere, in modo volutamente accentuato, come la presenza di due oggetti molto distanti dal centroide, che si discostano dal resto delle osservazioni, causino una variazione del coefficiente angolare della retta.

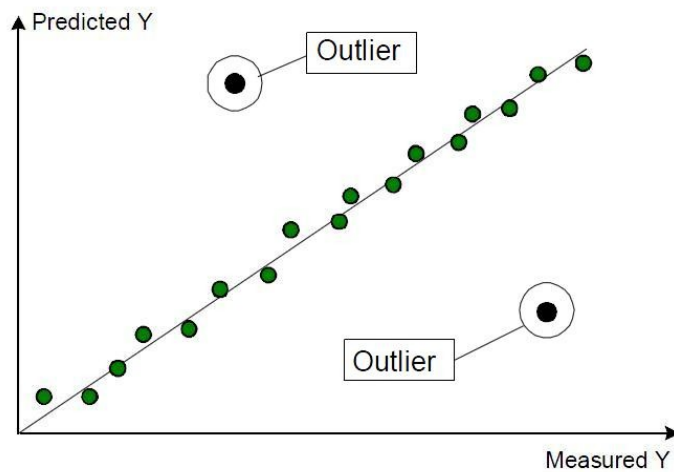


Figura 8.17 – Grafico “Predetto contro Misurato”: oggetti con residui elevati.

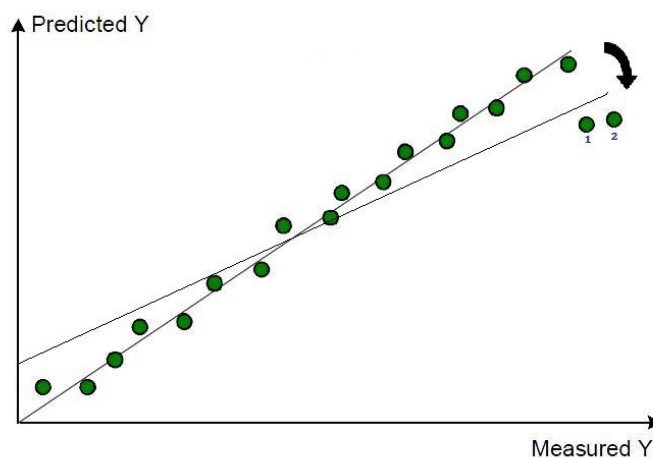


Figura 8.18 – Effetto leverage

Come per la varianza spiegata (Figura 8.15), è possibile mettere in grafico l'andamento della varianza residua (sia dei predittori che della variabile risposta) in funzione del numero di variabili latenti considerate (Figura 8.19).

Un buon modello usa un numero di variabili latenti che spiegano un'alta percentuale di varianza e la cui varianza residua tende allo zero.

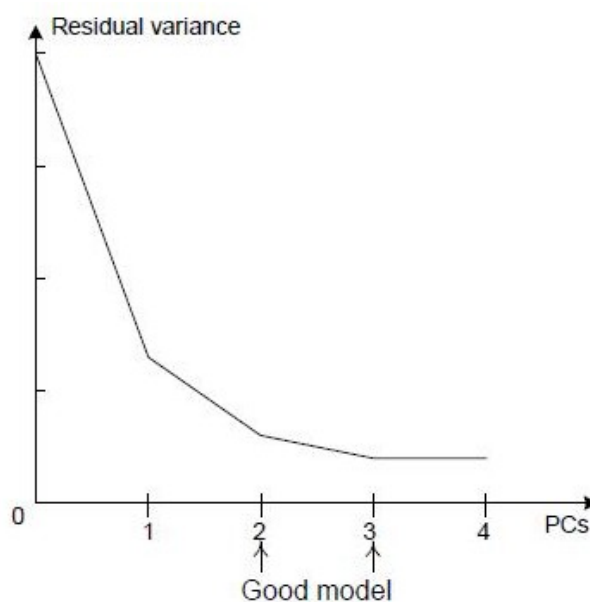


Figura 8.19 – Andamento della varianza residua in funzione delle variabili latenti (PCs)

Esistono una serie di grafici che permettono la visualizzazione dei residui, delle varianze, degli outliers e del leverage, mettendoli in relazione uno con l'altro o con altri parametri.

Di grande utilità è il plot diagnostico T^2 vs Q (Figura 8.20). L'indice T^2 rappresenta una variazione sistematica dei campioni ed il suo valore è legato al leverage posseduto dal campione. All'aumentare di T^2 , aumenta il leverage che un campione può possedere. T^2 descrive quanto il campione differisce dalla media e indica la distanza di un punto dal modello, nello spazio del modello. L'indice Q rappresenta una variazione disarmonica ed è relazionata alla varianza residua che un campione possiede. Q indica la distanza di un punto dallo spazio del modello. Il grafico mette quindi in relazione la varianza residua dei campioni (sia della X che della Y) con il loro leverage. Gli oggetti che presentano sia alta varianza residua che alto leverage, e che quindi si ritrovano nella parte in alto a destra nel grafico, sono da considerare pericolosi per la costruzione del modello in quanto, oltre ad essere descritti male dal modello, hanno un'alta influenza nella sua costruzione.

La visualizzazione grafica dell'andamento dell'errore in calibrazione e predizione è invece utile per decidere il numero di variabili latenti necessarie per la predizione. Più basso è l'errore in predizione, migliore è il modello costruito.

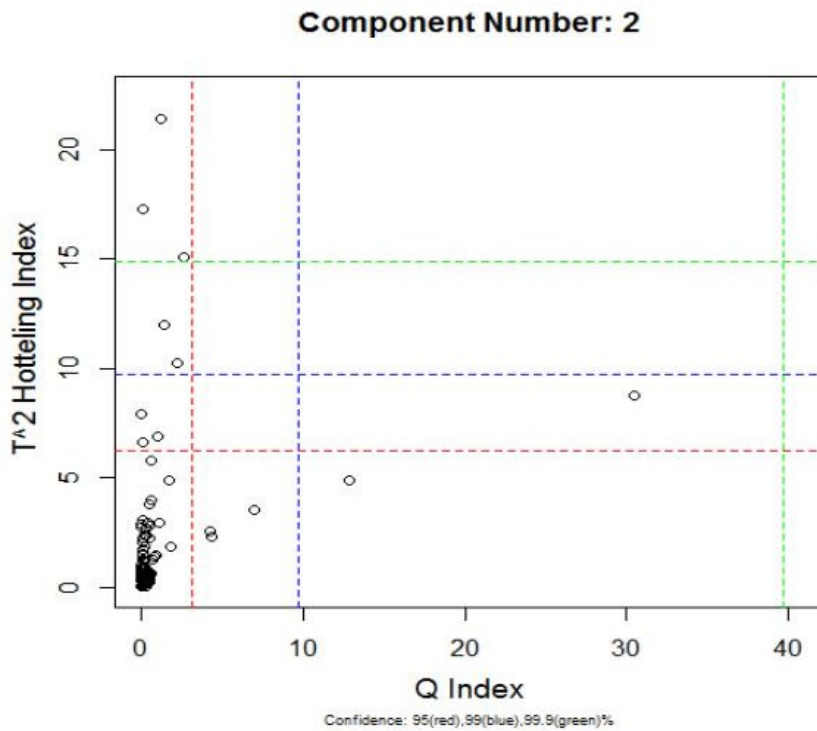


Figura 8.20 – Plot diagnostico T^2 vs Q

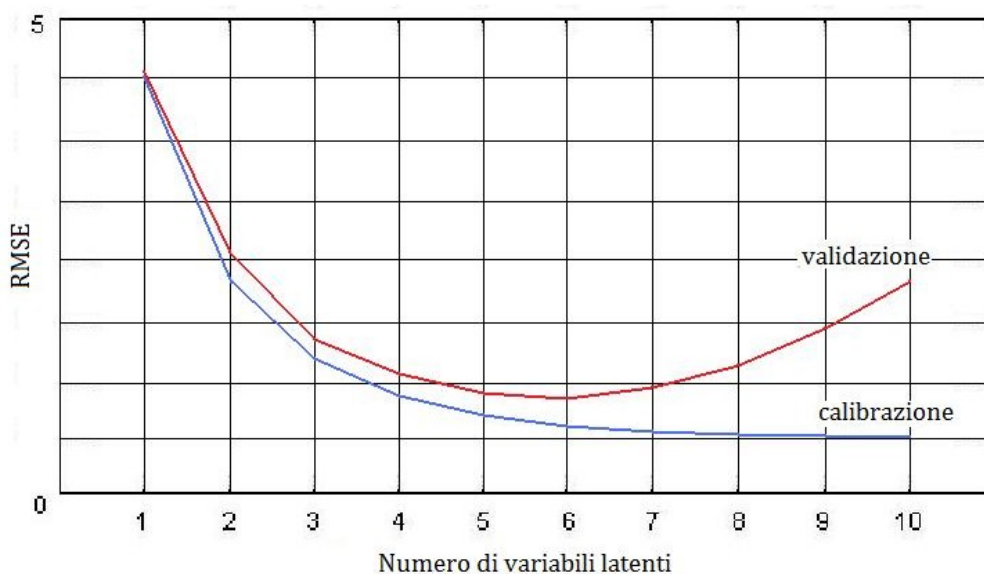


Figura 8.21 – Andamento degli errori in calibrazione e validazione rispetto al numero di variabili latenti

In Figura 8.21 si può vedere l'andamento monotono dell'errore in calibrazione dovuto al fatto che, incrementando il numero di variabili latenti, il modello tende a inglobare anche il rumore dei dati, adattandolo ad essi. L'errore in validazione, ovvero l'errore di predizione, ha un andamento differente in quanto la quantità di rumore dei dati del set esterno è diversa di quella dei dati del set di calibrazione, quindi la parte di modello che si adatta sul rumore dei dati di calibrazione provoca sicuramente un errore quando viene utilizzata con dati con rumore differente [Todeschini, 1998] [CAMO, 2006a] [CAMO, 2006b] [Soyeurt, 2011] [Ferrand, 2011].

8.4 Lo spettro infrarosso come matrice dei predittori

Grazie alle moderne tecniche di analisi spettroscopiche (infrarosso, raman...) è possibile registrare spettri di matrici alimentari in pochi secondi. Nel modello di predizione, le variabili della matrice X dei predittori sono date dalle lunghezze d'onda dello spettro dei campioni registrati.

A volte, nella costruzione di un modello, è necessario pretrattare i dati spettrali tramite trasformazioni che li rendono più adatti all'analisi, riducendone il rumore e rendendo più visibile l'informazione utile. Spesso inoltre, è utile effettuare una selezione delle lunghezze d'onda dello spettro in modo da eliminare le zone non informative, migliorando il risultato della variabile risposta. Inizialmente, grazie alle moderne tecniche di acquisizione degli spettri, che ne riducono il rumore, si riteneva che una selezione delle bande non fosse necessaria [Thomas, 1990] ma, successivamente, l'effetto positivo di una selezione efficiente è stato gradualmente riconosciuto, sia per migliorare l'abilità predittiva che per ridurre la complessità del modello [Thomas, 1994].

8.4.1 Pretrattamenti spettrali

Gli spettri possono essere pretrattati con le operazioni descritte nel paragrafo 8.2.2 (trasformazioni logaritmiche, centraggi e scalature), ma i pretrattamenti spettrali più comuni per la riduzione del rumore sono i seguenti:

- **Smoothing:** lo scopo dello smoothing (lisciatura) è quello di rendere uno spettro rumoroso più chiaro e rivelare dati informativi parzialmente oscurati dal rumore. Lo smoothing agisce rimpiazzando ogni osservazione con la media delle osservazioni adiacenti calcolata includendo l'osservazione stessa.

- Multiplicative Scatter Correction (MSC): è un metodo utilizzato per compensare gli effetti additivi (di deviazione) e/o moltiplicativi (di amplificazione) nei dati spettrali in modo tale da evitare che dominino il segnale informativo. Si basa sulla costruzione di una retta di regressione per ogni spettro, espressa come funzione del valore medio di ogni lunghezza d'onda di tutto l'intero set di spettri. I coefficienti di regressione a (intercetta) e b (coefficiente angolare) calcolati sono utilizzati per correggere il valore di ogni campione secondo la seguente formula:

$$X'_{ij} = \frac{X_{ij} - a}{b}$$

- Derivate: la trasformazione derivata viene effettuata nel caso in cui lo spettro ottenuto sia poco chiaro, con bande poco dettagliate. La funzione derivata esalta le differenze fra le bande che risultano parzialmente sovrapposte nello spettro non derivato consentendone una maggiore risoluzione. Le bande ottenute con la derivata sono tanto più pronunciate quanto più strette sono le bande dello spettro. In questo modo si ha una maggiore discriminazione tra le bande strette e le bande larghe. Le operazioni di derivata permettono che lo spostamento costante della linea di base venga eliminato dalla derivata prima e che l'assorbanza di fondo che aumenta linearmente con la lunghezza d'onda venga eliminata dalla derivata seconda. Le derivate di ordine superiore, meno utilizzate, possono consentire di rivelare fenomeni che non appaiono chiari con l'utilizzo di derivate di ordine inferiore. La trasformazione derivata richiede che lo spettro sia continuo. Se uno spettro viene espresso in termini di assorbanza (A) in funzione della lunghezza d'onda (λ), le derivate degli spettri saranno date da:

$$\text{ordine zero: } A = f(x)$$

$$\text{primo ordine: } \frac{\partial A}{\partial \lambda} = f'(x)$$

$$\text{secondo ordine: } \frac{\partial^2 A}{\partial \lambda^2} = f''(x)$$

e così via.

In Figura 8.22 è possibile vedere l'effetto delle operazioni di derivata su una semplice banda di assorbimento gaussiana.

Molti software chemiometrici commerciali utilizzano la cosiddetta derivata Savitzky-Golay il cui algoritmo agisce prima di tutto con una particolare operazione di smoothing sull'intervallo di variabili considerate, basata sulla costruzione di un polinomio con il metodo dei minimi quadrati. Il polinomio, che fornisce il valore della nuova variabile "livellata", può essere utilizzato per effettuare le operazioni di derivata. L'operazione di smoothing è stata inserita nell'algoritmo in quanto l'operazione derivata è fortemente influenzata dalla presenza di rumore che verrebbe amplificato dalla derivata.

L'algoritmo seleziona una finestra di variabili spettrali data da N punti attorno al punto centrale n, tale che il numero di punti considerati sia uguale a 2N+1, costruendo un polinomio (di grado $j < 2N+1$) che minimizza l'errore quadratico medio rispetto ai dati. L'algoritmo effettua prima di tutto un cambio di variabile, sostituendo i valori delle lunghezze d'onda dei punti considerati nel seguente modo:

$$z = \frac{\lambda_i - \bar{\lambda}}{h}$$

con h intervallo che separa il punto λ_i considerato ed il valore medio $\bar{\lambda}$.

Il polinomio è dato da:

$$A' = a_0 + a_1z + a_2z^2 \dots + a_jz^j$$

Il valore A' corrisponde al nuovo valore "livellato" della variabile centrale n.

L'operazione di derivata sul punto centrale n viene effettuata utilizzando il polinomio costruito. In questo modo, la derivata prima sarà data da:

$$\frac{dA'}{dz} = a_1 + 2a_2z + \dots + ja_jz^{j-1}$$

la derivata seconda sarà data da:

$$\frac{d^2 A'}{dz^2} = 2a_2 + \dots + (j-2)j a_j z^{j-2}$$

e così via [Savitzky, 1964] [Owen, 2000] [CAMO, 2006b].

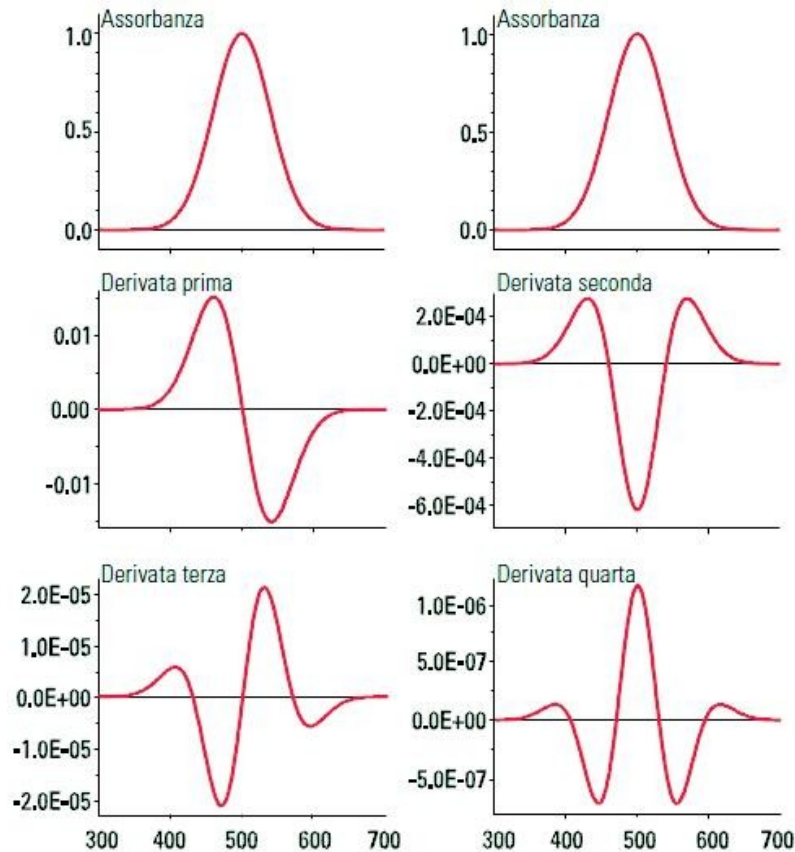


Figura 8.22 – Effetto delle operazioni di derivata su una gaussiana [Owen, 2000].

- Standard Normal Variate (SNV): è un trattamento spettrale che si basa sul centraggio di riga descritto nel paragrafo 8.2.2. Ogni spettro viene scalato individualmente secondo la formula:

$$X' = \frac{X_i - \bar{X}}{s}$$

dove X_i è il valore di un dato punto dello spettro, \bar{X} è il valore medio ed s è la deviazione standard dei valori dello spettro. Lo spettro viene quindi centrato sullo zero e la sua deviazione standard diventa unitaria.

L'operazione SNV permette di rimuovere gli effetti di scattering dallo spettro. La differenza con il trattamento MSC è che l'SNV agisce sugli spettri singoli mentre l'MSC utilizza la media dell'intero set di spettri.

- Averaging: è possibile effettuare la media di un sottogruppo di variabili spettrali per ottenere una riduzione del numero di variabili che consente di migliorare la precisione, ottenere risultati più stabili e più facilmente interpretabili.

8.4.2 Selezione delle variabili

Uno spettro di assorbimento di una sostanza pura presenta picchi riconoscibili dovuti all'assorbimento della radiazione da parte dei suoi gruppi chimici funzionali. Risulta quindi abbastanza semplice individuare i picchi e selezionare le bande di assorbimento corrispondenti. Quando viene invece registrato lo spettro di una matrice complessa, come per esempio il latte, la selezione delle lunghezze d'onda è abbastanza difficoltosa, in quanto, la presenza di più componenti, le loro interazioni chimico-fisiche ed altre fonti di deviazione della legge di Lambert-Beer, determinano uno spostamento od una sovrapposizione parziale o totale dei loro picchi di assorbimento, rendendo lo spettro di difficile comprensione. Oltretutto, l'impronta spettrale può essere cambiata dalle condizioni sperimentali, quali la temperatura. Una selezione delle variabili può migliorare le performance del modello e fornire un modello più robusto. Il suo obiettivo principale è la riduzione del rumore di fondo e della complessità del modello. I primi metodi di selezione di variabili si basavano sulla valutazione di tutte le possibili combinazioni delle variabili, il numero delle quali è dato da $2^p - 1$, con p numero di variabili totali. Il metodo è però semplice solo in un presenza di un numero ridotto di variabili mentre un numero troppo elevato di esse richiede tempi di elaborazione lunghissimi.

Il metodo di selezione storicamente più utilizzando è il metodo stepwise che si basa sulla costruzione di modelli di regressione in cui le singole lunghezze d'onda vengono mantenute o eliminate in base alla loro correlazione con la variabile risposta. Il metodo è dato dall'implementazione di due diverse tecniche, la *forward selection* e la *backward elimination*. Per entrambe le tecniche, si calcola la correlazione tra ogni variabile spettrale e la variabile risposta, applicando un test di verifica della significatività del coefficiente di correlazione, basato sul test F di Fisher, con una probabilità prefissata. La *forward selection* consiste nell'inserire, in un modello costituito da nessuna variabile spettrale, una variabile per volta in cui la prima variabile inserita è quella con la maggiore correlazione

con la variabile risposta (il maggiore valore F calcolato), la seconda variabile inserita è quella con la seconda maggiore correlazione con la variabile risposta e così via. Ogni qual volta viene inserita una variabile, si effettua una regressione tra le variabili presenti e la variabile risposta, valutandone il responso. L'inserimento termina quando l'aggiunta delle variabili determina un peggioramento del responso. La *backward elimination* consiste invece nell'eliminare, partendo da un modello in cui sono presenti tutte le variabili, una variabile per volta partendo da quella col minore valore di F. Ogni qual volta viene eliminata una variabile, viene effettuata una regressione con le variabili presenti nel modello. L'eliminazione prosegue fino a che il responso non subisce un peggioramento. Il metodo stepwise, in presenza di numerose variabili, non riesce a trovare i modelli migliori ma si ferma sui massimi relativi. È un metodo oggi superato da tecniche più efficaci, ma spesso è utilizzato per implementare altre tecniche.

Il metodo di selezione delle variabili basato sull'utilizzo degli algoritmi genetici (Capitolo 9) permette di selezionare solo le variabili informative per la predizione, riducendo così la possibilità di inserire nel modello variabili relazionate ad altra informazione, a rumore, e riducendo gli effetti della correlazione tra le variabili spettrali. Si ottengono così modelli parsimoniosi - ovvero contenenti il minor numero di variabili significative capaci di predire la variabile risposta - robusti ed affidabili anche in predizione su set esterno.

9 Gli algoritmi genetici

9.1 La teoria evolutiva

La computazione evolutiva utilizza i meccanismi dell'evoluzione biologica per affrontare problemi computazionali in svariati settori quando è necessario ricercare la soluzione tra un numero enorme di possibili alternative. Le regole della computazione evolutiva sono quelle tipiche della selezione naturale, con variazioni dovute ad incroci e/o mutazioni. L'obiettivo finale è quello di ottenere soluzioni di alta qualità per problemi complessi, che siano capaci di adattarsi al cambiamento dell'ambiente.

In biologia, il numero di possibili alternative è dato dall'insieme di tutte le sequenze genetiche e le soluzioni sono date da organismi altamente adattati che quindi possiedono una forte capacità di sopravvivere e di riprodursi. Questi organismi sono capaci di adattarsi all'ambiente e di competere o cooperare con gli altri organismi circostanti. Le condizioni mutevoli dell'ambiente fanno sì che i criteri di adattabilità cambino con conseguente evoluzione degli organismi. Le specie biologiche si evolvono nell'arco di millenni secondo variazioni casuali, come le mutazioni e le ricombinazioni, seguite dalla selezione naturale che si basa sulla conservazione delle variazioni genetiche favorevoli ed eliminazione di quelle sfavorevoli. Gli individui migliori saranno capaci di riprodursi trasmettendo alle generazioni future il loro materiale genetico.

L'evoluzione delle specie può essere quindi considerata una forma di ottimizzazione in cui la risposta da ottimizzare è la capacità di adattarsi all'ambiente e riprodursi. Bisogna però tener conto che ogni individuo possiede numerose caratteristiche e la loro importanza può essere estremamente differente. Gli esseri umani, per esempio, considerando come ottimo la lunghezza della vita e la sua qualità, possiedono caratteristiche che possono essere fondamentali (senza le quali un individuo non può sopravvivere, come l'assenza di patologie letali), importanti (come l'intelligenza, la forza e la bellezza) e non rilevanti (come il numero di piede o il colore degli occhi). Ogni variabile è quindi importante relativamente all'ottimo, ovvero allo scopo che ci si prefigge. L'uomo stesso ha applicato le teorie evolutive per i propri scopi, per esempio, selezionando, tra gli animali da latte, gli individui che hanno una produzione maggiore oppure selezionando i cavalli da corsa più veloci. In questi casi l'ottimo è dato da una specifica caratteristica da massimizzare.

Per capire i meccanismi alla base dell'evoluzione naturale bisogna definire alcuni concetti:

- la popolazione biologica è data da un gruppo di individui della stessa specie presenti nello stesso luogo, che si incrociano tra di loro;

- le unità ereditarie fondamentali degli organismi viventi sono i geni, ognuno dei quali definito dal tipo di basi azotate, e quindi dalla sequenza nucleotidica, che lo compongono.
- ogni variante di sequenza di un gene è chiamato allele;
- ogni individuo possiede una determinata sequenza di geni, il genotipo;
- ad ogni genotipo corrisponde un fenotipo, ovvero l'insieme di tutte le caratteristiche osservabili di un organismo vivente;
- l'insieme di tutti i genotipi di una popolazione costituisce il pool genico.

In base a questo si può dire che l'azione evolutiva avviene a livello genetico dell'individuo (genotipo) per cui tutto quello che i genitori apprendono nell'arco della loro vita non viene trasmesso geneticamente al figlio; la selezione naturale favorisce gli individui più efficienti in base alla propria combinazione genetica, selezionandone quindi il genotipo. Il processo evolutivo avviene tramite una ricombinazione genetica messa in atto da meccanismi biologici della riproduzione. Esiste inoltre la probabilità, seppur bassa, che avvengano mutazioni casuali del corredo genetico trasmesso. L'evoluzione avviene su intere popolazioni di individui tramite processi ciclici e generazionali che vengono determinati dall'ambiente e dall'interazione fra i vari individui della popolazione. Perché la popolazione possa evolversi è essenziale che sia caratterizzata da un'ampia variabilità genetica, in quanto vi è maggiore probabilità di adeguamento alle modifiche dell'ambiente. Una bassa variabilità genetica può condurre all'estinzione della popolazione.

L'aumento della variabilità genetica di una popolazione è causato dai seguenti processi:

- mutazione: data da una modificazione casuale della sequenza di basi azotate di un gene. Una mutazione modifica il genotipo di un individuo e può modificarne il fenotipo a seconda delle sue caratteristiche e delle interazioni con l'ambiente. La mutazione può avvenire in qualsiasi momento della vita di un organismo e può essere trasmessa per via ereditaria. La maggior parte delle mutazioni sono però svantaggiose, creando delle sequenze non funzionali.
- ricombinazione genetica: data dal rimescolamento dei geni dei genitori con creazione di nuove combinazioni di genotipi nella generazione successiva. La ricombinazione avviene nella fase di riproduzione.

9.2 Teoria evolutiva degli algoritmi genetici

Gli algoritmi genetici sono stati proposti da Holland negli anni '60, ma solo trenta anni dopo è stato possibile applicarli con tempi di elaborazione ragionevoli [Forrest, 1993] grazie all'avvento dei moderni computer. Gli algoritmi genetici sono quindi progettati per simulare ciò che avviene in natura, in modo tale da seguire le regole dell'evoluzione biologica; essi ricercano quindi delle soluzioni a fronte di condizioni mutevoli agendo, così come nell'evoluzione biologica, in parallelo su tutte le specie. Per fare ciò, c'è bisogno di una corrispondenza tra il materiale genetico e le condizioni sperimentali di un sistema in esame.

Si può dire che:

- materiale genetico = condizioni sperimentali.
L'adattabilità all'ambiente è funzione del materiale genetico così come il risultato di un esperimento è funzione delle condizioni sperimentali.
- gene = variabile.
Il materiale genetico è definito dai geni così come le condizioni sperimentali sono definite dal valore delle variabili studiate per l'esperimento.
- base azotata = bit.
L'informazione contenuta in ogni gene è definita dalla sequenza delle basi azotate: dato che esistono quattro diverse basi azotate, ogni gene è caratterizzato da una parola di lunghezza variabile scritta sulla base di un alfabeto di quattro lettere. Allo stesso modo, l'algoritmo genetico utilizza un codice binario per trasformare il valore di una variabile in una parola di lunghezza variabile scritta in bit (un alfabeto di due lettere, uno zero ed un uno).

In questo modo, le condizioni sperimentali possono essere codificate tramite una sequenza di numeri zero ed uno. Il termine inglese fitness, utilizzato negli algoritmi genetici, è sicuramente più adatto per indicare la funzione da massimizzare in quanto ingloba in sé numerosi significati, quali "adattamento", "adattabilità", "competitività", "successo biologico", "idoneità" [Mitchell, 1996] [Pearson, 2006] [Leardi, 2009].

9.2.1 La codifica delle variabili

La codifica delle variabili permette di assegnare un codice binario ad ogni condizione sperimentale del sistema in esame, permettendo di descrivere nello stesso codice più variabili alla volta, sia quantitative (per esempio la temperatura ed il tempo in cui avviene una reazione chimica) che qualitative (la presenza di un catalizzatore rispetto ad un altro o dell'agitazione di una miscela di reazione). La codifica delle variabili quantitative viene effettuata in base ad un intervallo di valori in cui la variabile può variare. All'interno dell'intervallo deve essere definita anche la differenza tra due livelli.

Se, per una reazione chimica, la variabile temperatura viene decisa variare tra 25°C e 60°C con un intervallo di 5°C, si otterranno otto diversi livelli di temperatura. Utilizzando due valori, zero ed uno, per il codice binario, ne viene che saranno necessari tre bit per descrivere la variabile temperatura, in quanto $8 = 2^3$. I codici a tre bit per ogni livello di temperatura saranno:

000 = livello 0 = 25°C

001 = livello 1 = 30°C

010 = livello 2 = 35°C

011 = livello 3 = 40°C

100 = livello 4 = 45°C

101 = livello 5 = 50°C

110 = livello 6 = 55°C

111 = livello 7 = 60°C

Se il tempo di reazione è variabile tra 10 e 40 minuti con un intervallo di 2 minuti, saranno necessari 4 bit per descrivere 16 diversi livelli temporali.

0000 = livello 0 = 10 min

0001 = livello 1 = 12 min

1110 = livello 15 = 38 min

1111 = livello 16 = 40 min

La codifica delle variabili qualitative viene invece effettuata utilizzando un solo bit con i due possibili valori zero ed uno. Per esempio, la codificazione per la presenza di agitazione della miscela di reazione sarà:

0 = livello 0 = nessuna agitazione
1 = livello 1 = presenza di agitazione

Se la reazione è possibile con due diversi catalizzatori A e B, la codifica sarà:

0 = livello 0 = catalizzatore A
1 = livello 1 = catalizzatore B

Se, per esempio, le condizioni sperimentali dell'esperimento chimico fossero date da $T=30^{\circ}\text{C}$ (codificato con 001), $t=20$ min (codificato con 0101), presenza di agitazione (codificato con 1) e presenza di catalizzatore A (codificato con 0), il codice binario sarebbe dato dalla stringa 001010110, formata dai 4 geni diversi corrispondenti alle 4 diverse variabili. Questa sequenza binaria, corrispondente al genotipo, viene definita cromosoma. Dopo aver codificato le condizioni sperimentali e aver generato il cromosoma corrispondente, viene misurata la risposta, o fitness (nell'esempio la resa della reazione chimica), che sarà associata al cromosoma [Leardi, 2009].

9.3 Passaggi degli algoritmi genetici

Gli algoritmi genetici seguono tre passi fondamentali:

1. creazione di una popolazione originaria;
2. riproduzione;
3. mutazione.

9.3.1 Creazione di una popolazione originaria

Il primo passaggio consiste nella decisione del numero n di individui di una popolazione. Ogni individuo possiede un cromosoma che viene determinato in maniera casuale. In questo modo ogni bit di ogni cromosoma assume un valore casuale che può essere 0 o 1. Se il cromosoma formato corrisponde ad una condizione sperimentale possibile, viene calcolato il suo fitness [Leardi, 2009].

9.3.2 Riproduzione

Nella fase di riproduzione si ha la ricombinazione dei cromosomi della popolazione creata nel primo passaggio. In natura, la probabilità che i migliori individui si riproducano è maggiore di quella degli individui peggiori. I genotipi migliori saranno ricombinati per generare nuovi genotipi mentre i genotipi peggiori andranno pian piano ad estinguersi. Gli algoritmi genetici utilizzano lo stesso principio riuscendo ad attribuire un valore di probabilità di riproduzione ad ogni cromosoma in funzione della loro risposta. La funzione utilizzata calcola la probabilità di riproduzione di ogni cromosoma (p_i) come il rapporto tra la risposta di quel cromosoma e la somma delle risposte della popolazione:

$$p_i = \frac{\text{resp}_i}{\sum \text{resp}}$$

Il valore di probabilità di ogni cromosoma è compreso tra zero ed uno.

La fase iniziale della riproduzione consiste nella copia dei cromosomi della popolazione iniziale N , a seconda della loro probabilità. Ordinando i cromosomi rispetto alla probabilità decrescente otterremo che i valori di p_i (con $i=1,2,\dots,n$) saranno ordinati in modo tale che p_1 corrisponda alla probabilità maggiore, p_2 alla seconda probabilità maggiore e così via. A questo punto vengono scelti n numeri casuali compresi tra 0,000 ed 1,000, che sceglieranno quali cromosomi verranno copiati a seconda dell'intervallo di probabilità in cui cadranno all'interno del range tra zero ed uno (Figura 9.1).

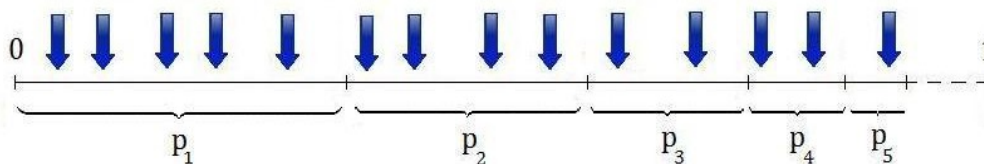


Figura 9.1 – Schema della selezione dei cromosomi più probabili.

Essendo l'intervallo di una probabilità maggiore più ampio dell'intervallo di una probabilità minore, è più probabile che i numeri random scelti rientrino nell'intervallo con probabilità maggiore. In Figura 9.1 si vede come i numeri casuali scelti (le frecce blu), siano presenti maggiormente negli intervalli di probabilità più ampi. Quando un numero cade all'interno di un intervallo, il cromosoma associato verrà copiato nella generazione successiva, dove è quindi più probabile trovare i cromosomi con un valore di p alto rispetto

ai cromosomi con un valore di p basso. I migliori individui verranno copiati più di una volta mentre i peggiori scompariranno. In questo modo la risposta media della generazione $N+1$ sarà statisticamente migliore della risposta media della generazione N .

A questo punto viene simulata la riproduzione. Gli n individui vengono fatti accoppiare formando $n/2$ coppie. Da ogni coppia vengono prodotti due nuovi individui (la prole) i cui cromosomi sono ottenuti tramite le tecniche di “crossover” a partire dai cromosomi genitori. Le tecniche di “crossover” più frequentemente applicate sono il “crossover singolo” ed il “crossover uniforme”. Nel “crossover singolo” viene selezionato in maniera casuale un punto di rottura dei cromosomi genitori, dividendoli in due parti: la parte a sinistra del punto di rottura contenente un numero x di geni, e la parte a destra del punto di rottura, contenente un numero y di geni, tali che $x+y=g$ con g numero totale di geni del cromosoma. Il primo cromosoma figlio sarà dato dall’unione dei geni x del primo genitore con i geni y del secondo genitore. Il secondo cromosoma figlio sarà dato dall’unione dei geni x del secondo genitore con i geni y del primo genitore (Figura 9.2).

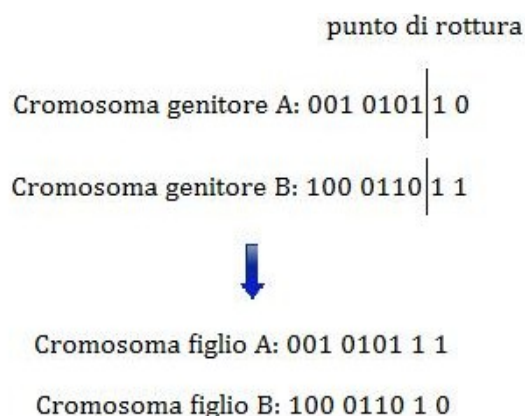


Figura 9.2 – Crossover singolo

Il limite di questo metodo è che l’ordine delle variabili all’interno del cromosoma risulta essere molto importante; la probabilità che due geni contigui nel cromosoma siano trasmessi ognuno ad un figlio diverso sarà bassa ed è data da $1/(g-1)$. Al contrario, il primo e l’ultimo gene saranno sempre trasmessi a figli diversi.

Nel “crossover uniforme”, ad ogni gene viene assegnato un numero casuale compreso tra zero ed uno. Il numero può rientrare o nell’intervallo tra 0,000 e 0,500 o nell’intervallo tra 0,500 ed 1,000. Nel primo caso, il gene del genitore A viene trasferito al primo figlio ed il

gene del genitore B viene trasferito al secondo figlio. Nel secondo caso avviene l'inverso, ovvero il gene del genitore B viene trasferito al primo figlio mentre il gene del genitore A viene trasferito al secondo figlio. In questo modo ogni gene di ogni genitore ha la stessa probabilità di essere trasferito ad un figlio o ad un altro [Leardi, 2009].

9.3.3 Mutazione

La mutazione sopperisce ad uno dei problemi derivanti dalle tecniche di crossover. Se i cromosomi genitori possiedono un gene identico, i cromosomi figli otterranno entrambi quello stesso gene e gli altri valori (alleli) di quel gene non saranno mai trasmessi alle generazioni successive, eliminando di fatto delle possibili condizioni sperimentali nei cromosomi figli.

Così, mentre il "crossover" agisce sui geni, lasciando i bit invariati, la mutazione fa variare i singoli bit di un gene, rendendo possibile la comparsa, nei cromosomi figli, di geni con una sequenza di bit non presente nei cromosomi genitori.

In natura, la mutazione avviene con una bassissima probabilità, mentre negli algoritmi genetici la probabilità di mutazione è dell'ordine dell'1-2%. Quindi, su una popolazione di n individui, se un cromosoma è formato da z bit (totale dei bit di tutti i geni) ed ha il 2% di probabilità che uno dei z_i bit venga mutato, il numero di mutazioni medio per generazione sarà dato da:

$$\text{mut} = n(0,02 \times z)$$

Alla conclusione dei tre passaggi base degli algoritmi genetici (creazione di una popolazione originaria, riproduzione e mutazione) si ottiene la seconda generazione, con un numero di individui identico a quello della prima generazione e con una risposta media superiore a quella della prima generazione. L'intero processo viene ripetuto più volte, con l'ottenimento di successive generazioni di risposta sempre migliore, fino a quando viene soddisfatto un criterio di arresto che può essere dato da un numero prefissato di generazioni, da un tempo di elaborazione predefinito o dall'ottenimento di un valore di risposta prefissato [Leardi, 2009].

9.4 Proprietà generali degli algoritmi genetici

La forza degli algoritmi genetici è la contemporanea applicazione di due strategie: l'exploration e l'exploitation. La prima si basa sull'esplorazione di punti del dominio sperimentale scelti a caso mentre la seconda cerca di raggiungere il massimo locale nell'intorno di un punto precedentemente esplorato. Mentre, normalmente, le tecniche classiche di ottimizzazione si basano esclusivamente sulla fase di exploitation, gli algoritmi genetici bilanciano i due aspetti cercandone il migliore compromesso. A fine elaborazione, gli algoritmi genetici non produrranno un'unica soluzione ottimale ma una popolazione di soluzioni estremamente buone di solito con risposta molto simile. La particolarità risiede anche nel fatto che le soluzioni estremamente buone fornite possono essere ottenute anche da condizioni sperimentali completamente differenti, lasciando all'operatore la scelta su quale condizione sperimentale sia migliore per lui, in base ai costi, al tempo o ad altri parametri.

Il tempo di elaborazione e la risposta ottenuta dagli algoritmi genetici dipende dalla grandezza della popolazione iniziale; una popolazione molto vasta conterrà una variabilità genetica molto ampia permettendone di esplorare una molteplicità di punti. In una popolazione piccola è invece molto probabile che i cromosomi dei diversi individui siano estremamente simili e quindi la variabilità genetica sia molto bassa. Il tempo di elaborazione per una popolazione vasta sarà perciò molto maggiore di quello necessario per una piccola popolazione perciò, nello stesso periodo di tempo, il numero di generazioni ottenute da essa sarà molto minore di quello ottenuto dalla popolazione più piccola, per cui un cromosoma molto buono avrà un effetto minore sull'elaborazione totale. L'aumento della grandezza di una popolazione ha quindi come effetto l'aumento dell'exploration ed una diminuzione dell'exploitation.

Nella fase di riproduzione, si è visto che la risposta media aumenta all'aumentare delle generazioni. È possibile però che il migliore cromosoma di una generazione sia in realtà peggiore del migliore cromosoma della generazione precedente. Questo è dovuto al fatto che i cromosomi genitori non vengono copiati nella generazione successiva ma finiscono per estinguersi. Per risolvere questo problema, gli algoritmi genetici utilizzano un processo chiamato "elitismo", copiando i k migliori cromosomi da una generazione alla successiva così come sono. Su una popolazione di n cromosomi, l'algoritmo produrrà $n-k$ cromosomi genitori ed $(n-k)/2$ coppie di cromosomi genitori, ottenendo $n-k$ cromosomi figli ai quali vanno aggiunti i k cromosomi, ottenendo così una seconda generazione di n individui, con i k cromosomi che possono anch'essi essere scelti come cromosomi genitori. Il caso

estremo è quello in cui si ha la massimizzazione dell'elitismo, cioè quando $k=2$. In questo caso vengono scelti i due migliori cromosomi e solo questi due, oltre a rimanere intatti nella popolazione, verranno selezionati per la riproduzione. Gli $n-k$ cromosomi non sono invece selezionati per la riproduzione e rimangono inalterati nella popolazione. I due cromosomi figli generati dai due cromosomi k , in base alla loro risposta, verranno o eliminati o mantenuti a scapito dei cromosomi con peggiore risposta, già presenti nella popolazione.

Il vantaggio dell'elitismo è che i k migliori cromosomi verranno conservati fino alla fine dell'elaborazione, permettendo inoltre tempi di elaborazione minori. Il rischio è invece dato dal fatto che, scegliendo un numero ridotto di k cromosomi, si riduce la capacità di exploration e l'elaborazione viene eseguita lavorando su una regione ridotta di cromosomi molto simili tra di loro. Una riproduzione che non considera l'elitismo produce un aumento della capacità della fase di exploration, ottenendo una variabilità genetica maggiore. Una riproduzione che invece massimizza l'elitismo produce un aumento della capacità della fase di exploitation, focalizzandosi su una determinata regione. Anche in questo caso gli algoritmi genetici cercano il migliore compromesso.

Anche nella fase di mutazione esiste un bilanciamento tra exploration e exploitation. Maggiore è il numero di mutazioni, maggiore è l'exploration; minore è il numero di mutazioni, maggiore è l'exploitation [Leardi, 2009].

9.5 Applicazioni degli algoritmi genetici

Gli algoritmi genetici sono diventati un importante strumento per l'ottimizzazione di problemi di varie aree disciplinari. Il loro utilizzo è rivolto a problemi la cui risposta è data da una soluzione ottimale globale o da una serie di soluzioni ottimali locali in quanto il sistema in esame è descritto da un insieme di variabili. Esempi sono dati da lavori in cui gli algoritmi genetici sono riusciti a trovare le conformazioni a più bassa energia di molecole organiche con risultati migliori di quelli ottenuti con altri metodi [Judson, 1993], o in cui la loro applicazione è stata utile nel campo del design molecolare per ottenere molecole con certe proprietà o attività [Venkatasubramanian, 1994a] [Venkatasubramanian, 1994b]. In altri lavori, gli algoritmi genetici hanno permesso di ricercare lo spazio conformazionale di piccole molecole tridimensionali per ottenere modelli farmacofori, di ottimizzare l'attacco di ligandi flessibili in siti proteici parzialmente flessibili e di sovrapporre automaticamente molecole flessibili [Jones, 1996] o di trovare strutture fisiche ipotetiche di

polietilentereftalato corrispondenti ad una certa combinazione di proprietà meccaniche e di contrazione [De Weijer, 1993] o, ancora, di risolvere problemi di gestione delle acque sotterranee quali l'ottimizzazione del pompaggio di acqua da una falda, lo sviluppo di un metodo di distribuzione dell'acqua con costi minimi e di un metodo di bonifica delle falde acquifere col minimo costo [McKinney, 1994].

9.6 Applicazione degli algoritmi genetici nella selezione di variabili spettrali

La selezione delle variabili per una calibrazione multivariata può essere considerata un problema di ottimizzazione. Gli algoritmi genetici applicati alla PLS si sono dimostrati essere efficienti per la selezione delle lunghezze d'onda degli spettri fornendo risultati migliori di quelli ottenuti dall'intero spettro di assorbimento [Leardi, 2000].

Bisogna inoltre tenere conto del fatto che le variabili spettrali presentano un'alta correlazione in quanto la variabile x_i è molto correlata con le variabili x_{i-1} e x_{i+1} . Per questo motivo si è soliti considerare regioni spettrali anziché singole lunghezze d'onda, ognuna delle quali formata da un dato numero di variabili adiacenti. Nel valutare quindi un modello bisogna considerare non solo il numero di lunghezze d'onda selezionate, ma anche il numero di regioni spettrali utilizzate.

Per poter adattare gli algoritmi genetici alla selezione di variabili spettrali, sono stati effettuati numerosi cambiamenti agli algoritmi più semplici, adattandoli allo scopo. Ciò che bisogna considerare è che non esiste un algoritmo ideale perciò la sua buona riuscita dipende dal grado di adattamento al problema considerato, con un buon bilanciamento tra le fasi di exploration e di exploitation [Leardi, 2009].

9.6.1 La corsa dell'algoritmo genetico

Prima di descrivere i passaggi degli algoritmi, è necessario definire come viene effettuata la codifica delle variabili spettrali: ogni spettro è descritto da un cromosoma composto da un numero di geni pari al numero di lunghezze d'onda dello spettro. Ogni gene è formato da un solo bit codificato 1 o 0 a seconda che quella variabile sia presa in considerazione o meno [Leardi, 1992]. Per esempio, per un insieme di 10 variabili, l'utilizzo delle variabili 1, 5, 8 e 9 produrrà il codice binario 1000100110.

Il primo passaggio effettuato dagli algoritmi genetici è dato dalla creazione di una popolazione iniziale che consiste nel 1) decidere il numero n di individui che ne faranno

parte; 2) creare un numero di cromosomi uguale al numero n di individui decisi per la popolazione; 3) valutare la risposta di tutti i cromosomi creati.

Negli algoritmi genetici semplici, i cromosomi iniziali vengono determinati in modo casuale, con una probabilità del 50% che una variabile sia selezionata o meno, e che quindi ogni gene del cromosoma sia codificato con 1 o con 0. Questo metodo ha però due svantaggi principali: 1) se il modello contiene numerose variabili, i tempi di computazione sono molto elevati; 2) è molto probabile che la selezione di circa metà delle lunghezze d'onda inglobi alcune variabili buone per la predizione; il risultato ottenuto sarà buono grazie alle poche variabili buone presenti, ma le risposte di diversi cromosomi saranno molto simili e sarà quindi molto difficile determinare quali variabili non sono buone per la predizione ed eliminarle dal modello finale.

Negli algoritmi genetici adattati alla selezione delle lunghezze d'onda di uno spettro, i cromosomi iniziali vengono determinati fissando un valore di probabilità di selezione $p = x_c/v$ con x_c numero medio di variabili che si è deciso di voler considerare in un cromosoma e v numero di variabili spettrali totali, ottenendo cromosomi contenenti un numero limitato di geni codificati 1 (variabili selezionate). Quindi, in ogni cromosoma costituito da v geni, esistono mediamente x_c geni codificati 1 e $v-x_c$ geni codificati 0. Il valore di default di x_c nel software è 5.

In questo modo si otterranno cromosomi con piccoli blocchi di variabili molto diversi tra un cromosoma e l'altro, permettendo tempi di elaborazione molto ridotti e la possibilità di rilevare facilmente quali siano le variabili rilevanti per la predizione. Le variabili non buone saranno facilmente eliminate e le variabili informative, divise in piccoli blocchi, inizieranno a crescere di numero con il progredire dell'elaborazione.

Una volta creato un numero prefissato di cromosomi, ne viene calcolata la risposta (la percentuale di varianza spiegata) tramite regressione PLS in cross-validazione. La qualità di un sottoinsieme di variabili è valutata sia in base alla risposta che in base al numero di variabili considerate. È quindi importante sapere qual è il cromosoma con la risposta migliore, ottenuto da un dato numero di variabili. I cromosomi con queste caratteristiche sono altamente informativi, perciò vengono bloccati ed entrano a far parte della popolazione dalla quale non possono essere eliminati finché un altro cromosoma, che utilizza al massimo lo stesso numero di variabili, non produce una risposta migliore.

Una volta creata la popolazione, si passa alla fase di riproduzione e mutazione. La fase di riproduzione utilizza il metodo "crossover uniforme" descritto nel paragrafo 9.3.2, in modo da poter evitare qualsiasi influenza dovuta all'ordine delle variabili, basando la scelta dei

cromosomi genitori in base al metodo di probabilità descritto nello stesso paragrafo e visualizzato in Figura 9.1.

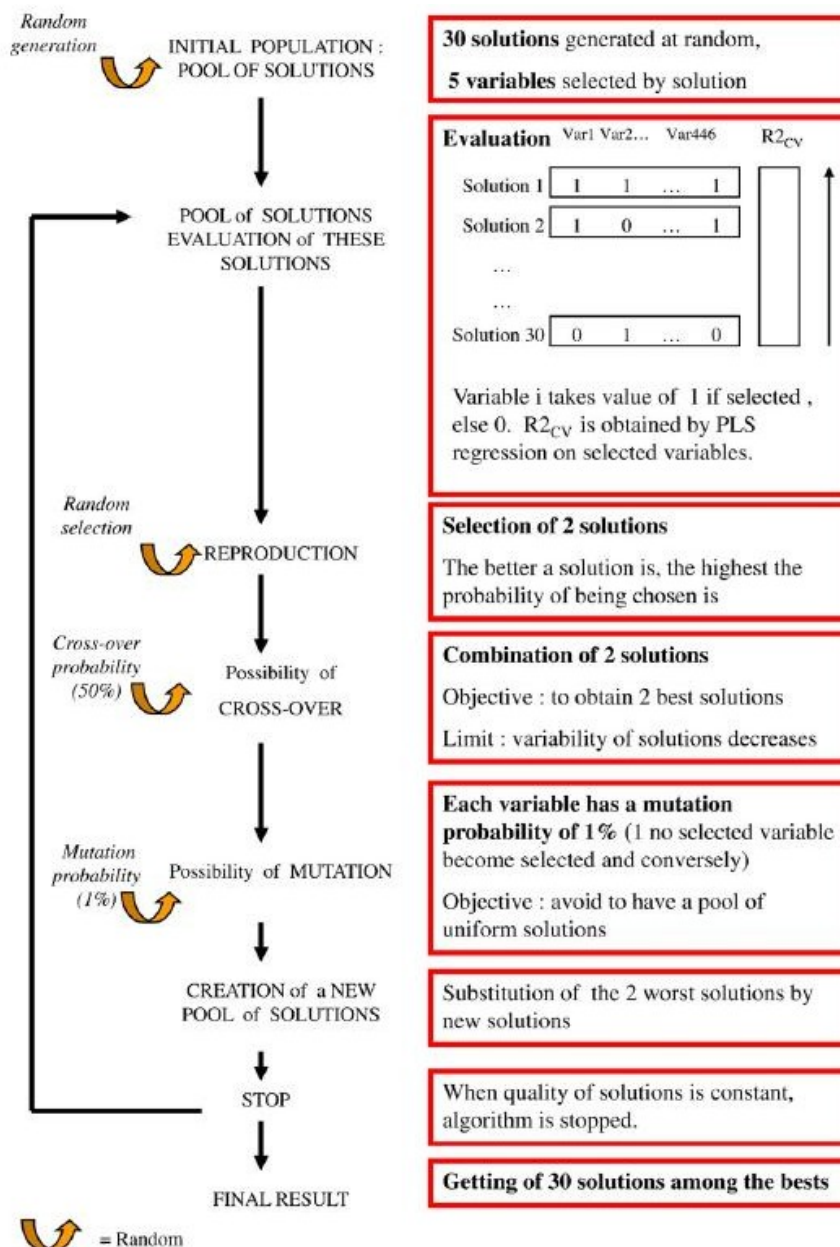


Figura 9.3 – Schema di una corsa dell’algoritmo genetico [Ferrand, 2011].

Una volta generati i cromosomi figli, l’algoritmo produce in loro una mutazione. Il valore di probabilità di default nel software è 0,01 per gene. Le risposte dei cromosomi figli vengono valutate e, se risultano essere migliori di quelli di altri cromosomi presenti nella popolazione, i loro cromosomi verranno conservati a discapito dei cromosomi della

popolazione con le risposte peggiori. È anche possibile che un cromosoma figlio produca una risposta migliore di un cromosoma bloccato, utilizzando al massimo lo stesso numero di variabili. In questo caso il cromosoma figlio viene bloccato e l'altro cromosoma viene invece sbloccato.

Esiste la probabilità che un gruppo di variabili considerate in un cromosoma sia un sottoinsieme di un gruppo di variabili considerate da un altro cromosoma. Se la risposta del primo è migliore della risposta del secondo, allora le variabili in più contenute nel secondo cromosoma non sono informative ma rappresentano semplicemente rumore, perciò questo cromosoma verrà eliminato.

Per poter incrementare le performance di un algoritmo genetico, esso viene alternato con cicli di selezione di variabili ottenuti col metodo della *backward elimination* (paragrafo 8.4.2), ottenendo così un cosiddetto algoritmo ibrido. Il metodo viene applicato al miglior cromosoma, con una frequenza di un ciclo ogni 100 valutazioni ed un ciclo alla fine del processo (se il numero di valutazioni non è un multiplo di 100) e, se la risposta ottenuta dal cromosoma generato è migliore della risposta precedente, il nuovo cromosoma rimpiazzerà l'originale.

La fase evolutiva procede fino a quando non viene soddisfatto un criterio di arresto che viene a decretare la fine di una corsa dell'algoritmo. In Figura 9.3 è mostrato lo schema di una corsa dell'algoritmo.

9.6.2 Implementazioni dell'algoritmo per ridurre il rischio di overfitting e l'autocorrelazione tra variabili

La presenza di correlazioni casuali è sicuramente il fattore principale che limita l'utilizzo degli algoritmi genetici [Jouan-Rimbaud, 1996] ed un modello costruito senza tenerne conto non è sicuramente un modello affidabile e robusto. Inoltre, osservando l'evoluzione delle risposte durante la corsa dell'algoritmo, si può vedere che il miglioramento avviene in modo repentino nelle fasi iniziali del processo ma subisce un decremento di velocità nelle fasi successive. Dopo che viene utilizzata la parte informativa dei dati per la costruzione del modello, c'è il rischio che si utilizzi anche il rumore per completare il modellamento. L'algoritmo genetico deve quindi essere fermato ad un certo punto dell'elaborazione, evitando così la presenza di overfitting.

Per questi motivi, l'algoritmo genetico per la selezione delle variabili spettrali è stato implementato con alcune operazioni da effettuare prima, durante e dopo la corsa dell'algoritmo.

Test di randomizzazione

Attraverso l'utilizzo di un test di randomizzazione è possibile valutare l'efficacia dell'algoritmo genetico nel risolvere il problema in esame prima ancora di iniziare la selezione delle variabili, e decidere quante valutazioni far effettuare all'algoritmo in ogni sua corsa. Il test si basa sul seguente metodo: l'ordine degli elementi della matrice Y della variabile risposta sperimentale viene definito in modo casuale così che ogni riga della matrice X dei predittori corrisponda ad un valore di Y che in realtà non è il suo. Così facendo, il data set non ha in sé informazione; perciò, se si riesce ad ottenere un modellamento da questi dati, allora il modello è ottenuto esclusivamente da rumore. L'algoritmo genetico utilizza il test in ogni corsa, valutando 100 cromosomi in ognuna di esse, prendendo poi in considerazione la risposta migliore. Alla fine del ciclo dell'algoritmo (dove un ciclo è dato dal numero totale delle corse), viene calcolata la media delle risposte migliori di ogni corsa. Al diminuire di questo valore, aumenterà l'affidabilità del data set. Se il valore è invece molto alto, vorrà dire che l'algoritmo genetico riesce a modellare i dati anche quando non è presente informazione. Un ottimo data set presenta un valore di risposte medie minore di 4. Di norma, un algoritmo genetico può essere utilizzato con sicurezza quando la risposta media ottenuta dal test non supera il valore 8 [Leardi, 1998].

La decisione sul quando bloccare una corsa dell'algoritmo genetico, e quindi evitare l'overfitting, è presa utilizzando lo stesso test: l'algoritmo effettua un numero prefissato di corse, normalmente 100, con 200 valutazioni ognuna, nella prima metà delle quali si utilizza il vettore Y originale e nella seconda metà si utilizza il vettore Y randomizzato. Vengono calcolate le risposte di tutti i cromosomi di ogni corsa e viene presa in considerazione la risposta migliore di ogni corsa. A questo punto viene fatta la media delle migliori risposte ottenute con la matrice Y originale e la media delle migliori risposte ottenute con la matrice Y randomizzata. Il miglior momento in cui fermare l'algoritmo genetico è quello in cui si ha la massima differenza tra le due medie calcolate; si cerca quindi il numero di valutazioni per cui si ha la massimizzazione della differenza tra le due medie. Questo corrisponde a dire che le corse con il vettore Y originale mostrano l'abilità

dell'algoritmo genetico di modellare l'informazione ed il rumore, mentre le corse con il vettore Y randomizzato mostrano l'abilità dell'algoritmo genetico di modellare solo il rumore. Il vettore differenza è l'abilità di modellare solo l'informazione [Leardi, 1998].

Riduzione del numero di variabili iniziali

La procedura della selezione delle variabili è apparentemente semplice ma in realtà molto complicata in quanto esiste il forte rischio di sovrastimare l'abilità predittiva del modello a causa della possibilità di correlazioni casuali. Questo fatto risulta essere ancora più evidente quando il rapporto variabili/oggetti è molto alto. L'operazione più semplice che si può effettuare è quella di ridurre il numero di variabili utilizzando, al posto di un dato numero di lunghezze d'onda, il valore medio ottenuto su finestre di lunghezze d'onda contigue. Normalmente, le performance dell'algoritmo genetico sono ottimali quando il numero delle variabili iniziali è minore di 200 [Leardi, 2009].

Numero di corse e selezione delle variabili

I parametri dell'algoritmo genetico vengono impostati in modo tale da ottenere la massimizzazione della fase di exploitation in modo da ottenere un miglioramento molto veloce delle risposte già nei primi stadi del processo e ridurre il rischio di overfitting. Lo svantaggio è che, dato che in una singola corsa (una sola fase evolutiva) solo una minima parte del dominio totale viene esplorata, i risultati di più corse possono essere piuttosto differenti. Il risultato finale può quindi essere influenzato dalla popolazione iniziale casualmente generata. Per risolvere questo problema, l'algoritmo effettua un numero elevato di corse indipendenti, normalmente 100, per poter esplorare buona parte del dominio di ricerca.

Per non perdere l'informazione portata dalle corse precedenti, la probabilità che ogni variabile venga selezionata all'inizio di ogni corsa viene calcolata sulla base del numero di volte che quella variabile è stata precedentemente selezionata. In questo modo, ogni nuova corsa si focalizza maggiormente sulle regioni spettrali più interessanti, senza eliminare del tutto la possibilità di esplorare tutte le altre zone.

Alla fine delle 100 corse vengono considerati i 100 migliori cromosomi (il migliore di ogni corsa) sui quali si calcola la frequenza di selezione di ogni singola variabile, ottenendo un grafico frequenza di selezione/variabile. Sui valori delle frequenze si applica un'operazione di smoothing su una finestra di tre lunghezze d'onda adiacenti ed infine si

effettua una selezione stepwise per la scelta delle variabili da utilizzare. La selezione stepwise si basa sul metodo *forward selection* descritto nel paragrafo 8.4.2 e consiste nel valutare la predizione in cross-validazione dei modelli costruiti inserendo una variabile alla volta, in base al loro valore della frequenza di selezione ottenuto dopo l'operazione di smoothing. L'aggiunta progressiva delle variabili determina, finché esiste informazione, una diminuzione dell'errore in cross-validazione RMSECV fino al raggiungimento del minimo globale. Però, il minimo di RMSECV può essere ottenuto anche a seguito di oscillazioni casuali attorno al valore di plateau e quindi corrispondere a un numero molto elevato di variabili. L'algoritmo utilizza quindi il test F di Fisher per determinare qual è il modello più parsimonioso, ovvero il modello che utilizza il minore numero di variabili ottenendo un errore RMSECV non significativamente differente dal minimo globale.

I passaggi sono i seguenti:

- si localizza il valore minimo di RMSECV;
- si utilizza un test F con $p < 0,1$ e gradi di libertà pari al numero di oggetti del training set sia al numeratore che al denominatore, e si seleziona un valore soglia corrispondente al più alto valore di RMSECV che non è significativamente differente dal valore minimo;
- si cercano le soluzioni con il minore numero di variabili che hanno un valore di RMSECV minore del valore soglia.

Questo metodo permette di costruire modelli che generano un RMSEP, errore in predizione su un set esterno, di solito leggermente inferiore utilizzando un numero minore di variabili e regione spettrali più piccole.

Il procedimento che comprende le 100 ripetizioni delle corse dell'algoritmo fino alla selezione del modello più parsimonioso costituisce un ciclo dell'algoritmo [Leardi, 2009].

9.6.3 Il modello finale

Per ridurre ulteriormente la variabilità del modello, l'intero processo finora descritto (ciclo dell'algoritmo) viene ripetuto 5 volte, ognuna delle quali indipendente dalle altre. Le regioni selezionate in ogni ripetizione vengono confrontate insieme allo spettro originale e, se una regione viene selezionata nella maggior parte delle ripetizioni, allora quella regione è significativa mentre si assume che le regioni selezionate un numero limitato di volte siano state scelte per caso.

Se si ripete l'intera procedura che porta all'ottenimento del modello finale (5 cicli di 100 corse) si ottengono modelli molto simili, le cui differenze riguardano per lo più la definizione dei limiti delle regioni spettrali. Il modello finale così ottenuto presenta un'elevata ripetibilità. In questo modo è stato risolto uno dei maggiori problemi degli algoritmi genetici (ottenimento di soluzioni assai diverse tra di loro, pur partendo dallo stesso set di dati) [Leardi, 2009].

Le condizioni di settaggio migliori dell'algoritmo genetico sono le seguenti:

- Massimo numero di variabili iniziali: 200;
- Grandezza della popolazione: 30 cromosomi;
- Numero di variabili selezionate per ogni cromosoma della popolazione iniziale: 5;
- Metodo di regressione: PLS;
- Responso: Varianza percentuale espressa in cross-validazione (5 gruppi di cancellazione e numero di variabili latenti determinate tramite cross-validazione);
- Massimo numero di variabili selezionate nello stesso cromosoma: 30;
- Probabilità di mutazione: 1%;
- Massimo numero di variabili latenti considerate: 15;
- Numero di corse: 100;
- *Backward elimination* ogni 100 valutazioni ed alla fine del processo (se il numero di valutazioni non è un multiplo di 100);
- Grandezza della finestra per lo smoothing: 3;
- Pretrattamento: nessuno;
- Scaling: autoscaling.

10 La spettroscopia infrarossa nell'analisi alimentare

10.1 Introduzione

Nell'industria di trasformazione alimentare, le fasi di produzione dell'alimento sono precedute dall'analisi sulle materie prime e sono seguite dalle valutazioni analitiche sull'alimento trasformato. In molti casi è tuttavia utile monitorare il processo durante il suo svolgimento o avere risultati sulla vita commerciale del prodotto (shelf life). Spesso è inoltre necessario che l'analisi sia effettuata in tempi molto rapidi e con costi ridotti, e ciò al fine di un controllo di qualità sia sulle materie prime che sui prodotti. In assenza di strumentazione adatta allo scopo, il procedimento per ottenere questi obiettivi consiste nel prelievo di un campione dal ciclo di produzione che, una volta giunto in laboratorio, deve subire un pretrattamento chimico o fisico per essere successivamente destinato all'analisi strumentale. Il processo complessivo è distruttivo, dispendioso in termini di tempo ed in genere scarsamente economico. Una possibile alternativa è data dall'impiego di strumentazione analitica on-line specifica (es: conduttimetri, spettrofotometri, mVmetri elettronici) che permettono di misurare velocemente, accuratamente ed economicamente uno specifico parametro di interesse tecnologico (es: pH, conducibilità, colore). Tuttavia, vi sono delle circostanze in cui sarebbe desiderabile acquisire – sempre in maniera veloce, economica e possibilmente non distruttiva – una molteplicità di parametri analitici tra loro non correlabili e per cui la misurazione con sensore dedicato – ove fosse possibile – diviene complessa e/o antieconomica. Tale tipo di problematiche può esser affrontato – e nei casi più fortunati risolto – adottando tecniche analitiche aspecifiche quali ad esempio spettroscopie infrarossa o Raman. La trattazione chemiometrica degli spettri acquisiti con tali tecniche permette – almeno in linea di principio – di ottenere risultati a basso costo, in maniera non distruttiva e pressoché istantanea, garantendo inoltre una totale integrazione della misurazione al ciclo produttivo.

10.2 La spettroscopia infrarossa

La radiazione infrarossa copre una regione molto ampia dello spettro elettromagnetico, compresa tra la regione del visibile e la regione delle microonde. La regione spettrale dell'infrarosso si estende quindi tra 700 nm e 10^6 nm e viene suddivisa in tre intervalli:

- il vicino infrarosso (Near InfraRed, NIR), il cui intervallo si estende dai 700 ai 2500 nm, ovvero da circa 14300 e 4000 cm^{-1} ;

- il medio infrarosso (Medium InfraRed, MIR), il cui intervallo si estende dai 2500 nm ai 50000 nm, ovvero tra 4000 e 200 cm^{-1} ;
- il lontano infrarosso (Far InfraRed, FIR), il cui intervallo si estende dai 50000 nm e 10^6 nm, ovvero tra 200 e 10 cm^{-1} .

Quando una molecola viene irradiata con una radiazione infrarossa, essa assorbe energia nei suoi livelli vibrazionali e rotazionali. Una molecola gassosa può subire transizioni vibro-rotazionali mentre una molecola allo stato liquido o solido subisce solo transizioni vibrazionali in quanto, nei solidi, non avvengono rotazioni, mentre nei liquidi le molecole si urtano prima di aver compiuto una rotazione completa. L'assorbimento di una particolare lunghezza d'onda è relazionato alla specifica energia di eccitazione, quindi l'assorbimento della radiazione IR avviene quando quest'ultima possiede frequenza uguale alla frequenza di una specifica vibrazione. La lunghezza d'onda alla quale avviene una vibrazione è associata ad un particolare tipo di legame chimico e ad un particolare modo di vibrazione, che può essere di stiramento (stretching) o di deformazione (bending) (Figura 10.1). Lo stretching può avvenire in fase (stretching simmetrico) o fuori fase (stretching asimmetrico) (Figura 10.2). L'energia di vibrazione è anche relazionata alla forza del legame ed alla massa del sistema molecolare [Reh, 2001].

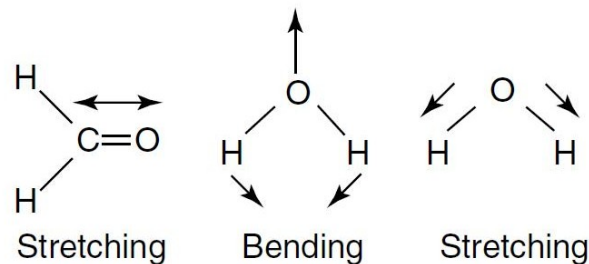


Figura 10.1 – Modi di vibrazione di stretching e bending [Stuart, 2004].

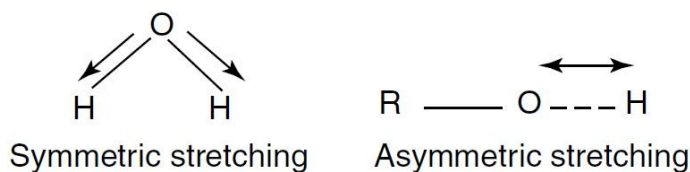


Figura 10.2 – Vibrazioni di stretching simmetriche ed asimmetriche [Stuart, 2004].

Perché una vibrazione dia un picco di assorbimento infrarosso, è necessario che la molecola subisca un cambiamento del momento di dipolo. Maggiore è la variazione del momento di dipolo, più intensa sarà la banda di assorbimento.

Come altri tipi di energia, anche l'energia vibrazionale è quantizzata; le molecole possono occupare livelli di energia discreti definiti da numeri interi (0, 1, 2...). A numero crescente corrisponde un livello di energia sempre più alto. Una transizione dal livello fondamentale al primo livello eccitato è definita transizione fondamentale mentre le transizioni dal livello fondamentale direttamente al secondo o al terzo livello vengono definite transizioni di primo e di secondo overtone rispettivamente (Figura 10.3).

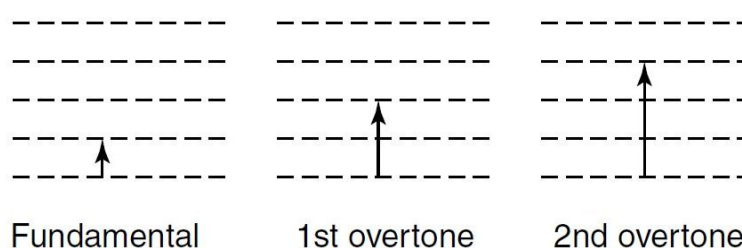


Figura 10.3 – Transizioni energetiche nell'infrarosso [Stuart, 2004].

Le frequenze fondamentali sono presenti sia nel medio che nel vicino infrarosso, ma le frequenze overtone si trovano solo nel vicino infrarosso. Le bande ottenute nello spettro NIR sono molto larghe e sono date da assorbimenti dei modi di vibrazione fondamentali e di overtone dei legami chimici C-H, N-H, e O-H e dalla transizione simultanea di due o più diversi modi vibrazionali che avvengono in un gruppo funzionale. Le bande sono quindi di difficile interpretazione. Lo spettro NIR è divisibile in tre regioni; nella prima, compresa tra 4000 e 6000 cm^{-1} circa, si collocano gli assorbimenti legati alla combinazione dei modi vibrazionali fondamentali; nella seconda, compresa tra 6000 e 7700 cm^{-1} circa, si collocano gli assorbimenti dovuti al primo overtone mentre nella terza, compresa tra 7700 e 14000 cm^{-1} circa si collocano gli assorbimenti dovuti al secondo e terzo overtone (Figura 10.4).

Lo spettro MIR presenta bande strette più facilmente interpretabili ed anch'esso è suddivisibile in tre zone principali: la zona di frequenze al di sopra dei 3000 cm^{-1} contiene bande associate ai gruppi ossidrilici ed amminici; la zona compresa tra 3000 e 1500 cm^{-1} contiene bande associate alla vibrazione di stretching delle catene aciliche, del carbonile e

degli alcheni; la zona al di sotto dei 1500 cm^{-1} è caratterizzata principalmente dalle vibrazioni di bending e da alcune di stretching delle catene aciliche (Figura 10.5).

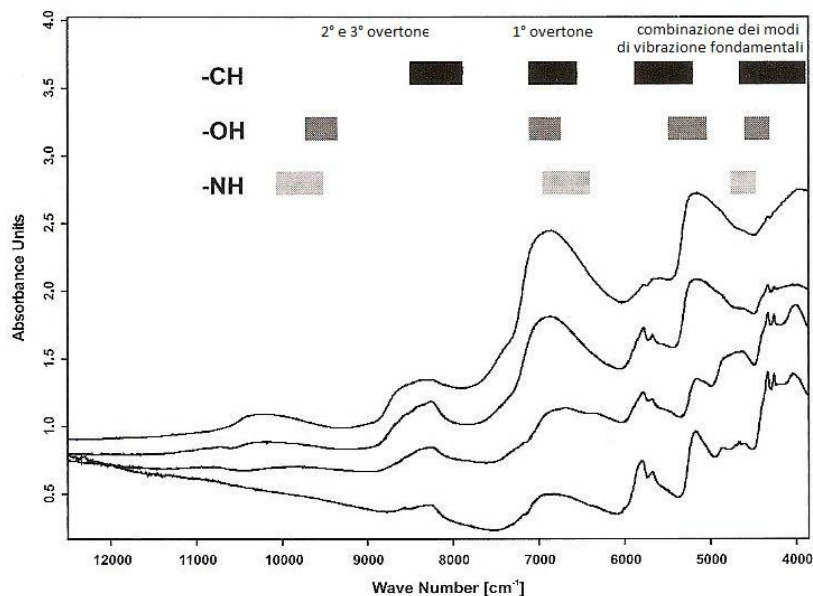


Figura 10.4 – Assorbimenti caratteristici degli spettri NIR [Niemoller, 2008].

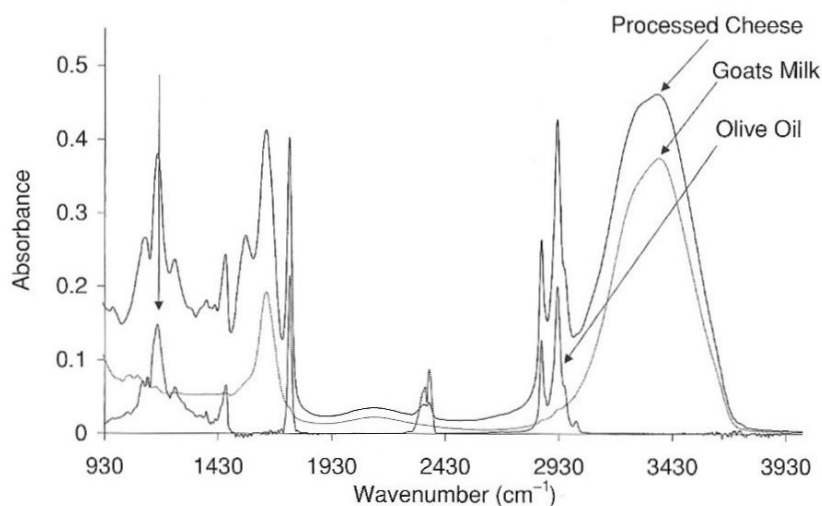


Figura 10.5 – Assorbimenti caratteristici degli spettri MIR [Fagan, 2008].

La regione infrarossa analiticamente più interessante è quindi il medio infrarosso in quanto presenta la maggiore informazione sulla struttura dei composti, con bande strette facilmente attribuibili. Il vicino infrarosso presenta bande spettrali molto larghe

difficilmente interpretabili mentre il lontano infrarosso presenta pochi picchi associabili a molecole organiche per cui non è adatto allo studio di molecole presenti in matrici alimentari [Cozzi, 2001a] [Fagan, 2008] [Abbas, 2012].

10.2.1 Spettroscopia MIR a Trasformata di Fourier

Con la spettroscopia MIR a Trasformata di Fourier è possibile registrare simultaneamente, con elevata velocità e grande sensibilità, uno spettro di una specifica matrice. La tecnica appare essere quindi applicabile nel campo del monitoraggio alimentare e dell'industria di trasformazione delle produzioni alimentari. Il componente più importante di uno spettrofotometro FT-MIR è senza dubbio l'interferometro di Michelson, costituito da un sistema di specchi che suddivide la radiazione elettromagnetica in due raggi distinti. Un raggio è indirizzato verso uno specchio fisso e l'altro verso uno specchio mobile che si muove a velocità costante. I due specchi riflettono i due raggi che si ricombinano dando luogo a fenomeni di interferenza costruttiva e distruttiva in base alla differenza del cammino ottico percorso dai due raggi (Figura 10.6). Dopo che il raggio ricombinato è passato attraverso il campione in esame, viene rilevato da un detector e viene registrato un interferogramma costituito da un'onda sinusoidale il cui periodo varia con la frequenza della radiazione che lo ha prodotto (Figura 10.7).

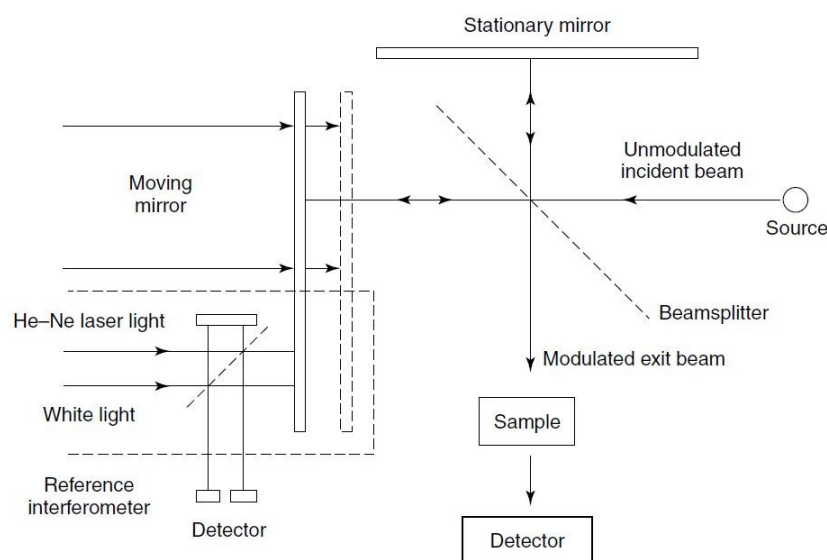


Figura 10.6 – Schema dell'interferometro di Michelson [Stuart, 2004].

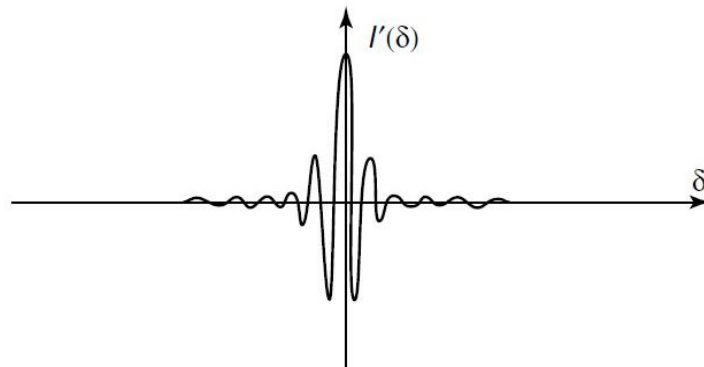


Figura 10.7 – Interferogramma ottenuto da una radiazione policromatica [Stuart, 2004].

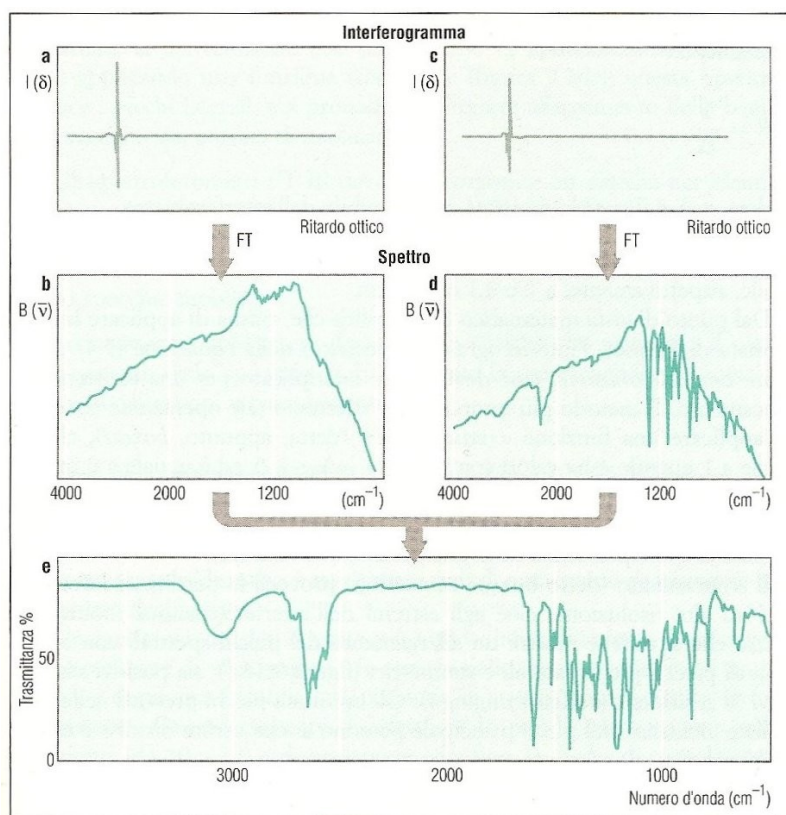


Figura 10.8 – Schema dell'acquisizione dello spettro FT-MIR [Cozzi, 2001a]

L'interferogramma descrive la variazione di intensità del segnale in funzione del cammino ottico e quindi del tempo, dato che il cammino ottico dipende dalla velocità di movimento dello specchio mobile. L'interferogramma viene poi trasformato digitalmente tramite l'algoritmo della Trasformata di Fourier:

$$B(\nu) = \int_{-\infty}^{+\infty} I(\delta) \cos(2\delta\nu) d\delta$$

dove $B(\nu)$ è l'intensità del segnale in funzione della frequenza, $I(\delta)$ è l'intensità della radiazione che giunge al rivelatore e δ è il ritardo ottico, ovvero la differenza tra i cammini ottici. Si ottiene così uno spettro dell'intensità del segnale $B(\nu)$ in funzione della frequenza. Lo spettrofotometro FT-MIR registra l'interferogramma sia della sorgente che del campione in esame, applicando ad entrambi la Trasformata di Fourier. Lo spettro finale, ottenuto dal rapporto tra lo spettro del campione e quello della sorgente, presenta picchi in trasmittanza % in funzione del numero d'onda (cm^{-1}). In questo modo si è ottenuto uno spettro registrato in modo simultaneo alle varie frequenze della regione spettrale considerata (Figura 10.8).

10.3 La spettroscopia infrarossa nell'analisi degli alimenti

Sia la spettroscopia FT-MIR che quella NIR sono state applicate ad una vasta gamma di prodotti alimentari per ottenere, tramite l'analisi multivariata, informazioni qualitative e/o quantitative dei loro componenti. Entrambe le spettroscopie presentano numerosi vantaggi quali l'alta precisione, la rapidità di misura ed i bassi costi, ma la spettroscopia FT-MIR ha – rispetto alla NIR – una sensibilità maggiore, permettendo di quantificare analiti presenti in basse concentrazioni. Lo svantaggio è invece dato dal fatto che, contrariamente al vicino infrarosso, il medio infrarosso presenta anche bande di assorbimento dell'acqua che risultano essere problematiche per l'analisi alimentare. Si ovvia a questo problema valutando comparativamente lo spettro del campione in relazione a quello dell'acqua, assunto quindi come bianco. Gli spettrometri FT-MIR permettono inoltre una maggiore trasferibilità della calibrazioni da strumento a strumento rispetto agli spettrometri NIR [Kondepati, 2008].

Gli obiettivi raggiunti in campo alimentare utilizzando la spettroscopia sia FT-MIR che NIR sono molteplici: determinazione dell'origine geografica, discriminazione delle specie, individuazione di adulterazioni, controllo della qualità alimentare, quantificazione di componenti chimiche, monitoraggio di processo, etc.

Da circa due decenni le pubblicazioni scientifiche inerenti analisi di alimenti mediante metodi spettroscopici aspecifici sono in costante crescita, al pari della tipologia di matrice alimentare considerata. Le matrici solide devono essere omogeneizzate o devono subire un processo di estrazione mentre quelle liquide o semiliquide possono essere analizzate senza alcun tipo di preparazione del campione, se non il mantenimento della temperatura costante [Van de Voort, 1992] [Reid, 2006] [Luykx, 2008] [Fagan, 2008] [Abbas, 2012].

L'autenticazione dell'origine di un alimento stabilisce se un campione appartiene o meno ad un'origine predefinita (geografica, temporale o di specie animale) e può essere utilizzata per i controlli di qualità ed autenticità. Il metodo, che si basa sullo sviluppo di un metodo chemiometrico di classificazione che utilizza parte o la totalità dello spettro infrarosso dell'alimento, è stato applicato a numerose matrici. Per quanto riguarda l'utilizzo del medio infrarosso, tra i numerosi lavori sono presenti studi sul vino, discriminato sia per zona geografica che per periodo di produzione [Picque, 2005], sui formaggi, discriminandone l'area di produzione per diverse varietà [Pillonel, 2003] [Karoui, 2005], sull'olio d'oliva, discriminato sulla base dell'origine geografica [Casale, 2010] e sul miele, discriminato per origine botanica [Ruoff, 2006]. In studi sulle carni [Al-Jowder, 1997] e sul latte scremato [Pappas, 2008] è stata invece discriminata l'origine della specie animale. Esempi dell'utilizzo del vicino infrarosso sono dati da studi sull'olio d'oliva, discriminato per origine geografica [Galtier, 2007] [Casale, 2008], sul miele, discriminato per origine geografica [Woodcock, 2007], sul vino bianco, anch'esso discriminato per origine geografica [Cozzolino, 2003] e sul tè, discriminato per varietà [He, 2007].

Numerosi sono gli studi sull'individuazione di adulterazioni che utilizzano porzioni infrarosse dello spettro elettromagnetico: per quanto riguarda il MIR, per esempio, è stato possibile individuare campioni di succo di mela adulterato con sciroppi di barbabietola e di canna [Sivakesava, 2001], dimostrare la capacità di classificare i succhi di mela sulla base della percentuale di succo di mela puro presente [Gomez-Carracedo, 2004], discriminare campioni di miele da campioni di miele adulterato con soluzioni di zucchero al 14% p/p [Kelly, 2004] e rilevare adulterazioni in campioni di olio di oliva [Marigheto, 1998]. Il NIR è stato utilizzato in Cina per rilevare latte in polvere adulterato con melamina, dopo un fatto di cronaca del 2008 che ha portato al ricovero di numerosi bambini [Dong, 2009] [Yuan, 2009], per discriminare l'olio di oliva adulterato con olio di soia, quantificandone anche il contenuto [Downey, 2002] e per discriminare il miele adulterato con fruttosio e glucosio [Downey, 2003].

In diversi studi, è stato possibile abbinare la spettroscopia infrarossa al monitoraggio di processo ed allo studio della shelf life di un prodotto. Esempi con l'utilizzo del MIR sono dati da lavori riguardo la fermentazione dei vini studiata in base al contenuto di zucchero, etanolo, glicerolo e composti fenolici [Di Egidio, 2010], riguardo la shelf life dei formaggi [Cattaneo, 2005] e riguardo la classificazione dei mosti di uva bianca nel processo di produzione del vino [Roussel, 2003]. Tra gli esempi di utilizzo del NIR sono presenti studi in cui si effettua la valutazione del malto tramite quantificazione dell'umidità e dell'azoto

totale nei processi di produzione della birra [Sileoni, 2010], la classificazione dei formaggi in base alla maturazione [Frankhuizen, 1992] e la valutazione della degradazione dell'olio di frittura [Ng, 2007].

La spettroscopia infrarossa è stata poi utilizzata, attraverso la costruzione di un modello predittivo, per la quantificazione di componenti di diversi alimenti. Esempi per il MIR sono dati dalla predizione dei saccaridi presenti nel miele [Tewari, 2004], degli acidi grassi del tessuto adiposo dei maiali (previa estrazione) [Ripoche, 2001], del contenuto di isomeri *cis* e *trans* degli acidi grassi nell'olio di oliva [Van de Voort, 1995], degli acidi grassi volatili contenuti nel formaggio svizzero [Koca, 2007] e del contenuto alcolico e di altri parametri per diversi marchi di birra belga [Polshin, 2011]. Per il NIR, esempi sono dati dalla quantificazione delle proteine e del grasso totali nella carne di manzo [Isaksson, 1996] [TØgersen, 2003], del contenuto di zucchero in frutti come le pesche ed i mandarini [Kawano, 1992] [Kawano, 1993], del contenuto totale di fenoli e flavonoidi nei chicchi di riso [Zhang, 2008] e del contenuto di polifenoli, del grado alcolico ed altri parametri nel vino [Urbano Cuadrado, 2005].

10.4 La spettroscopia infrarossa nell'analisi del latte

Il latte (bovino) è uno degli alimenti in cui l'abbinamento delle tecniche chemiometriche e di spettroscopia infrarossa è stato maggiormente sfruttato. La produzione di latte coinvolge numerosi passaggi che includono l'allevamento del bestiame, la mungitura, la conservazione, il trasporto e la ricezione da parte dell'azienda che effettuerà poi la lavorazione. Un'azienda di trasformazione di media grandezza può ricevere giornalmente diverse centinaia di campioni di latte da altrettanti conferitori. Con la spettroscopia infrarossa si può effettuare una rapida analisi dei campioni di latte, il che consente all'azienda di trasformazione di effettuare controlli istantanei sulla qualità del latte ricevuto. In questo modo è possibile, oltre a rilevare possibili adulterazioni del latte, avere informazioni sulla concentrazione dei principali componenti del latte. Un tale controllo può permettere una rimodulazione del prezzo del latte in base ai parametri di qualità, incentivando gli allevatori alla produzione di un latte migliore, e garantire ai consumatori, sempre più esigenti in campo alimentare, prodotti di alta qualità.

La determinazione della composizione del latte è essenziale per un'efficiente organizzazione dell'allevamento di animali da latte. L'implementazione di un sistema IR online nel sistema di mungitura automatico può rendere possibile la misurazione del

contenuto dei componenti del latte di animali individuali. Il valore del contenuto di grasso nel latte può essere utilizzato dagli allevatori per regolare la quantità di foraggio nella dieta; il contenuto di proteine può essere utilizzato come indicatore di un adeguato apporto energetico, mentre il contenuto di lattosio può essere utilizzato come indicatore della presenza di mastiti [Tsenkova, 2000].

L'utilizzo degli strumenti FT-IR on-line agevola il lavoro dei laboratori delle Associazioni Regionali per gli Allevatori consentendo loro di essere un punto di riferimento essenziale per i caseifici, gli enti universitari e di ricerca.

10.4.1 Spettro MIR del latte

Quando viene registrato lo spettro di una miscela di componenti quale può essere il latte, l'assorbanza ad una singola lunghezza d'onda è in teoria calcolata come una funzione additiva di ogni concentrazione dei componenti, in accordo con la legge di Beer. In questo modo si assume che non esistano fenomeni di interferenza tra i diversi componenti della miscela. Una semplice calibrazione che utilizza la legge di Beer non è però attuabile in quanto non si conosce l'intera composizione della miscela. È quindi necessario utilizzare metodi di regressione multipla abbinati ai metodi chemiometrici che permettono di costruire modelli di predizione della composizione di un analita in una miscela senza conoscere le quantità degli altri componenti.

Lo spettro del latte contiene quindi assorbimenti dei gruppi funzionali di tutti i suoi componenti e non è possibile identificare e quantificare un singolo componente dallo spettro. Infatti, nonostante lo spettro MIR contenga informazioni su tutti i componenti di un alimento, i segnali di un gruppo funzionale sono fortemente influenzati dai gruppi funzionali adiacenti [Reh, 2001] e possono subire spostamenti di frequenze. Inoltre, i picchi di assorbimento di ogni gruppo funzionale presenti nello spettro sono dati dalla sovrapposizione di picchi relativi a diversi componenti del latte, e i picchi dei componenti minoritari vengono coperti dai picchi dei componenti maggioritari.

Lo spettro del latte presenta quindi bande di assorbimento di tutti i suoi componenti e l'identificazione visiva dei picchi di assorbimento dei gruppi funzionali dei lipidi e dei singoli acidi grassi è di difficile attuazione. La conoscenza delle regioni di assorbimento dei gruppi funzionali dei composti lipidici puri potrebbe quindi non essere utile, ma rappresenta comunque un punto di partenza per l'identificazione delle regioni informative dello spettro.

In Figura 10.9 è mostrata la sovrapposizione degli spettri di assorbimento MIR di tre trigliceridi puri. Si può vedere che il legame estere, indicato con (1) in figura, assorbe tra 1700 e 1800 cm^{-1} , i gruppi metilici, (2) e (4) in figura, e metilenici, (2) e (3) in figura, assorbono tra 2800 e 3000 cm^{-1} mentre i doppi legami carbonio-carbonio, (5) in figura, assorbono nella regione compresa tra 3000 e 3050 cm^{-1} . L'ampiezza dei picchi è proporzionale al numero di gruppi funzionali presenti nella molecola. Si può infatti vedere che il trilinoleato (trigliceride contenente tre catene aciliche di linoleato) presenta il picco di assorbimento (3) - corrispondente alla vibrazione dei gruppi metilenici - più ampio dei picchi degli altri due trigliceridi. Nella regione compresa tra 1400 e 1500 cm^{-1} sono presenti gli assorbimenti delle vibrazioni di bending dei legami C-H dei gruppi metilenici mentre nella regione tra 1100 e 1300 cm^{-1} sono presenti gli assorbimenti dei legami C-O del glicerolo. Nella Figura 10.9 è mostrato anche lo spettro di assorbimento dell'acqua (linea verde) che coprirebbe parte dello spettro del grasso. Per questo motivo gli spettrofotometri MIR di matrici alimentari acquose, come il latte, forniscono uno spettro deconvoluto dai segnali dell'acqua [FIL-IDF, 2010].

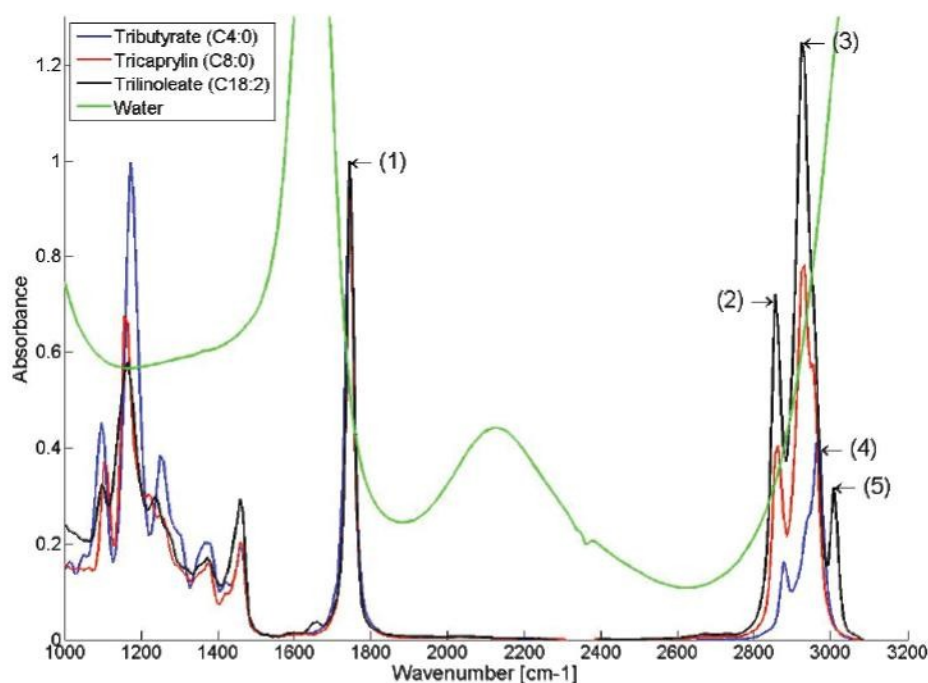


Figura 10.9 – Assorbimenti MIR di trigliceridi puri [FIL-IDF, 2010].

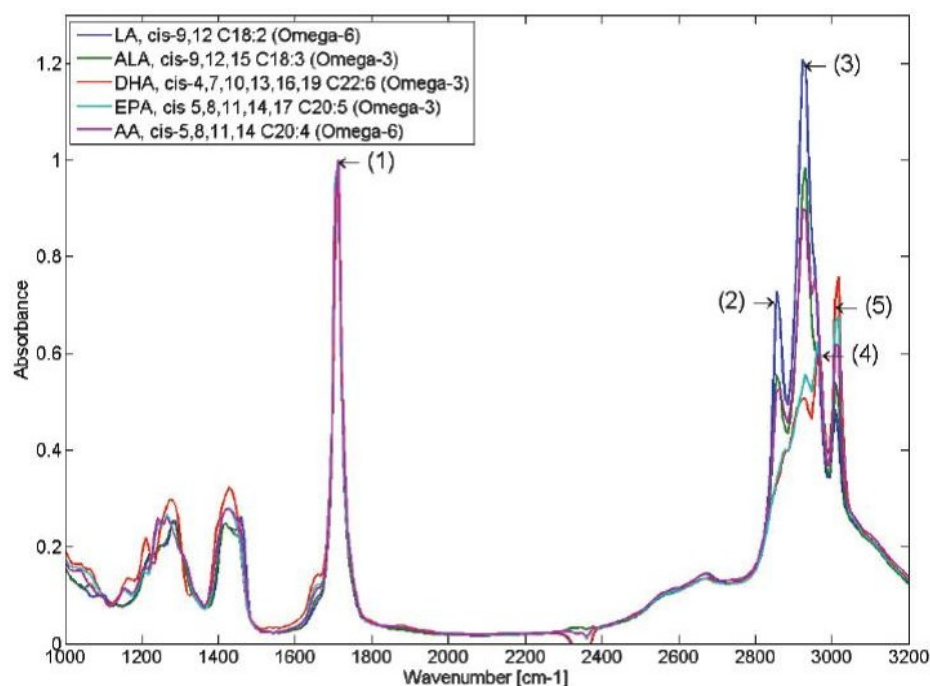


Figura 10.10 – Assorbimenti MIR di acidi insaturi puri [FIL-IDF, 2010].

In Figura 10.10 sono mostrati gli spettri di assorbimento MIR sovrapposti di alcuni acidi grassi insaturi. La differenza sostanziale rispetto agli acidi grassi saturi risiede nel picco (5), corrispondente al numero doppi legami presenti nella molecola, presente solo negli acidi polinsaturi. L'ampiezza del picco rispecchia il numero di doppi legami presenti nelle diverse molecole. Nella zona dell'impronta digitale, tra 1500 e 1000 cm^{-1} , è possibile vedere un andamento simile per gli acidi linoleico C18:2 *cis*-9, *cis*-12 e α -linolenico C18:3 *cis*-9, *cis*-12, *cis*-15 e per gli acidi arachidonico C20:4-*cis*-5, *cis*-8, *cis*-11, *cis*-14 ed eicosapentaenoico C20:5-*cis*-5, *cis*-8, *cis*-11, *cis*-14, *cis*-17, dovuto alla struttura identica in queste molecole tra il gruppo carbonilico ed il primo doppio legame [FIL-IDF, 2010].

10.4.2 Applicazioni della spettroscopia infrarossa nell'analisi del latte

Il primo tentativo di quantificare i componenti del latte tramite uno spettro infrarosso risale al 1957 [Goulden, 1957], mentre il primo impiego delle tecniche di regressione sui dati spettrali è del 1968 [Ben Gera, 1968]. Gli incoraggianti risultati pionieristici hanno determinato negli anni un continuo aumento degli studi riguardanti queste tecniche; ne sono esempi studi per la quantificazione del contenuto totale di grasso, proteine e lattosio nel latte in polvere [Baer, 1983a] [Baer, 1983b] [Frankhuizen, 1985] e per la determinazione online dell'umidità nel latte in polvere [Verbiese, 1981]. Metodi analitici

sono stati messi a punto per determinare il contenuto totale di grasso e di umidità sul latte liquido [Weaver, 1984] e ottimi risultati per la predizione del contenuto totale di grasso, proteine e lattosio sono stati ottenuti sia utilizzando spettroscopie NIR [Sato, 1987] [Carl, 1991] che MIR [Luinge, 1993] [Etzion, 2004]. Oggigiorno l'utilizzo del MIR o del NIR per la predizione del contenuto totale dei componenti maggioritari del latte bovino è ampiamente usato con ottimi risultati, quali coefficienti di correlazione tra valore predetto e misurato pari a 0,99 ed errori in predizione dello 0,1%. Sono state sviluppate recentemente anche metodiche ufficiali, quali la FIL-IDF 141C:2000, utilizzate quotidianamente nei laboratori accreditati per l'analisi del latte.

L'utilizzo della spettroscopia infrarossa come strumento di controllo nei processi industriali è stato riportato in studi sul monitoraggio della coagulazione del caglio [Laporte, 1998] [O'Callaghan, 2000] e sulle modificazioni del latte privo di lattosio durante la shelf life [Giardina, 2003], ma anche come strumento capace di discriminare campioni di latte in polvere adulterati con proteine vegetali in concentrazioni fino al 5% [Maraboli, 2002] e di classificare campioni di latte commerciale [Inon, 2004].

E' opportuno rimarcare che la messe di studi sinora ricordati è interamente relativa al latte bovino, mentre quasi assenti appaiono essere i contributi scientifici volti a definire analoghe proprietà su latte di diversa origine animale. E' possibile a riguardo ricordare gli studi di classificazione che hanno evidenziato la capacità delle tecniche chemiometriche applicate a dati spettrali di discriminare il latte di pecora da quello di capra per ottenere modelli applicabili al controllo sull'adulterazione del latte [Pappas, 2008].

Per quanto riguarda la predizione dei composti minoritari del latte, esistono alcuni lavori in cui viene determinato il contenuto di sottogruppi di proteine, quali le caseine, nel latte bovino [Hewavitharana, 1997] [Luginbühl, 2002], ma la maggior parte dei lavori, seppur presenti in numero ridotto, riguarda la quantificazione degli acidi grassi, unicamente nel latte bovino.

Non esistono invece modelli per la predizione di composti maggioritari e minoritari costruiti sul latte ovino e non risultano al momento disponibili validazioni su campioni di latte ovino di modelli costruiti su latte di altre specie.

L'applicazione di un modello calibrato su campioni di una matrice diversa non è infatti affidabile per la predizione del profilo acidico del latte ovino, a causa della diversa composizione di quest'ultimo e delle diverse interazioni tra i componenti.

Letteratura sulla predizione del profilo acidico del latte bovino

La predizione del profilo acidico del latte tramite spettroscopia FT-IR è stata effettuata, a livello di studi scientifici pubblicati, esclusivamente sul latte bovino [Soyeurt, 2006] [Rutten, 2009] [Soyeurt, 2011] [De Marchi, 2011] [Ferrand, 2011]. In tutti gli studi lo spettro MIR viene correlato ai valori di riferimento degli acidi grassi, determinati tramite estrazione chimica, derivatizzazione e caratterizzazione per via gascromatografica. In tre degli studi citati ([Soyeurt, 2006] [Soyeurt, 2011] [Ferrand, 2011]), il valore di grasso totale contenuto nei campioni di latte bovino è acquisito tramite metodica FIL-IDF 141C:2000, basata sulla predizione tramite spettroscopia medio-infrarossa a Trasformata di Fourier.

Riguardo la selezione delle variabili spettrali, tutti gli autori – eccetto Ferrand et al. – effettuano una selezione delle regioni spettrali in base all'identificazione dei picchi di assorbimento presenti nello spettro. Nel lavoro di Ferrand et al. la selezione delle regioni spettrali viene effettuata tramite algoritmi genetici.

I risultati dei diversi lavori non sono purtroppo confrontabili tra loro, in quanto vengono utilizzati diversi metodi di validazione e vengono riportati differenti parametri.

Nello specifico, nei lavori di Soyeurt et al. (2006) e De Marchi et al., viene scelto di effettuare una validazione interna col metodo della Cross Validazione (paragrafo 8.3.4) che utilizza lo stesso set di campioni sia per la calibrazione che per la validazione, metodo che presenta come limite una capacità predittiva ottimistica ed un elevato overfitting (ovvero il modellamento del rumore insieme all'informazione).

Nei restanti lavori viene invece effettuata una validazione su un set di campioni non utilizzato nella fase di calibrazione dei modelli (il cosiddetto test set). La validazione su set esterno permette di ottenere una reale valutazione dell'efficacia della predizione. I differenti autori utilizzano però diversi parametri di valutazione. Rutten et al. riportano i valori del coefficiente di determinazione R^2 e del bias in predizione. Il primo parametro non permette di valutare l'efficacia delle predizioni ma possiede solo valore indicativo. Il secondo parametro, invece, descrive solo una frazione dell'errore in predizione. Soyeurt et al. (2011) riportano invece solamente il valore R^2 per la validazione su set esterno. Ferrand et al. riportano il parametro Standard Error of Validation (SEv) (errore standard in validazione), e l'errore standard percentuale sulla media. I parametri sono affidabili per valutare la predizione. Gli algoritmi genetici utilizzati hanno selezionato mediamente 120 variabili su 446 totali e gli autori hanno ottenuti errori percentuali sulla media compresi tra

il 2 ed il 10% per gli acidi grassi saturi, tra il 5% ed il 15% per gli acidi grassi insaturi e tra l'1 ed il 10% per le classi principali di acidi grassi.

11 Parte sperimentale I: acquisizione dei dati di riferimento

11.1 Obiettivo della Tesi

Come riportato nell'introduzione, l'obiettivo del presente lavoro di Tesi è la costruzione di un modello di predizione degli acidi grassi e delle classi di acidi grassi nel latte ovino di razza Sarda tramite tecniche spettroscopiche a Trasformata di Fourier e con l'ausilio di tecniche chemiometriche.

Il modello costituirà un metodo veloce ed economico per la determinazione del profilo acidico del latte ovino che potrà essere utilizzato sia per scopi di ricerca, quali studi sull'alimentazione degli animali o sul miglioramento genetico, che per le analisi di routine previste nei controlli di qualità del latte ovino.

A tal scopo è stato necessario raccogliere un numero significativo di campioni di latte ovino, rappresentativi dell'ampia variabilità della composizione acidica determinata da fattori stagionali, geografici (pascolo, altitudine), alimentari e zootecnici.

Ogni campione di latte utilizzato è stato diviso in due aliquote:

- la prima è stata analizzata tramite Spettroscopia nel medio infrarosso a Trasformata di Fourier, per registrarne lo spettro di assorbimento, presso il laboratorio accreditato dell'ARA l'Associazione Regionale Allevatori della Sardegna;
- la seconda è stata analizzata, presso il laboratorio del Dipartimento per la Ricerca nelle Produzioni Animali (DIRPA, AGRIS Sardegna), tramite metodiche analitiche validate per la caratterizzazione del profilo acidico del latte, ottenendo così la miglior stima possibile del "valore vero".

I due set di dati ottenuti per ogni campione (valori di assorbanza e valori di riferimento degli acidi grassi) sono stati utilizzati nella costruzione dei modelli di calibrazione e nelle loro validazioni; in particolare, le funzioni di calibrazione sono state ottenute tramite regressione parziale con minimi quadrati (PLS) tra le concentrazioni di riferimento degli acidi grassi per ogni campione ed i relativi spettri infrarosso.

11.2 Campionamento

In modo da poter garantire la massima variabilità dei dati, sono stati effettuati diversi campionamenti in diverse condizioni spazio-temporali. I campioni di latte sono stati raccolti in un range temporale di due anni solari, rappresentativi di tre annate di lattazione

(periodo tra la primavera del 2011 e quella del 2013) in un areale che, per tipologia di latte, aree geografiche, profilo altimetrico del sito pascolato e natura della razione alimentare è stato ritenuto esser descrittivo delle più comuni condizioni di allevamento condotte nell'Isola (Figura 11.1).

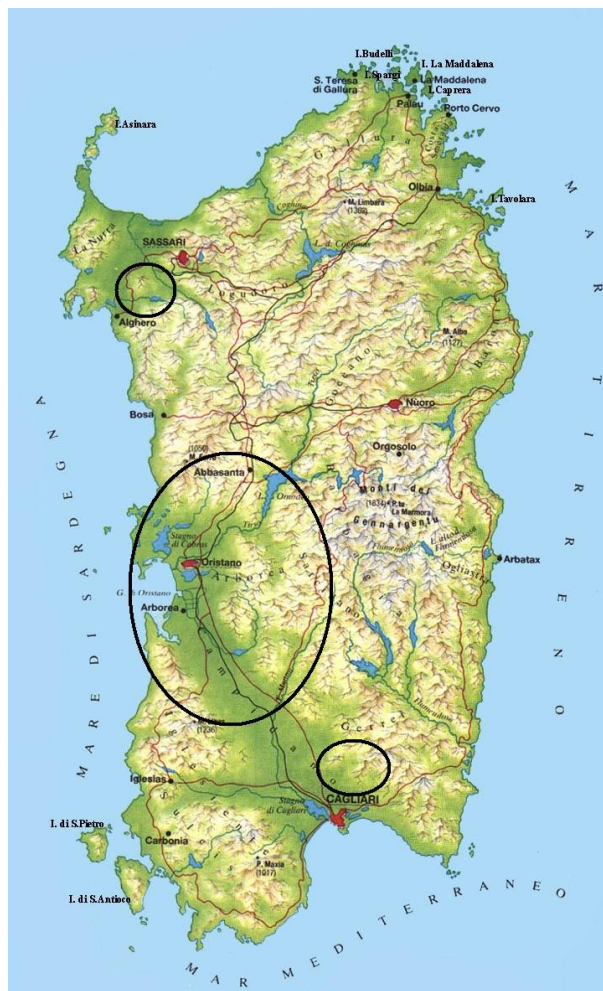


Figura 11.1 – Localizzazione degli allevamenti selezionati per il campionamento

In particolare, il campionamento è stato diviso in tre annate ed il totale dei campioni raccolti è suddivisibile in cinque sottogruppi:

- Primo gruppo (Bonassai 2011): 118 campioni individuali di latte ovino ottenuti da pecore di un gregge sperimentale del DIRPA, nel periodo compreso tra marzo e giugno del 2011;
- Secondo gruppo (Oristano 2012): 60 campioni di latte ovino di massa provenienti da altrettanti allevamenti della provincia di Oristano, del Medio Campidano e della

parte più occidentale della provincia di Nuoro nel periodo compreso tra febbraio e aprile 2012. I campioni di latte sono stati spediti presso il laboratorio del DIRPA dal Laboratorio di Analisi del Latte dell'ARA.

- Terzo gruppo (Monastir 2012): 26 campioni di latte ovino di massa raccolti in altrettanti allevamenti della zona di Monastir (Cagliari) nel periodo compreso tra maggio e giugno 2012. I campioni sono stati spediti presso il laboratorio del DIRPA dal Laboratorio di Analisi del Latte dell'ARA;
- Quarto gruppo (Autunno 2012): 20 campioni di latte ovino di massa raccolti in diversi allevamenti della provincia di Oristano nel novembre del 2012. I campioni sono stati spediti presso il laboratorio del DIRPA dal Laboratorio di Analisi del Latte dell'ARA;
- Quinto gruppo (Bonassai 2013): 26 campioni individuali di latte ovino ottenuti da pecore di un gregge sperimentale del DIRPA nel periodo compreso tra aprile e maggio del 2013. I 26 campioni sono suddivisibili in due gruppi in quanto il prelievo è avvenuto sia durante la mungitura della mattina che della sera;

Nell'arco dei tre anni di dottorato sono stati quindi raccolti un totale di 250 campioni.

11.3 Scelta del metodo di caratterizzazione del profilo acido del latte

Il metodo utilizzato per la caratterizzazione del profilo acido del latte ovino comprende una fase di estrazione dei lipidi dalla matrice latte, una fase di derivatizzazione (trans metilazione) dei lipidi estratti ed una fase di caratterizzazione dei metil esteri ottenuti tramite metodo gas-cromatografico.

Di seguito viene motivata la scelta del metodo utilizzato.

Estrazione della frazione lipidica dalla matrice latte

Nel paragrafo 7.1 si è parlato dei metodi analitici (ufficiali e sperimentali) generalmente utilizzati per l'estrazione della materia grassa dalla matrice latte finalizzata alla sua caratterizzazione. In un precedente lavoro di tesi di Laurea Specialistica in Chimica effettuato dal nostro gruppo di ricerca quattro metodi, due ufficiali (IDF 1D:1996/ISO1211:1999; IDF172:2001/ISO 14156:2001) e due sperimentali (Murphy, 1990; Hara and Radin, 1978) sono stati confrontati per valutarne le prestazioni relativamente alla capacità quantitativa di estrazione, alla quantità ed alla composizione

acidica delle varie classi lipidiche estratte. Le procedure di estrazione sono state applicate a quattro tipologie di latte ovino, differenti per quantità di grasso e composizione, in modo da verificarne la robustezza in funzione della variabilità spazio-temporale della matrice latte ovino.

I risultati non hanno individuato differenze significative tra i diversi metodi posti a confronto relativamente all'analisi quali-quantitativa degli acidi grassi provenienti dalla frazione gliceridica del latte. Sono invece state evidenziate differenze significative nella quantità di fosfolipidi estratti. Il metodo fisico [Murphy, 1990] produce un estratto lipidico con un contenuto di fosfolipidi significativamente inferiore rispetto a quello ottenuto con gli altri metodi di estrazione ($P < 0.01$). Il motivo risiede nel fatto che il metodo prevede la rottura della membrana dei globuli di grasso tramite shock termico, che avviene con un rapido passaggio dalla temperatura di -20°C alla temperatura di 60°C e successiva centrifugazione per separare la fase lipidica. I componenti della membrana, costituiti da proteine e fosfolipidi, non vengono in tal modo estratti, rimanendo così confinati nel corpo di fondo. Nonostante questo, il metodo fisico tuttavia offre il vantaggio, rispetto agli altri metodi, di essere quello più veloce, quello più economico quello più ecocompatibile in relazione alla non necessità d'impiego di solventi organici per l'estrazione.

Partendo da questi presupposti, il metodo fisico è sembrato essere quello più adatto per l'estrazione di un numero così elevato di campioni sia in termini di tempi di analisi, di consumo di reagenti e di ecocompatibilità. Il basso recupero di fosfolipidi, evidenziato nel precedente lavoro di tesi risulta esser in questo caso ininfluenza, in quanto il contenuto in fosfolipidi nel latte ovino non supera mai il punto percentuale sul totale del grasso estratto.

Per la determinazione del contenuto totale di grasso nei campioni di latte analizzati è stato utilizzato il metodo FIL-IDF 141C:2000 (vedi paragrafo 7.1), basato sulla valutazione spettrofotometrica FT-MIR dei campioni. Tale metodo è più rapido del metodo di riferimento per la determinazione del grasso totale (FIL-IDF 1D: 1996/ISO 1211:1999), usa meno campione e permette di non utilizzare il solvente per l'estrazione della materia lipidica.

Derivatizzazione della frazione lipidica estratta

I campioni di lipide ottenuti in seguito alla fase di estrazione sono stati trans metilati per ottenere gli esteri metilici. E' stato scelto un metodo di derivatizzazione in ambiente basico, FIL-IDF 182:1999/ISO 15884, che utilizza KOH come catalizzatore. Il metodo è rapido e permette l'esterificazione, in condizioni blande, dei lipidi estratti senza la

formazione di sottoprodotti indesiderati. Il fatto che il metodo non permetta la trans-metilazione degli acidi grassi liberi è stato ritenuto trascurabile in quanto solo lo 0,1% del grasso del latte è costituito da acidi grassi liberi.

Caratterizzazione analitica dei lipidi derivatizzati

I metil esteri così ottenuti sono stati separati, identificati e quantificati tramite gascromatografia con una metodica messa a punto e validata nei laboratori DIRPA.

11.4 Materiali e Metodi

11.4.1 Estrazione della frazione lipidica: metodo fisico [Murphy, 1990]

Apparecchiature

- Bagno maria.
- Provette da centrifuga in prolipopilene.
- Centrifuga termostata.

Procedimento

- Scaldare a bagno maria il campione di latte ad una temperatura di 40°C, omogeneizzare il campione di latte agitando dolcemente ed evitando la formazione di schiuma.
- Dopo raffreddamento a 20°C, campionare 30 ml di latte in provette da centrifuga da 50 ml precedentemente tarate.
- Mettere il latte in frigorifero a 4°C per 1 ora;
- Centrifugare a 3000 rpm a 2°C per 30 min;
- Mettere le provette in freezer per qualche minuto per fare solidificare la crema separata;
- Forare lo strato di crema (due piccoli fori opposti) e separare il latte scremato;
- Recuperare la crema e conservarla a - 20°C;
- Sciogliere la crema in bagno maria a 60°C per 15 min (shock termico);
- Centrifugare a 3000 rpm a 37°C per 30 min;
- Raccogliere i lipidi separati;

- Ripetere lo shock termico per altre due volte;
- Raccogliere e conservare i lipidi a -20°C fino a successiva analisi.

11.4.2 Transmetilazione basica: metodo FIL-IDF 182:1999/ISO 15884

La materia grassa estratta viene solubilizzata in n-esano (contenente gli standard interni), e poi viene dapprima trattata con una soluzione 2N di KOH in metanolo e quindi agitata al Vortex per 1 min a temperatura ambiente. La reazione viene interrotta dopo 5 minuti di riposo tramite l'aggiunta di $\text{NaHSO}_4 \cdot \text{H}_2\text{O}$. Dopo separazione delle fasi, il surnatante (fase idrocarburica) contenente gli esteri metilici degli acidi grassi, viene trasferito in un vial e quindi iniettato al gascromatografo. L'analisi viene effettuata in doppio.

Apparecchiature

- Bilancia analitica;
- Bagno maria;
- Vortex;
- Centrifuga;
- Provette in vetro con tappo a vite in teflon;
- Vials da 2 ml;
- Pipette tarate a doppia tacca da 1 e da 5 ml.

Reagenti

- n-esano;
- Metanolo anidro;
- KOH metanolica 2N;
- $\text{NaHSO}_4 \cdot \text{H}_2\text{O}$;
- Soluzione degli standard interni.

Preparazione reagente metilante

In un pallone da 100 ml pesare 11,2 g di KOH e portare a volume con MeOH.

Preparazione della soluzione degli standard interni

Si prepara una soluzione in esano contenente 1 mg/ml di ogni standard interno. Gli standard interni utilizzati sono i seguenti: il metil estere dell'acido pentanoico ($\text{MeC}_5:0$), il

metil estere dell'acido nonanoico (MeC9:0), il metil estere dell'acido tridecanoico (MeC13:0), il metil estere dell'acido nonadecanoico (MeC19:0). Analoga notazione verrà utilizzata per indicare d'ora in avanti tutti gli altri metil esteri.

Procedimento

- Pesare 50 mg di lipide estratto in una provetta in vetro pirex con tappo nero in teflon;
- Aggiungere 5 ml di soluzione esanica contenente gli standard interni, agitando fino a completa dissoluzione del lipide;
- Aggiungere 0,2 ml di reagente metilante;
- Agitare al Vortex per 1 minuto circa a temperatura ambiente;
- Lasciare riposare la soluzione di reazione per circa 5 minuti;
- Aggiungere circa 0,5 g di $\text{NaHSO}_4 \cdot \text{H}_2\text{O}$;
- Centrifugare per 3 minuti a temperatura ambiente a 900 rpm;
- Trasferire la fase idrocarburica (surnatante) contenente i metilesteri degli acidi grassi (FAME) in un vial.
- Conservare la soluzione in frigorifero a 4°C fino a successiva analisi gascromatografica.

11.4.3 Separazione e quantificazione dei FAME

Per questo lavoro di Tesi di dottorato è stato utilizzato un gascromatografo GC Varian 3900 con una colonna polare SP2560 della SUPELCO (100 m x 0,25 mm id x 0,2 μm di fase fissa) costituita da una fase stazionaria di bis(cianopropil)polisilossano. Il gas di trasporto utilizzato è l'elio, con un flusso costante di 1ml/min. L'iniezione avviene con rapporto di splitting 1:50. La colonna viene sottoposta ad una programmata di temperatura così impostata:

- 45°C per 4 min;
- da 45°C a 175°C con un incremento di 13°C /min;
- 175°C per 27 min;
- da 175°C a 215°C con un incremento di 4°C /min;
- 215°C per 35 min.

La durata totale del corsa gascromatografica è di 86 minuti.

La temperatura impostata per iniettore e FID è 290°C. Il software utilizzato per la gestione del gascromatografo e per l'integrazione dei cromatogrammi è lo Star Chromatography Workstation versione 6.0, Varian.

I metil esteri degli acidi grassi dei campioni incogniti vengono identificati per confronto dei tempi di ritenzione con quelli degli standard di riferimento.

La quantificazione dei metil esteri degli acidi grassi è stata ottenuta tramite l'utilizzo dei seguenti standard interni:

- MeC5:0: utilizzato per la quantificazione degli acidi grassi da 4 a 6 atomi di carbonio;
- MeC9:0: utilizzato per la quantificazione degli acidi grassi da 8 a 10 atomi di carbonio;
- MeC13:0: utilizzato per la quantificazione degli acidi grassi da 11 a 17 atomi di carbonio;
- MeC19:0: utilizzato per la quantificazione degli acidi grassi da 18 a 26 atomi di carbonio.

I picchi del cromatogramma vengono convertiti in un dato quantitativo (mg/g di lipide) secondo la seguente espressione:

$$C_i = \frac{A_i * C_{i(st)}}{A_{i(st)} * m_i * F_{i(st)}}$$

dove:

C_i = concentrazione dell'acido grasso incognito;

A_i = area dell'acido grasso incognito;

$C_{i(st)}$ = concentrazione dello standard interno;

m_i = peso del lipide in g;

$F_{i(st)}$ = fattore di correzione per la standardizzazione delle aree.

Dal valore di ogni metilestere (FAME) espresso in (mg/g di lipide) si ottiene il valore percentuale di ogni metil estere sulla somma dei metilesteri contenuti nel campione analizzato.

La quantificazione è stata fatta mediante funzione di calibrazione con standard interno a quattro punti.

11.4.3.1 Curva di calibrazione

Preparazione della soluzione madre di standard:

In un pallone da 25 ml si pesa una certa aliquota degli STD, esclusi gli STD interni, e si porta a volume con esano.

Gli STD utilizzati e la quantità pesata per ciascuno sono:

MeC4:0 (200mg), MeC6:0 (200mg), MeC7:0 (200mg), MeC8:0 (200mg), MeC10:0 (200mg), MeC12:0 (200mg), MeC14:0 (200mg), MeC4:1 *cis*-9 (50mg), MeC15:0 (200mg), MeC16:0 (200mg), MeC17:0 (200mg), MeC18:0 (200mg), MeC18:1 *trans*-9 (200mg), MeC18:1 *trans*-11 (200mg), MeC18:1 *cis*-9 (200mg), MeC18:2 *trans*-9 *trans*-12 (50mg), MeC18:2 *cis*-9 *cis*-12 (200mg), MeC20:0 (100mg), MeC18:3 *cis*-9 *cis*-12 *cis*-15 (100mg), MeC18:2 *cis*-9 *trans*-11 (25mg), MeC21:0 (50mg), MeC18:2 *trans*-10 *cis*-12 (25mg), MeC18:2 *cis*-9 *cis*-11 (25mg), MeC18:2 *trans*-9 *trans*-11 (25mg), MeC22:0 (100mg), MeC20:4 *cis*-5 *cis*-8 *cis*-11 *cis*-14 (50mg), MeC23:0 (50mg), MeC24:0 (25mg), MeC20:5 *cis*-5 *cis*-8 *cis*-11 *cis*-14 *cis*-17 (25mg), MeC26:0 (25mg), MeC22:5 *cis*-7 *cis*-10 *cis*-13 *cis*-16 *cis*-19 (25mg), MeC22:6 *cis*-4 *cis*-7 *cis*-10 *cis*-13 *cis*-16 *cis*-19 (25mg).

Preparazione della soluzione contenente gli standard interni

Viene preparata una soluzione in esano contenente 2 mg/ml di ciascun STD interno.

In particolare sono stati utilizzati i seguenti standard interni: MeC5:0, MeC9:0, MeC13:0, MeC19:0.

Preparazione dei 4 livelli di diluizione con cui si costruisce la curva di calibrazione

Diluizione 4

In un pallone da 10 ml si introducono 5 ml della soluzione madre e si porta a volume con la soluzione degli standard interni.

Diluizione 3

In un pallone da 20 ml si introducono 10 ml della soluzione madre e si porta a volume con esano (Soluzione A).

Dalla soluzione A si prelevano 5 ml e li si porta a volume in un pallone da 10 ml con la soluzione degli standard interni.

Diluizione 2

In un pallone da 20ml si introducono 10 ml della soluzione A e si porta a volume con esano (Soluzione B). Dalla soluzione B si prelevano 5 ml e li si porta a volume in un pallone da 10 ml con la soluzione degli standard interni.

Diluizione 1

In un pallone da 50 ml si introducono 10 ml della soluzione B e si porta a volume con esano (Soluzione C). Dalla soluzione C vengono presi 5 ml e si portano a volume in un pallone da 10 ml con la soluzione degli standard interni.

Ognuna delle quattro soluzioni a livelli crescenti di diluizione è introdotta in triplo nel gascromatografo. Si costruisce così per ogni acido grasso una curva di calibrazione a quattro punti (Tabella 11.1). La quantificazione di ciascun acido grasso nel campione incognito viene quindi effettuata mediante interpolazione con la curva di calibrazione (es. Figura 11.2) utilizzando l'equazione descritta nel paragrafo 11.4.3.

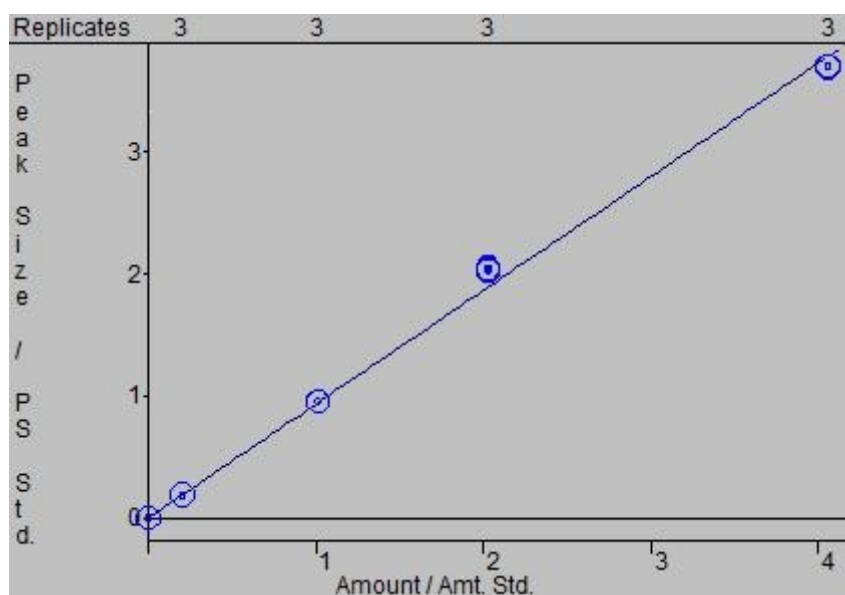


Figura 11.2 – Esempio di curva di calibrazione a quattro punti

Nella Figura 11.3 è mostrato un esempio di cromatogramma relativo alla separazione degli esteri metilici degli acidi grassi determinati in un campione di latte ovino. A partire dalla Figura 11.4 fino alla Figura 11.10 sono mostrate le espansioni del cromatogramma nelle diverse zone.

Tabella 11.1 - Concentrazione degli standard nelle quattro diluizioni

Standard	Standard interno	(g/25ml) Sostanza madre	mg/5ml Dil 1	mg/5ml Dil 2	mg/5ml Dil 3	mg/5ml Dil 4
MeC4:0	MeC5:0	0,2004	1,0020	5,0100	10,0200	20,0400
MeC6:0	MeC5:0	0,2002	1,0010	5,0050	10,0100	20,0200
MeC7:0	MeC5:0	0,1911	0,9555	4,7775	9,5550	19,1100
MeC8:0	MeC9:0	0,2030	1,0150	5,0750	10,1500	20,3000
MeC10:0	MeC9:0	0,2081	1,0405	5,2025	10,4050	20,8100
MeC12:0	MeC13:0	0,1996	0,9980	4,9900	9,9800	19,9600
MeC14:0	MeC13:0	0,2100	1,0500	5,2500	10,5000	21,0000
MeC14:1 <i>cis</i>-9	MeC13:0	0,0514	0,2570	1,2850	2,5700	5,1400
MeC15:0	MeC13:0	0,2014	1,0070	5,0350	10,0700	20,1400
MeC16:0	MeC13:0	0,2094	1,0470	5,2350	10,4700	20,9400
MeC17:0	MeC13:0	0,2010	1,0050	5,0250	10,0500	20,1000
MeC18:0	MeC19:0	0,2003	1,0015	5,0075	10,0150	20,0300
MeC18:1 <i>trans</i>-9	MeC19:0	0,2008	1,0040	5,0200	10,0400	20,0800
MeC18:1 <i>trans</i>-11	MeC19:0	0,2057	1,0285	5,1425	10,2850	20,5700
MeC18:1 <i>cis</i>-9	MeC19:0	0,2016	1,0080	5,0400	10,0800	20,1600
MeC18:2 <i>trans</i>-9, <i>trans</i>-12	MeC19:0	0,0164	0,0820	0,4100	0,8200	1,6400
MeC18:2 <i>cis</i>-9, <i>cis</i>-12	MeC19:0	0,2052	1,0260	5,1300	10,2600	20,5200
MeC20:0	MeC19:0	0,1000	0,5000	2,5000	5,0000	10,0000
MeC18:3 <i>cis</i>-9, <i>cis</i>-12, <i>cis</i>-15	MeC19:0	0,1020	0,5100	2,5500	5,1000	10,2000
MeC18:2 <i>cis</i>-9, <i>trans</i>-11	MeC19:0	0,0250	0,1250	0,6250	1,2500	2,5000
MeC21:0	MeC19:0	0,0468	0,2340	1,1700	2,3400	4,6800
MeC18:2 <i>trans</i>-10, <i>cis</i>-12	MeC19:0	0,0250	0,1250	0,6250	1,2500	2,5000
MeC18:2 <i>cis</i>-9, <i>cis</i>-11	MeC19:0	0,0250	0,1250	0,6250	1,2500	2,5000
MeC18:2 <i>trans</i>-9, <i>trans</i>-11	MeC19:0	0,0250	0,1250	0,6250	1,2500	2,5000
MeC22:0	MeC19:0	0,1012	0,5060	2,5300	5,0600	10,1200
MeC20:4 <i>cis</i>-5, <i>cis</i>-8, <i>cis</i>-11, <i>cis</i>-14	MeC19:0	0,0164	0,0820	0,4100	0,8200	1,6400
MeC23:0	MeC19:0	0,5060	0,2530	1,2650	2,5300	5,0600
MeC24:0	MeC19:0	0,3380	0,1690	0,8450	1,6900	3,3800
MeC20:5 <i>cis</i>-5, <i>cis</i>-8, <i>cis</i>-11, <i>cis</i>-14, <i>cis</i>-17	MeC19:0	0,2500	0,1250	0,6250	1,2500	2,5000
MeC26:0	MeC19:0	0,2800	0,1400	0,7000	1,4000	2,8000
MeC22:5 <i>cis</i>-7, <i>cis</i>-10, <i>cis</i>-13, <i>cis</i>-16, <i>cis</i>-19	MeC19:0	0,2500	0,1250	0,6250	1,2500	2,5000
MeC22:6 <i>cis</i>-4, <i>cis</i>-7, <i>cis</i>-10, <i>cis</i>-13, <i>cis</i>-16, <i>cis</i>-19	MeC19:0	0,2500	0,1250	0,6250	1,2500	2,5000

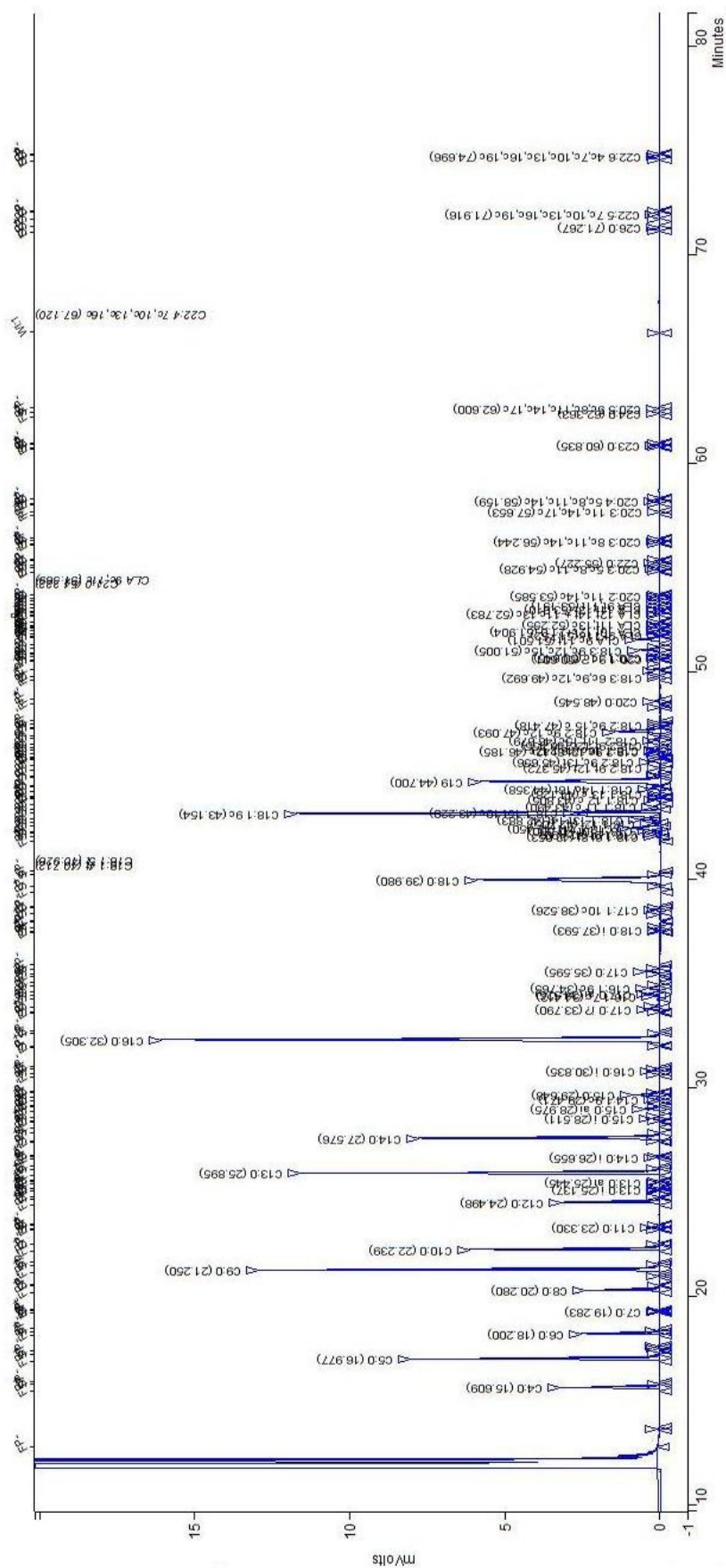


Figura 11.3 – Gascromatogramma relativo alla separazione degli esteri metilici determinati in un campione di latte.

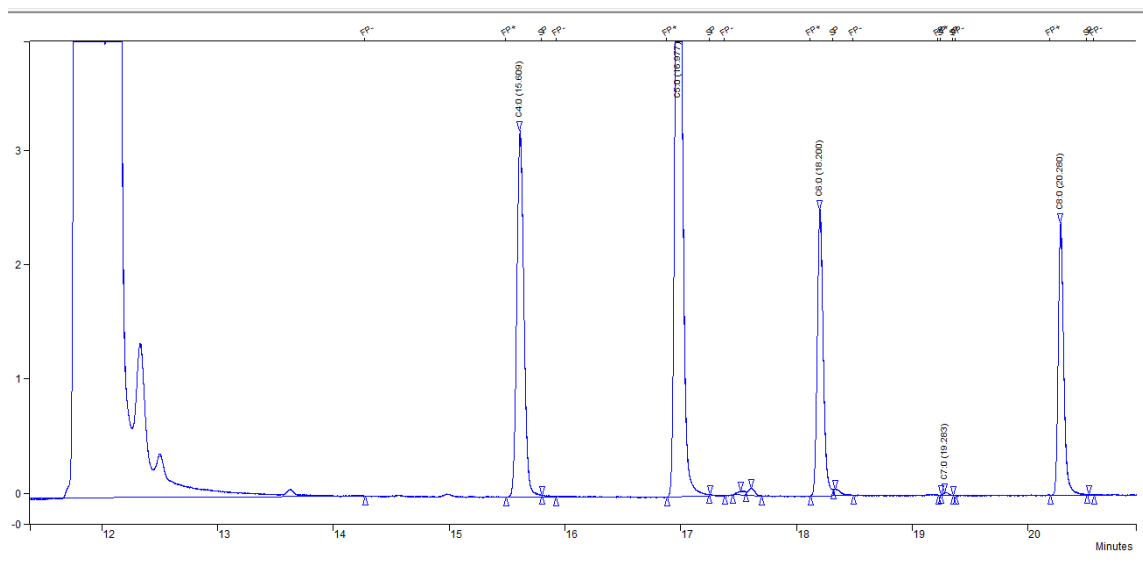


Figura 11.4 - Gascromatogramma espanso relativo alla separazione degli esteri metilici determinati in un campione di latte; porzione da MeC4:0 a MeC8:0.

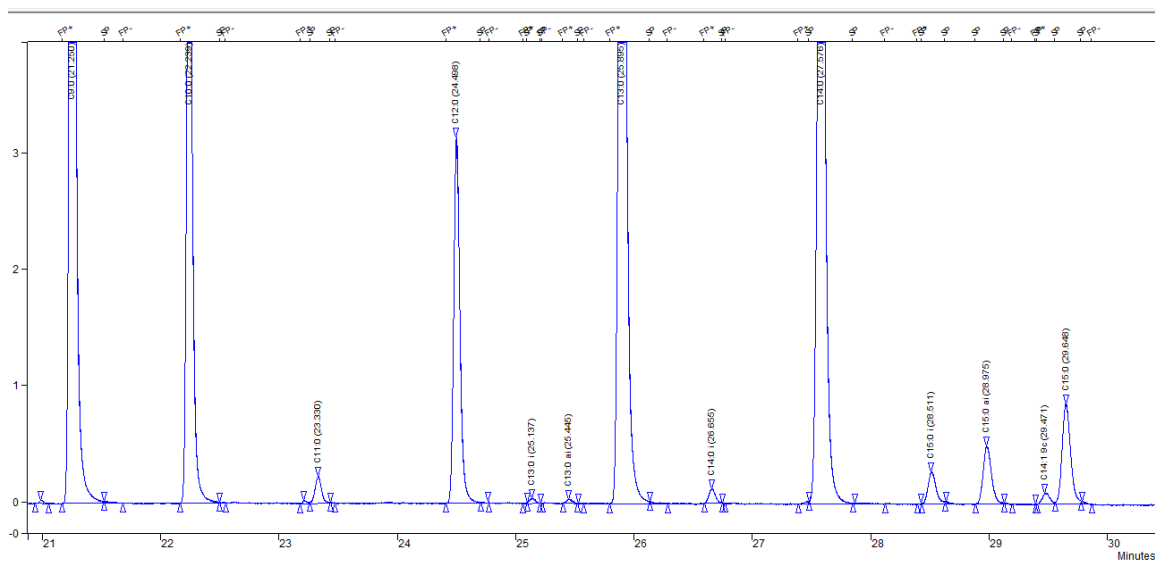


Figura 11.5 - Gascromatogramma espanso relativo alla separazione degli esteri metilici determinati in un campione di latte; porzione da MeC9:0 a MeC15:0.

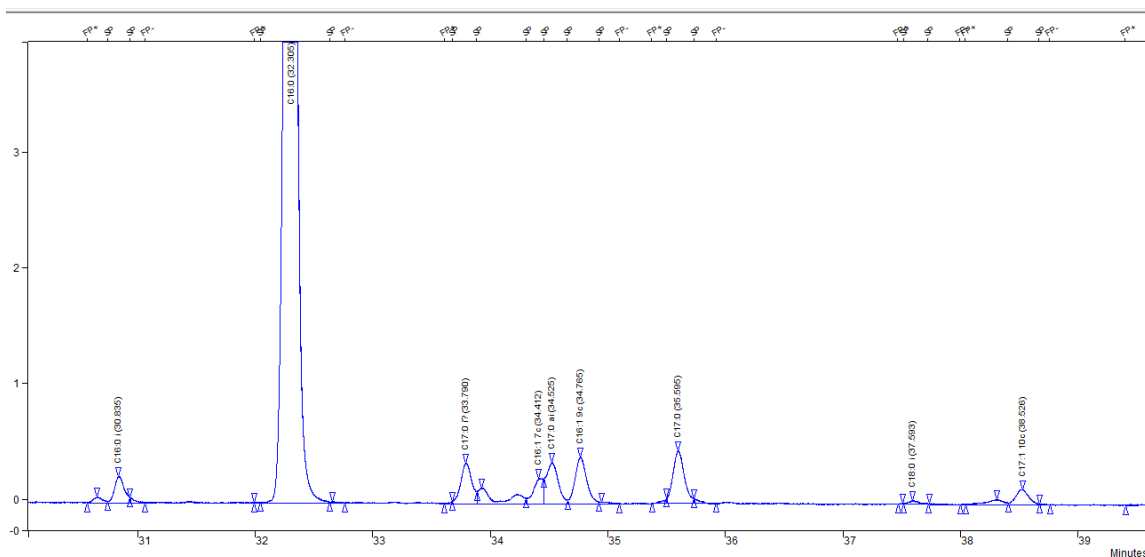


Figura 11.6 - Gascromatogramma espanso relativo alla separazione degli esteri metilici determinati in un campione di latte; porzione da MeC16:0 *iso* a MeC17:1 *cis*-10.

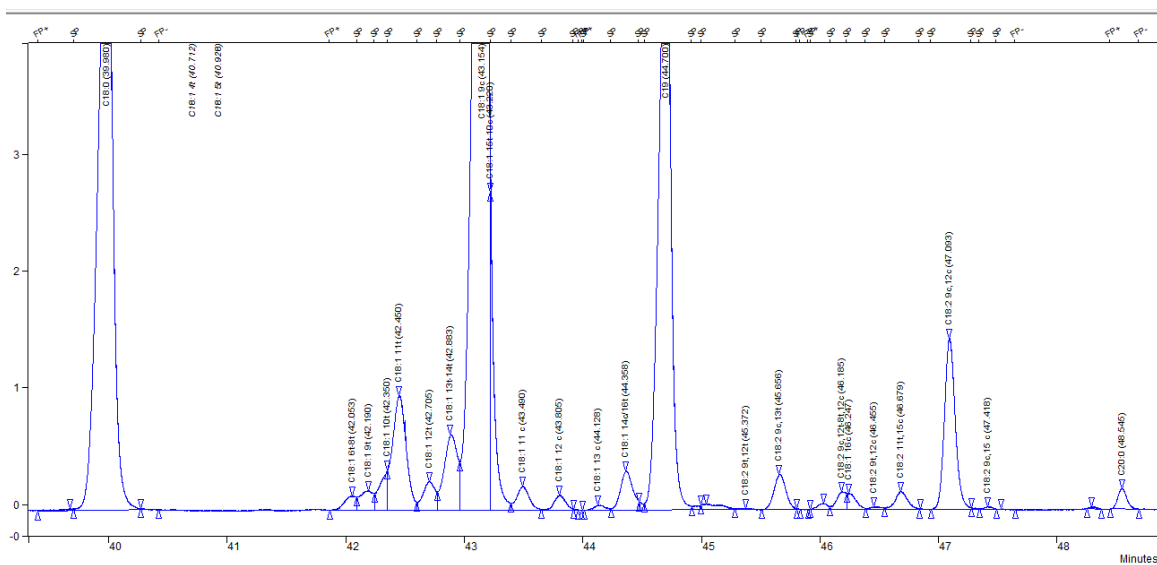


Figura 11.7 - Gascromatogramma espanso relativo alla separazione degli esteri metilici determinati in un campione di latte; porzione da MeC18:0 a MeC20:0.

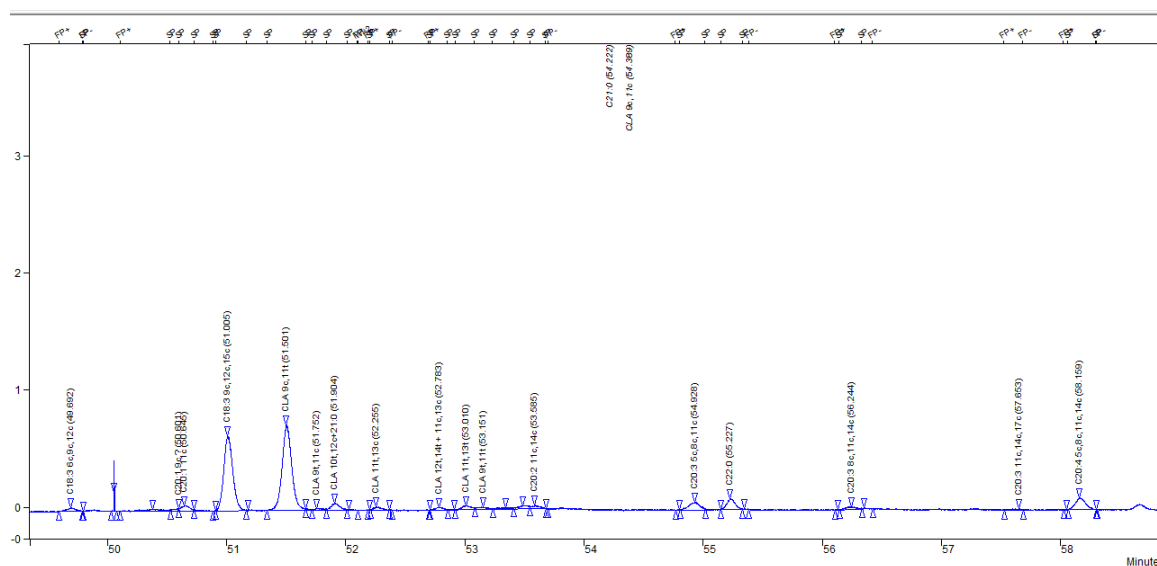


Figura 11.8 - Gascromatogramma espanso relativo alla separazione degli esteri metilici determinati in un campione di latte; porzione da MeC18:3 *cis*-6, *cis*-9, *cis*-12 a MeC20:4 *cis*-5, *cis*-8, *cis*-11, *cis*-14.

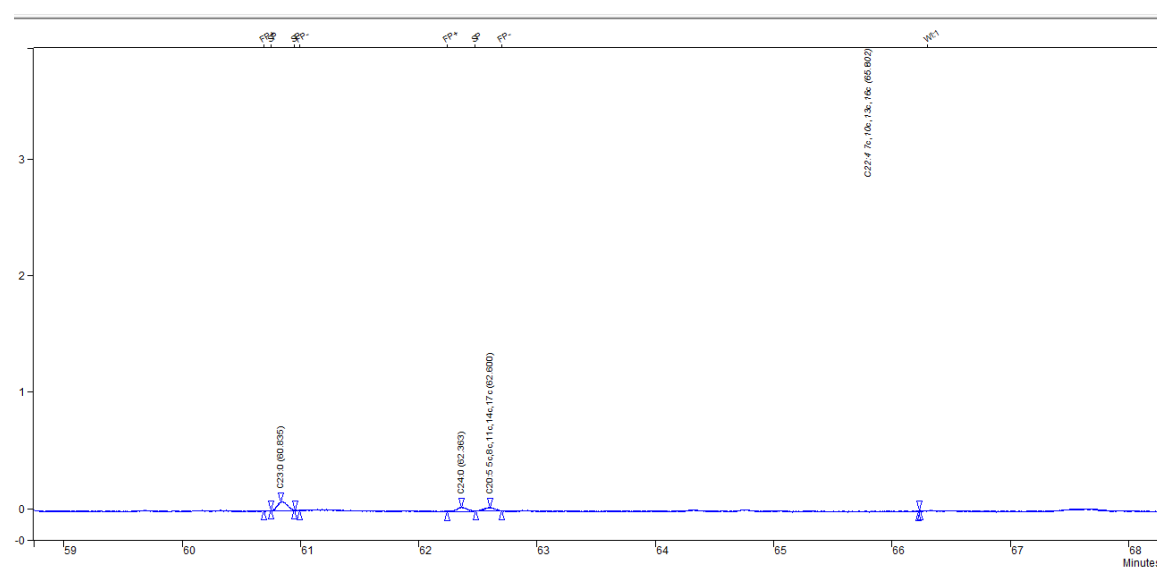


Figura 11.9 - Gascromatogramma espanso relativo alla separazione degli esteri metilici determinati in un campione di latte; porzione da MeC23:0 a MeC20:5 *cis*-5, *cis*-8, *cis*-11, *cis*-14, *cis*-17.

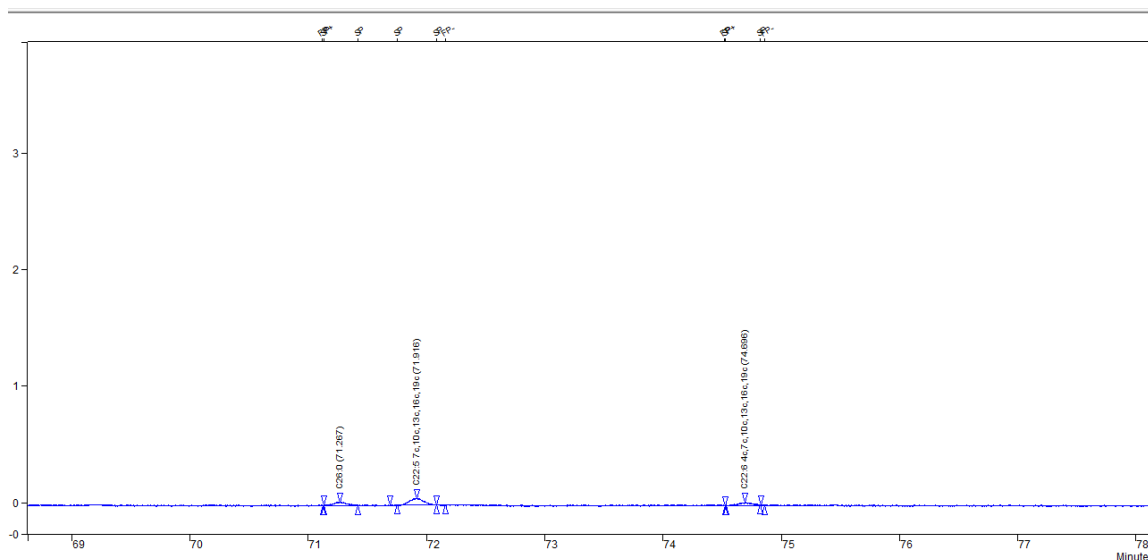


Figura 11.10 - Gascromatogramma espanso relativo alla separazione degli esteri metilici determinati in un campione di latte; porzione da MeC26:0 a MeC22:6 *cis*-4, *cis*-7, *cis*-10, *cis*-13, *cis*-16, *cis*-19.

11.4.4 Acquisizione dello spettro e della concentrazione di grasso nel latte

La determinazione del grasso totale nei campioni di latte analizzati è stata effettuata tramite metodica FIL-IDF 141C:2000 utilizzando uno spettrometro MilkoScan FT6000 (FOSS) presso il laboratorio dell'ARA di Oristano. Lo stesso strumento è stato impiegato per l'acquisizione dello spettro nel medio infrarosso dei campioni di latte ovino.

Strumentazione

Lo spettrometro MilkoScan FT6000 è uno spettrometro ad alte prestazioni nel medio infrarosso completamente automatico e rispondente alle direttive IDF ed AOAC. La tecnologia FT-IR ivi implementata consente virtualmente di determinare tutti i parametri di macrocomposizione del latte bovino. Nel 2003 il laboratorio ARA ha accreditato per la prima volta la misura dei parametri chimico-fisici e di macrocomposizione relativi al latte ovino. Lo spettrometro è dotato di un sistema di misurazione sequenziale che permette l'analisi routinaria dei campioni di latte ed un sistema di aspirazione che permette di:

- trasportare una quota rappresentativa del campione dalla provetta alla cuvetta;
- filtrare automaticamente il latte per proteggere omogeneizzatore e cuvetta e per evitare che copri sospesi causino blocchi durante l'analisi;

- omogeneizzare il campione onde evitare errori di campionamento associabili alle diverse dimensioni dei globuli di grasso;
- termostatare il campione;
- mantenere costante la pressione esercitata sulla cuvetta durante l'irradiazione infrarosso.

L'analisi può essere effettuata sia su campioni di latte a cui è stato aggiunto un conservante per evitare il deterioramento dei suoi costituenti, sia su campioni di latte fresco. Quest'ultimo deve essere mantenuto ad una temperatura di 4-5°C ed analizzato entro 72 ore dal prelievo.

Performance dello strumento MilkoScan FT6000:

- Capacità di analisi: dai 200 ai 500 campioni ogni ora;
- Volume di campione prelevato: 5 ml (programmabile tra 3,6 e 9,9 ml);
- Temperatura di lavoro: 37-42°C.

Acquisizione dello spettro

L'acquisizione dello spettro viene effettuata nel medio infrarosso, nella regione compresa tra 925,92 e 5011,54 cm^{-1} (tra 240 e 1290 pins; 1 cm^{-1} =3,858 pins). La risoluzione spettrale è pari a 3,858 cm^{-1} , pertanto l'intera scansione spettrale viene acquisita con 1060 punti.

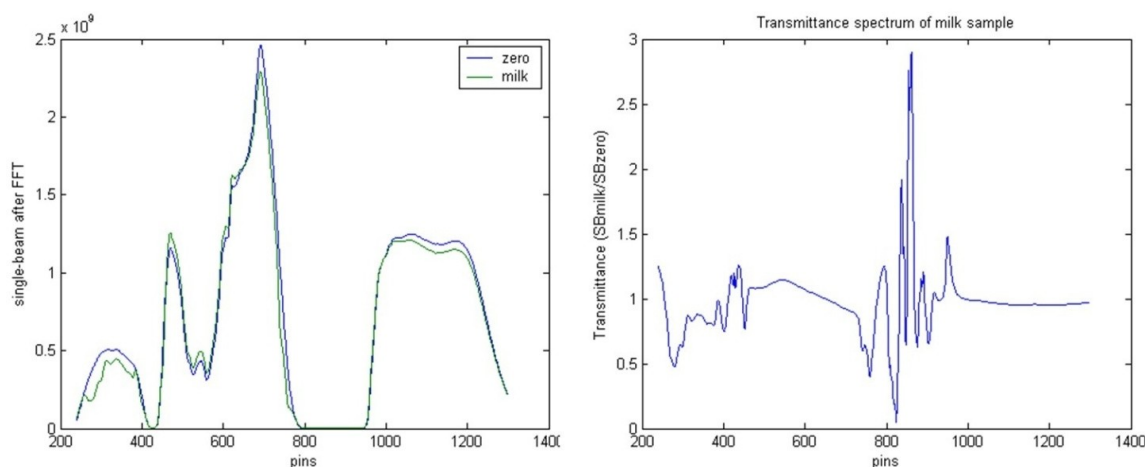


Figura 11.11 – A sinistra lo spettro IR del latte (linea verde) e dell'acqua (linea blu); a destra lo spettro in trasmittanza risultante dal filtraggio dello spettro dell'acqua da quello del latte.

Lo spettro del latte, registrato in trasmittanza, viene filtrato dai segnali IR relativi all'acqua in quanto, come si vede dalla Figura 11.11, l'acqua assorbe nel medio infrarosso e, ove venga acquisito lo spettro tal quale, i suoi segnali coprirebbero gli assorbimenti specifici della matrice latte. Infine, i valori di trasmittanza vengono convertiti in assorbanza. Lo spettro viene acquisito in doppio e i valori medi risultanti sono resi disponibili in un file importabile su altri software. In Figura 11.12 sono mostrati gli spettri FT-IR dei 250 campioni di latte ovino raccolti.

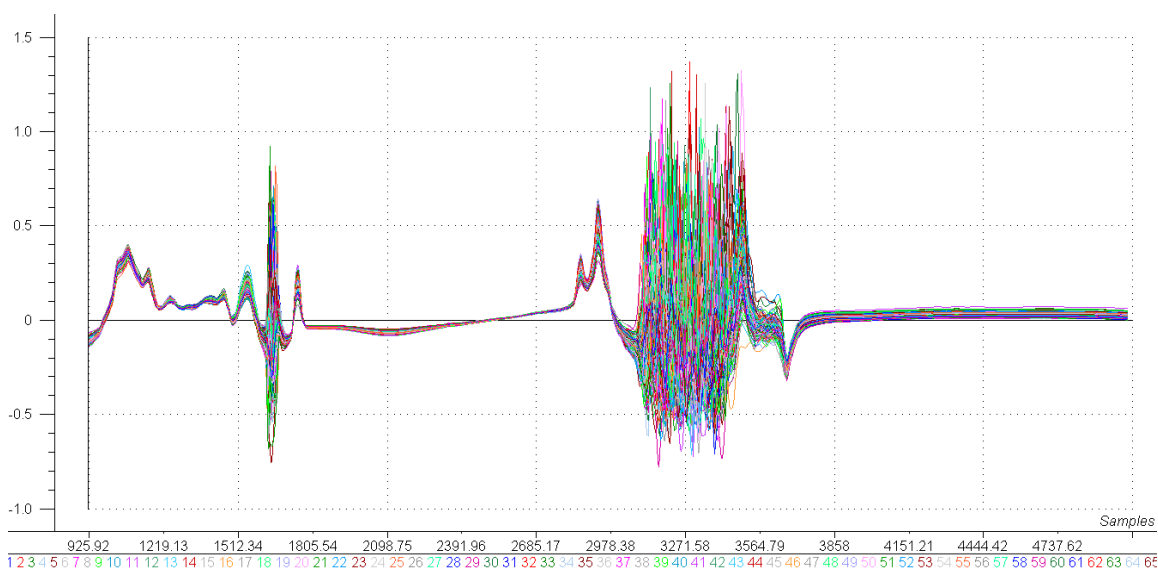


Figura 11.12 – Sovrapposizione di 250 spettri FT-IR (filtrati rispetto agli assorbimenti dell'acqua) di latte ovino.

Acquisizione del valore di grasso totale nei campioni di latte

La calibrazione utilizzata per la predizione del contenuto di grasso totale nel latte ovino tramite spettrometro MilkoScan FT6000 è basata sul metodo FIL-IDF 141C:2000.

Il metodo FIL-IDF 141C:2000 è un metodo validato sul latte bovino, applicabile a latte di diverse specie, per la determinazione del contenuto totale di grasso, di proteine, e di lattosio, basato sull'assorbimento degli specifici gruppi funzionali dei diversi componenti maggioritari del latte.

La precisione del metodo per il contenuto di grasso è la seguente: ripetibilità: < 0,04% nel 95% dei casi; riproducibilità: < 0,045% nel 95% dei casi.

Il valore di grasso determinabile varia tra il 2 ed il 15% (g grasso/100ml di latte).

Lo spettrometro Milkoscan FT6000 è implementato con calibrazioni costruite e validate nei laboratori FOSS sulla base di campioni di latte bovino ma non sono presenti calibrazioni costruite su latte ovino.

Nel Laboratorio del latte dell'ARA è stata costruita una calibrazione basata su campioni di latte ovino per la determinazione del grasso totale contenuto in esso. Il metodo è stato costruito e validato sulla base del metodo FIL-IDF 141C:2000 ed ha ottenuto l'accREDITAMENTO nel 2003.

La calibrazione è in continuo aggiornamento in quanto ogni 15 giorni vengono inseriti i dati di 5 campioni di latte ovino i cui dati di riferimento sono stati acquisiti tramite metodica IDF 1D:1996/ISO1211:1999 nei laboratori dell'Associazione Italiana Allevatori (AIA).

11.4.5 Espressione degli acidi grassi individuali in g/100g di latte

Nel sistema di pagamento in base alla qualità comunemente in uso per il latte vaccino, il quantitativo di acidi grassi e delle diverse classi di acidi grassi viene espresso non in percentuale sul totale di grasso presente nel campione, ma come peso/volume di latte, in particolare nell'unità di misura g /100g di latte. Stante ciò, si è quindi deciso di convertire i valori di riferimento dei metil esteri degli acidi grassi da valore percentuale sulla somma dei metil esteri contenuti nel campione analizzato, all'unità di misura g/100g di latte.

Tale conversione viene effettuata in accordo con quanto descritto dalla FIL-IDF in una sua recente pubblicazione [FIL-IDF, 2010]:

$$FA\left(\frac{\text{g}}{100 \text{ g latte}}\right) = \frac{0,950 \text{ g FAME}}{\text{g grasso}} \times FA\left(\frac{\text{g}}{100 \text{ g FAME}}\right) \times FAT\left(\frac{\text{g grasso}}{100 \text{ ml latte}}\right) \div \text{densità}\left(\frac{1,036 \text{ g latte}}{\text{ml latte}}\right)$$

dove con FA si intende la quantità di acido grasso, con FAME la sommatoria dei metil esteri degli acidi grassi, con FAT il contenuto totale di grasso nel latte determinato tramite spettroscopia infrarossa, mentre il coefficiente 0,95 è il contenuto medio di metil esteri su grammo di grasso (differenza attribuibile alla catena del glicerolo).

I risultati degli acidi grassi individuali sono stati utilizzati per calcolare le seguenti classi:

Acidi Grassi Saturi (SFA);

Acidi Grassi Insaturi (UFA);

Acidi Grassi Monoinsaturi (MUFA);

Acidi Grassi Polinsaturi (PUFA);

Acidi Grassi Omega 6 (PUFA ω -6);

Acidi Grassi Omega 3 (PUFA ω -3).

11.5 Risultati e discussione

L'analisi dei 250 campioni di latte ovino, effettuata in doppio, ha portato al riconoscimento dei metil esteri di 65 acidi grassi.

In Figura 11.13 è raffigurato l'istogramma rappresentativo dell'abbondanza media relativa dei 65 metilesteri quantificati nei 250 campioni di latte ovino. Si può vedere che i metil esteri presenti in quantità maggiore sono quelli degli acidi grassi C4:0, C6:0, C8:0, C10:0, C12:0, C14:0, C16:0, C18:0, C18:1 *trans*-11, C18:1 *cis*-9, C18:2 *cis*-9, *cis*-12, C18:3 *cis*-9, *cis*-12, *cis*-15 e del C18:2 *cis*-9, *trans*-11. Gli acidi grassi indicati, insieme alle relative classi di appartenenza, sono stati scelti per la costruzione dei modelli di predizione.

I dati analitici di riferimento ottenuti per il dataset in esame sono riportati nella Tabella 11.2.

Tabella 11.2 - Concentrazioni medie degli acidi grassi nei 250 campioni di latte (g/100g latte).

Metil estere (g/100g latte)	Media	CV%*	Range
C4:0	0,235	22	0,109 – 0,462
C6:0	0,166	22	0,067 – 0,317
C8:0	0,143	24	0,061 – 0,277
C10:0	0,401	27	0,154 – 0,822
C12:0	0,209	28	0,069 – 0,415
C14:0	0,521	26	0,179 – 1,045
C16:0	1,124	25	0,432 – 2,357
C18:0	0,511	33	0,137 – 1,256
C18:1 <i>trans</i>-11	0,185	42	0,026 – 0,409
C18:1 <i>cis</i>-9	0,796	33	0,232 – 2,089
C18:2 <i>cis</i>-9, <i>cis</i>-12	0,135	33	0,047 – 0,311
C18:3 <i>cis</i>-9, <i>cis</i>-12, <i>cis</i>-15	0,100	48	0,027 – 0,267
C18:2 <i>cis</i>-9, <i>trans</i>-11	0,086	42	0,016 – 0,188
Totale Saturi	3,570	23	1,411 – 7,300
Totale Insaturi	1,853	26	0,719 – 4,043
Totale Monoinsaturi	1,366	27	0,496 – 3,071
Totale Polinsaturi	0,487	30	0,175 – 0,972
Totale Acidi ω-6	0,174	30	0,064 – 0,396
Totale Acidi ω-3	0,162	41	0,040 – 0,388

* coefficiente di variazione percentuale della distribuzione

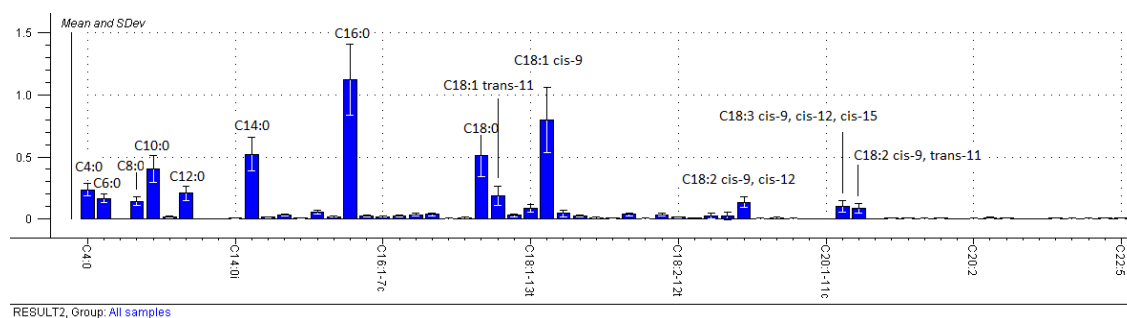


Figura 11.13 – Istogramma dei dati medi dei metilesteri dei 250 campioni di latte analizzati (medie e deviazioni standard della distribuzione)

Le Tabella 11.3 – Tabella 11.21 mostrano i dati di riferimento dei metil esteri degli acidi grassi - e delle classi di acidi grassi - ottenuti per i 250 campioni di latte (media, CV% e range, di tutte le distribuzioni), divisi a seconda del gruppo di campionamento. I dati sono raffigurati negli istogrammi a partire dalla Figura 11.14 fino alla Figura 11.32, in cui sono mostrati, per ogni gruppo di campionamento, il valore minimo, il 25° percentile, la mediana, il 75° percentile ed il valore massimo. La zona evidenziata tramite colorazione corrisponde quindi al secondo e terzo quartile, in cui sono presenti il 50% delle osservazioni. I diversi gruppi sono evidenziati nel seguente modo: gruppo blu = Bonassai 2011; gruppo rosso = Oristano 2012; gruppo verde = Monastir 2012; gruppo celeste = Autunno 2012; gruppo marrone = Bonassai 2013.

Si può vedere che il gruppo Bonassai 2011 presenta i range più ampi per tutti i metil esteri degli acidi grassi e delle classi considerate, probabilmente dovuto al fatto che i campioni di latte sono individuali e non di massa. In particolare, tale gruppo presenta elevate concentrazioni della maggior parte dei metil esteri degli acidi grassi saturi (C6:0, C8:0, C10:0, C12:0, C14:0, C16:0) e di alcuni acidi polinsaturi (C18:2 *cis*-9, *cis*-12 e C18:3 *cis*-9, *cis*-12, *cis*-15) e, di conseguenza, delle classi corrispondenti (somma dei metil esteri degli acidi grassi saturi, polinsaturi, ω -6 ed ω -3).

Anche il gruppo Oristano 2012 presenta elevati valori di diversi metil esteri di acidi grassi saturi (C4:0, C6:0, C8:0, C10:0, C12:0), ma si distingue dagli altri gruppi per le maggiori concentrazioni di C18:1 *trans*-11 e C18:2 *cis*-9, *trans*-11, che si riflette nelle corrispondenti classi (insaturi, monoinsaturi e polinsaturi).

Il gruppo Monastir 2012 è caratterizzato da basse concentrazioni medie per diversi metil esteri degli acidi grassi saturi a media e corta catena (C4:0, C6:0, C8:0, C10:0, C12:0) e

degli acidi grassi polinsaturi C18:2 *cis*-9, *cis*-12 e C18:3 *cis*-9, *cis*-12, *cis*-15, e quindi per le classi di ω -6 ed ω -3.

I gruppi Autunno 2012 e Bonassai 2013 presentano basse concentrazioni di diversi metil esteri degli acidi grassi saturi (C4:0, C6:0, C8:0, C10:0, C12:0, C14:0, C16:0) e dei metil esteri insaturi C18:1 *trans*-11, C18:2 *cis*-9, *cis*-12, C18:3 *cis*-9, *cis*-12, *cis*-15 e C18:2 *cis*-9, *trans*-11. I metil esteri degli acidi grassi maggiormente rappresentati dai due gruppi sono invece quelli del C18:0 e del C18:1 *cis*-9.

Tabella 11.3 – Valori medi del metil estere del C4:0 nei diversi sottogruppi.

C4:0	Media (g/100g latte)	CV%*	Range
Campioni Totali	0,235	22	0,109 – 0,462
Bonassai 2011	0,230	26	0,120 – 0,462
Oristano 2012	0,257	9	0,192 – 0,312
Monastir 2012	0,210	16	0,137 – 0,266
Autunno 2012	0,249	16	0,109 – 0,295
Bonassai 2013	0,222	25	0,125 – 0,320

* coefficiente di variazione percentuale della distribuzione

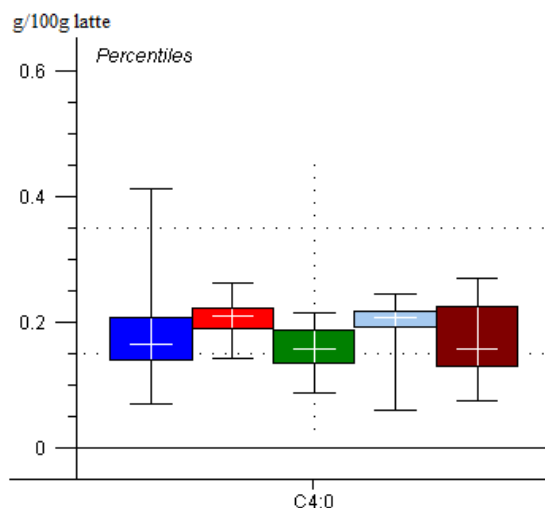


Figura 11.14 – Quartili, mediane e range per il C4:0 dei diversi sottogruppi (gruppo blu = Bonassai 2011; gruppo rosso = Oristano 2012; gruppo verde = Monastir 2012; gruppo celeste = Autunno 2012; gruppo marrone = Bonassai 2013).

Tabella 11.4 – Valori medi del metil estere del C6:0 nei diversi sottogruppi.

C6:0	Media (g/100g latte)	CV%*	Range
Campioni Totali	0,166	22	0,067 – 0,317
Bonassai 2011	0,171	23	0,097 – 0,317
Oristano 2012	0,182	9	0,142 – 0,225
Monastir 2012	0,142	18	0,105 – 0,206
Autunno 2012	0,148	19	0,067 – 0,186
Bonassai 2013	0,143	21	0,085 – 0,201

* coefficiente di variazione percentuale della distribuzione

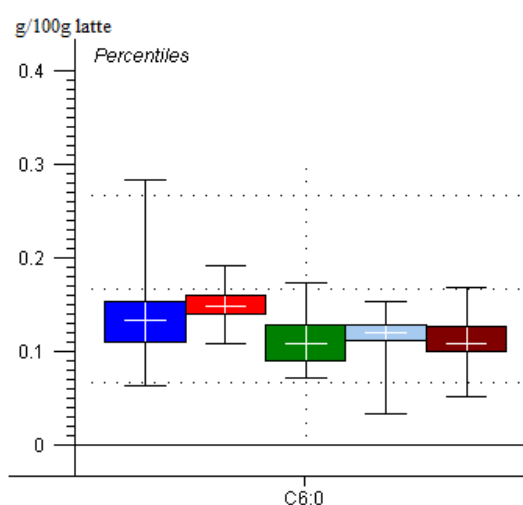


Figura 11.15 - Quartili, mediane e range per il C6:0 dei diversi sottogruppi (gruppo blu = Bonassai 2011; gruppo rosso = Oristano 2012; gruppo verde = Monastir 2012; gruppo celeste = Autunno 2012; gruppo marrone = Bonassai 2013).

Tabella 11.5 – Valori medi del metil estere del C8:0 nei diversi sottogruppi.

C8:0	Media (g/100g latte)	CV%*	Range
Campioni Totali	0,143	24	0,061 – 0,277
Bonassai 2011	0,147	24	0,088 – 0,277
Oristano 2012	0,162	14	0,105 – 0,221
Monastir 2012	0,120	23	0,081 – 0,197
Autunno 2012	0,120	22	0,061 – 0,169
Bonassai 2013	0,116	22	0,064 – 0,159

* coefficiente di variazione percentuale della distribuzione

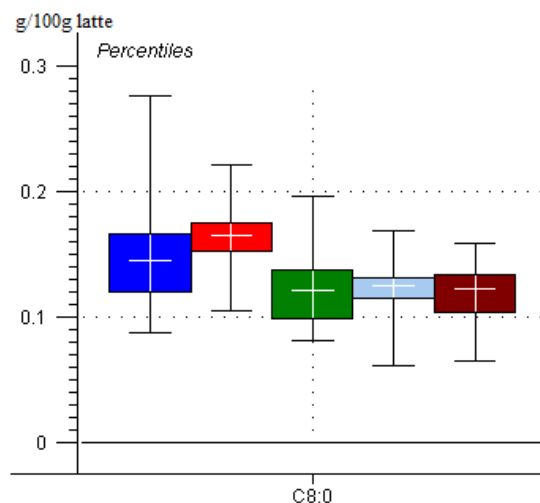


Figura 11.16 - Quartili, mediane e range per il C8:0 dei diversi sottogruppi (gruppo blu = Bonassai 2011; gruppo rosso = Oristano 2012; gruppo verde = Monastir 2012; gruppo celeste = Autunno 2012; gruppo marrone = Bonassai 2013).

Tabella 11.6 – Valori medi del metil estere del C10:0 nei diversi sottogruppi.

C10:0	Media (g/100g latte)	CV%*	Range
Campioni Totali	0,401	27	0,154 – 0,822
Bonassai 2011	0,432	26	0,234 – 0,822
Oristano 2012	0,443	17	0,255 – 0,701
Monastir 2012	0,331	24	0,218 – 0,575
Autunno 2012	0,309	24	0,154 – 0,473
Bonassai 2013	0,306	23	0,169 – 0,412

* coefficiente di variazione percentuale della distribuzione

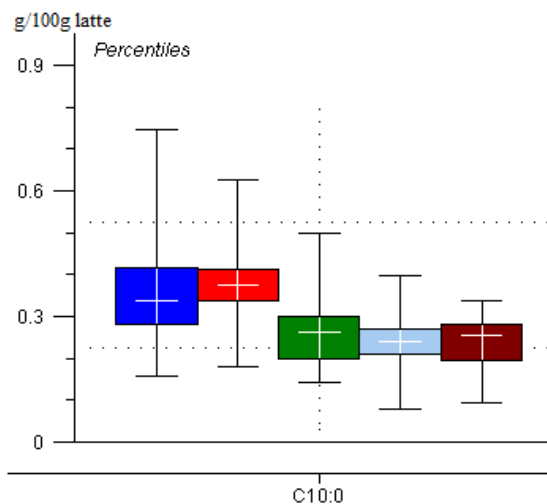


Figura 11.17 - Quartili, mediane e range per il C10:0 dei diversi sottogruppi (gruppo blu = Bonassai 2011; gruppo rosso = Oristano 2012; gruppo verde = Monastir 2012; gruppo celeste = Autunno 2012; gruppo marrone = Bonassai 2013).

Tabella 11.7 – Valori medi del metil estere del C12:0 nei diversi sottogruppi.

C12:0	Media (g/100g latte)	CV%*	Range
Campioni Totali	0,209	28	0,069 – 0,415
Bonassai 2011	0,225	28	0,109 – 0,415
Oristano 2012	0,228	17	0,149 – 0,393
Monastir 2012	0,174	22	0,118 – 0,293
Autunno 2012	0,168	22	0,069 – 0,242
Bonassai 2013	0,160	21	0,097 – 0,220

* coefficiente di variazione percentuale della distribuzione

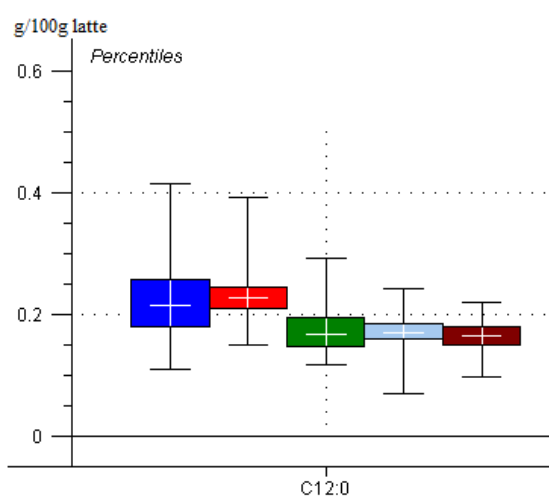


Figura 11.18 - Quartili, mediane e range per il C12:0 dei diversi sottogruppi (gruppo blu = Bonassai 2011; gruppo rosso = Oristano 2012; gruppo verde = Monastir 2012; gruppo celeste = Autunno 2012; gruppo marrone = Bonassai 2013).

Tabella 11.8 – Valori medi del metil estere del C14:0 nei diversi sottogruppi.

C14:0	Media (g/100g latte)	CV%*	Range
Campioni Totali	0,521	26	0,179 – 1,045
Bonassai 2011	0,569	29	0,236 – 1,045
Oristano 2012	0,509	11	0,320 – 0,698
Monastir 2012	0,478	17	0,355 – 0,653
Autunno 2012	0,461	18	0,179 – 0,568
Bonassai 2013	0,422	21	0,280 – 0,597

* coefficiente di variazione percentuale della distribuzione

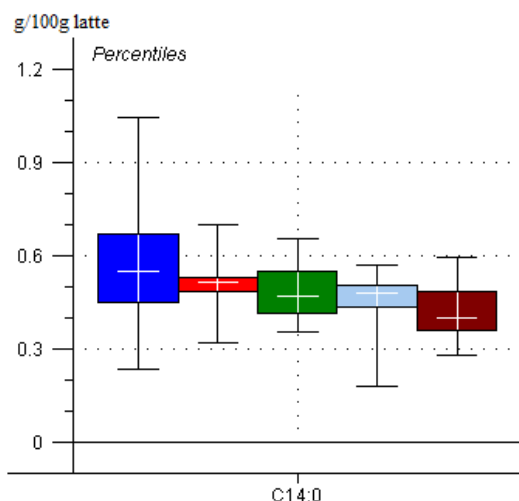


Figura 11.19 - Quartili, mediane e range per il C14:0 dei diversi sottogruppi (gruppo blu = Bonassai 2011; gruppo rosso = Oristano 2012; gruppo verde = Monastir 2012; gruppo celeste = Autunno 2012; gruppo marrone = Bonassai 2013).

Tabella 11.9 – Valori medi del metil estere del C16:0 nei diversi sottogruppi.

C16:0	Media (g/100g latte)	CV%*	Range
Campioni Totali	1,124	25	0,432 – 2,357
Bonassai 2011	1,228	28	0,487 – 2,357
Oristano 2012	1,031	11	0,782 – 1,331
Monastir 2012	1,024	17	0,745 – 1,414
Autunno 2012	1,105	18	0,432 – 1,386
Bonassai 2013	0,979	24	0,694 – 1,424

* coefficiente di variazione percentuale della distribuzione

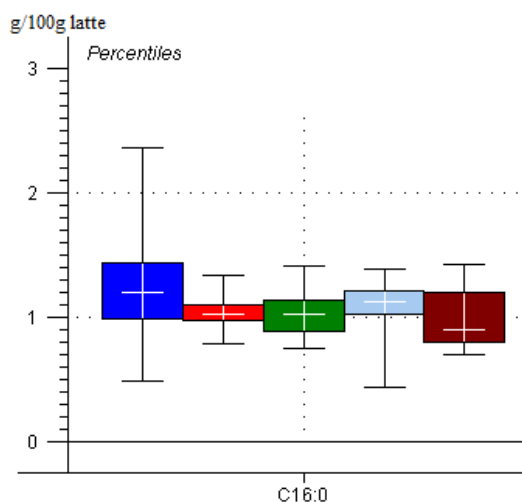


Figura 11.20 - Quartili, mediane e range per il C16:0 dei diversi sottogruppi (gruppo blu = Bonassai 2011; gruppo rosso = Oristano 2012; gruppo verde = Monastir 2012; gruppo celeste = Autunno 2012; gruppo marrone = Bonassai 2013).

Tabella 11.10 – Valori medi del metil estere del C18:0 nei diversi sottogruppi.

C18:0	Media (g/100g latte)	CV%*	Range
Campioni Totali	0,511	33	0,137 – 1,256
Bonassai 2011	0,507	39	0,210 – 1,256
Oristano 2012	0,478	28	0,137 – 0,891
Monastir 2012	0,495	22	0,249 – 0,716
Autunno 2012	0,575	19	0,240 – 0,685
Bonassai 2013	0,570	26	0,342 – 0,817

* coefficiente di variazione percentuale della distribuzione

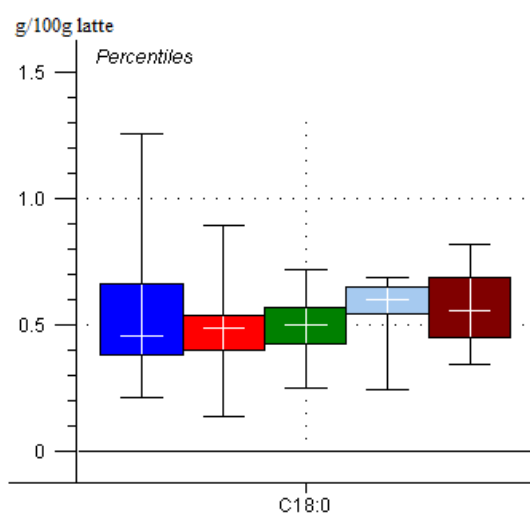


Figura 11.21 - Quartili, mediane e range per il C18:0 dei diversi sottogruppi (gruppo blu = Bonassai 2011; gruppo rosso = Oristano 2012; gruppo verde = Monastir 2012; gruppo celeste = Autunno 2012; gruppo marrone = Bonassai 2013).

Tabella 11.11 – Valori medi del metil estere del C18:1 *trans*-11 nei diversi sottogruppi.

C18:1 <i>trans</i>-11	Media (g/100g latte)	CV%*	Range
Campioni Totali	0,185	42	0,026 – 0,409
Bonassai 2011	0,175	36	0,062 – 0,409
Oristano 2012	0,279	15	0,173 – 0,399
Monastir 2012	0,134	16	0,090 – 0,178
Autunno 2012	0,122	37	0,026 – 0,232
Bonassai 2013	0,116	33	0,060 – 0,225

* coefficiente di variazione percentuale della distribuzione

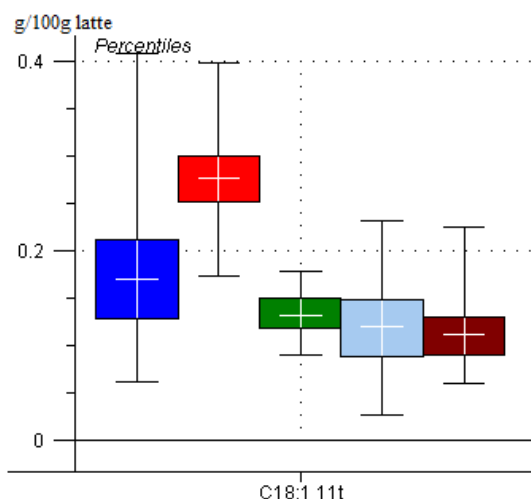


Figura 11.22 - Quartili, mediane e range per il C18:1 *trans*-11 dei diversi sottogruppi (gruppo blu = Bonassai 2011; gruppo rosso = Oristano 2012; gruppo verde = Monastir 2012; gruppo celeste = Autunno 2012; gruppo marrone = Bonassai 2013).

Tabella 11.12 – Valori medi del metil estere del C18:1 *cis*-9 nei diversi sottogruppi.

C18:1 <i>cis</i> -9	Media (g/100g latte)	CV%*	Range
Campioni Totali	0,796	33	0,232 – 2,089
Bonassai 2011	0,736	39	0,232 – 2,089
Oristano 2012	0,806	30	0,336 – 1,488
Monastir 2012	0,778	21	0,508 – 1,105
Autunno 2012	0,986	19	0,393 – 1,252
Bonassai 2013	0,919	26	0,578 – 1,280

* coefficiente di variazione percentuale della distribuzione

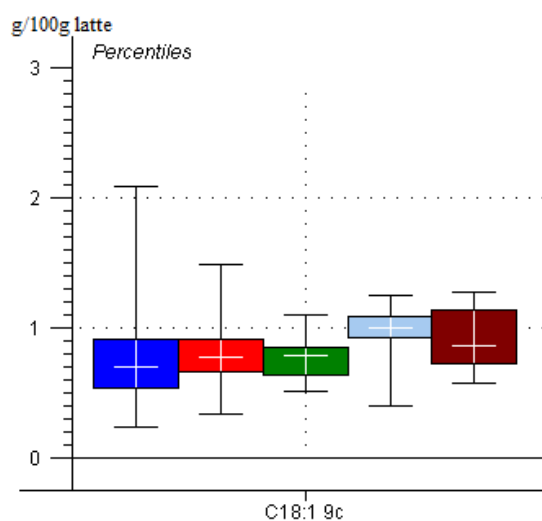


Figura 11.23 - Quartili, mediane e range per il C18:1 *cis*-9 dei diversi sottogruppi (gruppo blu = Bonassai 2011; gruppo rosso = Oristano 2012; gruppo verde = Monastir 2012; gruppo celeste = Autunno 2012; gruppo marrone = Bonassai 2013).

Tabella 11.13 – Valori medi del metil estere del C18:2 *cis*-9, *cis*-12 nei diversi sottogruppi.

C18:2 <i>cis</i> -9, <i>cis</i> -12	Media (g/100g latte)	CV%*	Range
Campioni Totali	0,135	33	0,047 – 0,311
Bonassai 2011	0,157	31	0,059 – 0,311
Oristano 2012	0,127	24	0,083 – 0,208
Monastir 2012	0,102	25	0,063 – 0,150
Autunno 2012	0,112	25	0,047 – 0,171
Bonassai 2013	0,107	28	0,072 – 0,179

* coefficiente di variazione percentuale della distribuzione

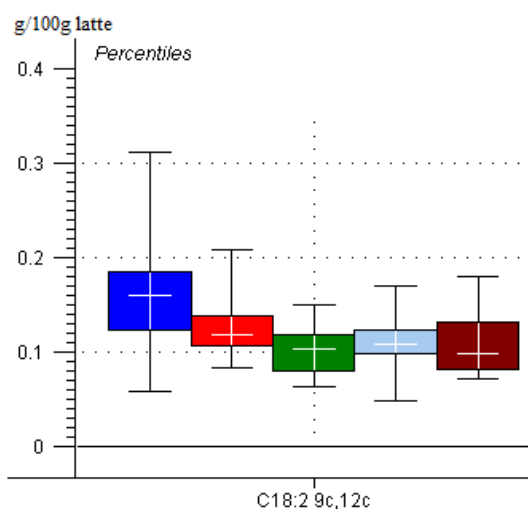


Figura 11.24 - Quartili, mediane e range per il C18:2 *cis*-9, *cis*-12 dei diversi sottogruppi (gruppo blu = Bonassai 2011; gruppo rosso = Oristano 2012; gruppo verde = Monastir 2012; gruppo celeste = Autunno 2012; gruppo marrone = Bonassai 2013).

Tabella 11.14 – Valori medi del metil estere del C18:3 *cis*-9, *cis*-12, *cis*-15 nei diversi sottogruppi.

C18:3 <i>cis</i> -9, <i>cis</i> -12, <i>cis</i> -15	Media (g/100g latte)	CV%*	Range
Campioni Totali	0,100	48	0,027 – 0,267
Bonassai 2011	0,135	32	0,049 – 0,267
Oristano 2012	0,085	19	0,052 – 0,129
Monastir 2012	0,061	25	0,039 – 0,106
Autunno 2012	0,056	27	0,027 – 0,089
Bonassai 2013	0,043	23	0,030 – 0,063

* coefficiente di variazione percentuale della distribuzione

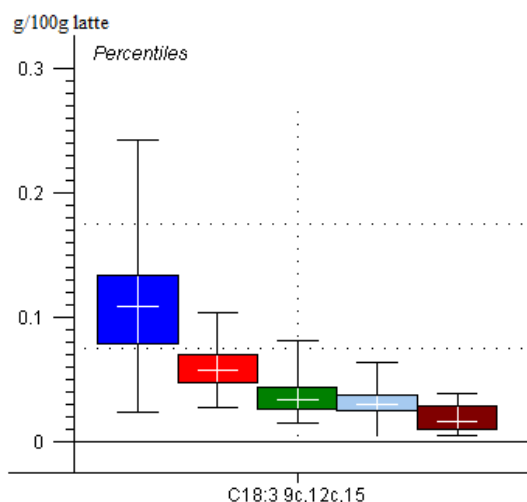


Figura 11.25 - Quartili, mediane e range per il C18:3 *cis*-9, *cis*-12, *cis*-15 dei diversi sottogruppi (gruppo blu = Bonassai 2011; gruppo rosso = Oristano 2012; gruppo verde = Monastir 2012; gruppo celeste = Autunno 2012; gruppo marrone = Bonassai 2013).

Tabella 11.15 – Valori medi del metil estere del C18:2 *cis*-9, *trans*-11 nei diversi sottogruppi.

C18:2 <i>cis</i> -9, <i>trans</i> -11	Media (g/100g latte)	CV%*	Range
Campioni Totali	0,086	42	0,016 – 0,188
Bonassai 2011	0,077	35	0,023 – 0,161
Oristano 2012	0,135	16	0,079 – 0,188
Monastir 2012	0,068	19	0,053 – 0,108
Autunno 2012	0,058	40	0,016 – 0,121
Bonassai 2013	0,056	32	0,021 – 0,101

* coefficiente di variazione percentuale della distribuzione

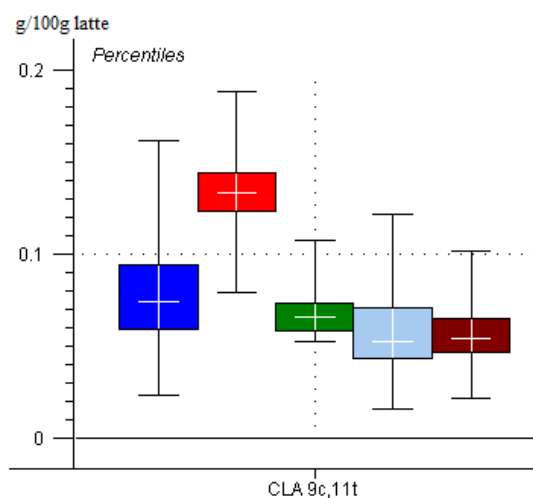


Figura 11.26 - Quartili, mediane e range per il C18:2 *cis*-9, *trans*-11 dei diversi sottogruppi (gruppo blu = Bonassai 2011; gruppo rosso = Oristano 2012; gruppo verde = Monastir 2012; gruppo celeste = Autunno 2012; gruppo marrone = Bonassai 2013).

Tabella 11.16 – Valori medi della somma dei metil esteri degli acidi grassi saturi nei diversi sottogruppi.

SFA	Media (g/100g latte)	CV%*	Range
Campioni Totali	3,570	23	1,411 – 7,300
Bonassai 2011	3,771	27	1,728 – 7,300
Oristano 2012	3,547	9	2,835 – 4,223
Monastir 2012	3,268	15	2,534 – 4,284
Autunno 2012	3,354	16	1,411 – 4,002
Bonassai 2013	3,181	21	2,160 – 4,333

* coefficiente di variazione percentuale della distribuzione

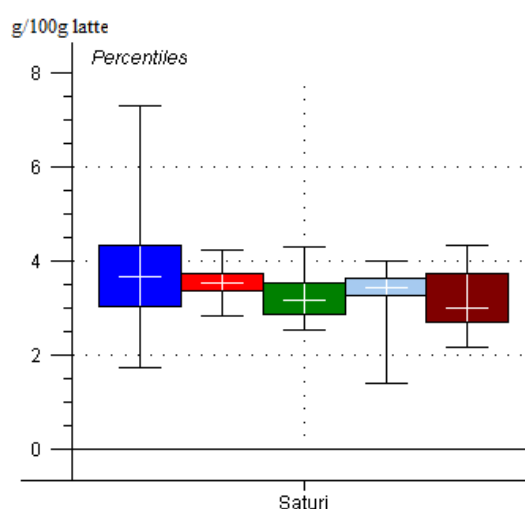


Figura 11.27 - Quartili, mediane e range per la somma degli acidi grassi saturi dei diversi sottogruppi (gruppo blu = Bonassai 2011; gruppo rosso = Oristano 2012; gruppo verde = Monastir 2012; gruppo celeste = Autunno 2012; gruppo marrone = Bonassai 2013).

Tabella 11.17– Valori medi della somma dei metil esteri degli acidi grassi insaturi nei diversi sottogruppi.

UFA	Media (g/100g latte)	CV%*	Range
Campioni Totali	1,853	26	0,719 – 4,043
Bonassai 2011	1,801	30	0,719 – 4,043
Oristano 2012	2,169	14	1,637 – 2,985
Monastir 2012	1,656	17	1,155 – 2,272
Autunno 2012	1,758	18	0,772 – 2,242
Bonassai 2013	1,633	24	1,116 – 2,261

* coefficiente di variazione percentuale della distribuzione

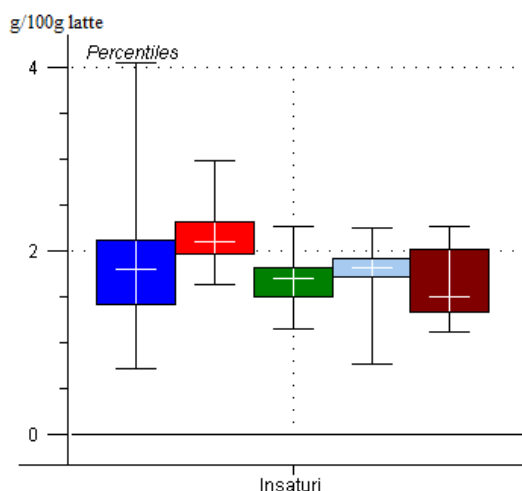


Figura 11.28 - Quartili, mediane e range per la somma degli acidi grassi insaturi dei diversi sottogruppi (gruppo blu = Bonassai 2011; gruppo rosso = Oristano 2012; gruppo verde = Monastir 2012; gruppo celeste = Autunno 2012; gruppo marrone = Bonassai 2013).

Tabella 11.18 – Valori medi della somma dei metil esteri degli acidi grassi monoinsaturi nei diversi sottogruppi.

MUFA	Media (g/100g latte)	CV%*	Range
Campioni Totali	1,366	27	0,496 – 3,071
Bonassai 2011	1,265	32	0,496 – 3,071
Oristano 2012	1,600	19	1,031 – 2,449
Monastir 2012	1,297	17	0,904 – 1,784
Autunno 2012	1,414	18	0,597 – 1,810
Bonassai 2013	1,322	25	0,862 – 1,847

* coefficiente di variazione percentuale della distribuzione

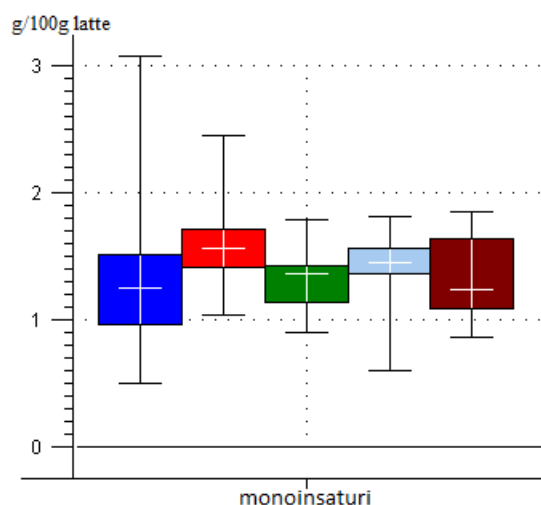


Figura 11.29 - Quartili, mediane e range per la somma degli acidi grassi monoinsaturi dei diversi sottogruppi (gruppo blu = Bonassai 2011; gruppo rosso = Oristano 2012; gruppo verde = Monastir 2012; gruppo celeste = Autunno 2012; gruppo marrone = Bonassai 2013).

Tabella 11.19 – Valori medi della somma dei metil esteri degli acidi grassi polinsaturi nei diversi sottogruppi.

PUFA	Media (g/100g latte)	CV%*	Range
Campioni Totali	0,487	30	0,175 – 0,972
Bonassai 2011	0,536	28	0,217 – 0,972
Oristano 2012	0,570	9	0,444 – 0,706
Monastir 2012	0,358	18	0,251 – 0,488
Autunno 2012	0,344	18	0,175 – 0,432
Bonassai 2013	0,311	24	0,221 – 0,439

* coefficiente di variazione percentuale della distribuzione

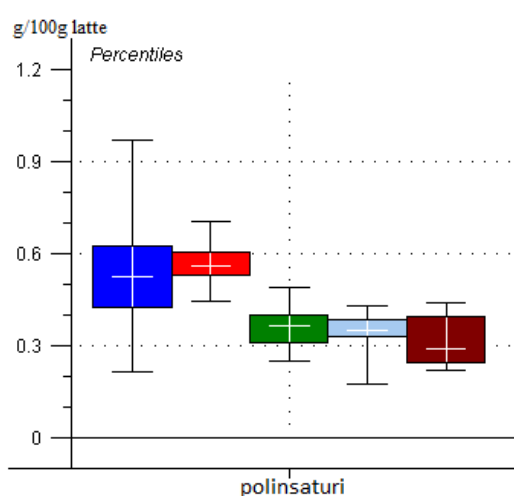


Figura 11.30 - Quartili, mediane e range per la somma degli acidi grassi polinsaturi dei diversi sottogruppi (gruppo blu = Bonassai 2011; gruppo rosso = Oristano 2012; gruppo verde = Monastir 2012; gruppo celeste = Autunno 2012; gruppo marrone = Bonassai 2013).

Tabella 11.20 – Valori medi della somma dei metil esteri degli acidi grassi ω -6 nei diversi sottogruppi.

PUFA ω -6	Media (g/100g latte)	CV%*	Range
Campioni Totali	0,174	30	0,064 – 0,396
Bonassai 2011	0,197	30	0,076 – 0,396
Oristano 2012	0,172	19	0,122 – 0,252
Monastir 2012	0,137	22	0,090 – 0,202
Autunno 2012	0,144	21	0,064 – 0,202
Bonassai 2013	0,137	28	0,094 – 0,215

* coefficiente di variazione percentuale della distribuzione

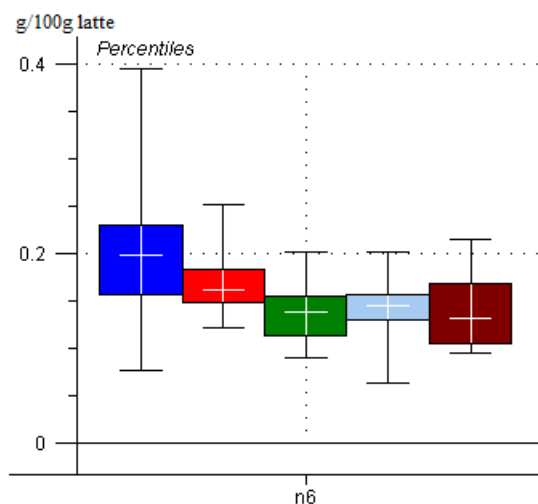


Figura 11.31 - Quartili, mediane e range per la somma degli acidi grassi ω -6 dei diversi sottogruppi (gruppo blu = Bonassai 2011; gruppo rosso = Oristano 2012; gruppo verde = Monastir 2012; gruppo celeste = Autunno 2012; gruppo marrone = Bonassai 2013).

Tabella 11.21 – Valori medi della somma dei metil esteri degli acidi grassi ω -3 nei diversi sottogruppi.

PUFA ω -3	Media (g/100g latte)	CV%*	Range
Campioni Totali	0,162	41	0,040 – 0,388
Bonassai 2011	0,199	30	0,095 – 0,388
Oristano 2012	0,178	20	0,108 – 0,271
Monastir 2012	0,099	21	0,064 – 0,152
Autunno 2012	0,095	27	0,040 – 0,141
Bonassai 2013	0,073	23	0,051 – 0,112

* coefficiente di variazione percentuale della distribuzione

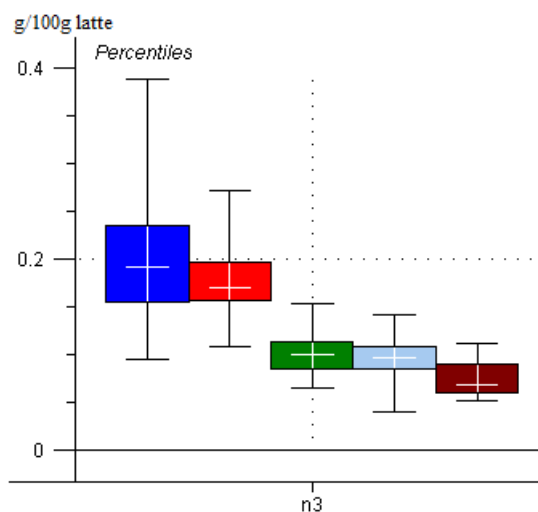


Figura 11.32 - Quartili, mediane e range per la somma degli acidi grassi ω -3 dei diversi sottogruppi (gruppo blu = Bonassai 2011; gruppo rosso = Oristano 2012; gruppo verde = Monastir 2012; gruppo celeste = Autunno 2012; gruppo marrone = Bonassai 2013).

12 Parte sperimentale II: validazione del metodo di riferimento per la caratterizzazione del profilo acidico del latte ovino

L'affidabilità di un modello di predizione basato sull'assorbimento infrarosso dei gruppi funzionali dei composti chimici presenti in una matrice è dipendente dai dati di riferimento utilizzati per la costruzione del modello. È quindi necessario che il metodo analitico utilizzato per l'acquisizione dei dati di riferimento sia caratterizzato quantomeno da buona precisione ed una sostanziale assenza di bias. Il metodo analitico utilizzato per la caratterizzazione del profilo acidico del latte ovino è stato quindi in tal senso testato per stimarne l'applicabilità al fine desiderato.

12.1 Precisione

La precisione è definita come il grado di accordo tra una serie di misure ripetute, eseguite sullo stesso campione. Essa si può descrivere in termini di:

ripetibilità (r): esprime l'accordo fra una serie di misure ottenute sullo stesso campione, da uno stesso operatore, applicando quello specifico metodo, nelle stesse condizioni sperimentali (stesso laboratorio e identica strumentazione), nell'ambito della stessa sessione di misura.

riproducibilità (R): esprime l'accordo fra una serie di misure ottenute sullo stesso campione da operatori diversi, applicando quello specifico metodo in diverse condizioni sperimentali e diverse giornate.

Dal momento che la riproducibilità tiene conto dell'indeterminazione che proviene da un maggior numero di fonti (diversità di operatori, reagenti, strumenti), il valore dell'incertezza ad essa associata è solitamente maggiore del valore di quello associato a misure di ripetibilità.

Normalmente la precisione si esprime in termini di s (stima della deviazione standard) o di CV% (coefficiente di variazione percentuale).

Un sistema adottato per stabilire il range di accettabilità delle performance di precisione utilizza il modello matematico sviluppato da William Horwitz [Horwitz, 1982]. Il modello descrive l'andamento dei coefficienti di variazione percentuale del parametro riproducibilità ($CV\%_R$) in funzione della concentrazione dell'analita. L'equazione sviluppata da Horwitz è costruita in base all'analisi di un'elevata serie di dati (oltre 7500

records) relativi ad un ampio range di concentrazione, analita e metodo analitico prodotti nel periodo tra il 1915 ed il 1995, nell'ambito di studi collaborativi curati dall'organizzazione AOAC International. Horwitz notò che - qualora si osservi una diminuzione della concentrazione di un analita di due ordini di grandezza, il coefficiente di variazione CV% raddoppia. Egli descrisse questo andamento tramite l'equazione seguente:

$$CV\%_R = 2^{1-0,5\log C}$$

ove $CV\%_R$ è il coefficiente di variazione percentuale di riproducibilità e C è la concentrazione dell'analita espressa come frazione decimale.

Esprimendo l'equazione in termini di deviazione standard s , si ottiene un'equazione che esprime la dipendenza della deviazione standard dalla concentrazione:

$$s_R = 0,02C^{0,8495}$$

Nella sua forma logaritmica, l'equazione diventa:

$$\log_{10} s_R = 0,8495 \log_{10} C - 1,699$$

Mettendo in grafico il logaritmo della deviazione standard di riproducibilità in funzione del logaritmo della concentrazione, il dataset utilizzato da Horwitz è descritto dall'andamento lineare riportato in Figura 12.1.

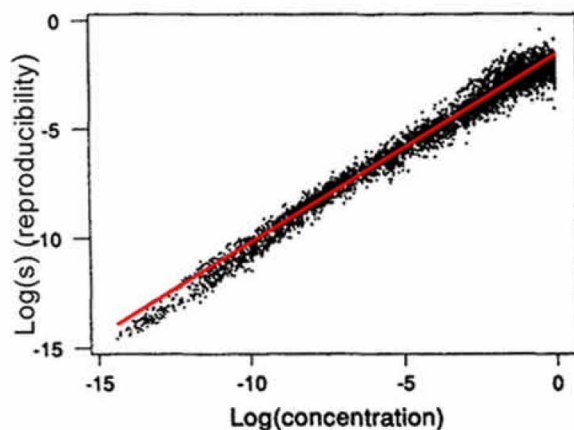


Figura 12.1 – Andamento del logaritmo della deviazione standard in funzione del logaritmo della concentrazione ottenuto dal set di dati di Horwitz.

L'analisi del dataset di Horwitz permette di ottenere le seguenti conclusioni:

- Il CV% è indipendente dalla natura dell'analita, dalla matrice e dal metodo analitico ma dipende soltanto dalla concentrazione;
- La maggior parte dei dati sperimentali è compreso nell'intervallo:

$$0,5s_R < s_R < 1,5s_R$$

per cui

$$0,5CV\%_R < CV\%_{\text{sperim}} < 1,5CV\%_R$$

Questo intervallo è definito nella pratica di laboratorio “banda di Horwitz”. Un CV% ottenuto sperimentalmente è ritenuto accettabile se inferiore al limite superiore della banda di Horwitz. Inoltre, viene utilizzato un secondo parametro, chiamato indice HorRat, (abbreviazione di Horwitz Ratio), pari al rapporto tra il CV% sperimentale ($CV\%_{\text{sperim}}$) ed il CV% calcolato tramite l'equazione di Horwitz. Se l'indice di HorRat è inferiore al valore 1,5 allora il $CV\%_{\text{sperim}}$ è ritenuto accettabile. Per cui, per valutare la riproducibilità, si utilizza il parametro HorRat_R :

$$\text{HorRat}_R = \frac{CV\%_{\text{sperim}}}{CV\%_R} < 1,5$$

Dato che la precisione di ripetibilità può essere stimata 0,66 volte la precisione di riproducibilità, è possibile scrivere la seguente equazione per valutare la ripetibilità:

$$CV\%_r = 0,66CV\%_R$$

Anche il parametro Horwitz può essere utilizzato per stimare la ripetibilità, per cui

$$\text{HorRat}_r = \frac{CV\%_{\text{sperim}}}{CV\%_r} < 1,5$$

12.2 Esattezza

L'esattezza è la misura dell'accordo tra il valore medio ottenuto da una serie di risultati e il valore di riferimento accettato. Essa rappresenta la misura della presenza (o meno..) di errori sistematici (bias). L'esattezza di un metodo analitico è generalmente stimata mediante confronto con materiali di riferimento certificati o con metodiche indipendenti o – al limite – tramite test di recupero su matrici additivate.

Secondo le definizioni ISO:

- Un materiale di riferimento (anche detto Materiale di Riferimento da Laboratorio LRM) è un materiale avente una o più proprietà sufficientemente ben stabilite da poter essere usato per la calibrazione di apparecchiature, la valutazione di metodi o per l'analisi quantitativa di materiali.
- Un materiale di riferimento certificato (anche detto Materiale Standard di Riferimento CRM) è un materiale avente uno o più valori certificati di alcune sue proprietà. Questi devono essere certificati attraverso una procedura tecnicamente riconosciuta valida. Essi devono essere accompagnati da un certificato d'analisi.

Solo un CRM può esser utilizzato nell'ambito della verifica dell'esattezza in una procedura di validazione analitica. Tuttavia, la generale scarsa disponibilità (ed il conseguente elevato costo) dei CRM fanno sì che spesso si adottino procedure di additivazione multipla di matrici reali o di LRM.

Affinché il metodo in validazione possa esser considerato esatto è necessario che il valore analitico ottenuto non sia statisticamente dissimile da quello riportato nel certificato del CRM. Tuttavia il numero dei CRM presenti in commercio è largamente inferiore a quello delle specie e delle matrici di interesse analitico. Ove quindi l'analisi dei CRM non sia praticabile, è prassi abbastanza consolidata stimare l'esattezza mediante prove di recupero. La procedura delle prove di recupero prevede l'analisi di n (con n compreso tra 2 e 4) aliquote di uno stesso campione, di cui $n-1$ arricchite con quantità note dell'analita. La dipendenza della concentrazione sperimentale dalla quantità di analita aggiunta al campione è espressa dalla relazione:

$$C_{\text{sperim}} = C_0 + \frac{q_a}{m}$$

dove C_{sperim} è la concentrazione sperimentale di analita dopo n aggiunte; C_0 è la concentrazione iniziale di analita nel campione; q_a è la quantità di standard nell'aggiunta; m è la quantità pesata del campione.

Diagrammando C_{sperim} contro q_a/m si ottiene, con il metodo dei minimi quadrati, una retta di regressione (funzione di recupero) la cui pendenza, opportunamente percentualizzata, rappresenta il valore del recupero. Per stabilire se il recupero ottenuto è quantitativo o meno, si applica un test di Student a due code, calcolando il t_{sperim} con la relazione

$$|t_{\text{sperim}}| = \left| (\bar{c} - \mu) \frac{\sqrt{n}}{s} \right|$$

dove \bar{c} rappresenta il dato di recupero ottenuto, μ rappresenta il dato di recupero teorico (100%), n rappresenta il numero di osservazioni fatte ed s è la stima della deviazione standard. Fissato un certo livello di fiducia p (solitamente 95%), si valuta il numero di gradi di libertà ($n-1$) e si confronta il t di Student sperimentale con il t di Student tabulato per il livello di fiducia considerato. Il recupero può considerarsi quantitativo se il t sperimentale è inferiore al t tabulato.

Tabella 12.1 - Intervalli di recupero medio % in base alla concentrazione secondo l'AOAC.

Concentrazione (frazione di massa)	Concentrazione	Recupero medio %
1	1 kg/kg	98-102
10^{-1}	100 g/kg	98-102
10^{-2}	10 g/kg	97-103
10^{-3}	1 g/kg	95-105
10^{-4}	100 mg/kg	90-107
10^{-5}	10 mg/kg	80-110
10^{-6}	1 mg/kg	80-110
10^{-7}	100 $\mu\text{g}/\text{kg}$	80-110
10^{-8}	10 $\mu\text{g}/\text{kg}$	60-115
10^{-9}	1 $\mu\text{g}/\text{kg}$	40-120

La mera valutazione statistica tramite t-test dei dati provenienti da misure di esattezza può talvolta condurre a conclusioni fuorvianti circa l'accettabilità dei valori ottenuti. Se è vero che la probabilità statistica di presenza di bias in dati che si discostano da un recupero

quantitativo è tanto maggiore quanto maggiore è il gap osservato, non si può peraltro ignorare che il livello di esattezza di una misura dipende in maniera capitale dal range di concentrazioni studiato. L'AOAC, nel suo manuale per il Peer Verified Methods Program, nota che l'intervallo tipico di recupero medio tende a crescere in maniera notevole al decrescere della concentrazione dell'analita in esame (Tabella 12.1) e definisce dei range di recupero che – pur oggettivando la presenza di errore sistematico nella misura (t-test negativo) – permette di poter ritenere “accettabile” un valore di recupero funzione della concentrazione dell'analita che sia compreso entro il range di cui in Tabella 12.1.

12.3 Prove di ripetibilità e riproducibilità

La ripetibilità è stata effettuata tramite il calcolo del coefficiente di variazione percentuale ($CV\%_r$) dei risultati espressi in percentuale sui FAME ottenuti da 10 repliche di uno stesso campione di latte ovino. In Tabella 12.2 sono riportati i coefficienti di variazione sperimentale ed i valori del parametro HorRat per la ripetibilità ottenuti per i 65 acidi grassi analizzati e le 6 classi di acidi grassi. Si può vedere che i coefficienti di variazione presentano l'andamento dettato dall'equazione di Horwitz e tendono a diminuire all'aumentare della concentrazione dell'analita. Infatti, gli acidi grassi saturi con numero di carboni pari, gli acidi vaccenico, oleico, linoleico hanno $CV\%$ inferiore agli altri acidi grassi. In generale, infatti, le categorie degli acidi grassi, con concentrazioni in percentuale dell'ordine di 10^{-1} , presentano valori di $CV\%_r$ tra lo 0,5% e l'1,3%; gli acidi grassi con concentrazioni in percentuale dell'ordine di 10^{-2} , presentano $CV\%_r$ tra l'1,0% e l'1,7%; gli acidi grassi con concentrazioni in percentuale dell'ordine di 10^{-3} presentano $CV\%_r$ tra l'1,3% ed il 5,4%; gli acidi grassi con concentrazioni in percentuale dell'ordine di 10^{-4} presentano $CV\%_r$ tra il 2,5% ed il 7,8%. L'unico acido grasso che pare avere un $CV\%_r$ mediamente più alto del $CV\%_r$ medio del range di concentrazione di cui fa parte è il C16:0 che presenta $CV\%_r$ pari a 2,7% nonostante la sua concentrazione sia dell'ordine di grandezza 10^{-1} . Il $CV\%_r$ del C16:0 rientra comunque nella banda di oscillazione di Horwitz, oggettivando comunque un valore HorRat_r minore di 1,5.

La riproducibilità è stata effettuata tramite il calcolo del coefficiente di variazione percentuale dei dati in percentuale ottenuti da 20 repliche di un campione di latte ovino, analizzate in 4 giornate differenti (5 repliche per giornata). Il coefficiente di variazione di riproducibilità è stato calcolato su 4 valori medi, ottenuti dalla media delle 5 repliche di ogni giornata.

Tabella 12.2 – Dati di ripetibilità.

Acido grasso	CV% _{r_sperim}	HorRat _r	Acido grasso	CV% _{r_sperim}	HorRat _r
C4:0	1,2	0,5	C18:2-9c,12t/8t 12c	4,8	1,3
C6:0	1,4	0,5	C18:1 16c	4,7	1,3
C7:0	4,6	0,9	C18:2 9t,12c	6,5	1,7
C8:0	1,1	0,4	C18:2 11t,15c	2,2	0,6
C10:0	1,4	0,5	C18:2 9c,12c	1,3	0,5
C11:0	2,0	0,5	C18:2 9c,15c	7,1	1,3
C12:0	1,7	0,6	C20:0	3,5	0,9
C13:0 i	4,5	0,9	C18:3 6c,9c,12c	5,8	1,1
C13:0 ai	4,4	0,8	C20:1 9c	7,6	1,4
C14:0 i	2,5	0,5	C20:1 11c	7,8	1,5
C14:0	0,5	0,2	C18:3 9c,12c,15c	1,4	0,5
C15:0 i	1,7	0,4	CLA 9c,11t	1,1	0,4
C15:0 ai	1,3	0,4	CLA 9t,11c	6,5	1,2
C14:1 9c	2,2	0,6	CLA 10t,12c	5,4	1,4
C15:0	1,7	0,4	CLA 11t,13c	3,5	0,7
C16:0 i	1,7	0,5	CLA 12t,14t + 11c 13c	5,5	1,0
C16:0	2,7	1,5	CLA 11t,13t	4,0	1,1
C17:0 i	3,8	1,0	CLA 9t,11t	7,2	1,4
C16:1 7c	4,3	1,1	C20:2 11c,14c	4,0	0,8
C17:0 ai	3,7	1,0	C20:3 5c,8c,11c	2,6	0,7
C16:1 9c	3,0	0,8	C22:0	5,2	1,4
C17:0	3,7	1,0	C20:3 8c,11c,14c	7,3	1,4
C18:0 i	4,2	0,8	C20:3 11c,14c,17c	4,4	0,8
C17:1 10c	2,4	0,6	C20:4 5c,8c,11c,14c	2,3	0,6
C18:0	1,1	0,4	C23:0	6,0	1,1
C18:1 11t	1,1	0,4	C24:0	6,6	1,3
C18:1 12t	1,8	0,5	C20:5 5c,8c,11c,14c,17c	6,2	1,2
C18:1 13t+14t	1,4	0,5	C22:5 7c,10c,13c,16c,19c	3,2	0,8
C18:1 9c	1,3	0,7	C22:6 4c,7c, 10c,13c,16c,19c	6,3	1,2
C18:1 15t+10c	4,8	1,3	Saturi	0,6	0,3
C18:1 11c	3,1	0,8	Insaturi	1,0	0,5
C18:1 12c	2,3	0,6	Monoinsaturi	1,0	0,6
C18:1 13c	2,3	0,6	Polinsaturi	1,1	0,6
C18:1 14c+16t	1,5	0,4	ω-6	1,3	0,5
C18:2 9t,12t	4,9	0,9	ω-3	1,2	0,4
C18:2 9c,13t	1,1	0,3			

In Tabella 12.3 sono riportati i coefficienti di variazione sperimentale ed i valori del parametro HorRat per la riproducibilità ottenuti per i 65 acidi grassi analizzati e le 6 categorie di acidi grassi. Si può vedere che anche in questo caso i coefficienti di variazione presentano l'andamento dettato dall'equazione di Horwitz e tendono a diminuire all'aumentare della concentrazione dell'analita. In generale, infatti, le categorie degli acidi grassi, con concentrazioni in percentuale dell'ordine di 10^{-1} , presentano valori di CV%_R al di sotto dello 0,7%; gli acidi grassi con concentrazioni in percentuale dell'ordine di 10^{-2} ,

presentano CV%_R tra lo 0,4% e l'1,7%; gli acidi grassi con concentrazioni in percentuale dell'ordine di 10⁻³ presentano CV%_R tra l'1,0% e l'8,0%; gli acidi grassi con concentrazioni in percentuale dell'ordine di 10⁻⁴ presentano CV%_R tra il 1,9% e l'11 %. Tre acidi grassi, presenti in basse concentrazioni percentuali (10⁻⁴), presentano coefficienti di variazione di riproducibilità troppo elevati e non rientrano nella banda di Horwitz.

Tabella 12.3 – Dati di riproducibilità.

Acido grasso	CV% _R sperim	HorRat _R	Acido grasso	CV% _R sperim	HorRat _R
C4:0	1,4	0,3	C18:2-9c,12t/8t 12c	7,8	1,4
C6:0	1,1	0,3	C18:1 16c	6,2	1,1
C7:0	2,3	0,3	C18:2 9t,12c	4,3	0,5
C8:0	1,1	0,3	C18:2 11t,15c	1,1	0,2
C10:0	0,9	0,2	C18:2 9c,12c	0,5	0,1
C11:0	0,6	0,1	C18:2 9c,15c	5,5	0,7
C12:0	0,4	0,1	C20:0	1,5	0,3
C13:0 i	3,1	0,4	C18:3 6c,9c,12c	4,0	0,5
C13:0 ai	1,9	0,2	C20:1 9c	3,3	0,4
C14:0 i	1,3	0,2	C20:1 11c	4,8	0,6
C14:0	0,4	0,1	C18:3 9c,12c,15c	0,6	0,2
C15:0 i	0,9	0,2	CLA 9c,11t	0,5	0,1
C15:0 ai	0,9	0,2	CLA 9t,11c	10,8	1,4
C14:1 9c	1,9	0,3	CLA 10t,12c	1,1	0,2
C15:0	0,5	0,1	CLA 11t,13c	2,5	0,3
C16:0 i	0,5	0,1	CLA 12t,14t + 11c 13c	35,4	4,4
C16:0	0,7	0,2	CLA 11t,13t	7,5	1,3
C17:0 i	1,1	0,2	CLA 9t,11t	2,9	0,4
C16:1 7c	4,3	0,8	C20:2 11c,14c	6,3	0,8
C17:0 ai	1,8	0,3	C20:3 5c,8c,11c	1,7	0,3
C16:1 9c	1,3	0,2	C22:0	3,3	0,6
C17:0	1,8	0,3	C20:3 8c,11c,14c	37,2	6,6
C18:0 i	7,2	0,9	C20:3 11c,14c,17c	4,0	0,5
C17:1 10c	0,7	0,1	C20:4 5c,8c,11c,14c	3,1	0,5
C18:0	0,5	0,1	C23:0	5,7	0,7
C18:1 11t	0,9	0,2	C24:0	7,1	0,9
C18:1 12t	0,5	0,1	C20:5 5c,8c,11c,14c,17c	7,9	1,4
C18:1 13t+14t	3,3	0,6	C22:5 7c,10c,13c,16c,19c	1,1	0,2
C18:1 9c	0,3	0,1	C22:6 4c,7c, 10c,13c,16c,19c	7,5	0,9
C18:1 15t+10c	7,3	1,3	Saturi	0,3	0,1
C18:1 11c	3,9	0,7	Insaturi	0,6	0,2
C18:1 12c	1,8	0,3	Monoinsaturi	0,5	0,2
C18:1 13c	5,9	0,7	Polinsaturi	1,6	0,4
C18:1 14c+16t	1,2	0,2	ω-6	1,5	0,4
C18:2 9t,12t	35,8	4,5	ω-3	0,7	0,2
C18:2 9c,13t	2,3	0,4			

12.4 Prove di recupero

La procedura delle prove di recupero è stata eseguita effettuando l'analisi in triplo di 4 aliquote di un campione di grasso estratto da latte ovino, di cui 3 arricchite con quantità note dell'analita, ovvero con soluzioni al 100%, al 50% ed al 25% di standard di analiti presenti nella matrice. Non potendo effettuare – per ovvie ragioni dettate dall'elevato numero di analiti – le prove di recupero su ciascun analita, si è deciso di effettuare dette prove su un certo numero degli stessi, selezionati in termine di rappresentatività, di range di peso molecolare, di numero di insaturazioni nella catena idrocarburica e di oggettiva importanza. In particolare sono stati sottoposti a prove di recupero gli esteri metilici dei seguenti nove acidi grassi: C4:0, C6:0, C8:0, C12:0, C18:0, C18:1 *trans*-11, C18:1 *cis*-9, C18:2 *cis*-9, *cis*-12, e C18:2 *cis*-9, *trans*-11. La caratterizzazione del profilo acidico è stata eseguita utilizzando i metodi già in precedenza illustrati. Le soluzioni di tali standard sono state addizionate nelle diverse aliquote a partire dallo stadio di derivatizzazione del grasso in quanto, avendo a disposizione come standard gli acidi grassi in forma metilata e non in forma trigliceridica, l'aggiunta nello stadio di estrazione avrebbe inficiato il recupero degli acidi grassi a corta catena parzialmente solubili nella fase acquosa della matrice latte.

Il *t* di Student sperimentale è stato confrontato con il *t* di Student calcolato per 3 gradi di libertà (4 osservazioni meno 1) ed un valore $\alpha = 0,05$ ($p = 95\%$). Come si vede dalla Tabella 12.4 tutti i recuperi appaiono essere quantitativi, eccezion fatta che per il metil estere dell'acido rumenico.

Tabella 12.4 – Dati delle prove di recupero

Metil estere	Range (g/Kg)	Recupero \pm SD * (%)	Range accettabilità AOAC (%)	t sperim	t tab
MeC4:0	10	99 \pm 2	97-103	0,604	2,353
MeC6:0	10	99 \pm 2	97-103	1,286	2,353
MeC8:0	10	99,9 \pm 0.2	97-103	0,528	2,353
MeC12:0	10	99 \pm 2	97-103	0,522	2,353
MeC18:0	100	100 \pm 2	98-102	0,187	2,353
MeC18:1 <i>trans</i> -11	10	100 \pm 4	97-103	0,261	2,353
MeC18:1 <i>cis</i> -9	100	101 \pm 2	98-102	0,528	2,353
MeC18:2 <i>cis</i> -9, <i>cis</i> -12	10	102 \pm 1	97-103	2,168	2,353
MeC18:2 <i>cis</i> -9, <i>trans</i> -11	10	105 \pm 1	97-103	9,539	2,353

* SD =deviazione standard

12.5 Variabilità sperimentale del metodo di riferimento

Si è visto che, per costruire un modello di predizione, si devono utilizzare tecniche di regressione che mettono in relazione un set di variabili indipendenti (nel nostro caso gli spettri FT-IR) con una variabile dipendente (nel nostro caso i valori di riferimento di un acido grasso). Questa dipendenza determina che l'errore ottenuto nel modello di predizione (RMSEP) non potrà essere inferiore alla variabilità sperimentale s_{ref}^2 del metodo di riferimento (paragrafo 8.3.5).

Si è quindi proceduto a calcolare la variabilità sperimentale per avere un'idea del minimo errore in predizione che si può teoricamente ottenere da ogni modello. Sono stati raccolti 15 campioni individuali di latte ovino ottenuti da pecore di un gregge sperimentale del Dipartimento per la Ricerca nelle Produzioni Animali (DIRPA, AGRIS Sardegna) su cui è stata applicata la metodica di riferimento (paragrafo 11.4) utilizzata per l'acquisizione dei "valori veri". Ogni campione è stato analizzato in doppio e le repliche sono state elaborate in tempi diversi e da operatori diversi in modo tale da massimizzarne la deviazione.

La variabilità sperimentale è stata calcolata per gli acidi grassi e le classi di acidi grassi di cui verranno costruiti i modelli di predizione, ovvero i C4:0, C6:0, C8:0, C10:0, C12:0, C14:0, C16:0, C18:0, C18:1 *trans*-11, C18:1 *cis*-9, C18:2 *cis*-9, *cis*-12, C18:3 *cis*-9, *cis*-12, *cis*-15, C18:2 *cis*-9, *trans*-11 e le classi di: Acidi Grassi Saturi (SFA), Acidi Grassi Insaturi (UFA), Acidi Grassi Monoinsaturi (MUFA), Acidi Grassi Polinsaturi (PUFA), Acidi Grassi Omega 6 (PUFA ω -6), Acidi Grassi Omega 3 (PUFA ω -3).

L'equazione utilizzata

$$s_{\text{ref}}^2 = \sqrt{\frac{\sum_i (s_i^2 \times \text{gdl}_i)}{\sum_i \text{gdl}_i}}$$

dove s_i^2 è la varianza tra le repliche di un campione e gdl_i i gradi di libertà, ha portato all'ottenimento dei risultati mostrati in Tabella 12.5.

Tabella 12.5 – Risultati della variabilità sperimentale del metodo di riferimento

Metil estere	Media (g/100g latte)	Variabilità sperimentale S_{ref}^2
MeC4:0	0,238	0,004
MeC6:0	0,170	0,004
MeC8:0	0,145	0,004
MeC10:0	0,41	0,01
MeC12:0	0,216	0,006
MeC14:0	0,523	0,008
MeC16:0	1,07	0,03
MeC18:0	0,456	0,008
MeC18:1 <i>trans</i> -11	0,251	0,005
MeC18:1 <i>cis</i> -9	0,848	0,015
MeC18:2 <i>cis</i> -9, <i>cis</i> -12	0,130	0,003
MeC18:3 <i>cis</i> -9, <i>cis</i> -12, <i>cis</i> -15	0,132	0,003
MeC18:2 <i>cis</i> -9, <i>trans</i> -11	0,098	0,003
SFA	3,57	0,01
UFA	1,79	0,01
MUFA	1,325	0,009
PUFA	0,464	0,005
PUFA ω -6	0,161	0,002
PUFA ω -3	0,154	0,003

13 Parte sperimentale III: analisi delle componenti principali

L'analisi delle componenti principali (PCA) (paragrafo 8.2.1) dei set di dati è stata effettuata con l'obiettivo di massimizzare la variabilità in essi contenuta permettendo così di inglobare la maggior parte dell'informazione in un numero ridotto di variabili. In questo modo è possibile raffigurare la maggior parte della variabilità del sistema di dati in grafici a due dimensioni che permettono di osservare l'andamento dei dati e rilevare possibili outliers. Le elaborazioni sono state eseguite per i due set di dati in possesso, ovvero il set contenente i valori delle assorbanze registrate tramite spettroscopia nel medio infrarosso; ed il set contenente i valori dei metil esteri degli acidi grassi analizzati con il metodo di riferimento, espressi nell'unità di misura (g/100g latte). Le PCA sono state eseguite tramite il software R versione 3.0.0 (The R Foundation For Statistical Computing). Prima di ogni elaborazione, è stata effettuata l'operazione di autoscaling (paragrafo 8.2.2) su ogni set di dati. Il numero massimo di componenti principali scelte per ogni elaborazione è 15.

13.1 PCA degli spettri FT-IR in assorbanza

La matrice utilizzata nell'elaborazione ha dimensione 250x1060 e contiene quindi 250 osservazioni, ognuna delle quali caratterizzata da uno spettro infrarosso registrato su 1060 numeri d'onda.

In Figura 13.1 è raffigurato l'andamento della varianza spiegata in funzione delle componenti principali. Le 15 componenti principali spiegano in totale il 95% della varianza del sistema mentre il numero di componenti principali ritenute significative (pari a 5), ne spiegano l'87,83%.

Tabella 13.1 – Valori di varianza spiegata percentuale per ogni componente principale

Componente Principale	Varianza spiegata (%)	Componente Principale	Varianza spiegata (%)	Componente Principale	Varianza spiegata (%)
PC1	46,73	PC6	1,76	PC11	0,56
PC2	24,95	PC7	1,18	PC12	0,52
PC3	9,10	PC8	0,79	PC13	0,39
PC4	3,79	PC9	0,71	PC14	0,38
PC5	3,26	PC10	0,58	PC15	0,33

In Figura 13.2 è raffigurato il grafico degli scores della PCA dei dati spettrali. Non è visibile una distinzione dei diversi gruppi di campioni ma è possibile notare che un campione, il 104, si discosta notevolmente dal resto del gruppo in una direzione che non segue quella della variabilità del sistema. In Figura 13.3 è riportato il plot diagnostico T^2 vs Q (paragrafo 8.3.5) per le prime due componenti principali. Il valore dell'indice T^2 è legato al leverage posseduto da un dato campione (ovvero all'influenza di un dato oggetto sulla pendenza della retta di regressione, e quindi sulla costruzione del modello) mentre l'indice Q è relazionato alla sua varianza residua. Alla luce di quanto emerso, il campione 104, un outlier pericoloso in quanto tale sia in T^2 che in Q , è stato rimosso dall'elaborazione.

Nell'ottica della predizione del profilo acidico del latte, si può notare che sono presenti alcuni campioni che si discostano dal gruppo totale aumentandone la variabilità. Questi campioni sono presenti in numero poco consistente (7 su 250) e, in quanto outliers in T^2 , avranno un'influenza molto grande sulla costruzione del modello. Il loro numero ridotto li rende pericolosi per l'affidabilità dei modelli costruiti.

Per questo motivo si è deciso di eliminare i campioni 5, 7, 18, 30, 59, 70, 80, che non verranno quindi considerati nella costruzione dei modelli di predizione degli acidi grassi nel latte ovino.

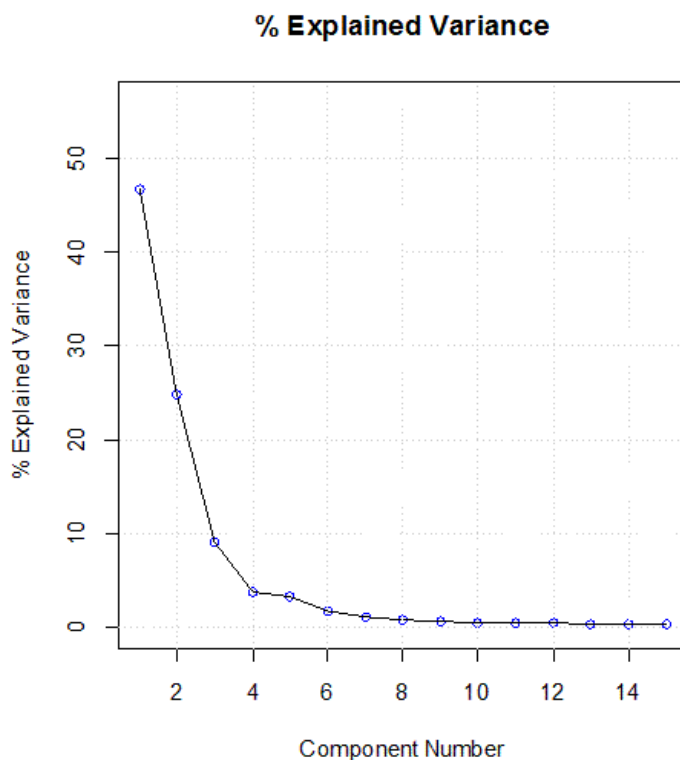


Figura 13.1 – Andamento della varianza spiegata in funzione delle componenti principali.

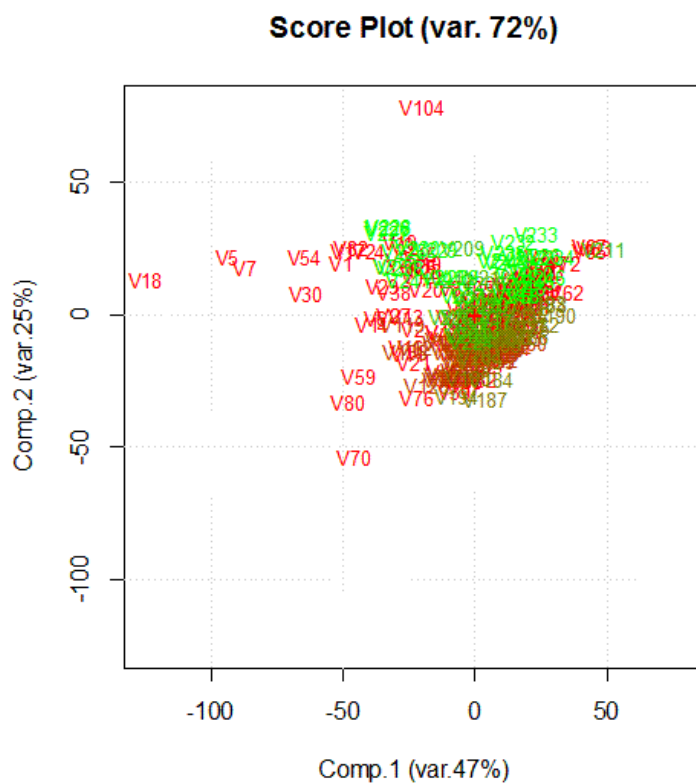


Figura 13.2 – Grafico degli scores

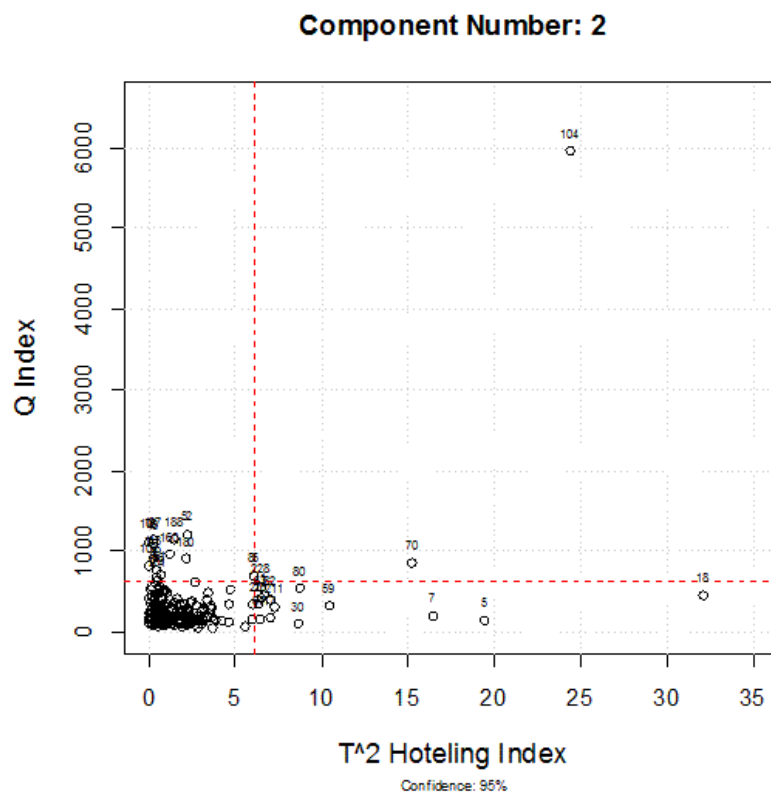


Figura 13.3 – Plot diagnostico T² vs Q

13.2 PCA dei valori di riferimento

13.2.1 PCA su 242 campioni e 71 variabili

La PCA dei valori di riferimento è stata effettuata previa rimozione degli outliers rilevati dalla PCA degli spettri. La matrice utilizzata nell'elaborazione ha dimensione 242x71 e contiene quindi 242 osservazioni, ognuna delle quali caratterizzata da 71 variabili (65 metil esteri di acidi grassi e 6 classi di metil esteri di acidi grassi).

In Figura 13.4 è raffigurato l'andamento della varianza spiegata in funzione delle componenti principali. Le 15 componenti principali spiegano in totale l'89,50% della varianza del sistema mentre il numero di componenti principali ritenute significative (5), ne spiegano il 71%.

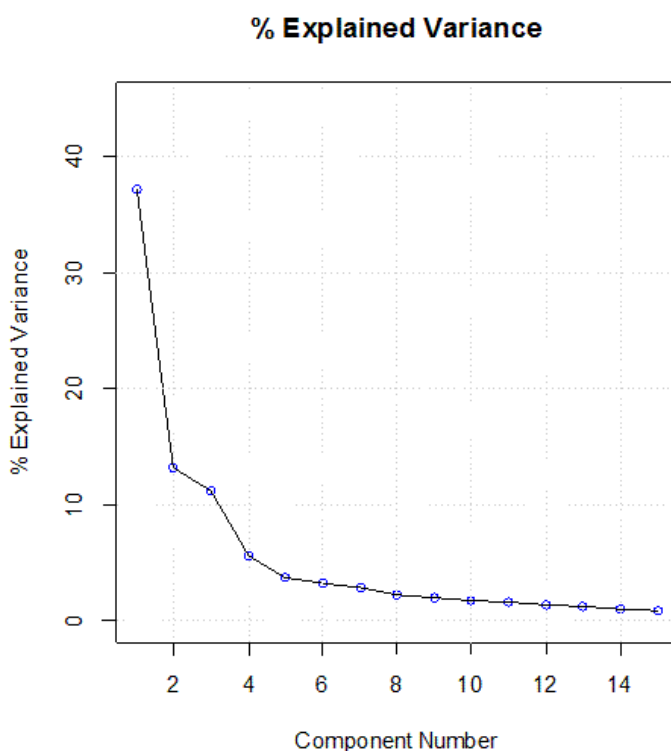


Figura 13.4 – Andamento della varianza spiegata in funzione delle componenti principali

Tabella 13.2 - Valori di varianza spiegata percentuale per ogni componente principale.

Componente Principale	Varianza spiegata (%)	Componente Principale	Varianza spiegata (%)	Componente Principale	Varianza spiegata (%)
PC1	37,10	PC6	3,24	PC11	1,66
PC2	13,24	PC7	2,87	PC12	1,43
PC3	11,21	PC8	2,30	PC13	1,28
PC4	5,67	PC9	2,06	PC14	1,01
PC5	3,78	PC10	1,72	PC15	0,94

In Figura 13.5 è mostrato il grafico degli scores in cui sono rappresentate le prime due componenti principali e in cui i 242 campioni sono stati evidenziati tramite colorazione in base alla loro origine (differenti gruppi del campionamento). Il grafico evidenzia una non completa separazione tra i gruppi, che migliora in parte comparando la seconda con la terza componente principale (Figura 13.6).



Figura 13.5 – Grafico degli scores

In Figura 13.7 si può vedere il grafico degli scores ottenuto dalla PCA dei gruppi Bonassai 2011 (rosso) e Bonassai 2013 (verde), appartenenti alla stessa area geografica ma raccolti in tempi diversi. La prima componente principale descrive la variabilità all'interno dei gruppi mentre la seconda componente discrimina il periodo. E' possibile inoltre notare una separazione del gruppo Bonassai 2013 in due ulteriori sottogruppi, che corrispondono ai campioni di latte raccolti durante la mungitura della mattina (sottogruppo sulla sinistra) ed i campioni di latte raccolti durante la mungitura della sera (sottogruppo sulla destra).

In Figura 13.8 è mostrato il grafico degli scores della PCA ottenuta tra i campioni dei gruppi Oristano 2012 e Monastir 2012, differenti per zona geografica. La discriminazione viene effettuata prevalentemente dalla prima componente principale, mentre la seconda descrive la variabilità all'interno dei gruppi.

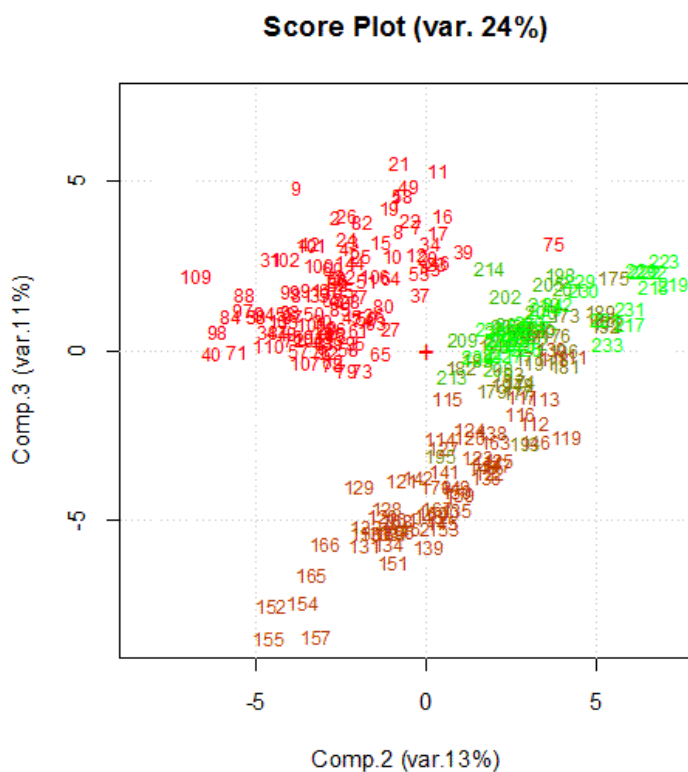


Figura 13.6 – Grafico degli scores (PC3 vs PC2) (rosso = Bonassai 2011; marrone scuro = Oristano 2012; marrone chiaro = Autunno 2012; verde scuro = Monastir 2012; verde chiaro = Bonassai 2013).



Figura 13.7 – Grafico degli scores gruppi Bonassai 2011 (rosso) e Bonassai 2013 (verde)

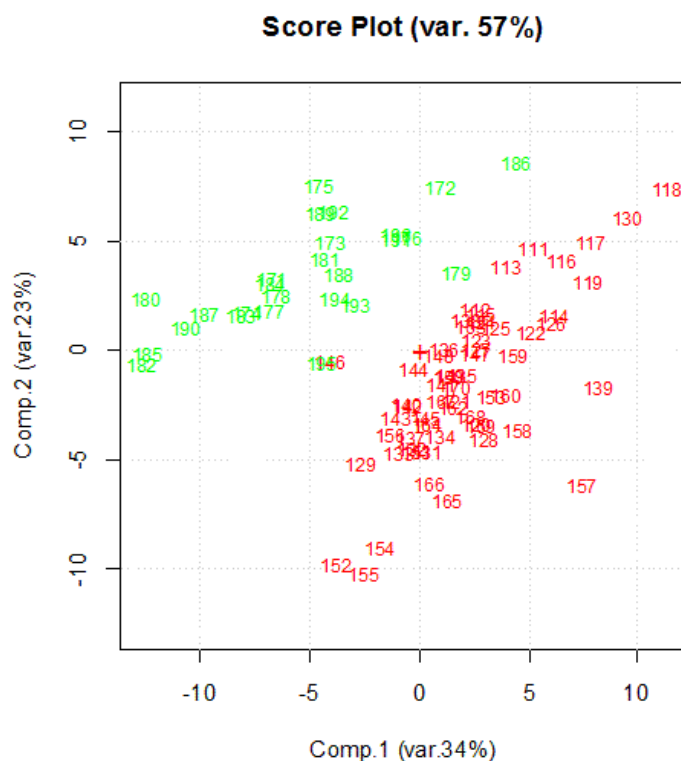


Figura 13.8 - Grafico degli scores ottenuto dalla PCA dei campioni dei gruppi Oristano 2012 (rosso) e Monastir 2012 (verde).

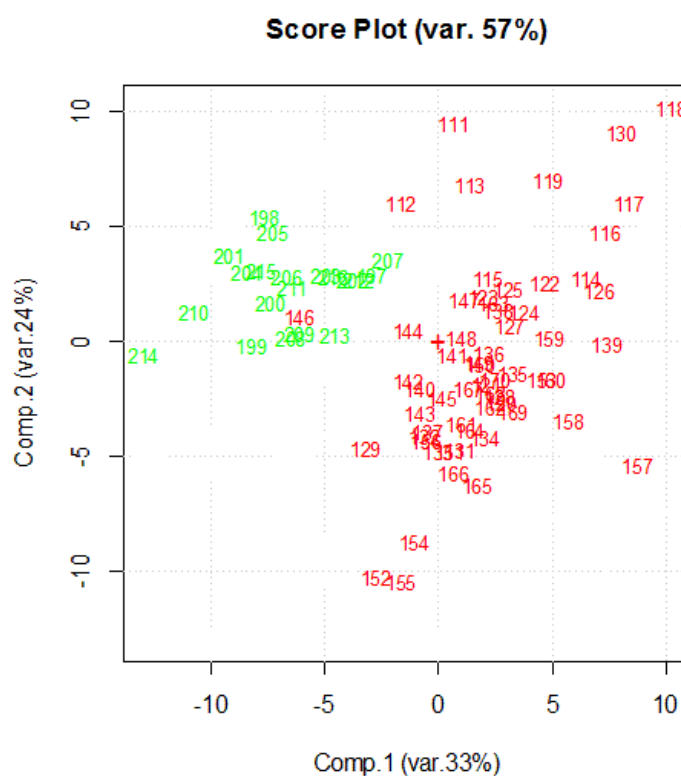


Figura 13.9 – Grafico degli scores ottenuto dalla PCA dei gruppi Oristano 2012 (rosso) e Autunno 2012 (verde).

In Figura 13.9 è mostrato il grafico degli scores della PCA ottenuta tra i campioni dei gruppi Oristano 2012 (rosso) e Autunno 2012 (verde), distinti per periodo temporale. Anche in questo caso la discriminazione viene effettuata lungo la prima componente principale mentre la seconda descrive la variabilità all'interno dei gruppi.

13.2.2 PCA su 242 campioni e 19 variabili

L'analisi delle componenti principali dei dati espressi in g/100 g di latte è stata effettuata considerando solo le 19 variabili (Tabella 13.3) di cui si vuole ottenere la predizione tramite spettroscopia infrarossa. La matrice utilizzata nell'elaborazione ha dimensione 242x19 e contiene quindi 242 osservazioni, ognuna delle quali caratterizzata da 19 variabili (13 metil esteri di acidi grassi e 6 classi di metil esteri di acidi grassi).

L'andamento della varianza spiegata è mostrato in Figura 13.10 mentre i valori per ogni componente principale sono mostrati nella Tabella 13.4. La varianza spiegata dalle 15 componenti principali è del 100% mentre il 95,16% è spiegato da sole 5 componenti.

Il grafico degli scores è mostrato in Figura 13.11. Anche qui è possibile notare una certa distinzione dei diversi gruppi, che risulta più evidente raffigurando la seconda con la terza componente principale (Figura 13.12), che però descrivono solo il 29% della varianza totale.

Tabella 13.3 – Legenda

Variabile		Variabile	
1	C4:0	11	C18:2 <i>cis</i> -9, <i>cis</i> -12
2	C6:0	12	C18:3 <i>cis</i> -9, <i>cis</i> -12, <i>cis</i> -15
3	C8:0	13	C18:2 <i>cis</i> -9, <i>trans</i> -11
4	C10:0	14	Acidi Grassi Saturi
5	C12:0	15	Acidi Grassi Insaturi
6	C14:0	16	Acidi Grassi Monoinsaturi
7	C16:0	17	Acidi Grassi Polinsaturi
8	C18:0	18	Acidi Grassi ω -6
9	C18:1 <i>trans</i> -11	19	Acidi Grassi ω -3
10	C18:1 <i>cis</i> -9		

Confrontando il grafico degli scores in Figura 13.12 con il grafico dei loadings in Figura 13.13 appare possibile differenziare i gruppi di campioni in base al contenuto dei metil esteri degli acidi grassi principali, così come fatto nel capitolo 11.5 dall'osservazione dei dati e degli istogrammi: si può notare che il gruppo Bonassai 2011 presenta alte

concentrazioni di MeC14:0 (6) e MeC16:0 (7) e della somma dei metil esteri degli acidi ω -6 (18) ed ω -3 (19) e quindi di C18:2 *cis*-9, *cis*-12 (11) e C18:3 *cis*-9, *cis*-12, *cis*-15 (12), mentre presenta basse concentrazioni della somma dei metil esteri di acidi insaturi (15), in particolare di monoinsaturi (16) e di C18:1 *trans*-11 (9). Il gruppo Oristano 2012 presenta le maggiori concentrazioni dei metil esteri degli acidi insaturi (15), in particolare di C18:1 *trans*-11 (9) e C18:2 *cis*-9, *trans*-11 (13) e basse concentrazioni dei metil esteri degli acidi C18:0 (8) e C18:1 *cis*-9 (10), che sono invece maggiormente presenti nei gruppi Bonassai 2013 e Autunno 2012. Gli esteri metilici degli acidi grassi a media e corta catena (2, 3, 4, 5) sono presenti in maggiore quantità nei gruppi Bonassai 2011 e Oristano 2012.

Tabella 13.4 – Varianza spiegata percentuale per ogni componente principale

Componente Principale	Varianza spiegata (%)	Componente Principale	Varianza spiegata (%)	Componente Principale	Varianza spiegata (%)
PC1	56,64	PC6	2,01	PC11	0,16
PC2	17,12	PC7	1,15	PC12	0,14
PC3	11,91	PC8	0,69	PC13	0,11
PC4	6,75	PC9	0,28	PC14	0,04
PC5	2,74	PC10	0,17	PC15	0,03

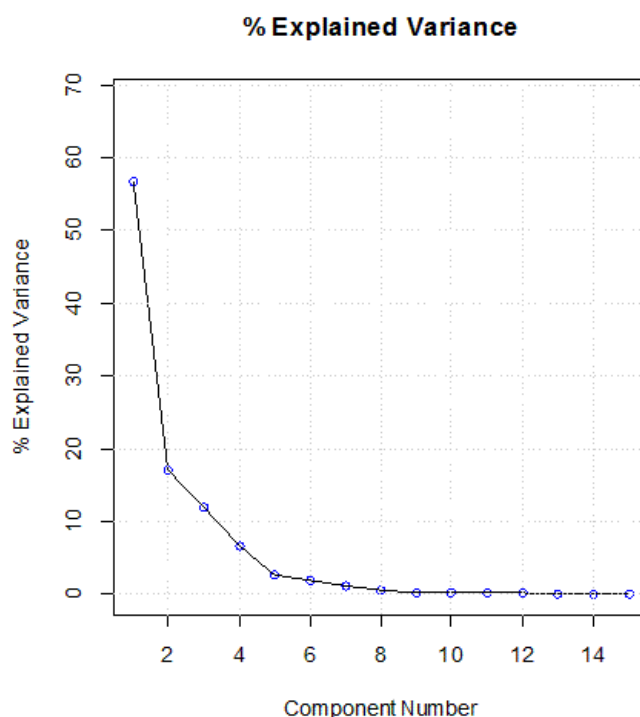


Figura 13.10 – Andamento della varianza spiegata in funzione delle componenti principali.

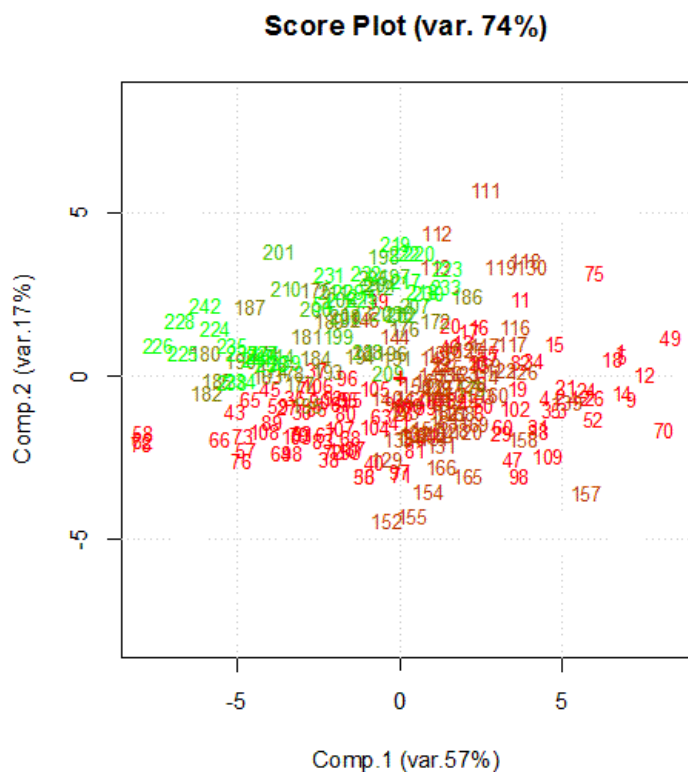


Figura 13.11 – Grafico degli scores

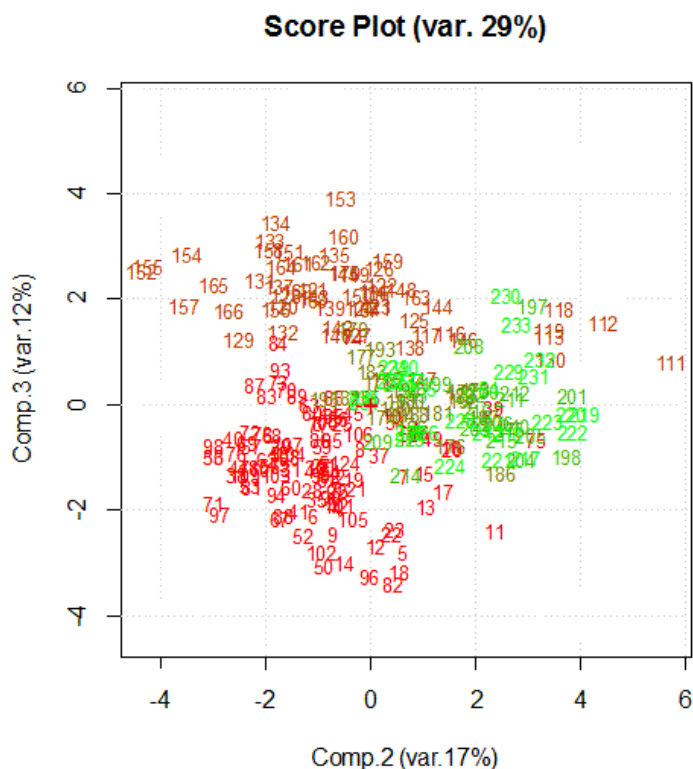


Figura 13.12 – Grafico degli scores (PC3 vs PC2) (rosso = Bonassai 2011; marrone scuro = Oristano 2012; marrone chiaro = Autunno 2012; verde scuro = Monastir 2012; verde chiaro = Bonassai 2013).

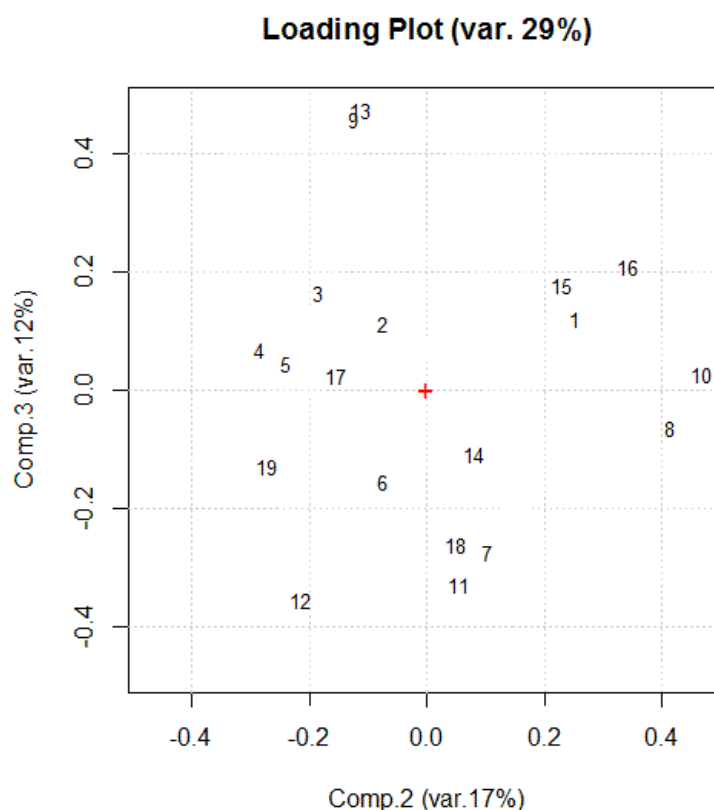


Figura 13.13 – Grafico dei loadings

13.3 PCA degli spettri FT-IR in assorbanza pretrattati con operazioni matematiche

Dato che l'analisi delle componenti principali dei dati spettrali originali non ha portato ad una distinzione dei diversi gruppi, sono state studiate le PCA degli spettri pretrattati con alcune operazioni matematiche. I pretrattamenti studiati sono lo Smoothing, la Multiplicative Scatter Correction (MSC), la Derivata prima e lo Standard Normal Variate (SNV) (paragrafo 8.4.1). Un solo pretrattamento, la Derivata prima, ha portato ai risultati sperati mentre gli altri pretrattamenti non hanno cambiato la visualizzazione dei campioni a partire dagli spettri originali o l'hanno addirittura peggiorata. Per questo motivo viene di seguito riportata esclusivamente la PCA degli spettri pretrattati con Derivata prima.

13.3.1 PCA degli spettri pretrattati con derivata prima

La PCA è stata effettuata sul set di dati dopo rimozione degli outliers indicati nei paragrafi precedenti.

La matrice utilizzata nell'elaborazione ha dimensione 242x1060 e contiene quindi 242 osservazioni, ognuna delle quali caratterizzata da uno spettro infrarosso registrato su 1060 numeri d'onda.

Prima dell'elaborazione, gli spettri sono stati pretrattati con un'operazione di derivata prima Savitzky-Golay impostata per operare con un polinomio di primo ordine su intervalli spettrali di 5 variabili (N=2) (paragrafo 8.4.1). L'operazione è stata effettuata tramite il software The Unscrambler 9.1 (CAMO) ed i dati sono stati poi importati sul software R utilizzato per le PCA.

Tabella 13.5 – Valori di varianza spiegata percentuale per ogni componente principale

Componente Principale	Varianza spiegata (%)	Componente Principale	Varianza spiegata (%)	Componente Principale	Varianza spiegata (%)
PC1	32,91	PC6	2,21	PC11	0,88
PC2	20,93	PC7	1,81	PC12	0,85
PC3	11,10	PC8	1,29	PC13	0,74
PC4	3,49	PC9	1,05	PC14	0,68
PC5	3,05	PC10	0,90	PC15	0,68

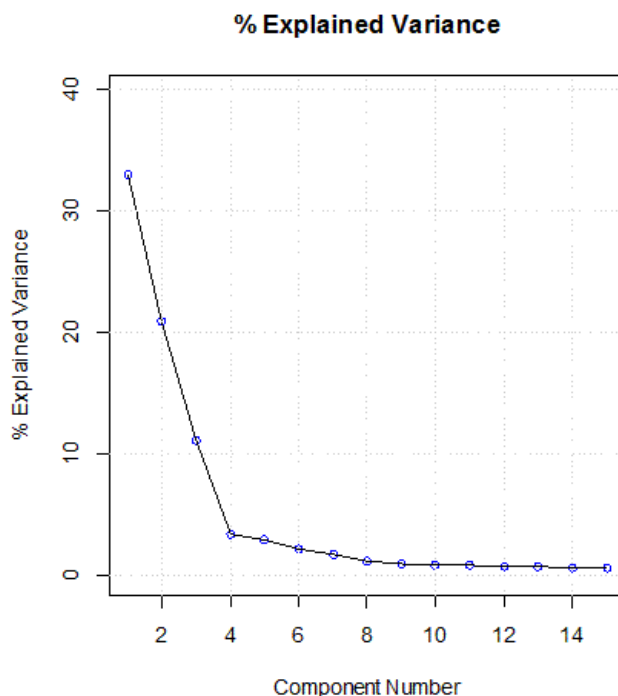


Figura 13.14 – Andamento della varianza spiegata in funzione delle componenti principali

In Figura 13.14 è raffigurato l'andamento della varianza spiegata in funzione delle componenti principali ed in Tabella 13.5 sono mostrati i valori di varianza spiegata per ogni componente principale. Le 15 componenti principali spiegano in totale l'82,60% della varianza del sistema mentre il numero di componenti principali ritenute significative (pari a 4), ne spiegano il 68,43%.

L'operazione di derivata prima ha permesso di distinguere in maniera chiara i diversi gruppi di campioni (Figura 13.15). L'effetto annata tra i campioni dei gruppi di Bonassai è osservabile nel grafico degli scores in Figura 13.16 ottenuto dalla PCA dei due gruppi; la discriminazione è osservabile lungo la seconda componente. Come per la PCA ottenuta dai dati di riferimento, è possibile notare la separazione del gruppo Bonassai 2013 in due sottogruppi che corrispondono ai campioni di latte raccolti durante la mungitura della mattina (nel grafico rappresentato dal gruppo a sinistra) ed i campioni di latte raccolti durante la mungitura della sera (nel grafico rappresentato dal gruppo a destra).

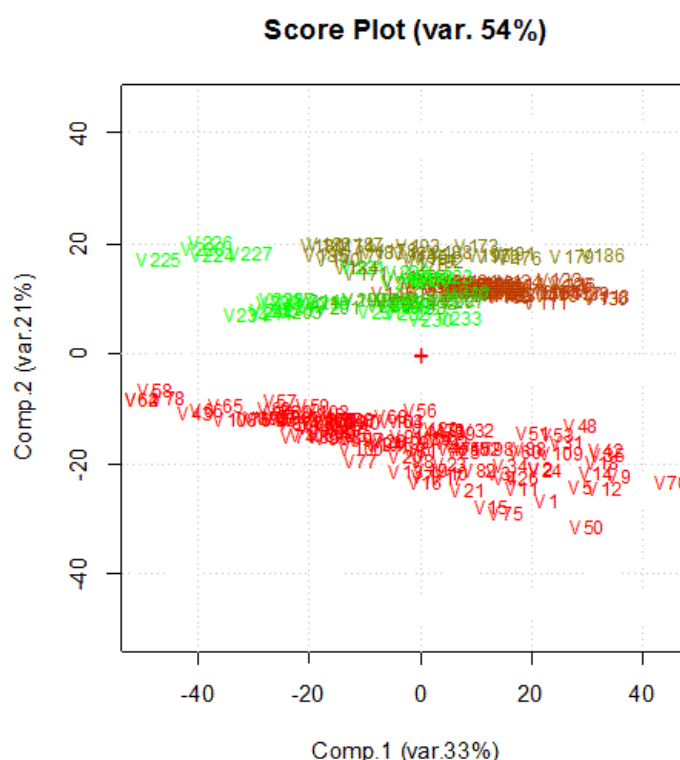


Figura 13.15 – Grafico degli scores degli spettri in derivata prima (rosso = Bonassai 2011; marrone scuro = Oristano 2012; marrone chiaro = Autunno 2012; verde scuro = Monastir 2012; verde chiaro = Bonassai 2013).

In Figura 13.17 è mostrato il grafico degli scores della PCA ottenuta tra i campioni dei gruppi Oristano 2012 (rosso) e Monastir 2012 (verde), differenti per zona geografica. La discriminazione viene effettuata dalla seconda componente principale, mentre la prima descrive la variabilità nei gruppi.

In Figura 13.18 è mostrato il grafico degli scores della PCA ottenuta tra i campioni dei gruppi Oristano 2012 (rosso) e Autunno 2012 (verde), distinti per periodo temporale. La discriminazione viene effettuata prevalentemente lungo la prima componente principale.

I grafici ottenuti nell'analisi delle componenti principali degli spettri pretrattati con derivata prima risultano essere molto simili ai grafici ottenuti dall'analisi delle componenti principali dei dati di riferimento dei metil esteri degli acidi grassi.

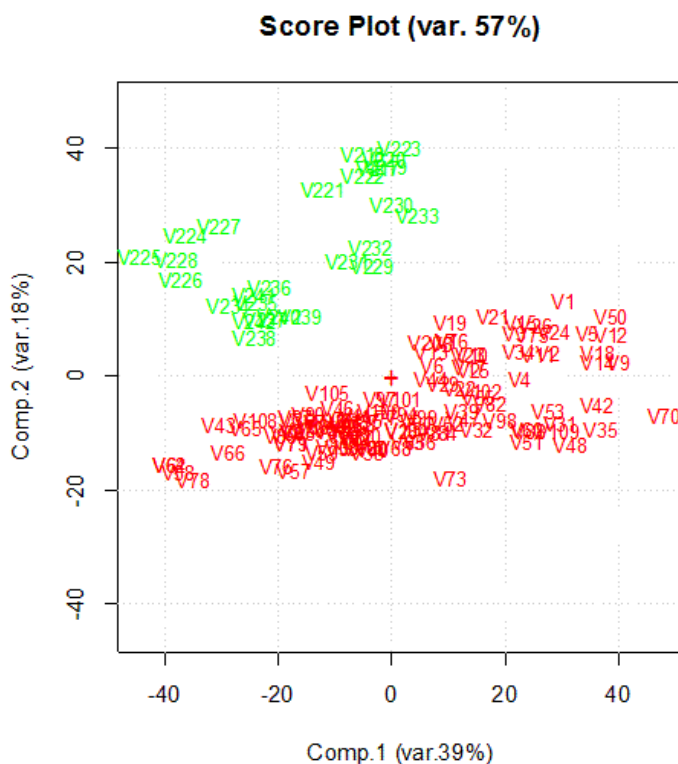


Figura 13.16 – Grafico degli scores dei gruppi Bonassai 2011 (rosso) e Bonassai 2013 (verde).

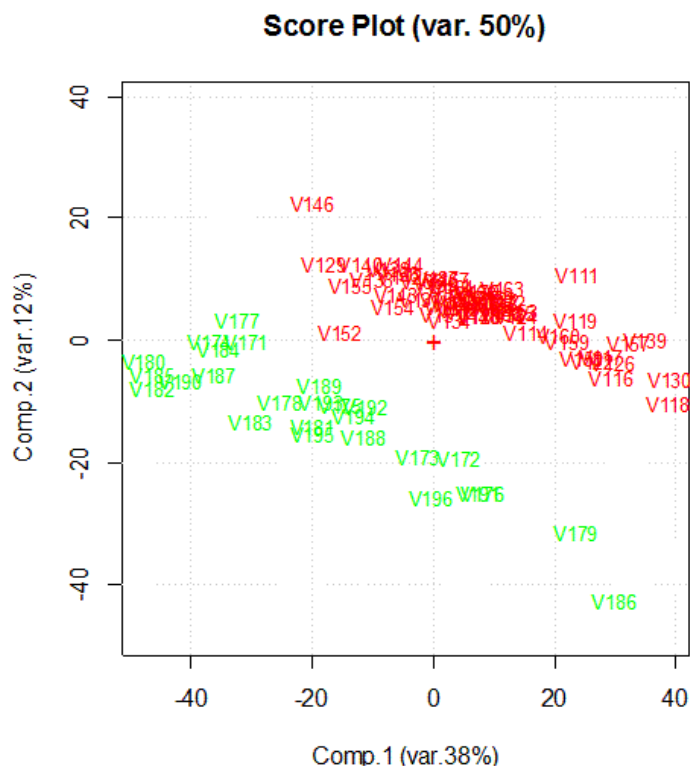


Figura 13.17 – Grafico degli scores dei gruppi Oristano 2012 (rosso) e Monastir 2012 (verde).

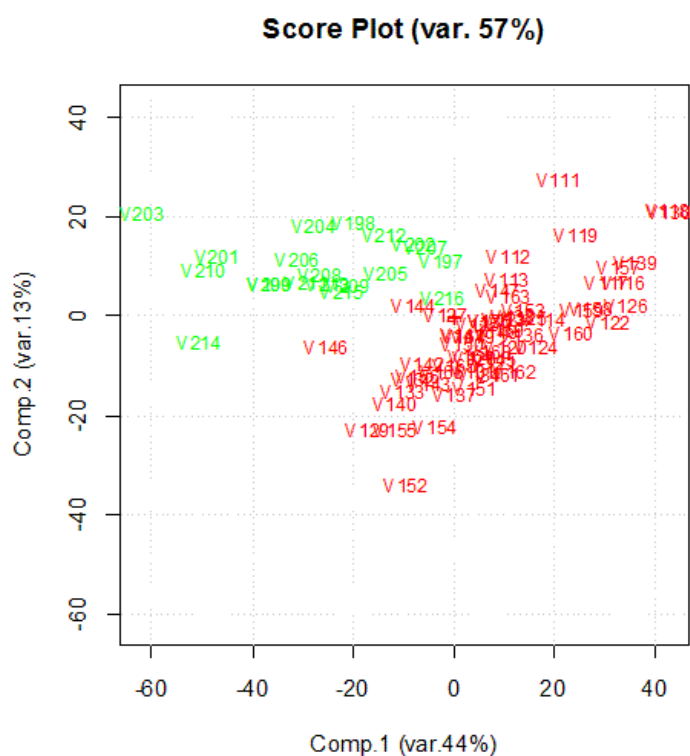


Figura 13.18 – Grafico degli scores dei gruppi Oristano 2012 (rosso) e Autunno 2012 (verde).

14 Parte Sperimentale IV: elaborazione dei modelli di predizione

Il set di dati utilizzato per la costruzione e la validazione dei modelli di predizione dei metil esteri degli acidi grassi e delle classi di metil esteri degli acidi grassi contiene 242 osservazioni (242 campioni di latte ovino), ognuna caratterizzata da uno spettro di assorbimento FT-IR nel medio infrarosso e dai valori di riferimento di ogni metil estere.

Per ogni composto è stato costruito un modello di predizione tramite regressione parziale con minimi quadrati (PLS) utilizzando gli spettri FT-IR dei campioni di latte come matrice dei predittori, ed i dati di riferimento del singolo composto come vettore della variabile risposta, per ottenere una predizione del singolo composto.

In particolare, è stato scelto di procedere effettuando dapprima una fase di calibrazione, valutata tramite una validazione interna (Cross Validazione, paragrafo 8.3.4) che ha permesso di verificare l'efficacia della fase di calibrazione. Successivamente, il modello costruito nella fase di calibrazione è stato validato su un set esterno di campioni (il cosiddetto test set) non considerato nella fase di calibrazione.

Per mantenere il concetto di "predizione", il test set esterno dovrebbe essere formato da campioni temporalmente successivi a quelli utilizzati per la calibrazione del modello. Il campionamento effettuato per questa Tesi di dottorato, non consente di suddividere il set totale di campioni in modo tale da avere, come set di validazione, campioni delle stesse aree geografiche e temporalmente successivi ai campioni del set di calibrazione. Per questi motivi, la suddivisione in training set (set di calibrazione) e test set (set di validazione), del set di 242 campioni, è stata effettuata in modo random. I campioni del training set sono stati utilizzati esclusivamente per la costruzione del modello (fase di calibrazione) mentre i campioni del test set sono stati utilizzati esclusivamente per la validazione del modello (fase di validazione). I due set di dati ottenuti sono stati utilizzati per tutti i modelli di predizione.

La validazione su set esterno random presenta un bias più piccolo rispetto a quello che si sarebbe ottenuto dalla validazione su set esterno costituito da campioni temporalmente successivi a quelli utilizzati per la costruzione del modello, in quanto i campioni del set esterno provengono dalle stesse popolazioni dei campioni del training set.

La suddivisione random è stata eseguita in modo tale da inserire circa il 60% dei campioni nel training set e circa il 40% nel test set. Il training set contiene pertanto 145 campioni e il test set ne contiene 97. Ogni elaborazione atta alla costruzione del modello è stata effettuata utilizzando esclusivamente i dati dei campioni inseriti nel training set. I dati dei

campioni inseriti nel test set sono invece stati utilizzati esclusivamente per validare i modelli costruiti sul training set.

14.1 Analisi dello spettro

Lo spettro FT-IR è stato registrato nella regione del medio infrarosso compresa tra 925,92 e 5011,54 cm^{-1} , con una risoluzione spettrale di 3,858 cm^{-1} , tramite spettrometro MilkoScan FT6000 (FOSS). Lo spettro è ottenuto filtrando lo spettro del latte dallo spettro dell'acqua, per cui non sono presenti gli assorbimenti legati ai gruppi ossidrilici della molecola d'acqua.

L'analisi dello spettro FT-IR, così come tutte le elaborazioni inerenti la calibrazione del modello, è stata eseguita sulla base degli spettri dei campioni inseriti nel gruppo del training set. In Figura 14.1 è mostrata la sovrapposizione degli spettri FT-IR dei campioni del training set. Le regioni a numeri d'onda superiori a circa 3000 cm^{-1} sono caratterizzate da un elevato rumore strumentale (zona tra 3050 e 3850 cm^{-1} , comunque non caratteristica di assorbimenti di gruppi funzionali tipici dei trigliceridi) o dalla sostanziale assenza di assorbimenti infrarossi (zona tra 3850 e 5012 cm^{-1}). Gli assorbimenti caratteristici dei gruppi funzionali dei trigliceridi sono per contro compresi nella zona tra 926 e 3050 cm^{-1} . In Figura 14.2 è mostrata la porzione di spettro considerata.

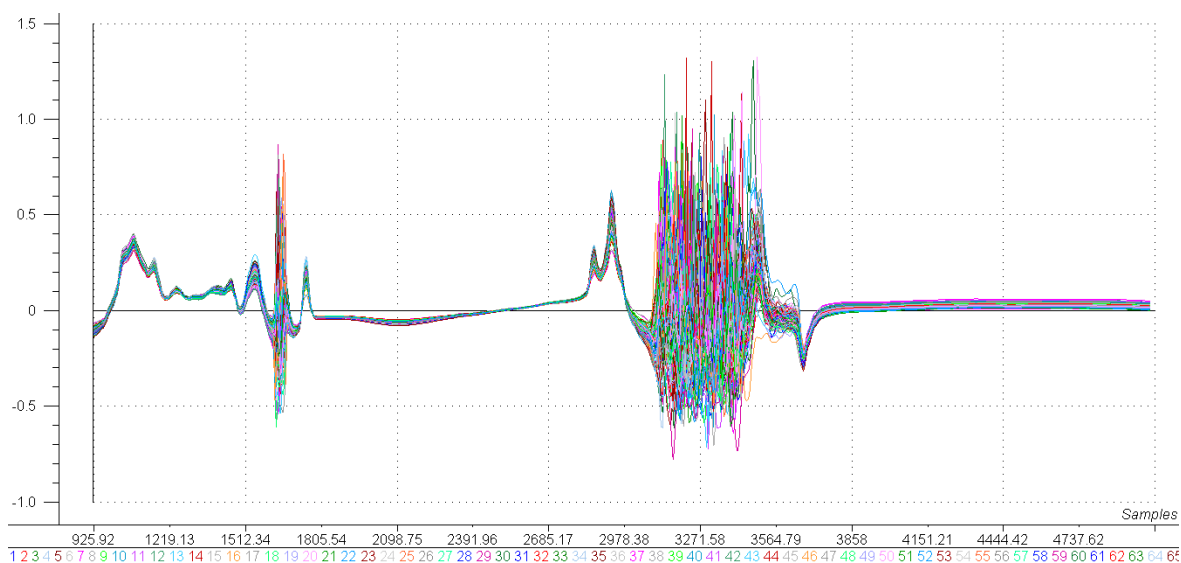


Figura 14.1 – Spettri sovrapposti dei 145 campioni di latte ovino inseriti nel training set (assorbanza vs numeri d'onda)

Nella zona dell'impronta digitale dello spettro (regione al di sotto di 1500 cm^{-1}), sono presenti numerosi assorbimenti IR: il picco di assorbimento nella regione compresa tra 1000 e 1050 cm^{-1} può essere attribuibile allo stiramento asimmetrico dei legami P-O-C presenti nel latte nei fosfolipidi e nelle fosfoproteine; la regione compresa tra 1050 e 1200 cm^{-1} è caratteristica dello stiramento del legame C-O, (maggiormente presente nel latte in molecole quali i trigliceridi ed il lattosio) ed è possibile attribuire i due segnali presenti agli stiramenti simmetrici ed asimmetrici dei legami C-O-C esterei; i due picchi poco intensi compresi nella regione tra 1200 e 1350 cm^{-1} sono attribuibili alle vibrazioni di wagging e twisting dei C-H metilenici mentre i due picchi successivi (regione tra 1350 e 1475 cm^{-1}) corrispondono ai piegamenti simmetrici ed asimmetrici dei C-H metilici ed al piegamento di scissoring dei C-H metilenici. La regione compresa tra 1250 e 1300 cm^{-1} è anche caratteristica dei segnali di stiramento del doppio legame fosforo-ossigeno, mentre la regione intorno ai 1400 cm^{-1} è anche caratteristica della vibrazione di rocking dei C-H di alcheni *cis*-disostituiti.

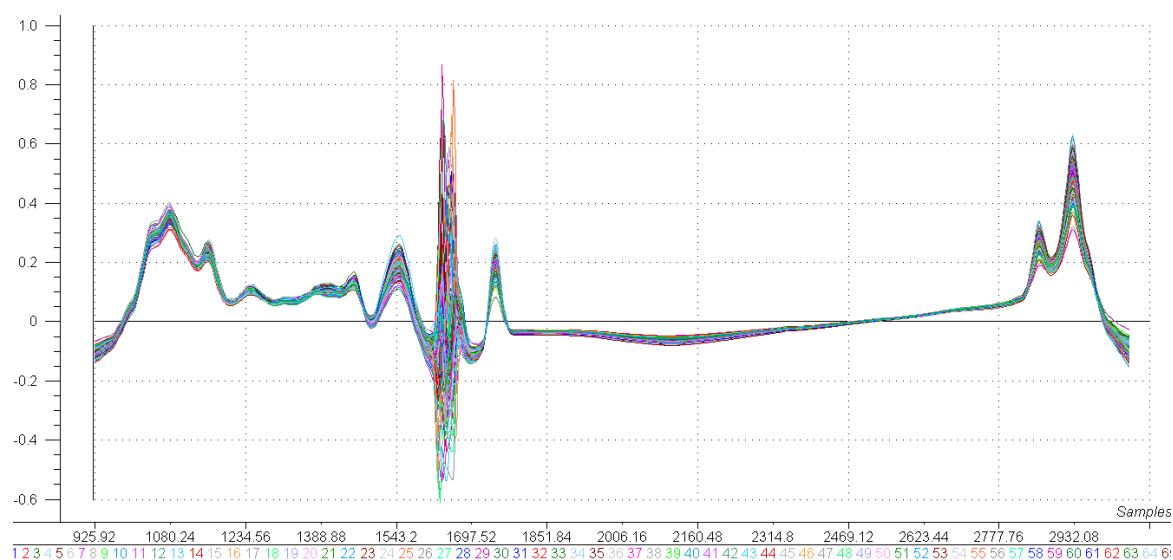


Figura 14.2 – Regione spettrale caratteristica degli assorbimenti dei gruppi funzionali dei trigliceridi

Il picco presente nella regione compresa tra 1500 e 1600 cm^{-1} , non attribuibile ad assorbimenti di gruppi funzionali dei trigliceridi, è caratteristica del piegamento del legame N-H delle ammidi e dello stiramento dei legami C-N ammidici. Il picco è quindi legato al contenuto proteico e di urea del latte.

La regione spettrale compresa tra 1600 e 1700 cm^{-1} presenta elevato rumore strumentale. Questa regione è caratteristica principalmente degli stiramenti dei legami carbonilici dei composti ammidici.

Tra 1730 e 1770 cm^{-1} è presente un picco attribuibile allo stiramento dei carbonili esterei, segnale quindi attribuibile ai trigliceridi.

La regione tra 1800 e 2800 cm^{-1} , povera in genere di assorbimenti infrarossi significativi, non presenta segnali IR riconducibili a gruppi funzionali presenti nei trigliceridi.

Nella regione tra 2800 e 3000 cm^{-1} sono invece presenti due intensi picchi di assorbimento, attribuibili alla vibrazione di stiramento simmetrica ed asimmetrica dei legami C-H dei gruppi metilici e metilenici.

Non è invece possibile attribuire segnali IR significativi riconducibili a vibrazioni legate alla presenza di insaturazioni alcheniche: in particolare, la regione della vibrazione dei piegamenti C-H su carboni ibridizzati sp^2 in alcheni in configurazione *cis* (tra 675 e 730 cm^{-1}) è al di fuori dell'intervallo spettrale considerato, mentre lo stesso tipo di vibrazione presente negli alcheni in configurazione *trans* (tra 900 e 1000 cm^{-1}) non evidenzia segnali di rilievo. Allo stesso modo, la regione degli stiramenti dei C-H olefinici (tra 3000 e 3050 cm^{-1}) non presenta segnali distinguibili. La regione tra 1600 e 1700 cm^{-1} , caratterizzata da elevato rumore strumentale, è anche caratteristica degli stiramenti dei doppi legami C=C in configurazione *cis* e *trans* (rispettivamente tra 1620 e 1660 cm^{-1} e 1670 e 1680 cm^{-1}), che non sono quindi identificabili.

14.2 Costruzione dei modelli di predizione

I modelli di predizione basati sull'assorbimento dei gruppi funzionali presenti nelle molecole sono stati elaborati per i seguenti acidi grassi: C4:0, C6:0, C8:0, C10:0, C12:0, C14:0, C16:0, C18:0, C18:1 *trans*-11, C18:1 *cis*-9, C18:2 *cis*-9, *cis*-12, C18:3 *cis*-9, *cis*-12, *cis*-15, C18:2 *cis*-9, *trans*-11 e per le seguenti classi di acidi grassi: Acidi Grassi Saturi (SFA), Acidi Grassi Insaturi (UFA), Acidi Grassi Monoinsaturi (MUFA), Acidi Grassi Polinsaturi (PUFA), Acidi Grassi Omega 6 (PUFA ω -6), ed Acidi Grassi Omega 3 (PUFA ω -3).

Tutte le elaborazioni effettuate per ogni variabile da predire sono state eseguite secondo i seguenti passaggi:

- Fase di calibrazione: costruzione di un modello di regressione ottenuto tramite PLS (regressione parziale con minimi quadrati) utilizzando i 145 campioni inseriti nel training set (caratterizzati dagli spettri FT-IR e dai valori di riferimento della variabile da predire). Il modello di regressione viene validato internamente tramite la Cross-Validazione Leave More Out - vedi paragrafo 8.3.4 - (riportata con la sigla CV) utilizzando 5 gruppi di cancellazione ed un numero massimo di variabili latenti pari a 15. Sulla matrice dei predittori (spettri FT-IR), è stata effettuata l'operazione di autoscaling (vedi paragrafo 8.2.2).
- Fase di validazione: il modello di regressione costruito nella fase di calibrazione è validato utilizzando i 97 campioni inseriti nel test set. Il numero di variabili latenti utilizzate per la predizione sul set esterno viene deciso nella fase di calibrazione, in base al valore minimo del RMSECV (Root Mean Square Error in Cross Validation). Anche la matrice dei predittori del test set ha subito l'operazione di autoscaling (sulla base dei parametri del training set).

Le variazioni effettuate nelle diverse elaborazioni eseguite per ogni composto da predire riguardano la selezione delle variabili spettrali, allo scopo di ricercare quelle maggiormente informative e quindi più adatte alla predizione di ogni singolo composto.

Le elaborazioni sono state effettuate tramite il software Matlab versione 4.2 (The MathWorks, Inc.) e Matlab versione 6.0 (The MathWorks, Inc.).

14.2.1 Costruzione dei modelli di predizione con l'utilizzo dello spettro FT-IR intero

Il primo passo effettuato consiste nella costruzione dei modelli di predizione di ogni acido grasso, utilizzando l'intero spettro FT-IR (contenente 1060 variabili) come matrice dei predittori.

Sono stati valutati i modelli di predizione basati sull'adozione dell'intero spettro FT-IR sia a partire dai dati originali che a partire dagli spettri pretrattati con operazioni matematiche. I pretrattamenti spettrali valutati sono lo Smoothing, la Multiplicative Scatter Correction (MSC), la Derivata Prima, e lo Standard Normal Variate (SNV) (vedi paragrafo 8.4.1). Nessuno dei pretrattamenti spettrali indicati ha prodotto risultati migliori di quelli ottenuti dalle elaborazioni con l'utilizzo dello spettro originale. Per questo motivo tutte le elaborazioni successivamente effettuate si basano sullo spettro FT-IR originale. Di seguito

vengono riportati esclusivamente i risultati ottenuti dalla costruzione dei modelli di predizione tramite l'utilizzo dello spettro FT-IR originale.

Tabella 14.1 – Risultati ottenuti dai modelli di predizione costruiti utilizzando l'intero spettro FT-IR

Acido grasso	Fase di Calibrazione			Fase di Validazione			
	LV	% Varianza Spiegata in CV	RMSECV	Err. % CV	RMSEP	bias	Err. %
C4:0	8	79,86	0,019	8,09	0,026	-0,002	11,06
C6:0	8	82,58	0,012	7,23	0,012	0,001	7,23
C8:0	8	78,10	0,014	9,79	0,014	0,001	9,79
C10:0	8	71,30	0,051	12,72	0,051	0,002	12,72
C12:0	7	71,01	0,028	13,40	0,027	-0,001	12,92
C14:0	7	83,79	0,049	9,40	0,056	-0,002	10,75
C16:0	8	85,01	0,092	8,19	0,105	-0,008	9,34
C18:0	7	70,19	0,081	15,85	0,083	0,013	16,24
C18:1 <i>trans</i>-11	15	57,80	0,049	26,49	0,052	0,002	28,11
C18:1 <i>cis</i>-9	8	65,81	0,138	17,34	0,133	-0,010	16,71
C18:2 <i>cis</i>-9, <i>cis</i>-12	7	62,38	0,023	17,04	0,027	-0,002	20,00
C18:3 <i>cis</i>-9, <i>cis</i>-12, <i>cis</i>-15	10	58,66	0,026	26,00	0,020	-0,001	20,00
C18:2 <i>cis</i>-9, <i>trans</i>-11	15	71,80	0,020	23,26	0,024	0,001	27,91
Acidi Grassi Saturi	7	94,93	0,147	4,12	0,163	-0,002	4,57
Acidi Grassi Insaturi	7	90,43	0,132	7,12	0,142	-0,005	7,66
Acidi Grassi Monoinsaturi	6	85,39	0,131	9,59	0,139	0,001	10,18
Acidi Grassi Polinsaturi	15	83,53	0,053	10,88	0,056	-0,005	11,50
Acidi Grassi ω-6	7	69,54	0,024	13,79	0,027	-0,001	15,52
Acidi Grassi ω-3	11	55,97	0,038	23,46	0,032	-0,001	19,75

LV = variabili latenti significative determinate in Cross Validazione ed utilizzate per la predizione esterna.

% Varianza spiegata in CV = percentuale di varianza spiegata dal modello ottenuto in Cross-Validazione.

RMSECV = errore quadratico medio in Cross-Validazione.

Err. % CV = errore relativo percentuale (in Cross Validazione) = RMSECV/Media valori predetti in CV.

RMSEP = errore quadratico medio in predizione su set esterno.

bias = bias in predizione sul set esterno.

Err. % = errore relativo percentuale (in predizione su set esterno) = RMSEP/Media valori predetti in validazione esterna.

In Tabella 14.1 sono mostrati i risultati ottenuti dai modelli di predizione costruiti, per ogni singolo acido grasso e per ogni singola classe di acidi grassi, utilizzando l'intero spettro FT-IR. Nella tabella citata vengono mostrati diversi parametri: il numero di variabili latenti significative (LV), stabilite nella fase di calibrazione e utilizzate per la predizione su set esterno; la varianza percentuale spiegata dal modello ottenuto in Cross Validazione; il Root Mean Square Error ottenuto in Cross Validazione (RMSECV); l'errore relativo percentuale in Cross Validazione, dato dal rapporto percentuale tra RMSECV e la media dei valori

predetti in Cross Validazione; il Root Mean Square Error in predizione sul set esterno (RMSEP); il bias ottenuto nella predizione su set esterno; e l'errore relativo percentuale in predizione su set esterno, dato dal rapporto percentuale tra RMSEP e la media dei valori predetti per il test set.

Considerando gli acidi grassi, i valori di varianza percentuale spiegata sono compresi tra il 58% per il C18:1 *trans*-11 e l'85% per il C16:0 mentre, considerando le classi di acidi grassi, esse variano tra il 56% per gli acidi grassi ω -3 ed il 95% per gli acidi grassi saturi. Confrontando il RMSECV (ed il relativo errore percentuale sulla media) con il RMSEP (ed il relativo errore percentuale sulla media), si può vedere che le differenze tra i due parametri non sono mai rilevanti e quindi si può concludere che l'errore in predizione sul test set è analogo a quello in cross validation sul training set. Si può supporre che un miglioramento della capacità predittiva potrà essere ottenuto effettuando un'opportuna selezione delle regioni spettrali rilevanti.

14.2.2 Costruzione dei modelli di predizione con l'utilizzo dello spettro FT-IR ridotto

Il secondo step volto all'ottimizzazione del modello di predizione ha portato a prendere in considerazione solo la porzione informativa dello spettro FT-IR, defalcandolo delle regione tra i 3050 e 5012 cm^{-1} in quanto non specificatamente informativa.

La regione spettrale utilizzata contiene quindi 550 variabili. Anche in questo caso il modello è stato costruito tramite regressione PLS sul training set con validazione interna (Cross-Validazione). I risultati ottenuti sono mostrati in Tabella 14.2 (Fase di Calibrazione) dove vengono mostrati: il numero di variabili latenti significative (LV); la varianza percentuale spiegata dal modello ottenuto in Cross Validazione; il Root Mean Square Error ottenuto in Cross Validazione (RMSECV); l'errore relativo percentuale in Cross Validazione, dato dal rapporto percentuale tra RMSECV e la media dei valori predetti in Cross Validazione.

È possibile affermare che la regione esclusa non possiede informazione utile agli scopi del lavoro e la sua rimozione fa migliorare in maniera netta la fase di calibrazione, come mostrato sia dal netto incremento in percentuale di varianza spiegata in Cross Validazione, che dalla diminuzione generalizzata dei valori di RMSECV.

Il modello di calibrazione ottenuto è stato quindi validato sul set di campioni incogniti (test set). I risultati sono mostrati sempre in Tabella 14.2 (Fase di Validazione) in cui vengono riportati il Root Mean Square Error in predizione sul set esterno (RMSEP); il bias ottenuto

nella predizione su set esterno; e l'errore relativo percentuale in predizione su set esterno, dato dal rapporto percentuale tra RMSEP e la media dei valori predetti per il test set. Dal confronto con i risultati ottenuti dalla validazione del modello costruito con l'intero spettro FT-IR (Tabella 14.1), si può vedere che vi è un netto miglioramento della predizione di tutti i composti.

La validità della scelta dell'esclusione della regione spettrale indicata è quindi confermata dai risultati ottenuti nella predizione su set esterno.

Tabella 14.2 – Risultati ottenuti dai modelli di predizione costruiti utilizzando lo spettro FT-IR ridotto

Acido grasso	Fase di Calibrazione			Fase di Validazione			
	LV	% Varianza Spiegata in CV	RMSECV	Err. % CV	RMSEP	bias	Err. %
C4:0	15	88,05	0,014	5,96	0,015	-0,002	6,38
C6:0	15	90,17	0,009	5,42	0,011	0,002	6,63
C8:0	15	85,97	0,011	7,69	0,013	0,003	9,09
C10:0	15	87,56	0,034	8,48	0,040	0,007	9,98
C12:0	15	85,76	0,019	9,09	0,020	0,003	9,57
C14:0	15	89,40	0,039	7,49	0,047	0,004	9,02
C16:0	12	90,29	0,074	6,58	0,091	-0,002	8,10
C18:0	15	82,23	0,062	12,13	0,062	0,009	12,13
C18:1 <i>trans</i>-11	15	77,83	0,036	19,46	0,039	-0,005	21,08
C18:1 <i>cis</i>-9	15	83,49	0,096	12,06	0,124	-0,013	15,58
C18:2 <i>cis</i>-9, <i>cis</i>-12	14	73,27	0,019	14,07	0,024	0,001	17,78
C18:3 <i>cis</i>-9, <i>cis</i>-12, <i>cis</i>-15	7	70,72	0,022	22,00	0,020	0,000	20,00
C18:2 <i>cis</i>-9, <i>trans</i>-11	11	76,12	0,018	20,93	0,017	-0,002	19,77
Acidi Grassi Saturi	13	97,58	0,101	2,83	0,122	0,017	3,42
Acidi Grassi Insaturi	13	94,62	0,099	5,34	0,121	-0,016	6,53
Acidi Grassi Monoinsaturi	13	90,70	0,105	7,69	0,118	-0,015	8,64
Acidi Grassi Polinsaturi	15	88,78	0,044	9,03	0,042	-0,004	8,62
Acidi Grassi ω-6	14	80,54	0,019	10,92	0,025	0,002	14,37
Acidi Grassi ω-3	14	74,29	0,029	17,90	0,027	-0,001	16,67

LV = variabili latenti significative determinate in Cross Validazione ed utilizzate per la predizione esterna.

% Varianza spiegata in CV = percentuale di varianza spiegata dal modello ottenuto in Cross-Validazione.

RMSECV = errore quadratico medio in Cross-Validazione.

Err. % CV = errore relativo percentuale (in Cross Validazione) = RMSECV/Media valori predetti in CV.

RMSEP = errore quadratico medio in predizione su set esterno.

bias = bias in predizione sul set esterno.

Err. % = errore relativo percentuale (in predizione su set esterno) = RMSEP/Media valori predetti in validazione esterna.

14.2.3 Costruzione dei modelli di predizione con l'utilizzo delle bande spettrali selezionate dall'analisi visiva dello spettro

L'analisi dello spettro effettuata nel paragrafo 14.1 ha portato all'identificazione dei picchi il cui assorbimento può essere attribuibile, almeno in parte, ai gruppi funzionali dei trigliceridi.

Sono quindi stati costruiti i modelli di predizione per ogni acido grasso e classe di acidi grassi utilizzando le regioni spettrali identificate.

Per gli acidi grassi C4:0, C6:0, C8:0, C10:0, C12:0, C14:0, C16:0, C18:0, e per la classe degli acidi grassi saturi, sono state utilizzate le seguenti regioni spettrali:

- tra 1050 e 1200 cm^{-1} , stiramento simmetrico e asimmetrico dei legami C-O-C esterei;
- tra 1200 e 1350 cm^{-1} , vibrazioni di wagging e twisting dei C-H metilenici;
- tra 1350 e 1475 cm^{-1} , piegamenti simmetrici ed asimmetrici dei C-H metilici e piegamenti di scissoring dei legami C-H metilenici;
- tra 1730 e 1770 cm^{-1} , stiramento del gruppo carbonilico negli esteri;
- tra 2800 e 3000 cm^{-1} , stiramento simmetrico e asimmetrico dei legami C-H dei gruppi metilici e metilenici;

per un totale di 175 variabili spettrali.

Per gli acidi grassi C18:1 *cis*-9, C18:2 *cis*-9, *cis*-12, C18:3 *cis*-9, *cis*-12, *cis*-15 e per le classi di acidi grassi ω -6, ed ω -3 sono state utilizzate le seguenti regioni spettrali:

- tra 1050 e 1200 cm^{-1} , stiramento simmetrico e asimmetrico dei legami C-O-C esterei;
- tra 1200 e 1350 cm^{-1} , vibrazioni di wagging e twisting dei C-H metilenici;
- tra 1350 e 1475 cm^{-1} , piegamenti simmetrici ed asimmetrici dei C-H metilici e piegamenti di scissoring dei legami C-H metilenici;
- tra 1730 e 1770 cm^{-1} , stiramento del gruppo carbonilico negli esteri;
- tra 2800 e 3000 cm^{-1} , stiramento simmetrico e asimmetrico dei legami C-H dei gruppi metilici e metilenici;
- tra 3000 e 3050 cm^{-1} , stiramento dei gruppi C-H su carboni sp^2 .

per un totale di 191 variabili spettrali.

Per gli acidi grassi C18:1 *trans*-11 e C18:2 *cis*-9, *trans*-11 e per le classi di acidi grassi insaturi, monoinsaturi, polinsaturi, sono state utilizzate le seguenti regioni spettrali:

- tra 960 e 970 cm^{-1} , piegamento dei C-H in carboni ibridizzati sp^2 in configurazione *trans*;
- tra 1050 e 1200 cm^{-1} , stiramento simmetrico e asimmetrico dei legami C-O-C esterei;
- tra 1200 e 1350 cm^{-1} , vibrazioni di wagging e twisting dei C-H metilenici;
- tra 1350 e 1475 cm^{-1} , piegamenti simmetrici ed asimmetrici dei C-H metilici e piegamenti di scissoring dei legami C-H metilenici;
- tra 1730 e 1770 cm^{-1} , stiramento del gruppo carbonilico negli esteri;
- tra 2800 e 3000 cm^{-1} , stiramento simmetrico e asimmetrico dei legami C-H dei gruppi metilici e metilenici;
- tra 3000 e 3050 cm^{-1} , stiramento dei gruppi C-H su carboni sp^2 .

per un totale di 194 variabili spettrali.

Una volta selezionate le regioni spettrali ritenute significative, la loro efficacia viene valutata in termini di validazione del modello così ottenuto (Tabella 14.3).

Nella tabella citata vengono mostrati: il numero di variabili latenti significative (LV) utilizzate per la predizione su set esterno; il Root Mean Square Error per il set esterno (RMSEP); il bias ottenuto nella predizione su set esterno e l'errore relativo percentuale in predizione su set esterno, dato dal rapporto percentuale tra RMSEP e la media dei valori predetti per il test set (tale media è identica alla media dei valori di riferimento).

Confrontando i dati della Tabella 14.3 con i dati in precedenza ottenuti nella validazione dei modelli costruiti utilizzando lo spettro FT-IR ridotto costituito da 550 variabili (Tabella 14.2), è possibile notare che l'efficacia della predizione è rimasta invariata o, in alcuni casi, è migliorata. Il miglioramento è maggiormente visibile per alcune classi, quali quelle degli acidi grassi saturi, insaturi e monoinsaturi.

La selezione delle bande tramite analisi visiva dello spettro FT-IR ha permesso di introdurre nei modelli solo le variabili spettrali più informative, eliminando nel contempo le regioni spettrali rumorose o poco informative. In tal modo si è notevolmente ridotto il numero di variabili spettrali, alleggerendo le dimensioni dei modelli di predizione.

Tabella 14.3 - Risultati ottenuti dai modelli di predizione costruiti utilizzando le bande spettrali selezionate visivamente dallo spettro FT-IR

Acido grasso	Fase di Validazione			
	LV	RMSEP	bias	Err. %
C4:0	14	0,014	-0,002	5,96
C6:0	15	0,010	0,001	6,02
C8:0	15	0,011	0,002	7,69
C10:0	15	0,035	0,006	8,73
C12:0	15	0,019	0,003	9,09
C14:0	13	0,044	0,006	8,45
C16:0	14	0,086	-0,007	7,65
C18:0	12	0,064	0,006	12,52
C18:1 <i>trans</i>-11	14	0,037	-0,003	20,00
C18:1 <i>cis</i>-9	12	0,112	-0,015	14,07
C18:2 <i>cis</i>-9, <i>cis</i>-12	15	0,024	0,001	17,78
C18:3 <i>cis</i>-9, <i>cis</i>-12, <i>cis</i>-15	15	0,019	-0,001	19,00
C18:2 <i>cis</i>-9, <i>trans</i>-11	12	0,016	-0,003	18,60
Acidi Grassi Saturi	13	0,092	0,012	2,58
Acidi Grassi Insaturi	15	0,099	-0,016	5,34
Acidi Grassi Monoinsaturi	15	0,098	-0,014	7,17
Acidi Grassi Polinsaturi	15	0,038	0,000	7,80
Acidi Grassi ω-6	15	0,024	0,002	13,79
Acidi Grassi ω-3	14	0,028	0,000	17,28

LV = variabili latenti significative utilizzate per la predizione esterna.

RMSEP = errore quadratico medio in predizione su set esterno.

bias = bias in predizione sul set esterno.

Err. % = errore relativo percentuale (in predizione su set esterno)

= RMSEP/Media valori predetti in validazione esterna.

14.2.4 Costruzione dei modelli di predizione con l'utilizzo delle bande spettrali selezionate dagli algoritmi genetici

Per ottenere una selezione delle regioni spettrali specifiche per ogni singolo acido grasso e classe di acidi grassi, è stata utilizzata la tecnica di selezione delle variabili tramite algoritmi genetici, precedentemente introdotta nel capitolo 9. Gli algoritmi genetici permettono di selezionare per ogni singolo composto le variabili spettrali realmente informative, eliminando le variabili legate a rumore o ad informazione differente da quella utile per la predizione. In questo modo viene drasticamente ridotto il rischio di overfitting e vengono costruiti modelli di predizione robusti, affidabili e semplici.

14.2.4.1 Selezione delle bande tramite algoritmi genetici

Come spiegato nel paragrafo 9.6.2, le performance degli algoritmi genetici sono ottimali quando il numero di variabili di partenza è minore di 200. Si è quindi deciso di effettuare una riduzione delle variabili spettrali tramite l'operazione Averaging (vedi paragrafo 8.4.1).

L'eliminazione della regione spettrale compresa tra 3050 e 5012 cm^{-1} effettuata precedentemente (vedi paragrafo 14.2.2), ha permesso di effettuare la riduzione - tramite l'operazione Averaging - a partire da uno spettro FT-IR contenente 550 variabili spettrali anziché le 1060 che costituiscono lo spettro originale. In questo modo si è potuto effettuare una riduzione dello spettro su una finestra di 3 variabili adiacenti, ottenendo in totale 183 variabili spettrali, utilizzate dagli algoritmi genetici come variabili di partenza.

Test di Randomizzazione

Il Test di Randomizzazione, preliminarmente descritto nel paragrafo 9.6.2, permette di valutare, per ogni variabile da predire, l'efficacia dell'algoritmo genetico nella selezione delle variabili spettrali del set di dati in possesso, e di stabilire il numero di valutazioni ottimali da far effettuare all'algoritmo genetico in ogni sua corsa. Il Test di Randomizzazione è stato eseguito sul training set.

Le condizioni di settaggio utilizzate per il Test di Randomizzazione sono le seguenti:

- Grandezza della popolazione: 30 cromosomi;
- Numero di variabili selezionate per ogni cromosoma della popolazione iniziale: 5;
- Metodo di regressione: PLS;
- Responso: Varianza percentuale espressa in Cross Validazione (5 gruppi di cancellazione e numero di variabili latenti determinate tramite Cross Validazione);
- Massimo numero di variabili selezionate nello stesso cromosoma: 30;
- Probabilità di Cross-Over: 50%;
- Probabilità di mutazione: 1%;
- Massimo numero di variabili latenti considerate: 15;
- Numero di corse: 100;
- Numero di valutazioni: 200 per ogni corsa;
- Pretrattamento: nessuno;
- Scaling: autoscaling.

Riguardo l'efficacia dell'algoritmo genetico nella selezione delle variabili spettrali del set di dati in possesso, il Test di Randomizzazione ha prodotto valori inferiori a 2,5 per tutti gli acidi grassi e per tutte le classi di acidi grassi. Tali valori, confrontati col valore soglia (pari a 4), stabiliscono che il data set in nostro possesso è informativo e l'algoritmo genetico è capace di modellare l'informazione.

I risultati del Test di Randomizzazione utilizzato per stabilire il numero di valutazioni ottimali da far effettuare all'algoritmo genetico in ogni sua corsa, sono mostrati in Tabella 14.4.

Tabella 14.4 – Risultati del Test di Randomizzazione

Acido grasso	Numero di valutazioni	Acido grasso	Numero di valutazioni
C4:0	200	C18:2 <i>cis</i>-9, <i>cis</i>-12	200
C6:0	150	C18:3 <i>cis</i>-9, <i>cis</i>-12, <i>cis</i>-15	200
C8:0	200	C18:2 <i>cis</i>-9, <i>trans</i>-11	200
C10:0	200	Acidi Grassi Saturi	140
C12:0	200	Acidi Grassi Insaturi	200
C14:0	200	Acidi Grassi Monoinsaturi	190
C16:0	170	Acidi Grassi Polinsaturi	200
C18:0	200	Acidi Grassi ω-6	200
C18:1 <i>trans</i>-11	200	Acidi Grassi ω-3	200
C18:1 <i>cis</i>-9	200		

Selezione delle variabili spettrali

Stabilito il numero di valutazioni da far effettuare all'algoritmo in ogni sua corsa, l'algoritmo genetico è stato utilizzato per la selezione delle variabili spettrali più adatte alla predizione di ogni variabile in esame. La selezione delle variabili spettrali è stata effettuata in base ai dati dei campioni del training set.

Le condizioni di settaggio utilizzate sono le seguenti:

- Grandezza della popolazione: 30 cromosomi;
- Numero di variabili selezionate per ogni cromosoma della popolazione iniziale: 5;
- Metodo di regressione: PLS;
- Responso: Varianza percentuale espressa in cross-validazione (5 gruppi di cancellazione e numero di variabili latenti determinate tramite cross-validazione);

- Massimo numero di variabili selezionate nello stesso cromosoma: 30;
- Probabilità di Cross-Over: 50%;
- Probabilità di mutazione: 1%;
- Massimo numero di variabili latenti considerate: 15;
- Numero di corse: 100;
- Numero di valutazioni: differente per ogni variabile risposta, in base ai risultati del Test di Randomizzazione (Tabella 14.4);
- *Backward elimination* ogni 100 valutazioni ed alla fine del processo (se il numero di valutazioni non è un multiplo di 100);
- Pretrattamento: nessuno;
- Scaling: autoscaling.

Per il C4:0, l'algoritmo genetico ha selezionato le variabili 10-12, 19-29 e 44-46 (per un totale di 17 variabili su 183, Figura 14.3), corrispondenti alle regioni spettrali tra 1030 e 1061 cm^{-1} , tra 1134 e 1258 cm^{-1} , e tra 1424 e 1454 cm^{-1} . Nelle regioni selezionate sono presenti segnali attribuibili allo stiramento asimmetrico dei legami C-O-C esterei, ed al piegamento asimmetrico dei legami C-H metilici. La prima regione è invece attribuibile allo stiramento dei legami P-O-C.

Per il C6:0, l'algoritmo genetico ha selezionato le variabili 16-23, 71-73 e 173 (per un totale di 12 variabili su 183, Figura 14.4), corrispondenti alle regioni spettrali tra 1100 e 1188 cm^{-1} , tra 1736 e 1767 cm^{-1} , e tra 2917 e 2924 cm^{-1} . Nelle regioni selezionate sono presenti picchi di assorbimento attribuibili allo stiramento asimmetrico dei legami C-O-C esterei, allo stiramento del gruppo carbonilico negli esteri, ed allo stiramento asimmetrico dei legami C-H dei gruppi metilici e metilenici.

Per il C8:0, l'algoritmo genetico ha selezionato le variabili 16-23, 71-73 e 172-173 (per un totale di 13 variabili su 183, Figura 14.5), corrispondenti alle regioni spettrali tra 1100 e 1188 cm^{-1} , tra 1736 e 1767 cm^{-1} , e tra 2905 e 2924 cm^{-1} . Nelle regioni selezionate sono presenti picchi di assorbimento attribuibili allo stiramento asimmetrico dei legami C-O-C esterei, allo stiramento del gruppo carbonilico negli esteri, ed allo stiramento asimmetrico dei legami C-H dei gruppi metilici e metilenici.

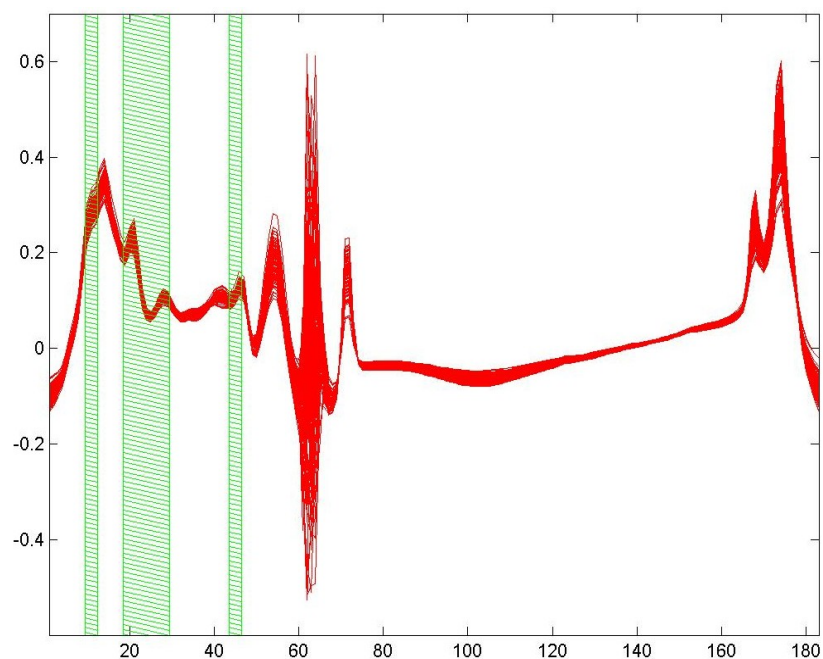


Figura 14.3 – Regioni spettrali selezionate (ombreggiate) dall’algoritmo genetico per la predizione del C4:0

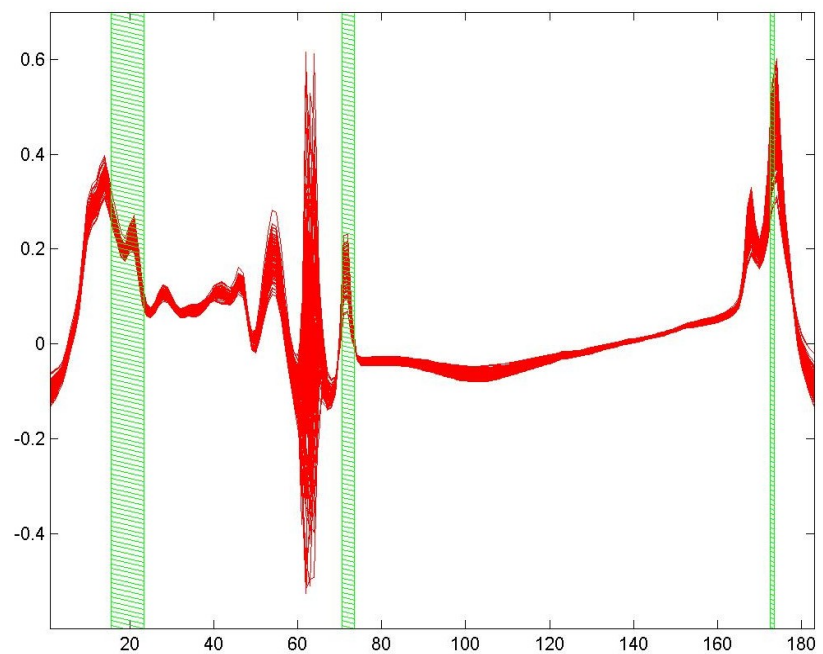


Figura 14.4 – Regioni spettrali selezionate (ombreggiate) dall’algoritmo genetico per la predizione del C6:0

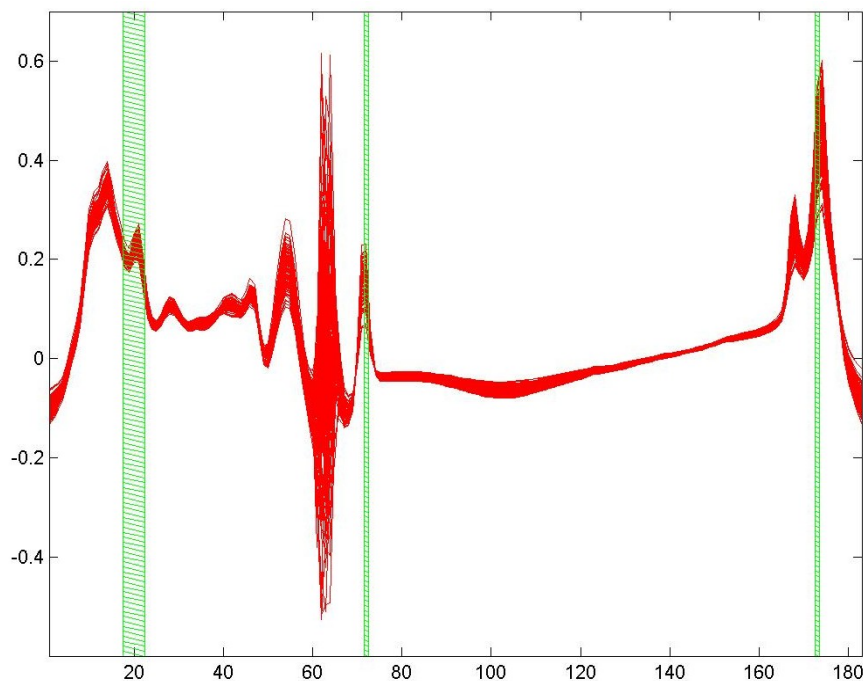


Figura 14.5 – Regioni spettrali selezionate (ombreggiate) dall’algoritmo genetico per la predizione del C8:0

Per il C10:0, l’algoritmo genetico ha selezionato le variabili 15-27, 70-73 e 166-177 (per un totale di 29 variabili su 183, Figura 14.6), corrispondenti alle regioni spettrali tra 1088 e 1235 cm^{-1} , tra 1725 e 1767 cm^{-1} , e tra 2836 e 2971 cm^{-1} . Nelle regioni selezionate sono presenti picchi di assorbimento attribuibili allo stiramento dei legami C-O-C e del gruppo carbonilico negli esteri, ed allo stiramento simmetrico ed asimmetrico dei legami C-H dei gruppi metilici e metilenici.

Per il C12:0, l’algoritmo genetico ha selezionato le variabili 14-27, 70-72 e 170-177 (per un totale di 25 variabili su 183, Figura 14.7), corrispondenti alle regioni spettrali tra 1076 e 1235 cm^{-1} , tra 1725 e 1755 cm^{-1} , e tra 2882 e 2971 cm^{-1} . Nelle regioni selezionate sono presenti assorbimenti attribuibili allo stiramento dei legami C-O-C esterei e a vibrazioni di wagging e twisting dei C-H metilenici; allo stiramento del gruppo carbonilico negli esteri; ed allo stiramento asimmetrico dei legami C-H dei gruppi metilici e metilenici.

Per il C14:0, l’algoritmo genetico ha selezionato le variabili 3-5, 14-27, 165-167 e 181-182 (per un totale di 22 variabili su 183, Figura 14.8), corrispondenti alle regioni spettrali tra 949 e 980 cm^{-1} , tra 1076 e 1235 cm^{-1} , tra 2824 e 2855 cm^{-1} , e tra 3009 e 3029 cm^{-1} . Nelle regioni selezionate sono presenti segnali attribuibili allo stiramento dei legami C-O-C esterei, alle vibrazioni di wagging e twisting dei C-H metilenici, ed allo stiramento simmetrico dei legami C-H metilici e metilenici. La prima regione selezionata è invece

riconducibile alle vibrazioni di piegamento dei gruppi C-H su carboni sp^2 in configurazione *trans*, mentre l'ultima regione è caratteristica degli stiramenti dei C-H alchenilici.

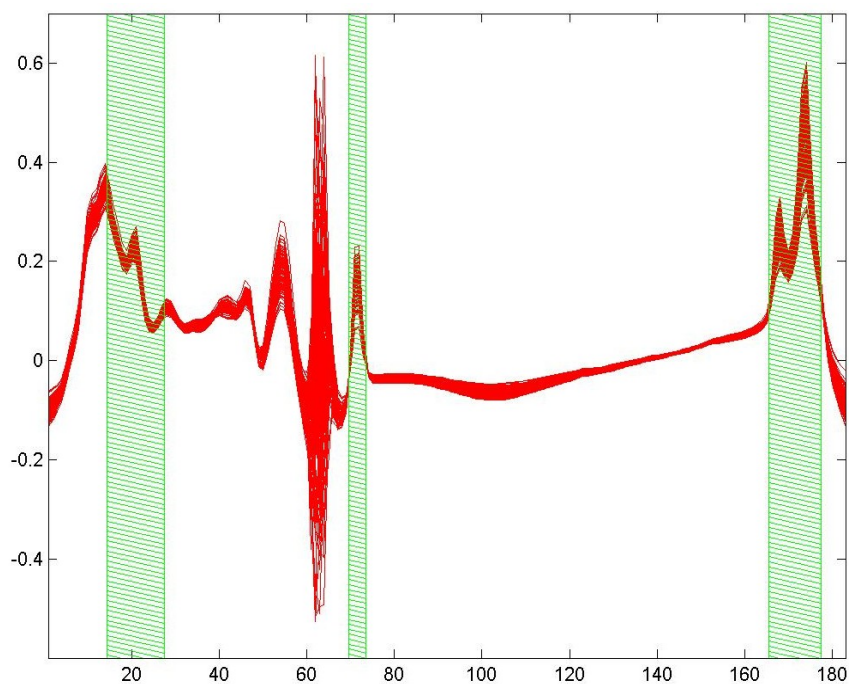


Figura 14.6 – Regioni spettrali selezionate (ombreggiate) dall'algoritmo genetico per la predizione del C10:0

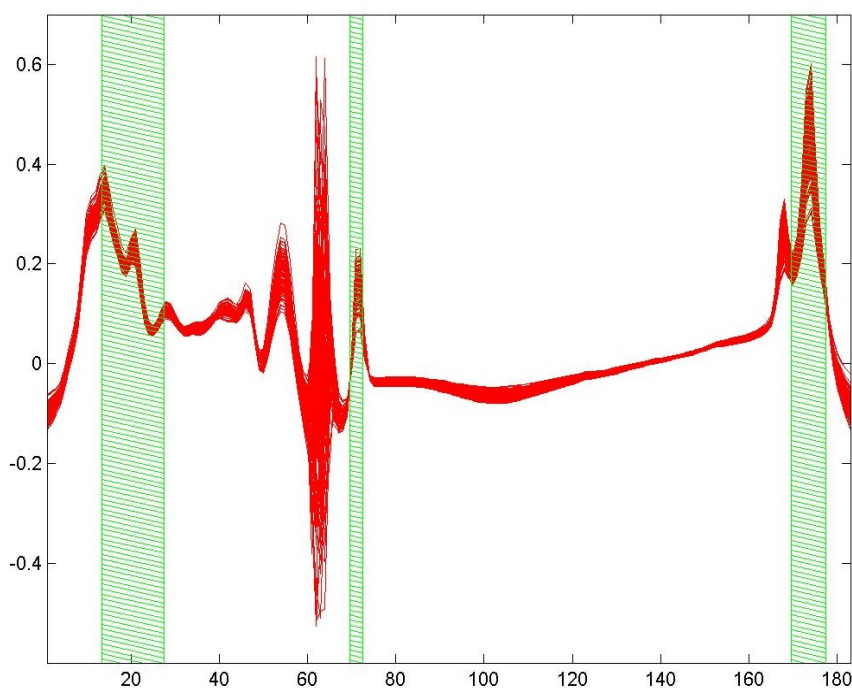


Figura 14.7 – Regioni spettrali selezionate (ombreggiate) dall'algoritmo genetico per la predizione del C12:0

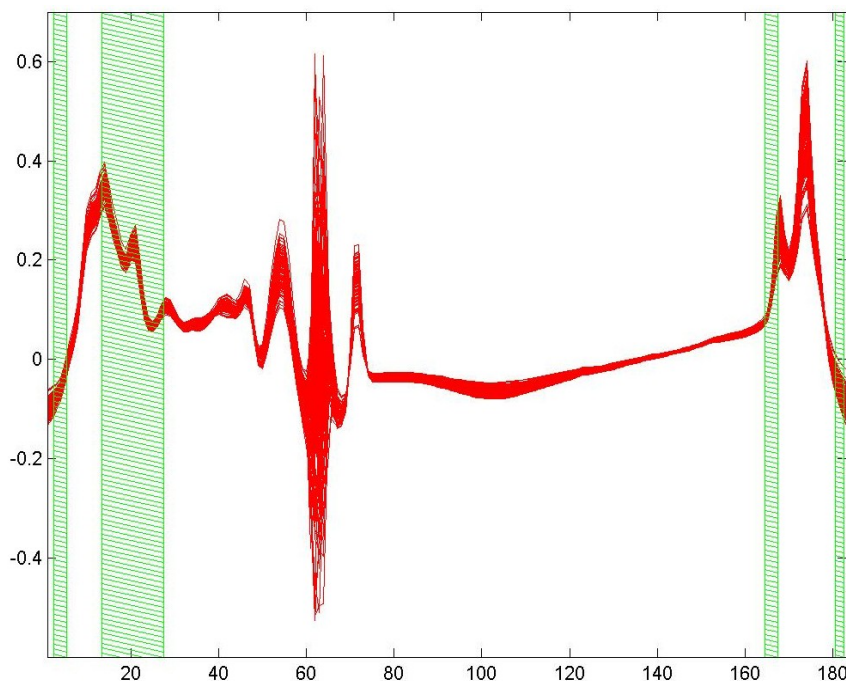


Figura 14.8 – Regioni spettrali selezionate (ombreggiate) dall’algoritmo genetico per la predizione del C14:0

Per il C16:0, l’algoritmo genetico ha selezionato le variabili 2-6, 165-167 e 172-174 (per un totale di 11 variabili su 183, Figura 14.9), corrispondenti alle regioni spettrali tra 937 e 992 cm^{-1} , tra 2824 e 2855 cm^{-1} , e tra 2905 e 2936 cm^{-1} . Nelle regioni selezionate sono presenti picchi di assorbimento attribuibili allo stiramento simmetrico ed asimmetrico dei legami C-H dei gruppi metilici e metilenici. La prima regione selezionata è invece caratteristica dei piegamenti dei C-H in carboni sp^2 in configurazione *trans*.

Per il C18:0, l’algoritmo genetico ha selezionato le variabili 23-29, 69-73, 166-174 (per un totale di 21 variabili su 183, Figura 14.10), corrispondenti alle regioni spettrali tra 1181 e 1258 cm^{-1} , tra 1713 e 1767 cm^{-1} , e tra 2836 e 2936 cm^{-1} . Nelle regioni selezionate sono presenti assorbimenti attribuibili allo stiramento asimmetrico dei legami C-O-C esterei, alle vibrazioni di wagging e twisting dei C-H metilenici, allo stiramento del gruppo carbonilico negli esteri, ed allo stiramento simmetrico ed asimmetrico dei legami C-H dei gruppi metilici e metilenici.

Per il C18:1 *trans*-11, l’algoritmo genetico ha selezionato le variabili 1-18, 28-31, 134-136, 177-182 (per un totale di 31 variabili su 183, Figura 14.11), corrispondenti alle regioni spettrali tra 926 e 1130 cm^{-1} , tra 1238 e 1281 cm^{-1} , tra 2465 e 2496 cm^{-1} , e tra 2963 e 3029 cm^{-1} . Nelle regioni selezionate sono presenti assorbimenti attribuibili ai piegamenti dei C-

H su carboni sp^2 in configurazione *trans*, agli stiramenti asimmetrici dei legami C-O-P, agli stiramenti simmetrici dei legami C-O-C estereo ed agli stiramenti dei legami C-H su carboni sp^2 . E' da notare che in questo caso l' algoritmo genetico ha selezionato anche la regione compresa tra 2465 e 2496 cm^{-1} , che visivamente non presenta picchi di assorbimento IR, e non è attribuibile a nessun assorbimento caratteristico dei trigliceridi. Per il C18:1 *cis*-9, l' algoritmo genetico ha selezionato le variabili 3-5, 20-28, 166-167, 176-178 (per un totale di 17 variabili su 183, Figura 14.12), corrispondenti alle regioni spettrali tra 949 e 980 cm^{-1} , tra 1146 e 1246 cm^{-1} , tra 2836 e 2855 cm^{-1} , e tra 2951 e 2982 cm^{-1} . Nelle regioni selezionate sono presenti assorbimenti attribuibili ai piegamenti dei C-H in carboni sp^2 in configurazione *trans*, agli stiramenti asimmetrici dei legami C-O-C esterei, alle vibrazioni di wagging e twisting dei C-H metilenici, ed agli stiramenti simmetrici ed asimmetrici dei C-H su metili e metileni.

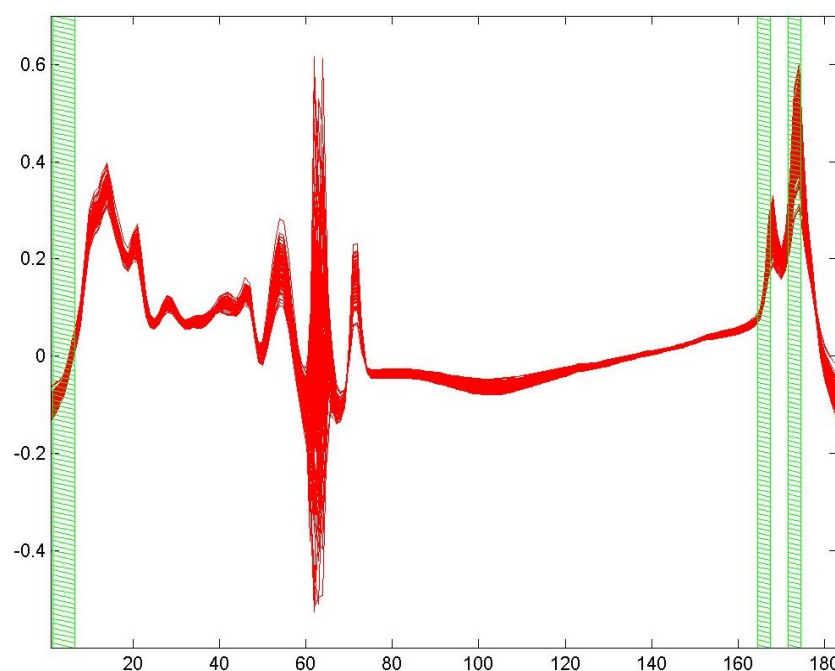


Figura 14.9 – Regioni spettrali selezionate (ombreggiate) dall' algoritmo genetico per la predizione del C16:0

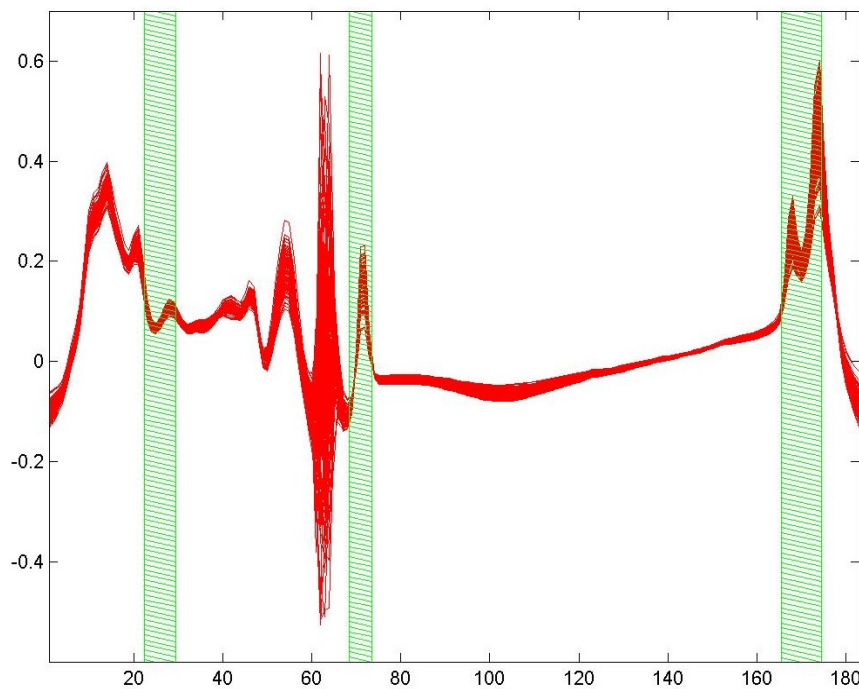


Figura 14.10 – Regioni spettrali selezionate (ombreggiate) dall’algoritmo genetico per la predizione del C18:0

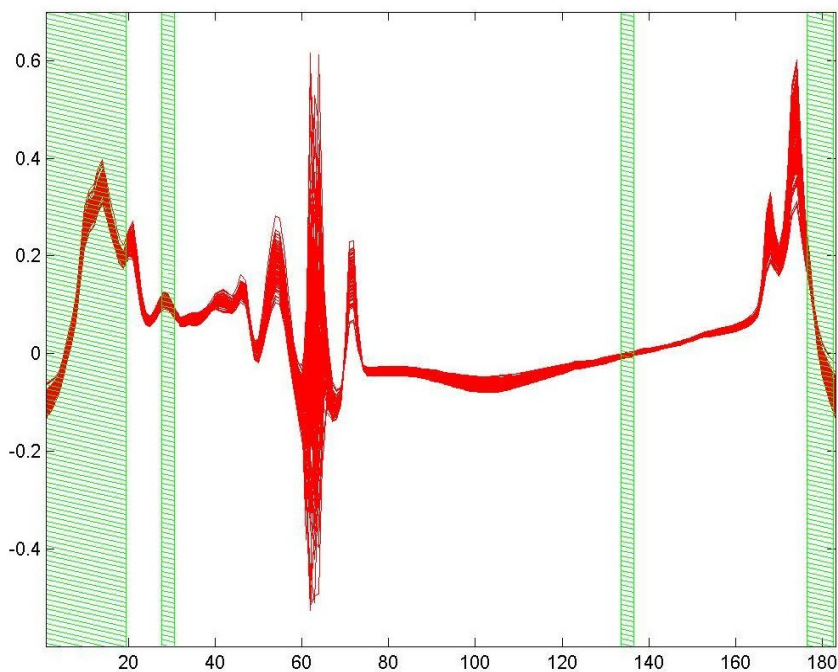


Figura 14.11 – Regioni spettrali selezionate (ombreggiate) dall’algoritmo genetico per la predizione del C18:1 *trans*-11

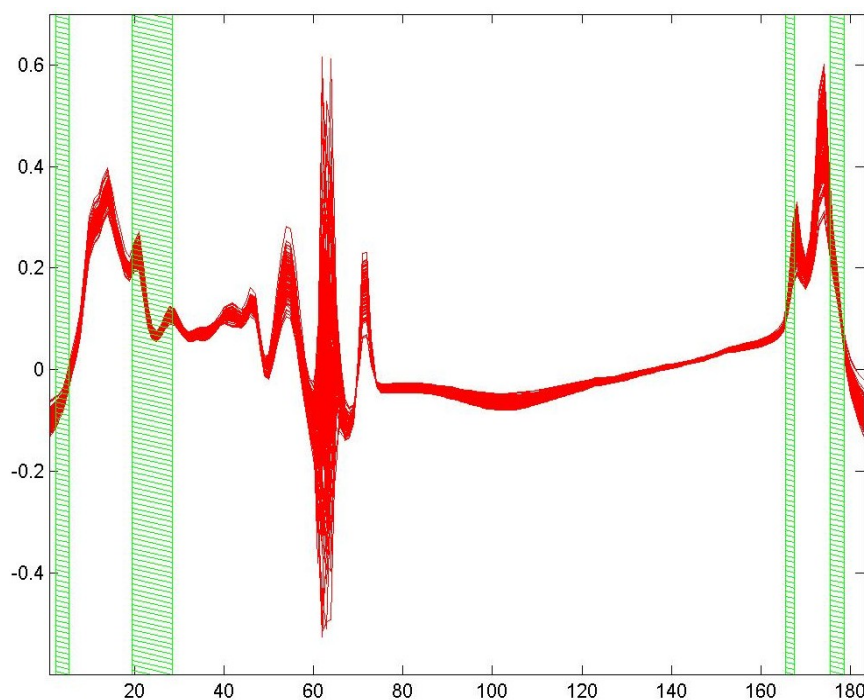


Figura 14.12 – Regioni spettrali selezionate (ombreggiate) dall’algoritmo genetico per la predizione del C18:1 *cis*-9

Per il C18:2 *cis*-9, *cis*-12 l’algoritmo genetico ha selezionato le variabili 2-3, 6-9, 17-20, 90-91, 97-101, 110-124, 144-146, 166-174, 180-181 (per un totale di 46 variabili su 183, Figura 14.13), corrispondenti alle regioni spettrali tra 937 e 957 cm^{-1} , tra 984 e 1026 cm^{-1} , tra 1111 e 1154 cm^{-1} , tra 1956 e 1975 cm^{-1} , tra 2037 e 2091 cm^{-1} , tra 2187 e 2357 cm^{-1} , tra 2581 e 2612 cm^{-1} , tra 2836 e 2936 cm^{-1} , e tra 2998 e 3017 cm^{-1} . Nelle regioni selezionate sono presenti assorbimenti attribuibili a stiramenti asimmetrici di legami C-O-C esterei, di legami C-H su metili e metileni (simmetrici ed asimmetrici), e di legami C-H su carboni ibridizzati sp^2 . La prime due regioni selezionate sono invece caratteristiche dei piegamenti dei C-H su carboni sp^2 in alcheni in configurazione *trans* e degli stiramenti dei P-O-C. Anche in questo caso l’algoritmo seleziona porzioni tangibili di spettro (tra 1956 e 1975 cm^{-1} , tra 2037 e 2091 cm^{-1} , tra 2187 e 2357 cm^{-1} e tra 2581 e 2612 cm^{-1}) che non appaiono esser immediatamente riconducibili ad assorbimenti relativi a gruppi funzionali presenti su trigliceridi.

Per il C18:3 *cis*-9, *cis*-12, *cis*-15 l’algoritmo genetico ha selezionato le variabili 77-79, 90-91, 122-124 (per un totale di 8 variabili su 183, Figura 14.14), corrispondenti alle regioni spettrali tra 1806 e 1836 cm^{-1} , tra 1956 e 1975 cm^{-1} , tra 2326 e 2357 cm^{-1} . Nessuna di tali

regioni pare esser caratteristica di assorbimenti riconducibili a gruppi funzionali dei trigliceridi.

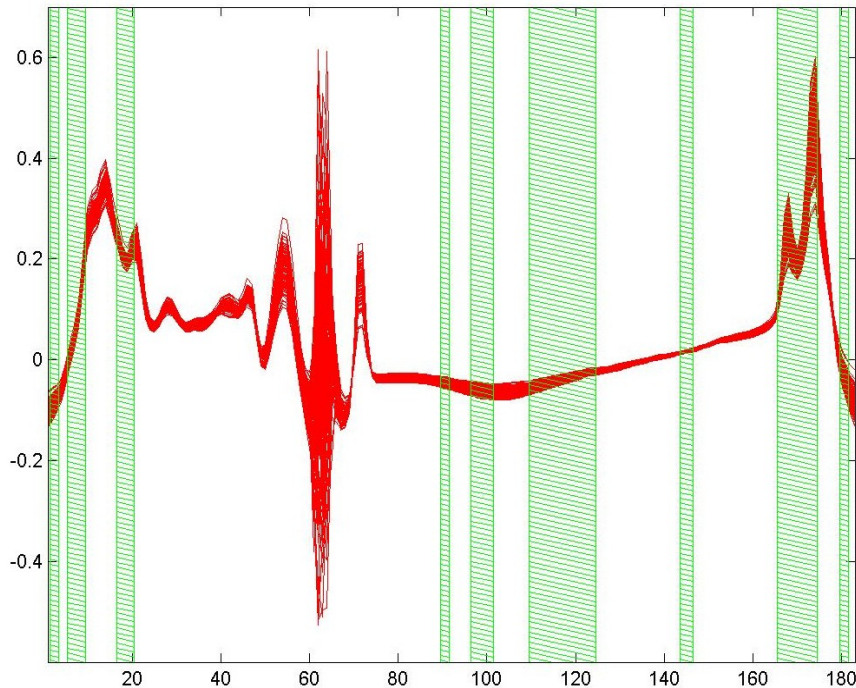


Figura 14.13 – Regioni spettrali selezionate (ombreggiate) dall’algoritmo genetico per la predizione del C18:2 *cis*-9, *cis*-12

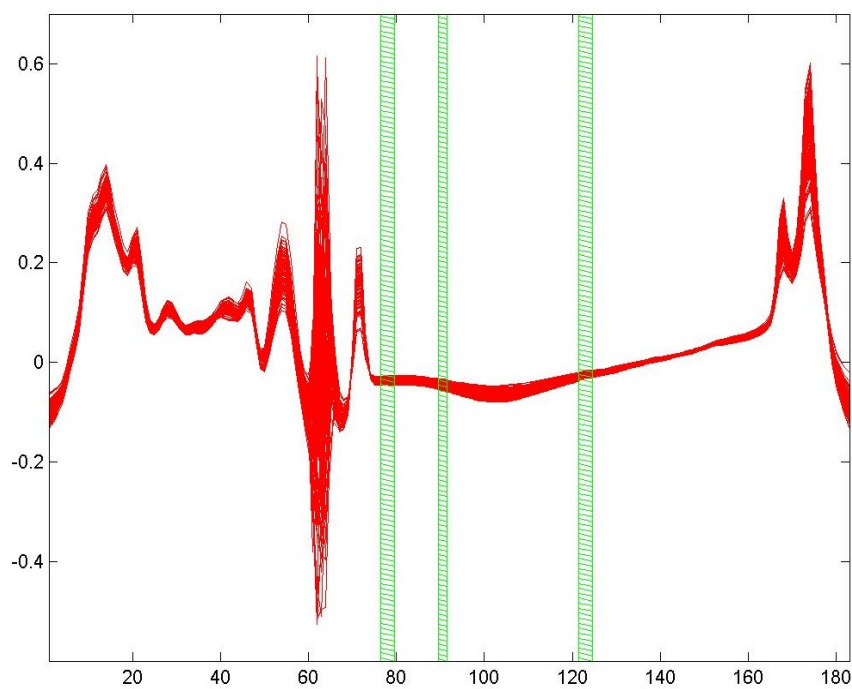


Figura 14.14 – Regioni spettrali selezionate (ombreggiate) dall’algoritmo genetico per la predizione del C18:3 *cis*-9, *cis*-12, *cis*-15

Per il C18:2 *cis*-9, *trans*-11 l'algoritmo genetico ha selezionato le variabili 3-6, 15-16, 28-30, 124, 178-182 (per un totale di 15 variabili su 183, Figura 14.15), corrispondenti alle regioni spettrali tra 949 e 992 cm^{-1} , tra 1088 e 1107 cm^{-1} , tra 1238 e 1269 cm^{-1} , tra 2350 e 2357 cm^{-1} , tra 2975 e 3029 cm^{-1} . Nelle regioni selezionate sono presenti assorbimenti attribuibili ai piegamenti dei C-H in carboni sp^2 in configurazione *trans*, agli stiramenti simmetrici dei legami C-O-C esterei, alle vibrazioni di wagging e twisting dei C-H metilenici ed agli stiramenti dei C-H in carboni sp^2 . Anche qui la regione tra 2350 e 2357 cm^{-1} non è riconducibile ad assorbimenti relativi a gruppi funzionali dei trigliceridi.

Per la classe degli acidi grassi saturi, l'algoritmo genetico ha selezionato le variabili 3-5, 26, 71-73, 164-168, 172-174 (per un totale di 15 variabili su 183, Figura 14.16), corrispondenti alle regioni spettrali tra 949 e 980 cm^{-1} , tra 1215 e 1223 cm^{-1} , tra 1736 e 1767 cm^{-1} , tra 2812 e 2866 cm^{-1} , tra 2905 e 2936 cm^{-1} . Nelle regioni selezionate sono presenti picchi di assorbimento attribuibili alle vibrazioni di wagging e twisting dei C-H metilenici, agli stiramenti dei legami C-O-C esterei, del gruppo carbonilico negli esteri, ed agli stiramenti simmetrici ed asimmetrici dei legami C-H su metili e metileni. Appare quantomeno bizzarro che venga inoltre selezionata a descrivere acidi grassi saturi la regione tra 949 e 980 cm^{-1} , tipica del piegamento fuori dal piano dei legami C-H in carboni ibridizzati sp^2 in configurazione *trans*!

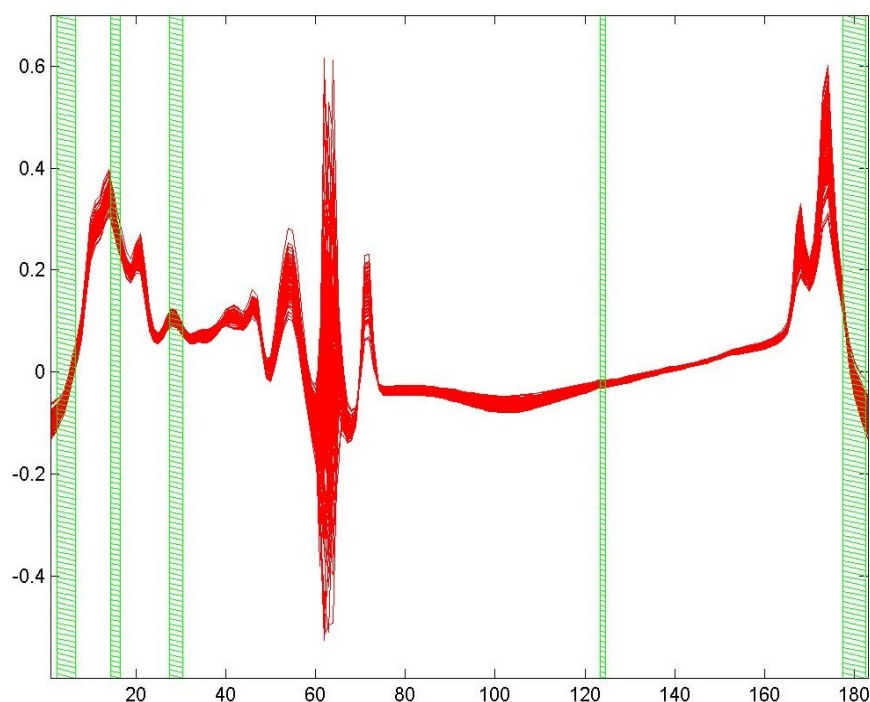


Figura 14.15 – Regioni spettrali selezionate (ombreggiate) dall'algoritmo genetico per la predizione del C18:2 *cis*-9, *trans*-11

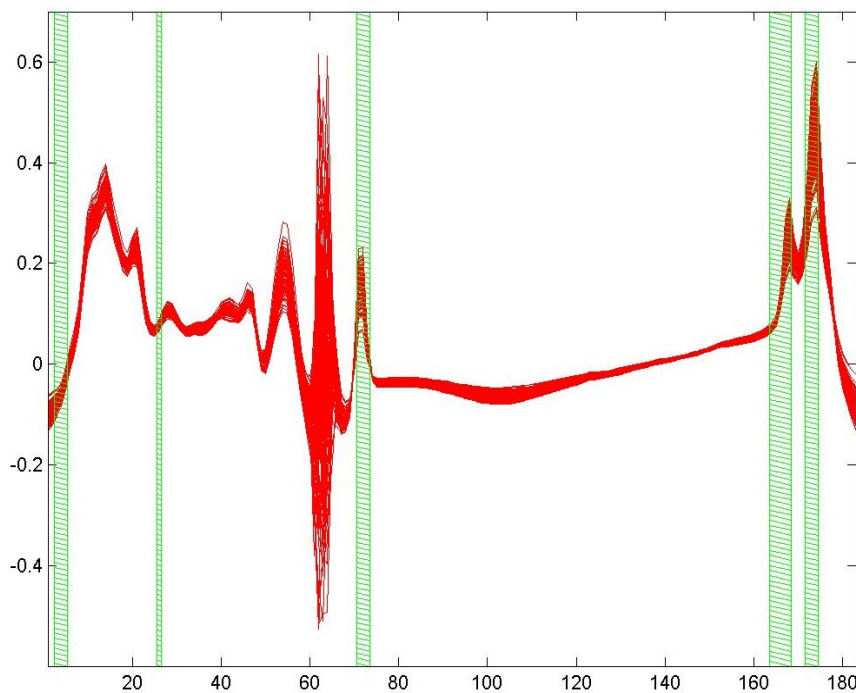


Figura 14.16 – Regioni spettrali selezionate (ombreggiate) dall’algoritmo genetico per la predizione della classe degli acidi grassi saturi

Per la classe degli acidi grassi insaturi, l’algoritmo genetico ha selezionato le variabili 17-20, 164-167, 172-174 (per un totale di 11 variabili su 183, Figura 14.17), corrispondenti alle regioni spettrali tra 1111 e 1154 cm^{-1} , tra 2812 e 2855 cm^{-1} , tra 2905 e 2936 cm^{-1} . Nelle regioni selezionate sono presenti assorbimenti attribuibili agli stiramenti asimmetrici dei legami C-O-C esterei ed agli stiramenti simmetrici ed asimmetrici dei legami C-H su metili e metilene. Non si può qui non rimarcare l’assenza delle variabili riconducibili alle zone spettrali di assorbimento di gruppi funzionali su composti insaturi.

Per la classe degli acidi grassi monoinsaturi, l’algoritmo genetico ha selezionato le variabili 19-22, 71, 164-167, 172 (per un totale di 10 variabili su 183, Figura 14.18), corrispondenti alle regioni spettrali tra 1134 e 1177 cm^{-1} , tra 1736 e 1744 cm^{-1} , tra 2812 e 2855 cm^{-1} . Nelle regioni selezionate sono presenti picchi di assorbimento attribuibili agli stiramenti asimmetrici dei legami C-O-C esterei, degli stiramenti dei C=O degli esteri, e dei C-H su metili e metilene (sia simmetrici che asimmetrici). Anche in questo caso l’algoritmo genetico non ha selezionato regioni relative ad assorbimenti di gruppi funzionali su composti insaturi.

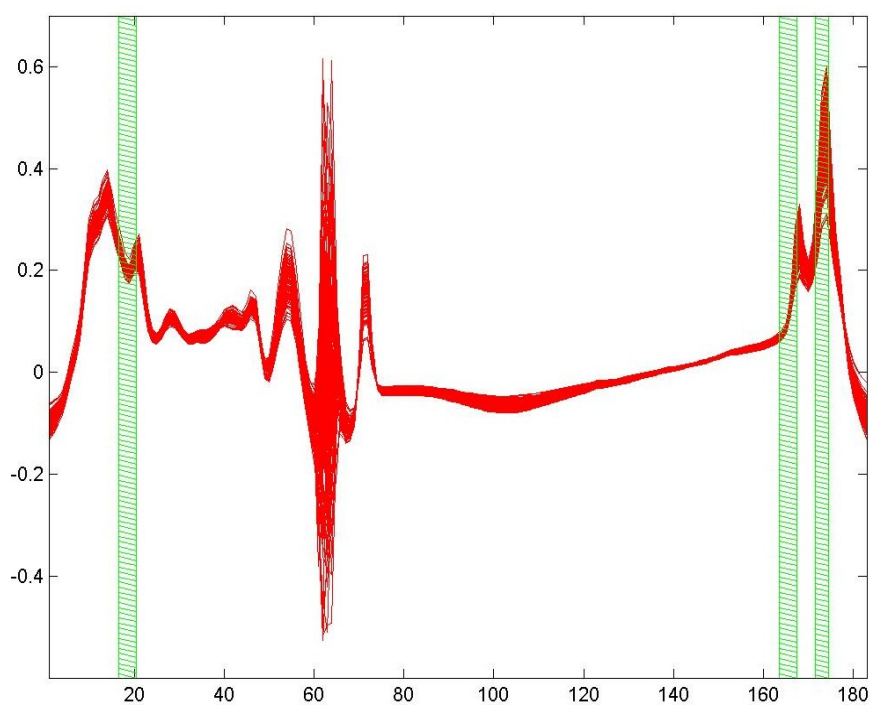


Figura 14.17 – Regioni spettrali selezionate (ombreggiate) dall’algoritmo genetico per la predizione della classe degli acidi grassi insaturi

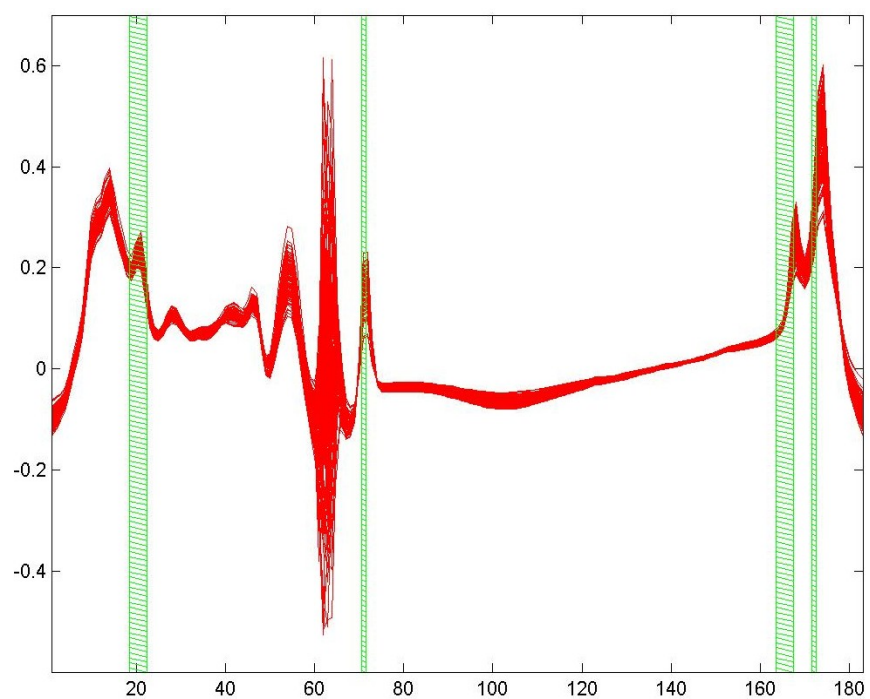


Figura 14.18 – Regioni spettrali selezionate (ombreggiate) dall’algoritmo genetico per la predizione della classe degli acidi grassi monoinsaturi

Per la classe degli acidi grassi polinsaturi, l'algoritmo genetico ha selezionato le variabili 1-16, 21-24, 28-34, 52-56, 71-73, 77-79, 88-91, 102-107, 150-179 (per un totale di 78 variabili su 183, Figura 14.19), corrispondenti alle regioni spettrali tra 926 e 1107 cm^{-1} , tra 1157 e 1200 cm^{-1} , tra 1238 e 1316 cm^{-1} , tra 1516 e 1570 cm^{-1} , tra 1736 e 1767 cm^{-1} , tra 1806 e 1836 cm^{-1} , tra 1933 e 1975 cm^{-1} , tra 2095 e 2160 cm^{-1} , tra 2650 e 2994 cm^{-1} .

Le numerose regioni selezionate comprendono sia zone di assorbimento riconducibili a stiramenti dei legami C-O-C e dei carbonili esterei, dei legami C-H metilici e metilenici, a vibrazioni di wagging e twisting dei legami C-H metilenici, ed ai piegamenti dei C-H olefinici in alcheni in configurazione *trans*. L'algoritmo seleziona anche zone spettrali apparentemente poco pertinenti, quale la regione tra 1516 e 1570 cm^{-1} , caratteristico del piegamento del legame N-H, e di stiramento del legame C-N, nelle ammidi, o del tutto non riconducibili a significative transizioni infrarosse, come quelle tra 1806 e 1836 cm^{-1} e tra 2095 e 2160 cm^{-1} .

Per la classe degli acidi grassi ω -6, l'algoritmo genetico ha selezionato le variabili 7-10, 17-19, 24-25, 54-56, 97-101, 110-112, 122-124, 167-174 (per un totale di 31 variabili su 183, Figura 14.20), corrispondenti alle regioni spettrali tra 995 e 1038 cm^{-1} , tra 1111 e 1142 cm^{-1} , tra 1192 e 1211 cm^{-1} , tra 1539 e 1570 cm^{-1} , tra 2037 e 2091 cm^{-1} , tra 2187 e 2218 cm^{-1} , tra 2326 e 2357 cm^{-1} , tra 2847 e 2936 cm^{-1} . Le regioni selezionate comprendono zone spettrali rappresentative di assorbimenti riconducibili a gruppi funzionali dei trigliceridi quali gli stiramenti dei legami C-O-C esterei (asimmetrici) e dei legami C-H su metili e metileni (simmetrici ed asimmetrici). Vengono inoltre selezionate regioni attribuibili agli stiramenti dei legami P-O-C, ed agli assorbimenti dei legami N-H e C-N ammidici. Sono parimenti selezionate altre regioni spettrali (tra 2037 e 2091 cm^{-1} , tra 2187 e 2218 cm^{-1} , tra 2326 e 2357 cm^{-1}) in cui non sono presenti evidenti assorbimenti infrarossi.

Infine, per la classe degli acidi grassi ω -3, l'algoritmo genetico ha selezionato le variabili 2-5, 77-79, 89-90, 168, 171-173 (per un totale di 13 variabili su 183, Figura 14.21), corrispondenti alle regioni spettrali tra 937 e 980 cm^{-1} , tra 1806 e 1836 cm^{-1} , tra 1944 e 1964 cm^{-1} , tra 2859 e 2866 cm^{-1} , tra 2894 e 2924 cm^{-1} . Le regioni selezionate sono quelle dei piegamenti dei legami C-H in carboni sp^2 in alcheni in configurazione *trans* e degli stiramenti simmetrici e asimmetrici dei legami C-H di metili e metileni. Sono infine selezionate due regioni (tra 1806 e 1836 cm^{-1} e tra 1944 e 1964 cm^{-1}) che non presentano evidenti assorbimenti.

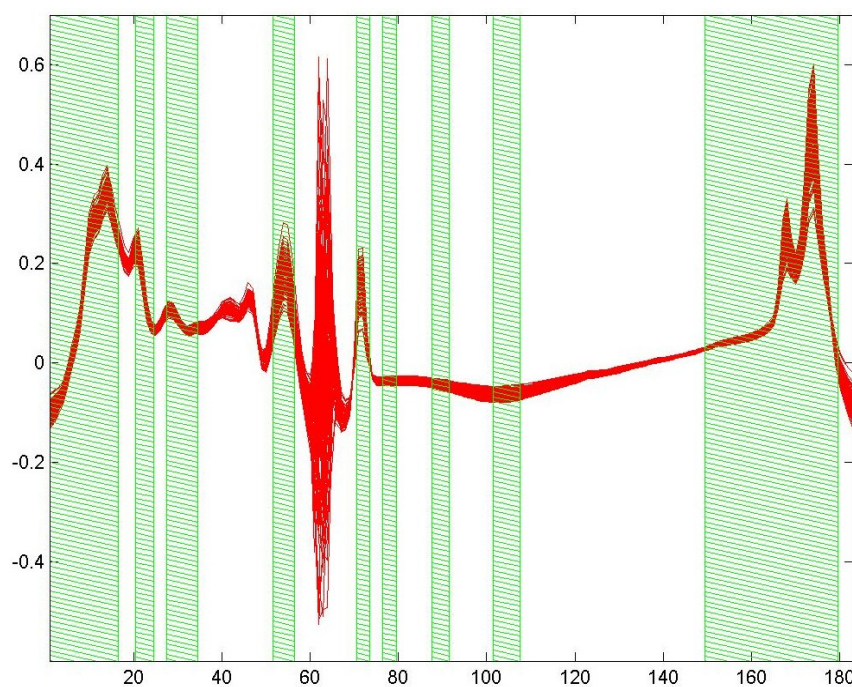


Figura 14.19 – Regioni spettrali selezionate (ombreggiate) dall’algoritmo genetico per la predizione della classe degli acidi grassi polinsaturi

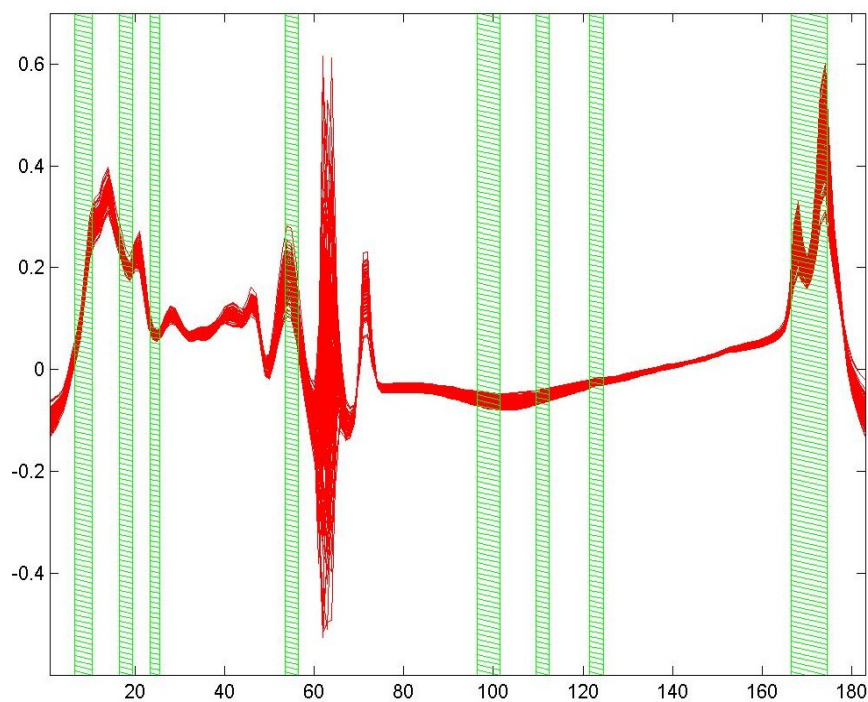


Figura 14.20 – Regioni spettrali selezionate (ombreggiate) dall’algoritmo genetico per la predizione della classe degli acidi grassi ω -6

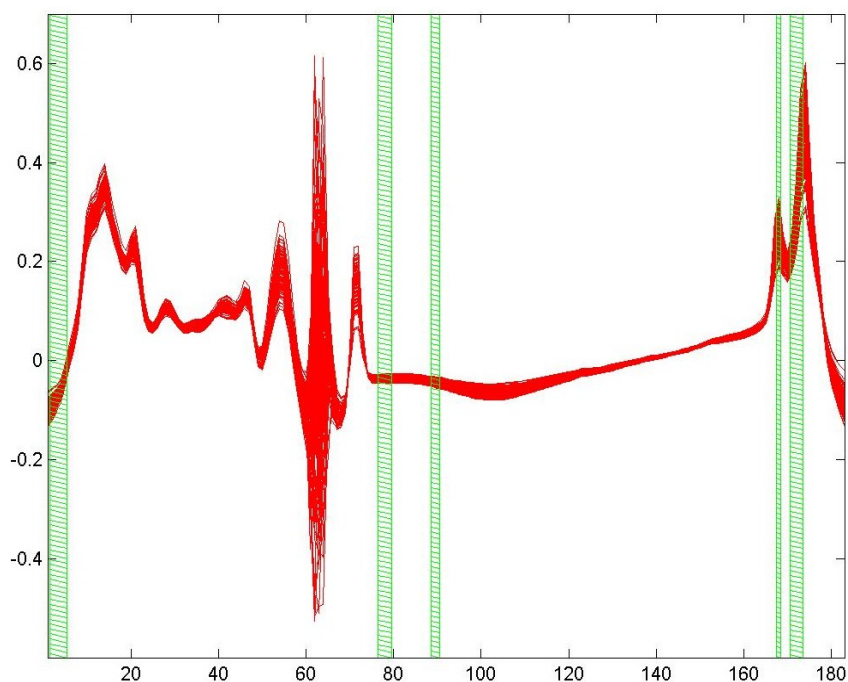


Figura 14.21 – Regioni spettrali selezionate (ombreggiate) dall’algoritmo genetico per la predizione della classe degli acidi grassi ω -3

Dall’analisi dei diversi grafici, si può dire che gli algoritmi genetici hanno selezionato maggiormente regioni caratteristiche di assorbimenti di gruppi funzionali presenti nei trigliceridi. In particolare, per la maggior parte degli acidi grassi, le regioni più significative sembrano essere quelle riferite allo stiramento dei C-O-C e dei carbonili esterei, ed agli stiramenti dei C-H metilici e metilenici. Appaiono poco, o per nulla, significative le regioni caratteristiche dei piegamenti dei C-H metilici e delle vibrazioni di wagging, twisting e scissoring dei C-H metilenici. Per alcuni acidi grassi, in particolare per i C18:2 *cis*-9, *cis*-12 e C18:3 *cis*-9, *cis*-12, *cis*-15, gli algoritmi genetici selezionano diverse regioni spettrali prive di assorbimenti, e ciò viene a riflettersi soprattutto nelle classi di acidi grassi polinsaturi ed ω -6. Ciò può essere attribuibile alla minore concentrazione di questi acidi grassi nel latte ovino, e quindi alla maggiore difficoltà di trovare una relazione tra gli assorbimenti spettrali ed i dati di riferimento, per cui gli algoritmi genetici fanno ricorso a zone spettrali apparentemente non informative. Non pare esserci invece una relazione tra le concentrazioni dei diversi acidi grassi e la selezione di regioni spettrali relative ad assorbimenti di gruppi funzionali non appartenenti al composto considerato.

14.2.4.2 Validazione dei modelli di predizione basati sulla selezione delle bande spettrali effettuata dagli algoritmi genetici

Una volta effettuata la selezione delle bande più adatte alla predizione di ogni specifico acido grasso e classe di acidi grassi, per ciascuna risposta sono stati costruiti i rispettivi modelli di predizione.

Ogni modello possiede quindi le regioni specifiche selezionate dall'algoritmo genetico per il composto/classe di composti che lo stesso deve predire. L'efficacia della predizione è valutata in termini di Root Mean Square Error in predizione (RMSEP, Tabella 14.5). La tabella citata contiene: il numero di variabili latenti significative (LV) utilizzate per la predizione su set esterno; il numero di variabili selezionate dall'algoritmo genetico a partire dallo spettro pretrattato con l'operazione Averaging (spettro contenente 183 variabili); il Root Mean Square Error in predizione sul set esterno (RMSEP); il bias ottenuto nella predizione su set esterno; e l'errore relativo percentuale in predizione su set esterno, dato dal rapporto percentuale tra RMSEP e la media dei valori predetti per il test set (tale media è identica alla media dei valori di riferimento). Dai risultati si può notare che l'algoritmo genetico seleziona un numero ridotto di variabili per tutti i composti, selezionando quelle più informative e ottenendo valori di RMSEP inferiori per la maggior parte degli acidi grassi rispetto a quelli ottenuti dai modelli di predizione costruiti senza una selezione di bande spettrali (Tabella 14.2).

Il numero di variabili significative selezionate per gli acidi grassi, riportate allo spettro ridotto, è mediamente pari a 62 (su uno spettro contenente 550 variabili), mentre per le classi di acidi grassi, esso è mediamente pari a 79 (su uno spettro contenente 550 variabili). La riduzione dei valori di RMSEP è più consistente per gli acidi grassi a media catena, per il C18:1 *cis*-9, e per le classi di acidi grassi saturi, insaturi e polinsaturi. Per gli altri composti, i valori di RMSEP sono rimasti pressoché invariati.

Nei casi in cui gli algoritmi genetici hanno selezionato regioni spettrali in cui sono presenti picchi di assorbimento teoricamente caratteristici di gruppi funzionali non inclusi nei trigliceridi (per esempio, per la classe di acidi grassi ω -6, il picco nella regione tra 1516 e 1570 cm^{-1} , caratteristico della vibrazione di piegamento del legame N-H, e di stiramento del legame C-N, nelle ammidi), o regioni spettrali in cui non sono presenti picchi di assorbimento, le elaborazioni dei modelli sono state ripetute con l'esclusione delle regioni spettrali in questione, ma i risultati della validazione su set esterno sono sempre peggiorati, a dimostrazione del fatto che le bande selezionate dagli algoritmi genetici sono informative.

Tabella 14.5 – Risultati ottenuti nella fase di validazione dei modelli costruiti utilizzando le bande selezionate dagli algoritmi genetici

Acido grasso	Fase di Validazione				
	LV	Variabili Selezionate (su 183)	RMSEP	bias	Err. %
C4:0	9	17	0,014	-0,002	5,96
C6:0	7	12	0,010	0,001	6,02
C8:0	11	13	0,011	0,001	7,69
C10:0	14	29	0,030	0,004	7,48
C12:0	11	25	0,017	0,000	8,13
C14:0	14	22	0,045	0,008	8,64
C16:0	8	11	0,089	-0,006	7,92
C18:0	11	21	0,066	0,011	12,92
C18:1 <i>trans</i>-11	9	31	0,036	-0,004	19,46
C18:1 <i>cis</i>-9	15	17	0,104	-0,012	13,07
C18:2 <i>cis</i>-9, <i>cis</i>-12	14	46	0,022	0,000	16,30
C18:3 <i>cis</i>-9, <i>cis</i>-12, <i>cis</i>-15	5	8	0,023	0,001	23,00
C18:2 <i>cis</i>-9, <i>trans</i>-11	12	15	0,020	-0,004	23,26
Acidi Grassi Saturi	15	15	0,102	0,013	2,86
Acidi Grassi Insaturi	10	11	0,108	-0,018	5,83
Acidi Grassi Monoinsaturi	10	10	0,117	-0,010	8,57
Acidi Grassi Polinsaturi	14	78	0,039	-0,004	8,01
Acidi Grassi ω-6	14	31	0,023	0,003	13,22
Acidi Grassi ω-3	10	13	0,028	-0,001	17,28

LV = variabili latenti significative utilizzate per la predizione esterna.

RMSEP = errore quadratico medio in predizione su set esterno.

bias = bias in predizione sul set esterno.

Err. % = errore relativo percentuale (in predizione su set esterno)

= RMSEP/Media valori predetti in validazione esterna.

Dai risultati ottenuti si può affermare che la selezione delle variabili spettrali effettuata tramite gli algoritmi genetici ha permesso di costruire dei modelli parsimoniosi, contenenti un numero ridotto di variabili spettrali, la cui validazione ha portato a predizioni più efficaci per la maggior parte dei composti. Anche quando la predizione risulta essere pressoché identica a quella ottenuta tramite i modelli costruiti senza una selezione delle regioni spettrali più significative, gli algoritmi genetici hanno permesso di costruire dei modelli più semplici, selezionando esclusivamente le regioni più adatte alla predizione, evitando quindi problemi di overfitting dovuti alla presenza di variabili correlate, di variabili non informative e di rumore.

14.3 Confronto dei modelli di predizione ottenuti tramite selezione delle regioni spettrali

Lo studio sulla selezione delle regioni spettrali più adatte alla predizione degli acidi grassi e delle classi di acidi grassi contenute nel latte ovino, ha portato alla costruzione di modelli di predizione robusti ed affidabili grazie alla riduzione del numero di variabili spettrali, che hanno permesso di migliorare l'efficacia della predizione di ogni singolo composto.

Nella Tabella 14.6 vengono riassunti i risultati della validazione su set esterno dei modelli di predizione costruiti (1) sulla base dello spettro FT-IR ridotto, senza selezione di bande spettrali (paragrafo 14.2.2), (2) sulla base della selezione delle regioni spettrali effettuata tramite osservazione dello spettro (paragrafo 14.2.3), (3) sulla base della selezione delle regioni spettrali effettuata tramite algoritmi genetici (paragrafo 14.2.4.2).

Nella tabella citata vengono mostrati, per ogni elaborazione: il numero di variabili spettrali utilizzate nella predizione (Var); e il Root Mean Square Error in predizione (RMSEP).

Dai risultati si può notare che entrambi i metodi di selezione delle variabili spettrali hanno portato ad una sostanziale riduzione del numero di lunghezze d'onda considerate. Gli algoritmi genetici hanno selezionato un numero di variabili fino a ventitre volte inferiore quello contenuto nello spettro ridotto, mentre la selezione delle bande tramite osservazione dello spettro ha permesso di ridurre di un terzo il numero di variabili iniziali. Grazie alla selezione delle bande, i valori di RMSEP hanno subito una diminuzione nella maggior parte dei casi, oppure sono rimasti invariati.

L'unico caso in cui gli algoritmi genetici selezionano un numero di variabili spettrali elevato è per la classe di acidi grassi polinsaturi, per la quale è preferibile il modello basato sulle variabili selezionate tramite osservazione dello spettro.

Gli algoritmi genetici hanno quindi permesso di ottenere modelli di predizione con il minor numero di variabili spettrali, mantenendo invariata, o migliorando, l'efficacia della predizione. Per gli scopi prefissi da questo lavoro di tesi, tutti gli errori in predizione sono da ritenersi accettabili.

Tabella 14.6 – Riepilogo dei risultati della fase di validazione dei modelli costruiti

Acido grasso	Nessuna selezione ^a		Selezione tramite osservazione dello spettro ^b		Selezione tramite Algoritmi Genetici ^c	
	Var	RMSEP	Var	RMSEP	Var	RMSEP
C4:0	550	0,015	175	0,014	51	0,014
C6:0	550	0,011	175	0,010	36	0,010
C8:0	550	0,013	175	0,011	39	0,011
C10:0	550	0,040	175	0,035	87	0,030
C12:0	550	0,020	175	0,019	75	0,017
C14:0	550	0,047	175	0,044	66	0,045
C16:0	550	0,091	175	0,086	33	0,089
C18:0	550	0,062	175	0,064	63	0,066
C18:1 <i>trans</i>-11	550	0,039	194	0,037	93	0,036
C18:1 <i>cis</i>-9	550	0,124	191	0,112	51	0,104
C18:2 <i>cis</i>-9, <i>cis</i>-12	550	0,024	191	0,024	138	0,022
C18:3 <i>cis</i>-9, <i>cis</i>-12, <i>cis</i>-15	550	0,020	191	0,019	24	0,023
C18:2 <i>cis</i>-9, <i>trans</i>-11	550	0,017	194	0,016	45	0,020
Acidi Grassi Saturi	550	0,122	175	0,092	45	0,102
Acidi Grassi Insaturi	550	0,121	194	0,099	33	0,108
Acidi Grassi Monoinsaturi	550	0,118	194	0,098	30	0,117
Acidi Grassi Polinsaturi	550	0,042	194	0,038	234	0,039
Acidi Grassi ω-6	550	0,025	191	0,024	93	0,023
Acidi Grassi ω-3	550	0,027	191	0,028	39	0,028

^a = risultati ottenuti nella fase di validazione dei modelli costruiti sullo spettro FT-IR ridotto contenente 550 variabili.

^b = risultati ottenuti nella fase di validazione dei modelli costruiti sulla base della selezione delle regioni spettrali effettuata tramite osservazione dello spettro FT-IR.

^c = risultati ottenuti nella fase di validazione dei modelli costruiti sulla base della selezione delle regioni spettrali effettuata tramite algoritmi genetici.

Var = numero di variabili spettrali utilizzate nella predizione. Per gli algoritmi genetici è indicato il numero di variabili selezionate, riportate allo spettro ridotto (contenente 550 variabili).

RMSEP = errore quadratico medio in predizione su set esterno.

14.4 Modelli di predizione selezionati

In base ai risultati descritti nel precedente paragrafo, vengono di seguito descritte le caratteristiche dei modelli di predizione selezionati.

Per tutti i composti, eccetto la classe di acidi grassi polinsaturi, sono stati preferiti i modelli basati sulle bande selezionate tramite algoritmi genetici.

Gli errori medi percentuali in predizione dei modelli per gli acidi grassi saturi (da C4:0 a C18:0) risultano variare tra il 6,0% ed il 12,9%.

Per gli acidi grassi C18:1 *trans*-11, C18:1 *cis*-9, e C18:2 *cis*-9, *cis*-12, i modelli predicono con un errore medio percentuale compreso tra il 13,1% ed il 19,5%.

Per il C18:3 *cis*-9, *cis*-12, *cis*-15 e per il C18:2 *cis*-9, *trans*-11 si hanno modelli caratterizzati da errori medi percentuali rispettivamente del 23,0 e 23,2%. Gli errori - apparentemente alti - sono accettabili, considerati sia gli scopi prefissi per l'utilizzo dei modelli, che le basse concentrazioni dei due acidi grassi, dell'ordine dei 100 ppm.

Per le classi di acidi grassi saturi, insaturi, monoinsaturi e polinsaturi, i modelli predicono con un errore medio percentuale compreso tra il 2,9% e l'8,6%.

I modelli di predizione per le classi di acidi ω -6 ed ω -3 predicono le due classi con errori medi percentuali rispettivamente di 13,2% e 17,3%.

Dai valori degli errori in predizione (RMSEP), confrontati con i valori di varianza sperimentale calcolata per il metodo di riferimento (paragrafo 12.5), si può evidenziare che, nonostante il fatto che l'errore in predizione del modello sia significativamente maggiore (da 3 a 13 volte) rispetto all'errore sperimentale del metodo di riferimento, si deve rimarcare che per gli scopi pratici per i quali i modelli in questione saranno utilizzati, la precisione ottenuta è comunque più che accettabile.

A partire dalla Figura 14.22, fino alla Figura 14.40 sono mostrati i grafici "Predetto contro misurato", ottenuti per la validazione di ogni modello scelto, che mostrano, per ogni acido grasso, la correlazione tra i dati di riferimento dei campioni del set esterno e i valori predetti dal modello per gli stessi campioni.

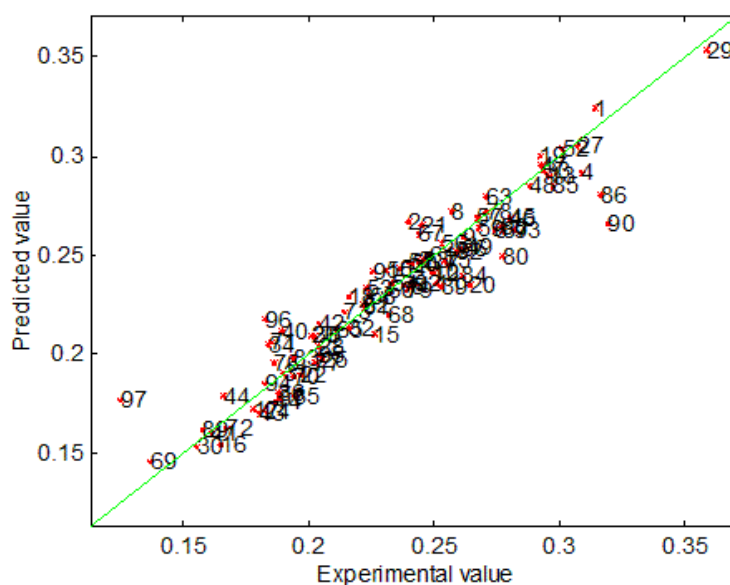


Figura 14.22 – Grafico “Predetto contro Misurato” per il C4:0, ottenuto dalla validazione del modello costruito in base alla selezione delle variabili spettrali tramite algoritmi genetici

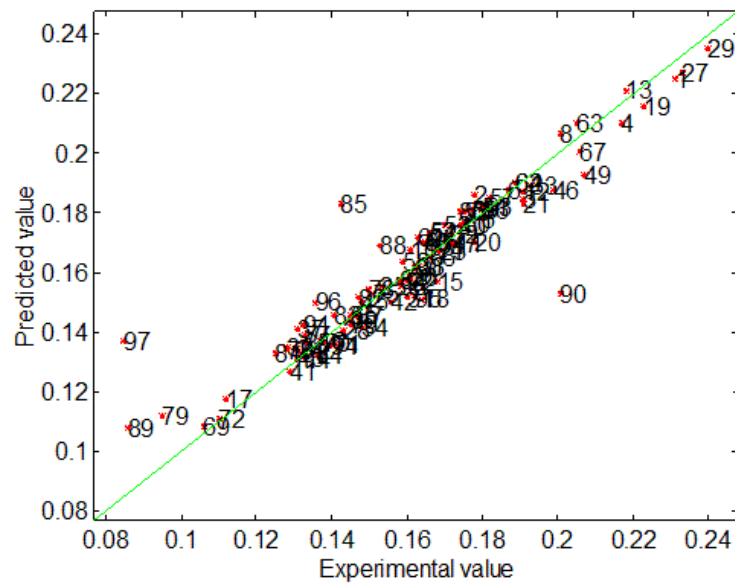


Figura 14.23 – Grafico “Predetto contro Misurato” per il C6:0, ottenuto dalla validazione del modello costruito in base alla selezione delle variabili spettrali tramite algoritmi genetici

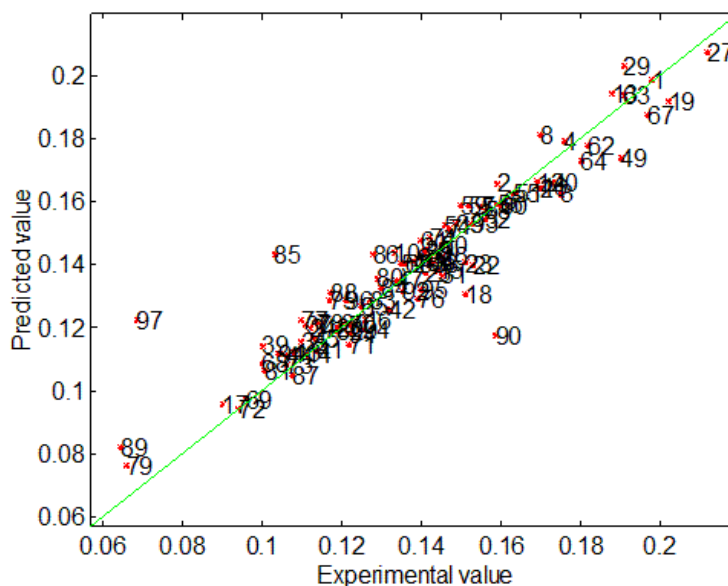


Figura 14.24 – Grafico “Predetto contro Misurato” per il C8:0, ottenuto dalla validazione del modello costruito in base alla selezione delle variabili spettrali tramite algoritmi genetici

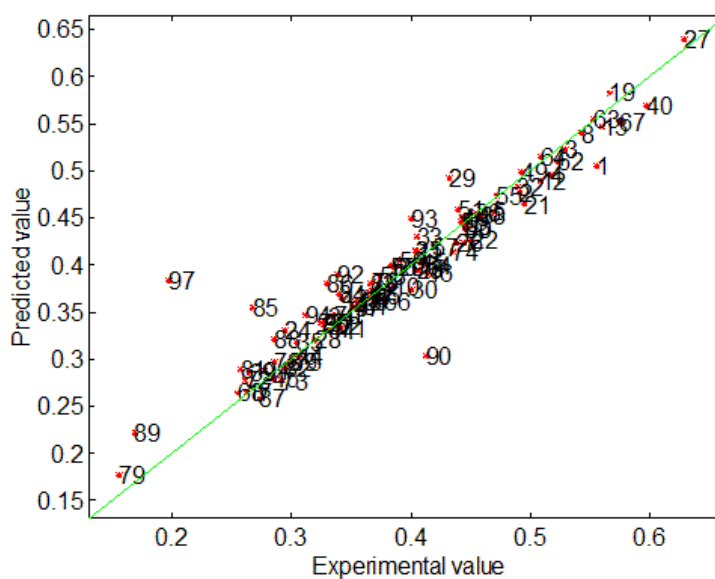


Figura 14.25 – Grafico “Predetto contro Misurato” per il C10:0, ottenuto dalla validazione del modello costruito in base alla selezione delle variabili spettrali tramite algoritmi genetici

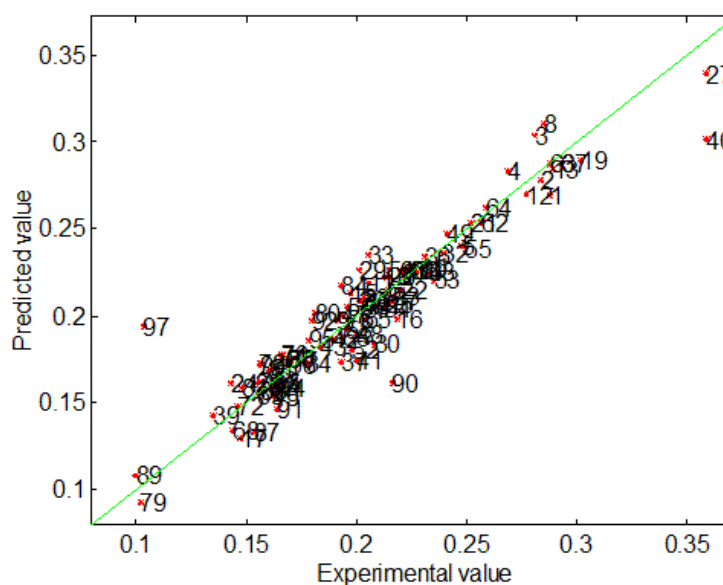


Figura 14.26 – Grafico “Predetto contro Misurato” per il C12:0, ottenuto dalla validazione del modello costruito in base alla selezione delle variabili spettrali tramite algoritmi genetici

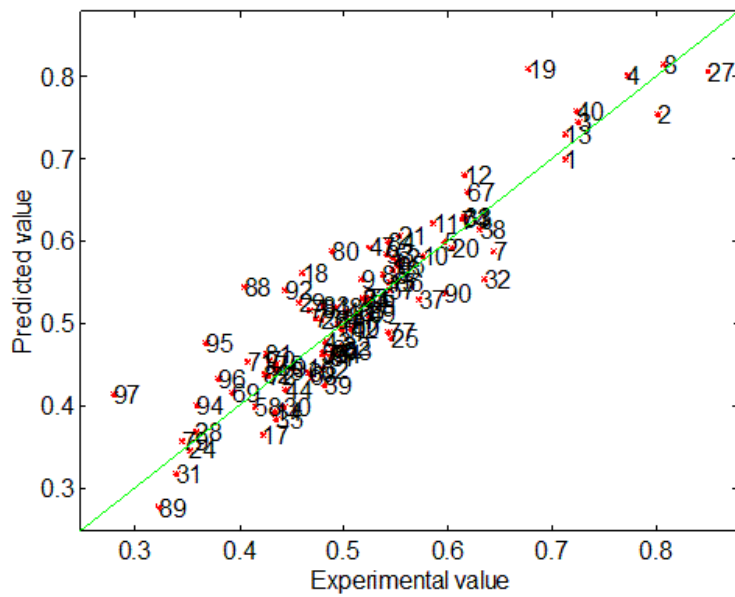


Figura 14.27 – Grafico “Predetto contro Misurato” per il C14:0, ottenuto dalla validazione del modello costruito in base alla selezione delle variabili spettrali tramite algoritmi genetici

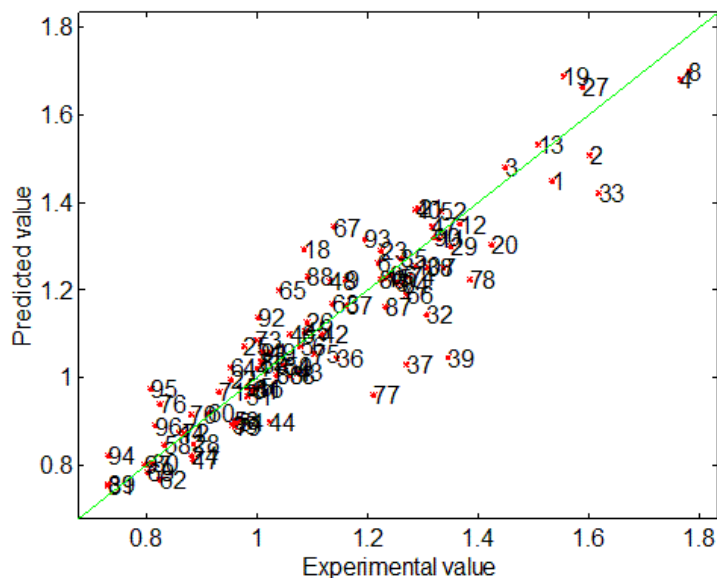


Figura 14.28 – Grafico “Predetto contro Misurato” per il C16:0, ottenuto dalla validazione del modello costruito in base alla selezione delle variabili spettrali tramite algoritmi genetici

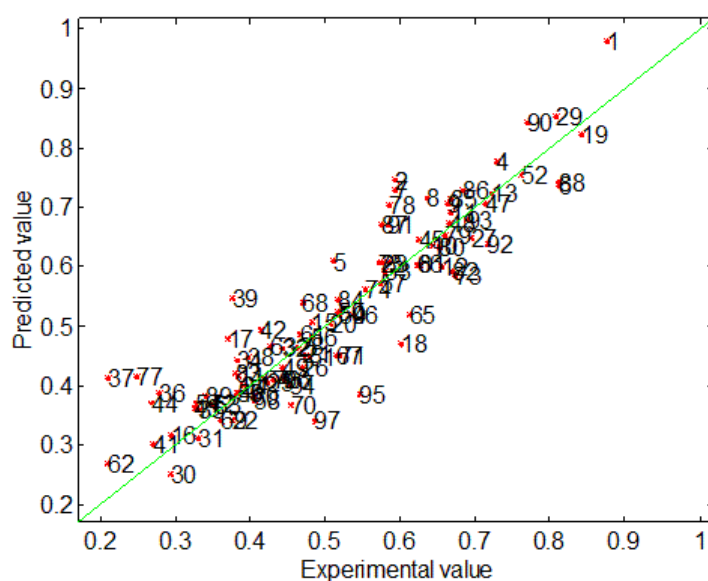


Figura 14.29 – Grafico “Predetto contro Misurato” per il C18:0, ottenuto dalla validazione del modello costruito in base alla selezione delle variabili spettrali tramite algoritmi genetici

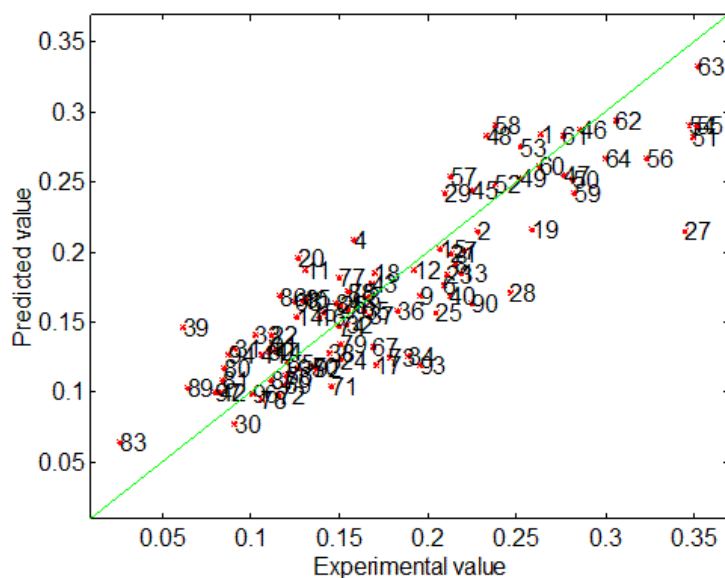


Figura 14.30 – Grafico “Predetto contro Misurato” per il C18:1 *trans*-11, ottenuto dalla validazione del modello costruito in base alla selezione delle variabili spettrali tramite algoritmi genetici

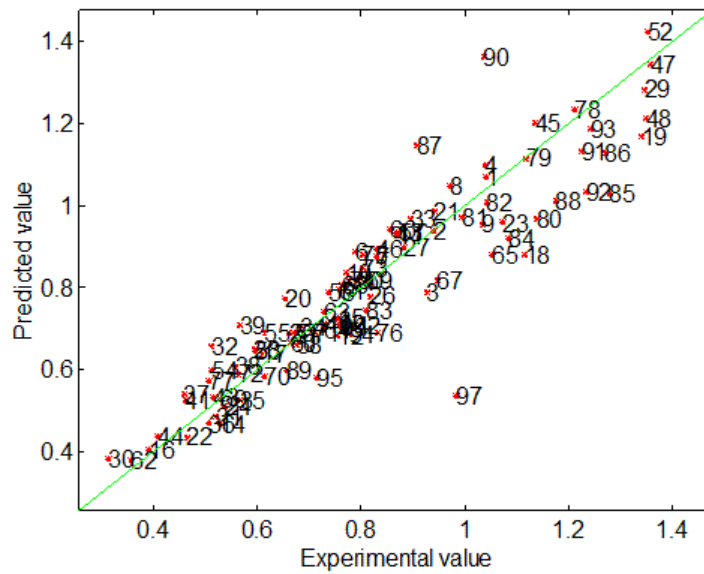


Figura 14.31 – Grafico “Predetto contro Misurato” per il C18:1 *cis*-9, ottenuto dalla validazione del modello costruito in base alla selezione delle variabili spettrali tramite algoritmi genetici

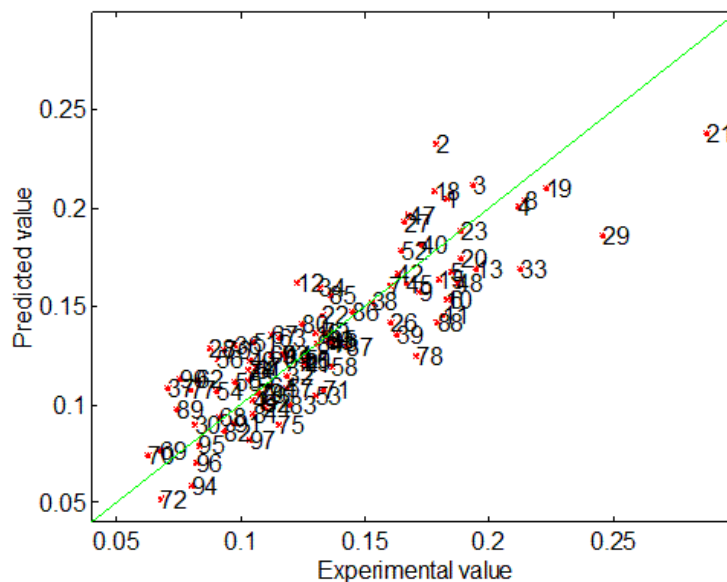


Figura 14.32 – Grafico “Predetto contro Misurato” per il C18:2 *cis*-9, *cis*-12 ottenuto dalla validazione del modello costruito in base alla selezione delle variabili spettrali tramite algoritmi genetici

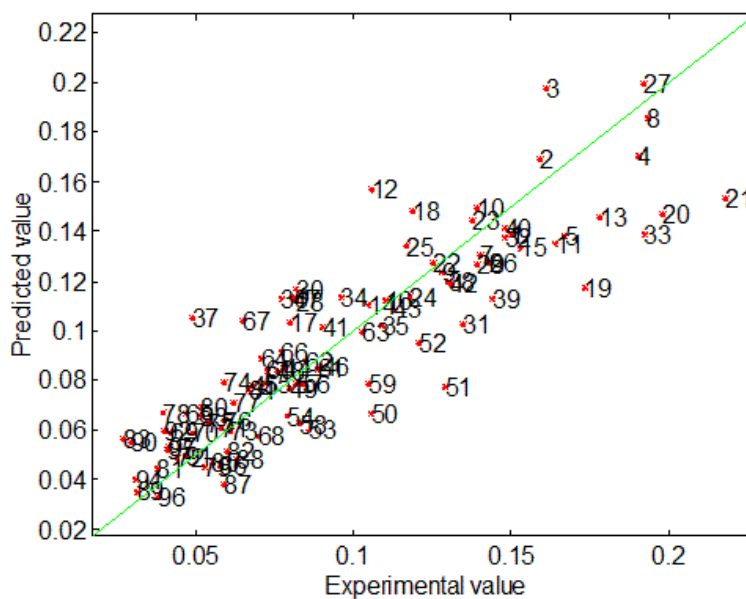


Figura 14.33 – Grafico “Predetto contro Misurato” per il C18:3 *cis*-9, *cis*-12, *cis*-15 ottenuto dalla validazione del modello costruito in base alla selezione delle variabili spettrali tramite algoritmi genetici

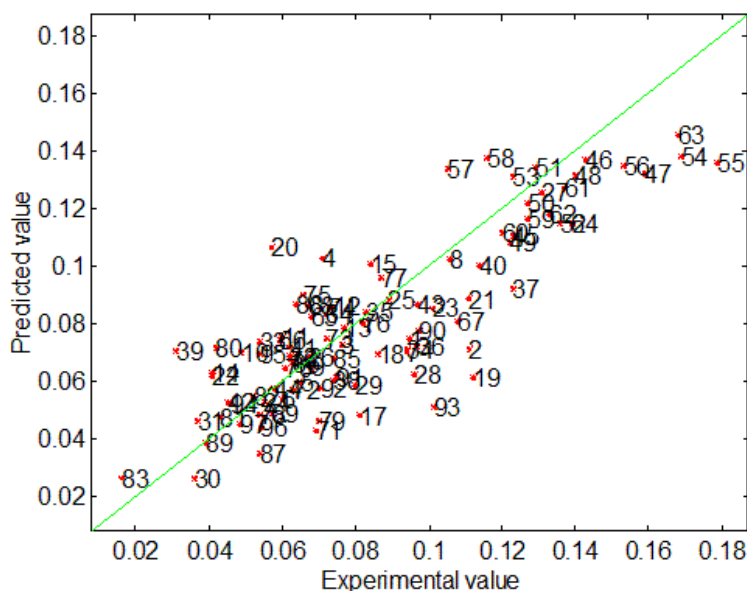


Figura 14.34 – Grafico “Predetto contro Misurato” per il C18:2 *cis*-9, *trans*-11 ottenuto dalla validazione del modello costruito in base alla selezione delle variabili spettrali tramite spettro algoritmi genetici

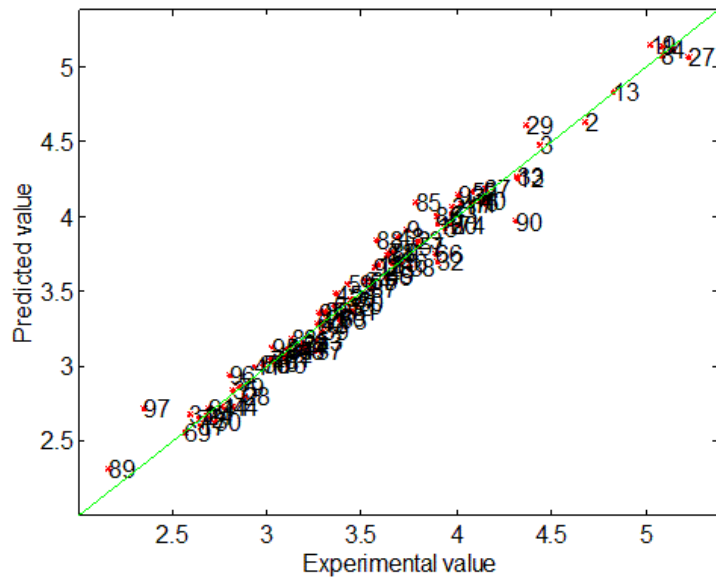


Figura 14.35 – Grafico “Predetto contro Misurato” per la classe di acidi grassi saturi ottenuto dalla validazione del modello costruito in base alla selezione delle variabili spettrali tramite algoritmi genetici

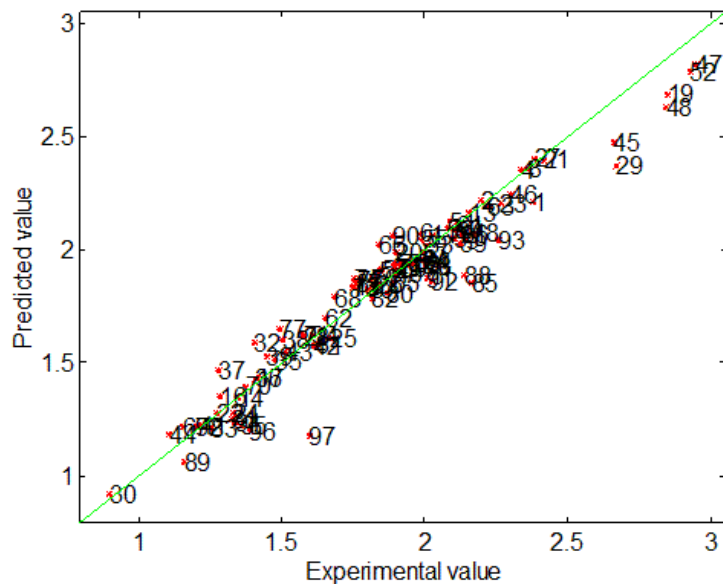


Figura 14.36 – Grafico “Predetto contro Misurato” per la classe di acidi grassi insaturi ottenuto dalla validazione del modello costruito in base alla selezione delle variabili spettrali tramite algoritmi genetici

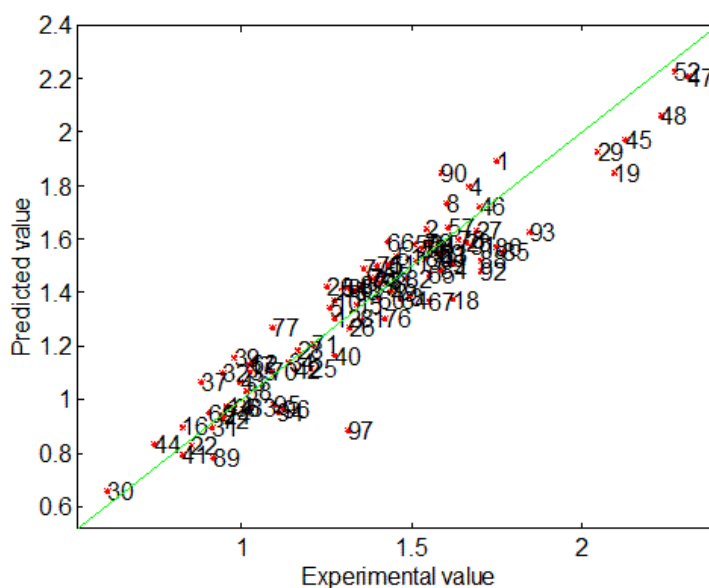


Figura 14.37 – Grafico “Predetto contro Misurato” per la classe di acidi grassi monoinsaturi ottenuto dalla validazione del modello costruito in base alla selezione delle variabili spettrali tramite spettro algoritmi genetici

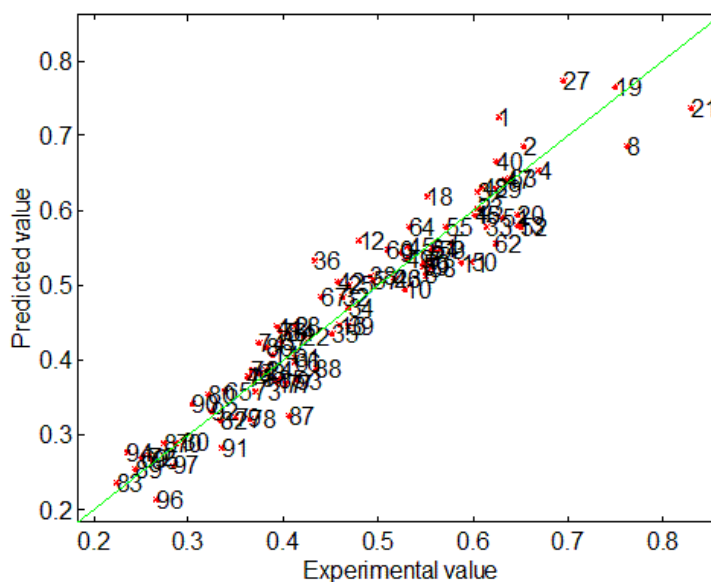


Figura 14.38 – Grafico “Predetto contro Misurato” per la classe di acidi grassi polinsaturi ottenuto dalla validazione del modello costruito in base alla selezione delle variabili spettrali tramite osservazione dello spettro

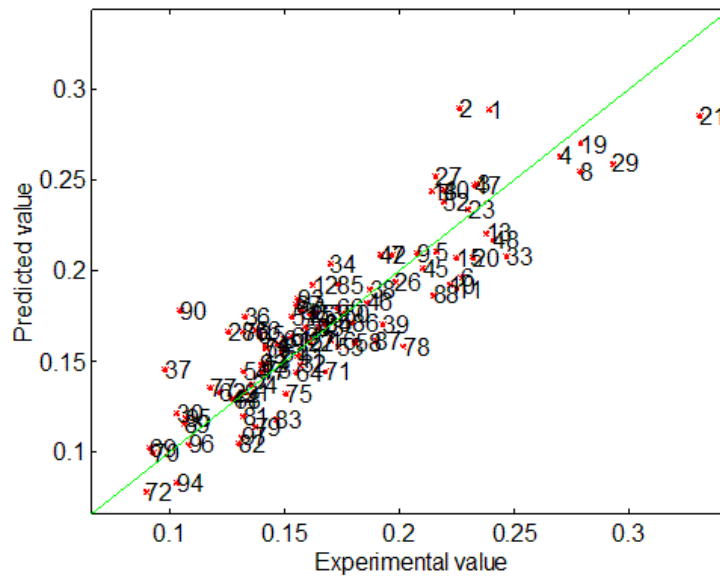


Figura 14.39 – Grafico “Predetto contro Misurato” per la classe di acidi grassi ω -6 ottenuto dalla validazione del modello costruito in base alla selezione delle variabili spettrali tramite algoritmi genetici

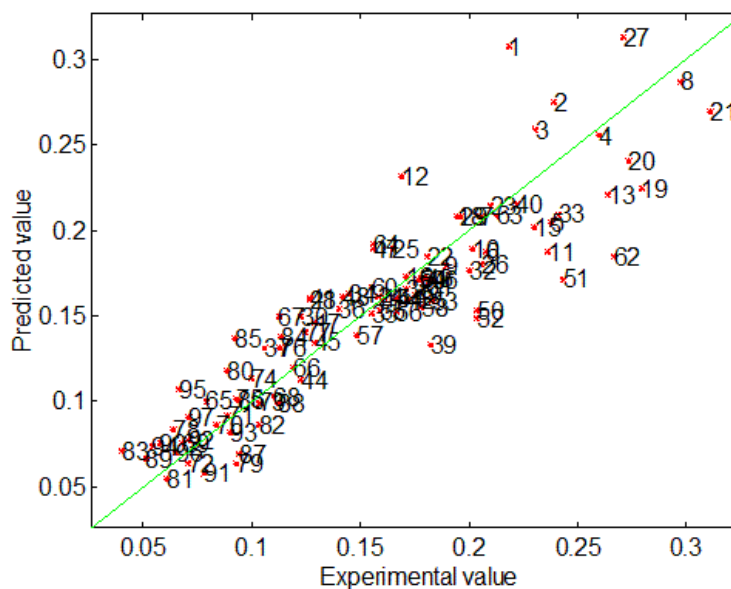


Figura 14.40 – Grafico “Predetto contro Misurato” per la classe di acidi grassi ω -3 ottenuto dalla validazione del modello costruito in base alla selezione delle variabili spettrali tramite algoritmi genetici

In definitiva, le tecniche di regressione multivariata hanno permesso di costruire modelli di predizione di composti minoritari del latte ovino tramite spettroscopia nel medio infrarosso a Trasformata di Fourier. Una selezione delle bande spettrali ha permesso inoltre di selezionare le regioni spettrali più informative. La selezione è stata effettuata sia tramite l'identificazione dei picchi di assorbimento relazionati a gruppi funzionali dei trigliceridi, sia tramite l'utilizzo degli algoritmi genetici. Entrambi i metodi hanno portato ad una selezione di un numero ridotto di regioni spettrali e a modelli di predizione con errori in predizione su set esterno inferiori ai modelli costruiti senza una selezione delle variabili spettrali.

In particolare, l'utilizzo degli algoritmi genetici ha permesso la selezione di un numero ridotto di variabili che meglio predicono la variabile incognita, evitando di selezionare variabili non informative o legate a rumore, e quindi evitando problemi di overfitting. I modelli costruiti sono quindi affidabili per le future predizioni. Inoltre, la tecnica degli algoritmi genetici per la selezione delle variabili non necessita di conoscenze specifiche in campo spettroscopico, e permette di selezionare le variabili informative in modo automatico.

I modelli di predizione verranno inseriti nello spettrometro MilkoScan FT6000 situato nel laboratorio di analisi del latte dell'Associazione Regionale Allevatori della Sardegna (ARA) in modo da introdurre i composti minoritari del latte ovino tra i parametri di qualità che determineranno il prezzo del latte. Gli errori di tutti i modelli in predizione sono infatti congruenti con il loro utilizzo nel campo dei controlli di qualità del latte ovino, permettendo l'introduzione di un sistema di pagamento a griglia, così come già avviene per i parametri di tenore di grasso e proteine.

15 Conclusioni

L'attività di ricerca che ha caratterizzato questa Tesi di dottorato rientra nel contesto degli studi relativi al miglioramento della qualità del latte ovino di razza Sarda. La produzione di latte ovino, e la sua trasformazione casearia, è l'attività economica principale della Regione Sardegna. Diventa pertanto fondamentale lavorare per il miglioramento della qualità della materia prima latte, anche in vista di un'auspicabile valorizzazione del comparto.

Il pagamento del latte ovino in Sardegna è attualmente determinato dalla valutazione di alcuni parametri igienico-sanitari, quali la conta batterica e la conta delle cellule somatiche, e dalla valutazione di alcuni parametri di qualità, quali il contenuto di grasso e di proteine. Tali controlli vengono effettuati dal laboratorio dell'Associazione Regionale Allevatori della Sardegna (ARA).

È noto però che il latte è un alimento funzionale, in quanto contiene diversi composti che possono avere effetti positivi sulla salute dell'uomo e tra questi particolarmente interessanti risultano essere gli acidi grassi. Alla luce di questo fatto diventa auspicabile allargare il lotto dei parametri di qualità, tenendo quindi conto anche del tenore di composti che - seppur non maggioritari - concorrono a valorizzare la materia prima latte. Diventa tuttavia importante che le analisi relative ai controlli di qualità siano effettuate con tecniche analitiche veloci ed economiche affinché i costi analitici non gravino ulteriormente sui costi di produzione del latte. In questo senso le tecniche analitiche online impieganti la spettroscopia infrarossa a Trasformata di Fourier paiono esser promettenti a riguardo, ancorché al momento prevalentemente utilizzate in questo campo quasi esclusivamente per il latte bovino e solo per la predizione delle concentrazioni dei composti maggioritari. Alla luce di queste considerazioni, il lavoro di ricerca effettuato in questa Tesi di dottorato è stato pertanto orientato alla messa a punto e alla validazione di tecniche di analisi utili per la determinazione degli acidi grassi nel latte ovino, funzionali a poter sviluppare un metodo spettroscopico online a Trasformata di Fourier nel medio infrarosso in grado di poter predire in maniera veloce, economica e sufficientemente accurata il contenuto di analiti anche non maggioritari in tale matrice.

Uno dei passaggi chiave dello studio è stato quello di adottare gli opportuni metodi di regressione finalizzati a correlare adeguatamente il dataset ottenuto utilizzando metodi analitici tradizionali (gascromatografia), ufficiali e validati, su un gran numero di campioni

di latte ovino raccolti in selezionate zone della Sardegna durante i tre anni della Tesi, con gli spettri FT-MIR raccolti sugli stessi campioni.

L'adozione di opportune tecniche chemiometriche di regressione ha permesso di ridurre la dimensionalità del dataset spettrale, inglobando la maggior parte dell'informazione in un contenuto numero di componenti principali.

A riguardo, i risultati sono stati ottenuti utilizzando la tecnica di regressione PLS (Partial Least Square Regression o regressione parziale con minimi quadrati). Sono state trovate così quelle dimensioni che meglio permettono di descrivere la relazione tra gli spettri FT-IR ed i dati di riferimento.

I modelli di predizione costruiti nel presente lavoro di tesi sono stati validati su un set di campioni di latte ovino non utilizzato nella fase di calibrazione del modello.

L'efficacia di predizione, realizzata sia per ogni singolo analita che per classi di acidi grassi, è stata ottimizzata in funzione delle variabili spettrali che denotavano il maggior contenuto di informazione. La selezione delle stesse è stata effettuata con due diversi metodi: 1) l'individuazione dei picchi di assorbimento, caratteristici dei gruppi funzionali degli acidi grassi, nello spettro FT-IR del latte ovino; 2) l'adozione degli algoritmi genetici. Gli algoritmi genetici hanno permesso di selezionare solo le bande informative che meglio si correlano con i "valori veri" del composto da predire, evitando di selezionare le regioni spettrali associate a rumore o a differente informazione.

Sono stati così ottenuti dei modelli di predizione parsimoniosi (capaci quindi di predire un composto in maniera affidabile utilizzando il minor numero di variabili spettrali), efficaci (errori medi in predizione compresi tra il 2,9 ed il 23,2%), ed in grado di utilizzare lo spettro FT-MIR del latte per ottenere, in maniera rapida, economica e non distruttiva, la valutazione del profilo degli acidi grassi nel latte ovino.

Le funzioni di calibrazione ottenute saranno nel futuro inserite tra le analisi di controllo effettuate dall'ARA Sardegna, in maniera tale da implementare il lotto dei parametri di qualità utilizzati per monitorare la qualità del latte e quindi per migliorare, sia a vantaggio dei produttori che dell'intero sistema produttivo, la natura del sistema di pagamento del latte ovino in Sardegna.

16 Bibliografia

Abbas O, Dardenne P., Baeten V. (2012). Near-Infrared, Mid-Infrared, and Raman Spectroscopy, *in*: Chemical Analysis of Food: Techniques and Applications. *Elsevier Inc.* pp. 59-89.

Addis M., Cabiddu A., Pinna G., Decandia M., Piredda G., Pirisi A., Molle G. (2005). Milk and Cheese Fatty Acid Composition in Sheep Fed Mediterranean Forages with Reference to Conjugated Linoleic Acid *cis-9,trans-11*. *Journal of Dairy Science* 88, 3443-3454.

Addis M., Cabiddu A., Decandia M., Spada S., Acciaro M., Pirisi A., Sitzia M., Costa E., Cannas A., Molle G. (2009). Effects of different fat-enriched concentrates on fatty acid profile of cheese from grazing dairy sheep. *Italian Journal of Animal Science* 8, 378-380.

Ahrens E.H., Hirsch J., Insull W., Tsaltas T.T., Blomstrand R., Peterson M.L. (1957). The influence of dietary fats on serum-lipid levels in man. *Lancet* 1, 943-953.

Al-Jowder O., Kemsley E.K., Wilson R.H. (1997). Mid-infrared spectroscopy and authenticity problems in selected meats: a feasibility study. *Food Chemistry* 59, 195-201.

Alais C. (2010). *Scienza del latte*, 3^a edizione. *Tecniche Nuove*.

Ambring A., Johansson M., Axelsen M., Gan L., Strandvik B., Friberg P. (2006). Mediterranean-inspired diet lowers the ratio of serum phospholipid n-6 to n-3 fatty acids, the number of leukocytes and platelets, and vascular endothelial growth factor in healthy subjects. *American Journal of Clinical Nutrition* 83, 575-581.

Anderson K.M., Castelli W.P., Levy D. (1987). Cholesterol and mortality. 30 Years of followup from the Framingham Study. *Journal of American Medical Association* 257, 2176-2180.

Antongiovanni M., Lercker G., Secchiari P. (2007). *Metodi di analisi per lo studio della frazione lipidica del latte, dei prodotti di origine animale e degli alimenti zootecnici*. Edizioni Plus, Pisa University Press.

ARA (2011). <http://www.ara.sardegna.it/index.aspx?m=81&f=3&idf=287> (ultimo accesso 04/11/2013).

Ascherio A., Katan M.B., Zock P.L., Stampfer M.J., Willett W.C. (1999). Trans fatty acids and coronary heart disease. *The New England Journal of Medicine*. 340, 1994-1998.

Marco Caredda. Messa a punto e validazione di metodi analitici specifici online impieganti strumentazione a Trasformata di Fourier per la determinazione di macro e microanaliti in latte ovino. Tesi di dottorato in Scienze chimiche, Ciclo XXVI, Università degli studi di Sassari.

Baer R.J., Frank J.F., Loewenstein M. (1983a). Compositional analysis of nonfat dry milk using near infrared diffuse reflectance spectroscopy. *Journal of the Association of Official Analytical Chemists* 66, 858-863.

Baer R.J., Frank J.F., Loewenstein M., Birth G.S. (1983b). Compositional analysis of whey powders using near infrared diffuse reflectance spectroscopy. *Journal of Food Science* 48, 959-961.

Bang H.O., Dyerberg J., Nielsen A. (1971). Plasma lipid and lipoprotein pattern in Greenlandic West-coast Eskimos. *Lancet* 1, 1143-6.

Banks W., Clapperton J.L., Ferrie M.E. (1976). Effect of feeding fat to dairy cows receiving a fat-deficient basal diet. II. Fatty acid composition of the milk fat. *Journal of Dairy Research* 43, 219-227.

Banks W., Dalgleish D.G. (1990). Milk and milk processing, *in: Dairy microbiology of milk vol. 1: The microbiology of milk, 2nd edition* (ed. Robinson R. K.). London, Elsevier Science Publishers Ltd, pp.1-35.

Bauman D.E., Griinari J.M. (2003). Nutritional regulation of milk fat synthesis. *Annual Review of Nutrition* 23, 203-227.

Bauman D.E., Lock A.L., Corl B.A., Ip C., Salter A.M., Parodi P.W. (2005). Milk fatty acids and human health: potential role of conjugated linoleic acid and trans fatty acids, *in: Ruminant Physiology: Digestion, Metabolism and Impact of Nutrition on Gene Expression, Immunology and Stress.* (ed. Sejrsen K., Hvelplund T., Nielsen M.O.), Wageningen Academic Publishers, Wageningen, The Netherlands, Academic Publishers, pp. 523-555.

Bauman D.E., Lock A.L. (2006). Conjugated Linoleic Acid: Biosynthesis and Nutritional Significance, *in: Advanced Dairy Chemistry, Lipids, vol. 2, 3rd edition* (ed. Fox P.F.). Chapman and Hall, London, UK, pp.93-136.

Bauman D.E., Tyburczy C., O'Donnell A.M., Lock A.L. (2011). Milk Lipids: Conjugated Linoleic Acid. *Encyclopedia of Dairy Sciences (Second Edition)*, pp. 660-664.

Belury M.A., Nickel K.P., Bird C.E., Wu Y. (1996). Dietary conjugated Linoleic acid modulation of phorbol ester skin tumor promotion. *Nutrition and Cancer* 26, 149-157.

Ben Gera I., Norris K.H. (1968). Influence of fat concentration on the absorption spectrum of milk in the near infrared region. *Israel journal of agricultural research* 18, 117-124.

Bernhart F.W. (1961). Correlations between growth-rate of the suckling of various species and the percentage of total calories from protein in the milk. *Nature* 191, 358-360.

Boehm G., Stahl B. (2003). Oligosaccharides *in*: Functional dairy products (ed. Mattila-Sandholm T., Saarela M.). *Woodhead Publishing Ltd and CRC Press LLC*, pp. 203-243.

Bouchoux A., Gesan-Guiziu G., Perez J., Cabane B. (2010). How to squeeze a sponge: casein micelles under osmotic stress, a SAXS study. *Biophysical Journal* 99, 3754-3762.

Buccioni A., Serra A., Mele M., Petacchi F., Antongiovanni M., Secchiari P. (2004). A rapid method of extraction for the milk fat aimed at determining fatty acid composition. *Progress in nutrition* 6, 39-43.

Cabiddu A., Decandia M., Molle G., Pinna G., Addis M., Spada S., Pirisi A., Piredda G. (2003). Effect of different pastures on CLA content in milk and sheep cheese. *Italian Journal of Animal Science* 2, 518-520.

Calder P.C. (2009). Polyunsaturated fatty acids and inflammatory processes: New twists in an old tale. *Biochimie* 91, 791–795.

CAMO Software AS (2006a). The Unscrambler Appendices: Method References. <http://www.camo.com/downloads/U9.6%20pdf%20manual/The%20Unscrambler%20Method%20References.pdf> (ultimo accesso 11/11/2013).

CAMO Software AS (2006b). The Unscrambler Methods. <http://www.camo.com/downloads/U9.6%20pdf%20manual/The%20Unscrambler%20Methods.pdf> (ultimo accesso 11/11/2013).

Carl R.T. (1991). Quantification of the fat-content of milk using a partial-least-squares method of data-analysis in the near-infrared. *Fresenius Journal of Analytical Chemistry* 339, 70–71.

Casale M., Casolino C., Ferrari G., Forina, M. (2008). Near infrared spectroscopy and class modelling techniques for the geographical authentication of Ligurian extra virgin olive oil. *Journal of Near Infrared Spectroscopy* 16, 39–47.

Casale M., Sinelli N., Oliveri P., Di Egidio V., Lanteri S. (2010). Chemometrical strategies for feature selection and data compression applied to NIR and MIR spectra of extra virgin olive oils for cultivar identification. *Talanta* 80, 1832–1837.

Cattaneo T.M.P., Giardina C., Sinelli N., Riva M., Giangiacomo R. (2005). Application of FT-NIR and FT-IR spectroscopy to study the shelf-life of Crescenza cheese. *International Dairy Journal* 15, 693–700.

Chilliard Y., Ferlay A., Looor J., Rouel J., Martin B. (2002). Trans and conjugated fatty acids in milk from cows and goats consuming pasture or receiving vegetable oils or seeds. *Italian Journal of Animal Science* 1, 243-254.

Chin S.F., Liu W., Storkson J.M., Ha Y.L., Pariza M.W. (1992). Dietary sources of conjugated dienoic isomers of linoleic acid, a newly recognized class of anticarcinogens. *Journal of Food Composition and Analysis* 5, 185-197.

Chouinard P.Y., Corneau L., Saebo A., Bauman D.E. (1999). Milk yield and composition during abomasal infusion of conjugated linoleic acids in dairy cows. *Journal of Dairy Science* 82, 2737-2745.

Christie W.W. (1990). Gas Chromatography and Lipids. A practical guide. *The Oily Press, Bridgwater, Somerset*.

Christie W.W. (1993). Preparation of ester derivatives of fatty acids for chromatographic analysis, *in: Advances in Lipid Methodology*, Two (ed. Christie W.W.). *The Oily Press, Dundee*, pp. 69-111.

Christie W.W. (1995). Composition and structure of milk lipids, *in: Advanced Dairy Chemistry, Lipids*, vol. 2, 2nd edition (Fox P.F.). *Chapman and Hall, London, UK*, pp.1-36.

Colsenet R., Soderman O., Mariette F. (2005). Effect of casein concentration in suspensions and gels on poly(ethylene glycol)s NMR self-diffusion measurements. *Macromolecules* 38, 9171–9179.

Contu M. (2013). La qualità del latte ovino in Sardegna. Analisi e prospettive. ARA Sardegna <http://www.ara.sardegna.it/index.aspx?m=81&f=3&idf=319> (ultimo accesso 04/11/2013).

Coppa M., Ferlay A., Leroux C., Jestin M., Chilliard Y., Martin B., Andueza D. (2010). Prediction of milk fatty acid composition by near infrared reflectance spectroscopy. *International Dairy Journal* 20, 182–189.

Corradini C. (1995). Chimica e tecnologia del latte. *Tecniche Nuove*.

Cozzi, R., Protti P., Ruaro T. (2001a). Analisi Chimica Strumentale, seconda edizione, vol. B: Metodi ottici. *Zanichelli*.

Marco Caredda. Messa a punto e validazione di metodi analitici aspecifici online impieganti strumentazione a Trasformata di Fourier per la determinazione di macro e microanaliti in latte ovino. Tesi di dottorato in Scienze chimiche, Ciclo XXVI, Università degli studi di Sassari.

Cozzi, R., Protti P., Ruaro T. (2001b). *Analisi Chimica Strumentale*, seconda edizione, vol. C: Metodi Cromatografici, Metodi di misura e trattamento dei dati. *Zanichelli*.

Cozzolino D., Smyth H.E., Gishen M. (2003). Feasibility study on the use of visible and near-infrared spectroscopy together with chemometrics to discriminate between commercial white wines of different varietal origins. *Journal of Agricultural and Food Chemistry* 51, 7703–7708.

Creamer L.K., Berry G.P., Mills O.E. (1977). A study of the dissociation of β -casein from the bovine casein micelle at low temperature. *New Zealand Journal of Dairy Science & Technology* 12, 58–66.

Creamer L.K., MacGibbon A.K.H. (1996). Some recent advances in the basic chemistry of milk proteins and lipids. *International Dairy Journal* 6, 539–568.

Cunnane, S.C. (1992). What is the Nutritional and Clinical Significance of α -Linolenic Acid in Human? *in: Essential Fatty Acids and Eicosanoids* (ed. Sinclair A., Gibson R.), *AOCS Press, Champaign, Illinois*, pp.379–382.

Dalgleish D.G. (2011). On the structural models of casein micelles: review and possible improvements. *Soft Matter* 7, 2265–72.

Dalgleish D.G., Corredig M. (2012). The Structure of the Casein Micelle of Milk and Its Changes During Processing. *Annual Review of Food Science and Technology* 3, 2.1–2.19.

Davies A.M.C., Fearn T. (2006). Back to basics: calibration statistics. *Spectroscopy Europe* 18, 31-32.

De la Fuente M.A., Luna P., Juarez M. (2006). Chromatographic techniques to determine conjugated linoleic acid isomers. *Trends in Analytical Chemistry* 25, 917-926.

De Marchi M., Penasa M., Cecchinato A., Mele M., Secchiari P., Bittante G. (2011). Effectiveness of mid-infrared spectroscopy to predict fatty acid composition of Brown Swiss bovine milk. *Animal* 5, 1653–1658.

Denke M.A., Grundy S.M. (1992). Comparison of effects of lauric acid and palmitic acid on plasma lipids and lipoproteins. *American Journal of Clinical Nutrition* 56, 895-898.

Department Of Health, (1994). Nutritional aspects of cardiovascular disease. *Report on Health and Social Subjects* 46, HMSO, London.

Marco Caredda. Messa a punto e validazione di metodi analitici aspecifici online impieganti strumentazione a Trasformata di Fourier per la determinazione di macro e microanaliti in latte ovino. Tesi di dottorato in Scienze chimiche, Ciclo XXVI, Università degli studi di Sassari.

De Weijer A.P., Lucasius C.B., Buydens L.M.C., Kateman, G. (1993). Using Genetic Algorithms for an Artificial Neural Network Model Inversion. *Chemometrics and Intelligent Laboratory Systems* 20, 45–55.

Dhiman T.R., Anand, G.R., Satter, L.D., Pariza, M.W. (2000). Conjugated linoleic acid content of milk from cows fed different diets. *Journal of Dairy Science* 82, 2146-2156.

Di Egidio V., Sinelli N., Giovanelli G., Moles A., Casiraghi E. (2010). NIR and MIR spectroscopy as rapid methods to monitor red wine fermentation. *European Food Research and Technology* 230, 947-955.

Diplock A.T., Aggett P.J., Ashwell M., et al. (1999). Scientific concepts of functional foods in Europe: consensus document. *British Journal of Nutrition* 81(suppl. 1), S1.

Dong Y.W., Tu Z.H., Zhu D.Z., Liu Y.W., Wang Y.N., Huang J.L., Sun B.L., Fan Z.N. (2009). Feasibility of using NIR spectroscopy to detect melamine in milk. *Guang Pu Xue Yu Guang Pu Fen Xi/Spectroscopy and Spectral Analysis* 29, 2934-2938.

Dooper M.M., Wassink L., M'Rabet L., Graus Y.M. (2002). The modulatory effects of prostaglandin-E on cytokine production by human peripheral blood mononuclear cells are independent of the prostaglandin subtype. *Immunology* 107, 152–159.

Downey G., McIntyre P., Davies A.N. (2002). Detecting and quantifying sunflower oil adulteration in extra virgin olive oils from the eastern Mediterranean by visible and near infrared spectroscopy. *Journal of Agricultural and Food Chemistry* 50, 5520-5525.

Downey G., Fouratier V., Kelly J.D. (2003). Detection of honey adulteration by addition of fructose and glucose using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy* 11, 447–456.

Etzion Y., Linker R., Cogan U., Shmulevich I. (2004). Determination of Protein Concentration in Raw Milk by Mid-Infrared Fourier Transform Infrared/Attenuated Total Reflectance Spectroscopy. *Journal of Dairy Science* 87, 2779–2788.

Fagan C.C., O'Donnell C.P. (2008). Application of Mid-Infrared Spectroscopy to Food Processing Systems, in: *Nondestructive Testing of Food Quality*, (Ed. Irudayaraj J., Reh C.), IFT Press and Blackwell Publishing, pp. 119-142.

FAOSTAT (2011); <http://faostat3.fao.org/home/index.html>

Ferrand M., Huquet B., Barbey S., Barillet F., Faucon F., Larroque H., Leray O., Trommenschlager J.M., Brochard M. (2011). Determination of fatty acid profile in cow's

milk using mid-infrared spectrometry: Interest of applying a variable selection by genetic algorithms before a PLS regression. *Chemometrics and Intelligent Laboratory Systems* 106, 183-189.

FIL-IDF (2010). New applications of Mid Infra-Red Spectrometry for the analysis of milk and milk products. *Bulletin of the international dairy federation* 447, 2010.

FIL-IDF (2012). The World Dairy Situation 2012. *Bulletin of the IDF* No. 458/2012.

FitzGerald R.J. (2003). Caseinophosphopeptides (CPPs) as functional ingredients, *in*: Functional dairy products (ed. Mattila-Sandholm T., Saarela M.). *Woodhead Publishing Ltd and CRC Press LLC*, pp. 187-202.

Folch J., Lees M., Sloane-Stanley G.H. (1957). A Simple Method for the Isolation and Purification of Total Lipides from Animal Tissues. *Journal of Biological Chemistry* 226, 497-509.

Fonden R., Saarela M., Mättö J., Mattila-Sandholm T. (2003). Lactic acid bacteria (LAB) in functional dairy products *in*: Functional dairy products (ed. Mattila-Sandholm T., Saarela M.). *Woodhead Publishing Ltd and CRC Press LLC*, pp. 244-261.

Forrest S. (1993). Genetic Algorithms: Principles of Natural Selection Applied to Computation. *Science* 261, 872–878.

Fox P.F., McSweeney P.L.H. (1998). Dairy Chemistry and Biochemistry. *Blackie Academic & Professional*.

Frankhuizen R., Van der Veen N.G. (1985). Determination of major and minor constituents in milk powders and cheese by near infrared reflectance spectroscopy. *Netherlands Milk and Dairy Journal* 39, 191-207.

Frankhuizen R. (1992). Near-infrared analysis of dairy products, *in*: Handbook of near infrared analysis. (Ed. Burns D.A., Ciurczak E.W.). *Marcel Dekker Inc., New York*, Chapter 22.

Galtier O., Dupuy N., Le Dréau Y., Ollivier D., Pinatel C., Kister J., Artaud J. (2007). Geographic origins and compositions of virgin olive oils determined by chemometric analysis of NIR spectra. *Analytica Chimica Acta* 595, 136-144.

Gaullier J.M., Halse J., Høy K., Kristiansen K., Fagertun H., Vik H., Gudmundsen O. (2005). Supplementation with conjugated linoleic acid for 24 months is well tolerated by

and reduces body fat mass in healthy, overweight humans. *Journal of Nutrition* 135, 778-784.

Gewirtz A.T., Collier-Hyams L.S., Young A.N. (2003). Lipoxin A4 analogs attenuate induction of intestinal epithelial proinflammatory gene expression and reduce the severity of dextran sodium sulfate-induced colitis. *Journal of Immunology* 168, 5260–5267.

Giardina C., Cattaneo T.M.P., Barzagli S. (2003). Study of modifications in delactosated milk during shelf-life by NIR and FT-IR Spectroscopy. *Milchwissenschaft* 58, 363–366.

Giovannucci E., Goldin B. (1997). The role of fat, fatty acids, and total energy intake in the etiology of human colon cancer. *American Journal of Clinical Nutrition* 66, 1564S–1571S.

Goldman D.W., Pickett W.C., Goetzl E.J. (1983). Human neutrophil chemotactic and degranulating activities of leukotriene B₅ (LTB₅) derived from eicosapentaenoic acid. *Biochemical and Biophysical Research Communications* 117, 282–288.

Gomez-Carracedo M.P., Andrade J.M., Fernandez E., Prada D., Muniategui S. (2004). Evaluation of the pure apple juice content in commercial apple beverages using FT-MIR-ATR and potential curves. *Spectroscopy Letters* 37, 73–93.

Gordon D.J., Rifkind B.M. (1989). High-density lipoprotein - the clinical implications of recent studies. *The New England Journal of Medicine* 321, 1311–1316.

Goulden J.D.S. (1957). Diffuse reflection spectra of dairy products in the near infrared region. *Journal of Dairy Research* 24, 242-251.

Griinari M., Bauman D.E. (1999). Biosynthesis of Conjugated Linoleic Acid and its Incorporation into Meat and Milk in Ruminants, *in: Advances in Conjugated Linoleic Acid Research* vol.1, (ed. Yurawecz M.P., Mossoba M.M.). *AOCS Press, Champaign, Illinois*, pp.180-200.

Gurr M.I. (1999). Lipids in Nutrition and Health: a Reappraisal. *The Oily Press Lipid Library*.

Gurr M.I., Harwood L.L., Frayn K.N. (2002). Lipid Biochemistry, 5th edition. *Blackwell Science, Oxford*.

Hair Jr J.F., Black W.C., Babin B.J., Anderson R.E. (2009). Multivariate Data Analysis, 7th Edition. *Prentice Hall*.

Marco Caredda. Messa a punto e validazione di metodi analitici aspecifici online impieganti strumentazione a Trasformata di Fourier per la determinazione di macro e microanaliti in latte ovino. Tesi di dottorato in Scienze chimiche, Ciclo XXVI, Università degli studi di Sassari.

Hayes K.C., Khosla P. (1992). Dietary fatty acids thresholds and cholesterolaemia. *Faseb Journal* 6, 2600–2607.

Hara A., Radin N.S. (1978). Lipid Extraction of Tissues with a Low-Toxicity Solvent. *Analytical Biochemistry* 90, 420-426.

Harris W.S., Mozaffarian D., Rimm E., Kris-Etherton P., Rudel L.L., Appel L.J., Engler M.M., Engler M.B., Sacks F. (2009). Omega-6 fatty acids and risk for cardiovascular disease: A science advisory from the American Heart Association Nutrition Subcommittee of the Council on Nutrition, Physical Activity, and Metabolism; Council on Cardiovascular Nursing; and Council on Epidemiology and Prevention. *Circulation* 119, 902–907.

He Y., Li X.L., Deng X.F. (2007). Discrimination of varieties of tea using near infrared spectroscopy by principal component analysis and BP model. *Journal of Food Engineering* 79, 1238–1242.

Hegsted D.M., Ansman L.M., Johnson J.A., Dallal G.E. (1993). Dietary fat and serum lipids: an evaluation of the experimental data. *American Journal of Clinical Nutrition* 57, 875–883.

Heid H.W., Keenan T.W. (2005). Intracellular origin and secretion of milk fat globules. *European Journal of Cell Biology* 84, 245–58.

Hewavitharana A.K., Van Brakel B. (1997). Fourier transform infrared spectrometric method for the rapid determination of casein in raw milk. *Analyst* 122, 701–704.

Hilliam M. (1996). Functional foods: the western consumer viewpoint. *Nutrition Review* 54, S189-S194.

Hocquette J.F., Richardson R.I, Prache S., Medale F., Duffy G., Scollan N.D. (2005). The future trends for research on quality and safety of animal products. *Italian Journal of Animal Science* 4, 49-72.

Holt C. (1994). The biological function of casein: Research reviews, *in: Hannah Research Yearbook, Hannah Research Institute, Scotland*, pp. 60-68.

Horwitz W. (1982). *Analytical Chemistry* 54, 67A.

Howe G.R., Aronson K.J., Benito E., Castelleto R., Cornee J., DuVy S., Gallagher R.P., Iscovich J.M., Deng-ao J., Kaaks R., Kune G.A., Lee H.P., Lee M., Miller A.B., Peters R.K., Potter J.D., Riboli E., Slattery M.L., Trichopoulos D., Tuyns A., Tzonou A., Watson L.F., Whittemore A.S., Wu-Williams A.H., Shu Z. (1997). The relationship between

dietary fat intake and risk of colorectal cancer: evidence from the combined analysis of 13 case-control studies. *Cancer Causes Control*. 8, 215–228.

Huppertz T., Kelly A.L. (2006). Physical chemistry of milk fat globule, *in*: Advanced Dairy Chemistry, Lipids, vol. 2, 3rd edition (Fox P.F.). *Springer, New York*.

Inon F.A., Garrigues S., Guardia M. (2004). Nutritional parameters of commercially available milk samples by FTIR and chemometric techniques. *Analytica Chimica Acta* 513, 401–412.

Ip C. (1990). Quantitative assessment of fat and calories as risk factors in mammary carcinogenesis in an experimental model, *in*: Recent Progress in Research on Nutrition and Cancer (ed. Mettlan C.J., Aoki T.). *Wiley-Liss, New York*, pp. 107-117.

Ip C., Chin S.F., Scimeca J.A., Pariza M.A. (1991). Mammary cancer prevention by conjugated dienic derivatives of linoleic acid. *Cancer Research* 51, 6118-6124.

Ip C. (1997). Review of the effects of *trans* fatty acids, oleic acid, n-3 polyunsaturated fatty acids, and conjugated linoleic acid on mammary carcinogenesis in animals. *American Journal of Clinical Nutrition* 66, 1523S-1529S.

Ip M.M., Masso-Welch P.A., Ip C. (2003). Prevention of mammary cancer with conjugated linoleic acid: role of stroma and the epithelium. *Journal of Mammary Gland Biology and Neoplasia* 8, 103-118.

Isaksson T., Nilsen B.N., TØgersen G., Hammond R.P., Hildrum K.I. (1996). On-line, proximate analysis of ground beef directly at a meat grinder outlet. *Meat Science* 43, 245–253.

ISMEA, Istituto di Servizi per il Mercato Agricolo Alimentare (2006). La qualità come strategia per l'agro-alimentare italiano. *Ministero delle politiche agricole, alimentari e forestali*.

ISMEA (2012). http://www.sardegnaagricoltura.it/documenti/14_43_20130218144025.pdf (ultimo accesso 04/11/2013).

ISTAT (2011); <http://agri.istat.it/jsp/dawinci.jsp?q=plL020000010000012000&an=2011&ig=1&ct=155&id=69A|8A|9A|11A> (ultimo accesso 04/11/2013).

ISTAT (2012). http://www.istat.it/it/files/2013/09/Tavole_di_dati.zip?title=Prodotti+di+qualit%C3%A0+

Marco Caredda. Messa a punto e validazione di metodi analitici aspecifici online impieganti strumentazione a Trasformata di Fourier per la determinazione di macro e microanaliti in latte ovino. Tesi di dottorato in Scienze chimiche, Ciclo XXVI, Università degli studi di Sassari.

[DOP%2C+IGP+e+STG+-+18%2Fset%2F2013+-+Tavole+di+dati.zip](#) (ultimo accesso 04/11/2013).

Jenkins T.C. (1993). Lipid metabolism in the rumen. *Journal of Dairy Science* 76, 3851-3863.

Jennes R., Sloan R.E. (1970). The composition of milk of various species. *Dairy Science Abstracts* 32, 599.

Jensen R.G., Ferris A.M., Lammi-Keefe C.J. (1991). The composition of milk fat. *Journal of Dairy Science* 74, 3228-3243.

Jiang J., Bjoerck L., Fonden R., Emanuelson M. (1996). Occurrence of conjugated cis-9, trans-11-octadecadienoic acid in bovine milk: Effects of feed and dietary regimen. *Journal of Dairy Science* 79, 438-445.

Jones G., Willett P., Glen R.C. (1996). Genetic Algorithms for Chemical Structure Handling and Molecular Recognition, *in: Genetic Algorithms in Molecular Modeling* (ed. Devillers J.). *Academic Press: New York*, pp 211–242.

Jouan-Rimbaud D., Massart D.L., de Noord O.E. (1996). Random Correlation in Variable Selection for Multivariate Calibration with a Genetic Algorithm. *Chemometrics and Intelligent Laboratory Systems* 35, 213–220.

Judson R.S., Jaeger E.P., Treasurywala A.M., Peterson M.L. (1993). Conformational Searching Methods for Small Molecules. II. Genetic Algorithm Approach. *Journal of Computational Chemistry* 14, 1407–1414.

Karoui R., Bosset J.-O., Mazerolles G., Kulmyrzaev A., Dufour E. (2005). Monitoring the geographic origin of both experimental French Jura hard cheeses and Swiss Gruyère and L'Étivaz PDO cheeses using mid-infrared and fluorescence spectroscopies: A preliminary investigation. *International Dairy Journal* 15, 275–286.

Kawano S., Watanabe H., Iwamoto M. (1992). Determination of sugar content in intact peaches by near infrared spectroscopy with fiber optics in interaction mode. *Journal of the Japanese Society for Horticultural Science* 61, 445–451.

Kawano S., Fujiwara T., Iwamoto M. (1993). Nondestructive determination of sugar content in satsuma mandarin using near infrared (NIR) transmittance. *Journal of the Japanese Society for Horticultural Science* 62, 465–470.

Marco Caredda. Messa a punto e validazione di metodi analitici specifici online impieganti strumentazione a Trasformata di Fourier per la determinazione di macro e microanaliti in latte ovino. Tesi di dottorato in Scienze chimiche, Ciclo XXVI, Università degli studi di Sassari.

Keenan T.W., Dylewski D.P., Woodford T.A., Ford R.H. (1983). Origin of the milk fat globules and the nature of the milk fat globule membrane, *in*: Developments in Dairy Chemistry (Fox P.F.). *Applied Science Publishers, New York*, pp. 83–118.

Keenan T.W., Dylewski D.P. (1995). Intracellular origin of milk lipid globules and the nature and structure of the milk fat globule membrane, *in*: Advanced Dairy Chemistry, Lipids, vol. 2, (ed. Fox P.F.). *Chapman and Hall, London*, pp. 89–130.

Kelly J.F.D., Downey G., Fouratier V. (2004). Initial study of honey adulteration using mid-infrared (MIR) spectroscopy and chemometrics. *Journal of Agricultural and Food Chemistry* 52, 33–39.

Keys A., Anderson J.T., Grande F. (1957). Prediction of serum-cholesterol response of man to changes in fats in the diet. *Lancet* 273, 959–966.

Khanal R.C. (2004). Potential health benefits of conjugated linoleic acid (CLA): A review. *Asian-Australasian Journal of Animal Sciences* 17, 1315-1328.

Knekt P., Jarvinen R., Seppanen R., Pukkala E., Aromaa A. (1996). Intake of dairy products and the risk of breast cancer. *British Journal of Cancer* 73, 687–691.

Koca N., Rodriguez-Saona L.E., Harper W.J., Alvarez V.B. (2007). Application of Fourier Transform Infrared Spectroscopy for Monitoring Short-Chain Free Fatty Acids in Swiss Cheese. *Journal of Dairy Science* 90, 3596–3603.

Kolonel L.N. (2001). Fat, meat and prostate cancer. *Epidemiologic Reviews* 23, 72-81.

Kondepati V.R., Heise H.M. (2008). The potential of mid and near infrared spectroscopy for reliable monitoring of bioprocesses. *Current Trends in Biotechnology and Pharmacy* 2, 117-132.

Korhonen H., Pihlanto-Leppala A., Rantamaki P. and Tupasela T. (1998). Impact of processing on bioactive proteins and peptides. *Trends in Food Science & Technology* 8, 307-319.

Kritchevsky D. (1999). Conjugated linoleic acid and experimental atherosclerosis in rabbits, *in*: Advances in Conjugated Linoleic Acid Research, Vol. 1 (ed. Yurawecz M.P., Mossoba M.M., Kramer J.K.G., Pariza M.W., Nelson G.J.), *American Oil Chemists' Society Press, Champaign, Illinois*, pp. 397-403.

Kritchevsky D. (2003). Conjugated linoleic acid in experimental atherosclerosis. *in: Advances in Conjugated Linoleic Acid Research*, Vol. 2 (ed. Sebedio J-L., Christie W.W., Adlof R.O.), *American Oil Chemists' Society Press, Champaign, Illinois*, pp. 292-301.

Kurushima H., Hayashi K., Shingu T., Kuga Y., Ohtani H., Okura Y., Tanaka K., Yasunobu Y., Nomura K., Kajiyama G. (1995). Opposite effects on cholesterol metabolism and their mechanisms induced by dietary oleic acid and palmitic acid in hamsters. *Biochimica et Biophysica Acta* 1258, 251-256.

Kushi L., Giovannucci E. (2002). Dietary fat and cancer. *American Journal of Medicine* 113, 63S-70S.

Laporte M.F., Martel R., Paquin P. (1998). The near infrared optic probe for monitoring coagulation rennet coagulation in cow's milk. *International Dairy Journal* 8, 659-666.

Larsen T.M., Toubro S., Astrup A. (2003). Efficacy and safety of dietary supplements containing CLA for the treatment of obesity: Evidence from animal and human studies. *Journal of Lipid Research* 44, 2234-2241.

Lavillonniere F., Bougnoux P. (1999). Conjugated linoleic acid (CLA) and the risk of breast cancer, *in: Advances in conjugated linoleic acid research*, vol. 1 (ed. Yurawecz M.P., Mossoba M.M., Kramer J.K.G., Pariza M.W., Nelson, G.), *AOCS Press, Champaign*, pp. 276-282.

Learidi R., Boggia R., Terrile M. (1992). Genetic Algorithms as a Strategy for Feature Selection. *Journal of Chemometrics* 6, 267-281.

Learidi R., Lupianez Gonzalez A. (1998). Genetic Algorithms Applied to Feature Selection in PLS Regression: How and When to Use Them. *Chemometrics and Intelligent Laboratory Systems* 41, 195-207.

Learidi R. (2000). Application of Genetic Algorithm-PLS for Feature Selection in Spectral Data Sets. *Journal of Chemometrics* 14, 643-655.

Learidi R. (2009). Genetic Algorithms. *Comprehensive Chemometrics* 1, 631-653.

Learidi R. (2013). Methods of Treatment of Data, *in: Handbook of Water Analysis*, Third Edition (Ed. Nollet L.M.L., De Gelder L.S.P.), *CRC Press*.

Lee T.H., Mencia-Huerta J.M., Shih C. (1984). Characterization and biologic properties of 5,12-dihydroxy derivatives of eicosapentaenoic acid, including leukotriene-B5 and the double lipoxigenase product. *Journal of Biological Chemistry* 258, 2383-2389.

Marco Caredda. Messa a punto e validazione di metodi analitici aspecifici online impieganti strumentazione a Trasformata di Fourier per la determinazione di macro e microanaliti in latte ovino. Tesi di dottorato in Scienze chimiche, Ciclo XXVI, Università degli studi di Sassari.

Lee K.N., Kritchevsky D., Pariza M.W. (1994). Conjugated linoleic acid and atherosclerosis in rabbits. *Atherosclerosis* 108, 19-25.

Lewis R.A., Austen K.F., Soberman R.J. (1990). Leukotrienes and other products of the 5-lipoxygenase pathway: biochemistry and relation to pathobiology in human diseases. *New England Journal of Medicine* 323, 645-655.

Liew C., Schut H.A.J., Chin S.F., Pariza M.W., Dashwood R.H. (1995). Protection of conjugated linoleic acids against 2-amino-3-methylimidazo[4,5-f]quinoline-induced colon carcinogenesis in the F344 rat: a study of inhibitory mechanisms. *Carcinogenesis* 16, 3037-3043.

Liou Y.A., King J., Zibrik D., Innis S.M. (2007). Decreasing linoleic acid with constant α -linolenic acid in dietary fats increases (n-3) eicosapentaenoic acid in plasma phospholipids in healthy men. *Journal of Nutrition* 137, 945-952.

Lock A.L., Horne C.A.M., Bauman D.E., Salter A.M. (2005). Butter naturally enriched in conjugated linoleic acid and vaccenic acid alters tissue fatty acids and improves the plasma lipoprotein profile in cholesterol-fed hamsters. *Journal of Nutrition* 135, 1934-1939.

Lopez C. (2011). Milk fat globules enveloped by their biological membrane: Unique colloidal assemblies with a specific composition and structure. *Current Opinion in Colloid & Interface Science* 16, 391-404.

Lovegrove J., Jackson K. (2003). Coronary heart disease, *in*: Functional dairy products (ed. Mattila-Sandholm T., Saarela M.). *Woodhead Publishing Ltd and CRC Press LLC*, pp. 54-93.

Lovegrove J.A., Lovegrove S.S., Lesauvage S.V., Brady L.M., Saini N., Minihiane A.M., Williams C.M. (2004). Moderate fish-oil supplementation reverses low-platelet, long-chain n-3 polyunsaturated fatty acid status and reduces plasma triacylglycerol concentrations in British Indo-Asians. *American Journal of Clinical Nutrition* 79, 974-82.

Luginbühl W. (2002). Evaluation of Designed Calibration Samples for Casein Calibration in Fourier Transform Infrared Analysis of Milk. *LWT - Food Science and Technology* 35, 554-558.

Luinge H.J., Hop E., Lutz E.T.G., Vanhemert J.A., Dejong E.A.M. (1993). Determination of the fat, protein and lactose content of milk using Fourier-transform infrared spectrometry. *Analytica Chimica Acta* 284, 419-433.

Marco Caredda. Messa a punto e validazione di metodi analitici aspecifici online impieganti strumentazione a Trasformata di Fourier per la determinazione di macro e microanaliti in latte ovino. Tesi di dottorato in Scienze chimiche, Ciclo XXVI, Università degli studi di Sassari.

Luykx D.M.A.M., van Ruth S.M. (2008). An overview of analytical methods for determining the geographical origin of food products. *Food Chemistry* 107, 897–911.

MacGibbon A.K.H. and Taylor M.W. (2006). Composition and structure of bovine milk lipids, *in*: *Advanced Dairy Chemistry, Lipids*, vol. 2, 3rd edition (ed. P.F. Fox). *Springer, New York*.

Maraboli A., Cattaneo T.M.P., Giangiacomo R. (2002). Detection of vegetable proteins from soy, pea and wheat isolates in milk powder by near infrared spectroscopy. *Journal of Near Infrared Spectroscopy* 10, 63–69.

Marigheto N.A., Kemsley E.K., Defernez M., Wilson R.H. (1998). A comparison of mid-infrared and Raman spectroscopies for the authentication of edible oils. *Journal of the American Oil Chemists Society* 75, 987-992.

Marino C., Galloni P. (2011). Microwaves: Exposure and Potential Health Consequences, *in*: *Encyclopedia of Environmental Health*, (Ed. Nriagu J.O.), *Elsevier, Burlington*, pp. 765-773.

Martin J.C., Valeille K. (2002). Conjugated linoleic acids: all the same or to everyone its own function? *Reproduction Nutrition Development* 42, 525-536.

Martini M., Altomonte I., Salari F. (2013). Evaluation of the fatty acid profile from the core and membrane of fat globules in ewe's milk during lactation. *LWT - Food Science and Technology* 50, 253-258.

McKinney D.C., Lin M.D. (1994). Genetic algorithm solution of groundwater management models. *Water Resources Research* 30, 1897–1906.

McMahon D., Oomen B.S. (2008). Supramolecular structure of casein micelles. *Journal of Dairy Science* 91, 1709–1721.

McPherson A.V., Kitchen B.J. (1983). Review of the progress of dairy science. The bovine milk fat globule membrane – its information, composition, structure and behavior in milk and dairy products. *Journal of Dairy Research* 50, 107.

Mele M., Serra A., Comba F., Buccioni A., Conte G., Secchiari P., Priolo A., Biondi L., Salem H. B. (2007). Effect of the inclusion of soybean oil in the diet of dairy goats on meat fatty acid composition of their suckling kids. *Options Méditerranéennes. Série A, Séminaires Méditerranéens* 74, 177-182.

Marco Caredda. Messa a punto e validazione di metodi analitici aspecifici online impieganti strumentazione a Trasformata di Fourier per la determinazione di macro e microanaliti in latte ovino. Tesi di dottorato in Scienze chimiche, Ciclo XXVI, Università degli studi di Sassari.

Mensink R.P. (1993). Effects of individual saturated fatty acids on serum lipids and lipoprotein concentrations. *American Journal of Clinical Nutrition* 57, 711S–714S.

Mensink R.P., Zock P.L., Kester A.D.M., Katan M.B. (2003). Effects of dietary fatty acids and carbohydrates on the ratio of serum total to HDL cholesterol and on serum lipids and apolipoproteins: a meta-analysis of 60 controlled trials. *American Journal of Clinical Nutrition* 77, 1146–1155.

Miles E.A., Allen E., Calder P.C. (2002). In vitro effects of eicosanoids derived from different 20-carbon Fatty acids on production of monocyte-derived cytokines in human whole blood cultures. *Cytokine* 20, 215–223.

Miller G.J., Miller N.E. (1975). Plasma-high-density-lipoprotein concentration and development of ischaemic heart disease. *Lancet* 1, 7879, 16–19.

Mitchell M. (1996). An introduction to genetic algorithms (Complex Adaptive Systems). *MIT Press*.

Moloney F., Yeow T., Mullen A., Nolan J.J., Roche H.M. (2004). Conjugated linoleic acid supplementation, insulin sensitivity, and lipoprotein metabolism in patients with type 2 diabetes mellitus. *American Journal of Clinical Nutrition* 80, 887-895.

Moore J.H., Christie W.W. (1979). Lipid metabolism in the mammary gland of ruminant animals. *Progress in Lipid Research* 17, 347-395.

Mori T.A., Hodgson J.M. (2013). Health Effects of Omega-6 Polyunsaturated Fatty Acids, *in: Encyclopedia of Human Nutrition*, 3rd Edition, (ed. Allen L.H., Prentice A., Caballero B.), *Academic Press*, pp. 209-214.

Motoyama K.R., Curb J.D., Kadowaki T., El-Saed A., Abbott R.D., Okamura T., Evans R.W., Nakamura Y., Sutton-Tyrrell K., Rodriguez B.L., Kadota A., Edmundowicz D., Willcox B.J., Choo J., Katsumi N., Otake T., Kadowaki S., Kuller L.H., Ueshima H., Sekikawa A. (2009). Association of serum n-6 and n-3 polyunsaturated fatty acids with lipids in 3 populations of middle-aged men. *American Journal of Clinical Nutrition* 90, 49-55.

Murphy J.J., McNeill G.P., Connolly J.F., Gleeson P.A. (1990). Effect on cow performance and milk fat composition of including full fat soybean and rapeseed in the concentrate mixture for lactating dairy cows. *Journal of Dairy Research* 57, 295-306.

Ng C.L., Wehling R.L., Cuppett S.L. (2007). Method for determining frying oil degradation by near-infrared spectroscopy. *Journal of Agricultural and Food Chemistry* 55, 593–597.

Nicolai B.M., Beullens K., Bobelyn E., Peirs A., Saeys W., Theron K.I., Lammertyn J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. *Postharvest Biology and Technology* 46, 99-118.

Niemoller A., Behmer D. (2008). Use of Near Infrared Spectroscopy in the Food Industry, *in: Nondestructive Testing of Food Quality*, (Ed. Irudayaraj J., Reh C.), *IFT Press and Blackwell Publishing*, pp. 119-142.

Noone E.J., Roche H.M., Nugent A.P., Gibney M.J. (2002). The effect of dietary supplementation using isomeric blends of conjugated linoleic acid on lipid metabolism in healthy human subjects. *British Journal of Nutrition* 88, 243-251.

Nudda A., Battaccone G., Usai M.G., Fancellu S., Pulina G. (2006). Supplementation with extruded linseed cake affects concentrations of conjugated linoleic acid and vaccenic acid in goat milk. *Journal of Dairy Science* 89, 277-282.

O'Callaghan D.J., O'Donnell C.P., Payne F.A. (2000). On-line sensing techniques for coagulum setting in renneted milks. *Journal of Food Engineering* 43, 155–165.

Ochoa J.J., Farquharson A.J., Grant I., Moffat L.E., Heys S.D., Wahle K.W. (2004). Conjugated linoleic acids (CLAs) decrease prostate cancer cell proliferation: different molecular mechanisms for cis-9, trans-11 and trans-10, cis-12 isomers. *Carcinogenesis* 25, 1185-1191.

Ono T., Obata T. (1989). A model for the assembly of bovine casein micelles from F2 and F3 subunits. *Journal of Dairy Research* 56, 3, 453-461.

Owen T. (2000). Fondamenti di spettroscopia UV-visibile: Principi. *Agilent Technologies 2000*.

Palmquist D.L., Beaulieu A.D, Barbano D.M. (1993). Feed and animal factors influencing milk fat composition. *Journal of Dairy Science* 76, 1753-1771.

Pappas C.S., Tarantilis P.A., Moschopoulou E., Moatsou G., Kandarakis I., Polissiou M.G. (2008). Identification and differentiation of goat and sheep milk based on diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS) using cluster analysis. *Food Chemistry* 106, 1271–1277.

Marco Caredda. Messa a punto e validazione di metodi analitici specifici online impieganti strumentazione a Trasformata di Fourier per la determinazione di macro e microanaliti in latte ovino. Tesi di dottorato in Scienze chimiche, Ciclo XXVI, Università degli studi di Sassari.

Park Y.W., Juarez M., Ramos M., Haenlein G.F.W. (2007). Physico-chemical characteristics of goat and sheep milk. *Small Ruminant Research* 68, 88-113.

Pariza M.W., Hargraves W.A. (1985). A beef derived mutagenesis modulator inhibits initiation of mouse epidermal tumors by 7,12-dimethylbenz(a)anthracene. *Carcinogenesis* 6, 591-593.

Parodi, P.W. (1977). Conjugated octadecadienoic acids of milk fat. *Journal of Dairy Science* 60, 1150–1153.

Parodi P.W. (2004). Milk fat in human nutrition. *Australian Journal of Dairy Technology* 59, 3-58.

Parodi P.W. (2006). Nutritional Significance of Milk Lipids, *in: Advanced Dairy Chemistry, Lipids*, vol. 2, 3rd edition (ed. Fox P.F.). Springer, New York, pp. 601-639.

Parodi, P.W. (2009). Has the association between saturated fatty acids, serum cholesterol and coronary heart disease been over emphasized? *International Dairy Journal* 19, 345-361.

Patel J.V., Tracey I., Hughes E.A., Lip G.Y. (2010). Omega-3 polyunsaturated acids and cardiovascular disease: notable ethnic differences or unfulfilled promise? *Journal of Thrombosis and Haemostasis* 8, 2095–2104.

Pearson H. (2006). Genetics: what is a gene?. *Nature* 441, 398–401.

Pekkanen J., Linn S., Heiss G., Suchindran C.M., Leon A., Rifkind B.M., Tyroler H.A. (1990). Ten-year mortality from cardiovascular disease in relation to cholesterol level among men with and without preexisting cardiovascular disease. *New England Journal of Medicine* 322, 1700–1707.

Picque D., Cattenoz T., Corrieu G., Berger J.L. (2005). Discrimination of red wines according to their geographical origin and vintage year by the use of mid-infrared spectroscopy. *Sciences des Aliments* 25, 207-220.

Pillonel L., Luginbuhl W., Picque D., Schaller E., Tabacchi R., Bosset J.O. (2003). Analytical methods for the determination of the geographic origin of Emmental cheese: mid- and near-infrared spectroscopy. *European Food Research & Technology* 216, 174-178.

Pirisi A., Piredda G., Carta A. (2002). Resa di trasformazione del latte ovino: applicabilità alla produzione industriale di un'equazione ricavata mediante micro caseificazioni. *XV congresso nazionale SIPAOC, Cagliari*.

Pirisi A., Lauret A., Dubeuf J.P. (2007). Basic and incentive payments for goat and sheep milk in relation to quality. *Small Ruminant Research* 68, 167-178.

Pirisi A., Comunian R., Urgeghe P.P., Scintu M.F. (2011). Sheep's and goat's dairy products in Italy: Technological, chemical, microbiological, and sensory aspects. *Small Ruminant Research* 101, 102-112.

Polshin E., Aernouts B., Saeys W., Delvaux F., Delvaux F.R, Saison D., Hertog M., Nicolaï B.M., Lammertyn J. (2011). Beer quality screening by FT-IR spectrometry: Impact of measurement strategies, data pre-processings and variable selection algorithms. *Journal of Food Engineering* 106, 188–198.

Raynal-Ljutovac K., Park Y.W., Gaucheron F., Bouhallab S. (2007). Heat stability and enzymatic modifications of goat and sheep milk. *Small Ruminant Research* 68, 207-220.

Recio I., de la Fuente M.A., Juarez M., Ramos M. (2009). Bioactive Components in Sheep Milk, *in: Bioactive Components in Milk and Dairy Products* (ed. Park Y. W.). *Wiley-Blackwell*, pp.83-104.

Reh C. (2001). In-line and off-line FTIR measurement, *in: Instrumentation and sensors for the food industry*, 2nd ed. (Ed. Kress-Rogers E., Brimelow C.J.B). *Woodhead and CRC Press, Cambridge and Boca Raton, FL*, pp. 213-232.

Reid L.M., O'Donnell C.P., Downey G. (2006). Recent technological advances for the determination of food authenticity. *Trends in Food Science & Technology* 17, 344–353.

Ripoche A, Guillard AS. (2001). Determination of fatty acid composition of pork fat by Fourier transform infrared spectroscopy. *Meat Science* 58, 299-304.

Roberfroid M.B. (1997). The European concept of functional food. *VTT Symposium* 177, 161-170.

Roche H.M., Noone E., Nugent A., Gibney M.J. (2001). Conjugated linoleic acid: A novel therapeutic nutrient? *Nutrition Research Review* 14, 173-187.

Roussel S., Bellon-Maurel V., Roger J.M., Grenier P. (2003). Authenticating white grape must variety with classification models based on aroma sensors, FT-IR and UV spectrometry. *Journal of Food Engineering* 60, 407–419.

Marco Caredda. Messa a punto e validazione di metodi analitici aspecifici online impieganti strumentazione a Trasformata di Fourier per la determinazione di macro e microanaliti in latte ovino. Tesi di dottorato in Scienze chimiche, Ciclo XXVI, Università degli studi di Sassari.

Ruoff K., Kunzli R., Iglesias MT., Bogdanov S., Bosset, JO., der Ohe K., der Ohe W., Amado R. (2006). Authentication of the botanical and geographical origin of honey by mid-infrared spectroscopy. *Journal of Agricultural and Food Chemistry* 54, 6873–6880.

Rutten M.J.M, Bovenhuis H., Hettinga K.A., van Valenberg H.J.F. van Arendonk J.A.M. (2009). Predicting bovine milk fat composition using infrared spectroscopy based on milk samples collected in winter and summer. *Journal of Dairy Science* 92, 6202–6209.

Sadocchi S. (1980). Manuale di analisi statistica multivariata. *Franco Angeli Editore*.

Sato T., Yoshino M., Farukawa S., Someya Y., Yano N., Uozumi J., Iwamoto M. (1987). Analysis of milk constituents by the near infrared spectrophotometric method. *Japan Journal of Zootechnical Science* 58, 698–706.

Savitzky A., Golay M.J.E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry* 36, 1627–1639.

Schmidt D.G. (1982). Association of caseins and casein micelle structure, *in*: Developments in Dairy Chemistry, Proteins, vol. 1, (ed. Fox P.F.). *Chapman and Hall, London*, pp. 61–86.

Schønberg S., Krokan H.E. (1995). The inhibitory effect of conjugated dienoic derivatives (CLA) of linoleic acid on the growth of human tumor cell lines is in part due to increased lipid peroxidation. *Anticancer Research* 15, 1241-1246.

Scintu, M.F., Piredda G. (2007). Typicity and biodiversity of goat and sheep milk products. *Small Ruminant Research* 68, 221-231.

Serhan C.N., Jain A., Marleau S. (2003). Reduced inflammation and tissue damage in transgenic rabbits overexpressing 15-lipoxygenase and endogenous anti-inflammatory lipid mediators. *Journal of Immunology* 171, 6856–6865.

Shultz T.D., Chewa B.P., Seamana W.R., Luedecke L.O. (1992). Inhibitory effect of conjugated dienoic derivatives of linoleic acid and β -carotene on the in vitro growth of human cancer cells. *Cancer Letters* 63, 125-133.

Signorelli F., Contarini G., Annicchiarico G., Napolitano F., Orrù L., Catillo G., Haenlein G.F.W., Moioli B. (2008). Breed differences in sheep milk fatty acid profile: Opportunities for sustainable use of animal genetic resources. *Small Ruminant Research* 78, 24-31.

Sileoni V., Perretti G., Marte L., Marconi O., Fantozzi P. (2010). Near-Infrared spectroscopy for proficient quality evaluation of malt and maize in beer industry. *Journal of the Institute of Brewing* 116, 134-140.

Simopoulos A.P. (1995). Trans Fatty Acids, *in: Handbook of Lipids in Human Nutrition* (ed. Spiller G.A.), *CRC Press*, pp. 91-99.

Simopoulos A.P., Cleland L.G. (2003). Omega-6/Omega-3 Essential Fatty Acid Ratio: The scientific evidence. *World Review of Nutrition and Dietetics, Karger Publisher*, vol. 92.

Simopoulos A.P. (2008). The importance of the ratio of Omega-6/Omega-3 essential fatty acid. *Biomedicine & Pharmacotherapy* 56, 365-379.

Sivakesava S., Irudayaraj J.M.K., Korach R.L. (2001). Detection of adulteration in apple juice using mid infrared spectroscopy. *Applied Engineering in Agriculture* 17, 815–820.

Soyeurt H., Dardenne P., Dehareng F., Lognay G., Veselko D., Marlier M., Bertozzi C., Mayeres P., Gengler N. (2006). Estimating Fatty Acid Content in Cow Milk Using Mid-Infrared Spectrometry. *Journal of Dairy Science* 89, 3690–3695.

Soyeurt H., Dehareng F., Gengler N., McParland S., Wall E., Berry D.P., Coffey M. (2011). Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries. *Journal Of Dairy Science* 94, 1657-1667.

Stamler J., Wentworth D., Neaton J.D. (1986). Is relationship between serum cholesterol and risk of premature death from coronary heart disease continuous and graded? *Journal of the American Medical Association* 256, 2823–2828.

Stampfer M.J., Sacks F.M., Salvini S., Willett W.C., Hennekens C.H. (1991). A prospective study of cholesterol, apolipoproteins, and the risk of myocardial infarction. *The New England Journal of Medicine* 325, 373–381.

Steinberg D. (2006). An interpretive history of the cholesterol controversy, part V: the discovery of statins and the end of the controversy. *Journal of Lipid Research* 47, 1339–1351.

Stuart B.H. (2004). *Infrared Spectroscopy: Fundamentals and Applications (Analytical Techniques in the Sciences)*. *John Wiley & Sons Ltd, Chichester*.

Terpstra A. H. (2004). Effect of conjugated linoleic acid on body composition and plasma lipids in humans: An overview of the literature. *American Journal of Clinical Nutrition* 79, 352-361.

Marco Caredda. Messa a punto e validazione di metodi analitici aspecifici online impieganti strumentazione a Trasformata di Fourier per la determinazione di macro e microanaliti in latte ovino. Tesi di dottorato in Scienze chimiche, Ciclo XXVI, Università degli studi di Sassari.

Tewari J., Irudayaraj J. (2004). Quantification of saccharides in multiple floral honeys using Fourier transform infrared micro-attenuated total reflectance spectroscopy. *Journal of Agricultural and Food Chemistry* 52, 3237–3243.

Thomas E.V., Haaland D.M. (1990). Comparison of Multivariate Calibration Methods for Quantitative Spectral Analysis. *Analytical Chemistry* 62, 1091–1099.

Thomas E.V. (1994). A Primer on Multivariate Calibration. *Analytical Chemistry* 66, 795a–804a.

Thompson H.J., Zhu Z.J., Banni S., Darcy K., Loftus T., Ip C. (1997). Morphological and biochemical status of the mammary gland as influenced by conjugated linoleic acid: Implication for a reduction in mammary cancer risk. *Cancer Research* 57, 5067.5072.

Tilley S.L., Coffman T.M., Koller B.H. (2001). Mixed messages: modulation of inflammation and immune responses by prostaglandins and thromboxanes. *Journal of Clinical Investigation* 108, 15–23.

TØgersen G., Arnesen J.F., Nilsen B.N., Hildrum K.I. (2003). On-line prediction of chemical composition of semi-frozen ground beef by noninvasive NIR spectroscopy. *Meat Science* 63, 515–523.

Todeschini R. (1998). Introduzione alla chemiometria. *Edises*.

Toomey S., Roche H., Fitzgerald D., Belton O. (2003). Regression of pre-established atherosclerosis in the apoE(-/-) mouse by conjugated linoleic acid. *Biochemical Society Transactions* 31, 1075-1079.

Tricon S., Burdge G.C., Kew S., Banerjee T., Russell J.J., Jones E.L., Grimble R.F., Williams C.M., Yaqoob P., Calder P.C. (2004). Opposing effects of *cis*-9, *trans*-11 and *trans*-10, *cis*-12 conjugated linoleic acid on blood lipids in healthy humans. *American Journal of Clinical Nutrition* 80, 614-620.

Tsenkova R., Atanassova S., Itoh K., Ozaki Y., Toyoda K. (2000). Near infrared spectroscopy for biomonitoring: Cow milk composition measurement in a spectral region from 1.100 to 2.400 nanometers. *Journal of Animal Science* 78, 515-522.

UE (2013).

http://ec.europa.eu/agriculture/quality/door/list.html;jsessionid=pL0hLqqLXhNmFQyF11b24mY3t9dJQPflg3xbL2YphGT4k6zdWn34!-370879141?recordStart=0&recordPerPage=100&filter.dossierNumber=&filter.comboName=&filterMin.milestone_mask=&filterMin.milestone=&filterMax.milestone_mask=&filt

[erMax.milestone=&filter.country=&filter.category=&filter.type=&filter.status=](#) (ultimo accesso 28/10/2013).

Urbano Cuadrado M., Luque de Castro M.D., Pérez Juan P.M., Gómez-Nieto M.A. (2005). Comparison and joint use of near infrared spectroscopy and Fourier transform mid infrared spectroscopy for the determination of wine parameters. *Talanta* 66, 218-224.

Urquhart P., Parkin S.M., Rogers J.S., Bosley J.A., Nicolaou A. (2002). The effect of conjugated linoleic acid on arachidonic acid metabolism and eicosanoid production in human saphenous vein endothelial cells. *Biochimica et Biophysica Acta* 1580, 150-160.

Van de Voort F.R. (1992). Fourier Transform Infrared Spectroscopy applied to food analysis. *Food Research International* 25, 397-403.

Van de Voort F.R., Ismail A.A., Sedman J. (1995). A rapid, Automated Method for the Determination of cis and trans Content of Fats and oils by Fourier Transform Infrared Spectroscopy. *Journal of the American Oil Chemists' Society* 72, 873-880.

Venkatasubramanian V., Chan K., Caruthers J.M. (1994a). Computer-Aided Molecular Design Using Genetic Algorithms. *Computers & Chemical Engineering* 18, 833-844.

Venkatasubramanian V., Chan K., Caruthers J.M. (1994b). On the Performance of Genetic Search for Large-Scale Molecular Design, *in: Proceedings of the Fifth International Symposium on Process Systems Engineering (PSE'94). Kyongu: S. Korea*, pp 1001-1006.

Verbise Y. (1981). Mesure d'humidité en continu par absorption infrarouge dans les industries agro-alimentaires et alimentaires (On-line moisture determination by infrared absorption in food industry). *Industries Agricoles et Alimentaires* 9, 759-768.

Verschuren W.M.M., Jacobs D.R., Bloemberg B.P.M., Kromhout D., Menotti A., Aravanis C., Blackburn H., Buzina R., Dontas A.S., Fidanza F., Karvonen M.J., Nedeljkovic S., Nissinen A., Toshima H. (1995). Serum total cholesterol and long-term coronary heart disease mortality in different cultures. Twenty-Five-year follow-up of the Seven Countries Study. *Journal of the American Medical Association* 274, 131-136.

Walstra, P., Jenness R. (1984). Dairy chemistry and physics. *John Wiley & Sons, Inc., New York, NY*.

Wang Y.W., Jones P.J.H. (2004). Dietary conjugated linoleic acid and body composition. *American Journal of Clinical Nutrition* 79, 1153S-1158S.

Marco Caredda. Messa a punto e validazione di metodi analitici aspecifici online impieganti strumentazione a Trasformata di Fourier per la determinazione di macro e microanaliti in latte ovino. Tesi di dottorato in Scienze chimiche, Ciclo XXVI, Università degli studi di Sassari.

Weaver R.W.V. (1984). Near infrared reflectance analysis applied to dairy products. Abstracts of Papers; Challenges to contemporary dairy analytical techniques; *Royal Society of Chemistry: Reading, U.K.*

Webb B.H., Johnson A.H., Alford J.A. (1974). *Fundamentals of Dairy Chemistry*, AVI Publ. Co., Westport.

Webb E.C. (1992). International Union of Biochemistry and Molecular Biology. Nomenclature Committee Enzyme Nomenclature 1992. *Academic Press*.

Weber P.C. (1989). Are we what we eat? Fatty acids in nutrition and in cell membranes: cell functions and disorders induced by dietary conditions. *Svanoybukt, Norway: Svanoy Foundation (Report no. 4)*, pp. 9–18.

White S.L., Bertrand J.A., Wade M.R., Washburn S.P., Green J.T., Jenkinst C. (2001). Comparison of fatty acid content of milk from Jersey and Holstein cows consuming pasture or a total mixed ration. *Journal of Dairy Science* 84, 2295-2301.

Willett W.C. (2001). Diet and Cancer: One View at the Start of the Millennium. *Cancer Epidemiology, Biomarkers & Prevention* 10, 3-8.

Willett W.C. (2002). Dietary fat plays a major role in obesity: no. *Obesity Review* 3, 59-68.

Woodcock T., Downey G., Kelly J.D., O'Donnell C.P. (2007). Geographical classification of honey samples by near-infrared spectroscopy: A feasibility study. *Journal of Agricultural and Food Chemistry* 55, 9128–9134.

World Health Organization (2003). Diet nutrition and the prevention of chronic diseases. *Report of the WHO/FAO Joint Expert Consultation, WHO, Technical Report Series 916*.

Yu Y., Correll P.H., Vanden Heuvel J.P. (2002). Conjugated linoleic acid decreases production of pro-inflammatory products in macrophages: Evidence for a PPAR gamma-dependent mechanism. *Biochimica et Biophysica Acta* 1581, 89-99.

Yuan S.L., He Y., Ma T.Y., Wu D., Nie P.C. (2009). Fast determination of melamine content in milk base on Vis/NIR spectroscopy method. *Guang Pu Xue Yu Guang Pu Fen Xi/Spectroscopy and Spectral Analysis* 29, 2939-2942.

Yurawecz M.P., Mossoba M.M., Kramer J.K.G., Pariza M.W., Nelson G.J. (1999). *Advances in Conjugated Linoleic Acid Research*, vol.1, AOCS Press, Champaign, Illinois, USA pp. 215-223.

Marco Caredda. Messa a punto e validazione di metodi analitici aspecifici online impieganti strumentazione a Trasformata di Fourier per la determinazione di macro e microanaliti in latte ovino. Tesi di dottorato in Scienze chimiche, Ciclo XXVI, Università degli studi di Sassari.

Zhang R.H., Mustafa A.F., Zhao X. (2006). Effects of feeding oilseeds rich in linoleic and linolenic fatty acids to lactating ewes on cheese yield and on fatty acid composition of milk and cheese. *Animal Feed Science and Technology* 127, 220-233.

Zhang C., Shen Y., Chen J., Xiao P., Bao J. (2008). Nondestructive prediction of total phenolics, flavonoid contents, and antioxidant capacity of rice grain using near-infrared spectroscopy. *Journal of Agricultural and Food Chemistry* 56, 8268-8272.

Ringraziamenti

I miei più sentiti e sinceri ringraziamenti al Prof. Gavino Sanna che con grande professionalità ha supportato scientificamente questo lavoro di Tesi, e per i suoi preziosi insegnamenti.

Ringrazio il Dott. Giovanni Piredda e la Dott.ssa M. Francesca Scintu per aver messo a mia disposizione i laboratori dell'Agenzia Agris Sardegna e per la disponibilità dimostrata nella risoluzione di problemi scientifici ed organizzativi.

Ringrazio la Dott.ssa Margherita Addis per la sua assistenza e per gli insegnamenti sulla chimica del latte.

Ringrazio il Dott. Ignazio Ibba, responsabile del Laboratorio di Analisi del Latte dell'ARA Sardegna, per la disponibilità dimostrata, e per avermi introdotto nel mondo dei controlli di qualità della Sardegna in campo lattiero-caseario.

Ringrazio il Prof. Riccardo Leardi per i preziosi insegnamenti di chemiometria e di algoritmi genetici, per la sua disponibilità e la sua pazienza.

Ringrazio i colleghi ed amici di Agris Sardegna, in particolare il dott. Carlo Piga (che per primo mi ha introdotto nei meandri della chemiometria), la dott.ssa Myriam Fiori, la dott.ssa Claudia Zazzu, la dott.ssa Cristina De Candia, la dott.ssa Margherita Addis, il dott. Marco Serra, e i Sig.ri Daniele Petretto, Adriano Porcu e Marzia Picinelli, che con la loro presenza hanno creato un clima di lavoro sereno.

Un grazie va ai miei zii Francesca, Nicola e Giacomina per il loro sostegno e la loro presenza; a Stefano e Marta che, con Elisa e Lucia son riusciti a strapparmi più di un sorriso; ed alla mia nuova famiglia che mi fa sempre sentire a casa.

Il ringraziamento più grande va però ad Ilaria, per la sua presenza, i suoi incoraggiamenti, e la sua pazienza. Grazie di esserci.

