



**UNIVERSITÀ DEGLI STUDI DI PADOVA**  
**DIPARTIMENTO DI INGEGNERIA INDUSTRIALE**

**TESI DI LAUREA MAGISTRALE IN**  
**INGEGNERIA CHIMICA E DEI PROCESSI INDUSTRIALI**

**MONITORAGGIO DELLA QUALITÀ DI PRODOTTI**  
**ALIMENTARI LUNGO LA FILIERA DI LAVORAZIONE:**  
**UN CASO DI STUDIO SULLO SPECK**

*Relatore: Prof. Massimiliano Barolo*

*Correlatori: Prof. Enrico Novelli*

*Ing. Pierantonio Facco*

*Laureando: LUCIO BOSCHETTI*

ANNO ACCADEMICO 2011 - 2012



*Est modus in rebus.*



# Riassunto

In questa Tesi si propone l'utilizzo della spettroscopia infrarossa (*near infrared spectroscopy*, NIRS) e dell'analisi d'immagine per il monitoraggio della qualità di un salume tipico del Tirolo, lo speck, lungo i principali stadi della filiera di lavorazione. È stata condotta un'analisi esplorativa preliminare molto dettagliata per approfondire la conoscenza del processo produttivo, evidenziando le correlazioni tra le proprietà più rilevanti della carcassa e del prodotto stagionato. Quindi, sono state utilizzate analisi NIRS e immagini digitali per stimare i maggiori indicatori di qualità dello speck (attività dell'acqua, umidità, sostanza secca, contenuto di NaCl e ceneri), con ottimi risultati soprattutto a partire dagli spettri NIR. La notevole rapidità ed accuratezza di queste acquisizioni suggerisce la possibilità di utilizzarle in una applicazione in linea, in sostituzione delle analisi chimiche dirette, costose in termini di tempi e risorse. Inoltre, l'utilizzo delle immagini digitali ha consentito di costruire una tecnica per la classificazione della qualità molto rapida ed efficace, evitando di ricondursi al giudizio sensoriale, soggettivo e non ripetibile, di un ispettore.



# Indice

Introduzione . . . . .	1
<b>1 Dati sulla lavorazione dello speck: struttura della filiera, misure e strumentazione . . . . .</b>	<b>5</b>
1.1 Dati disponibili . . . . .	5
1.2 Analisi del sangue sull'animale vivo . . . . .	6
1.3 Dati di filiera . . . . .	6
1.4 Analisi spettroscopiche NIR . . . . .	10
1.5 Analisi Instron . . . . .	13
1.6 Analisi colorimetriche . . . . .	13
1.7 Immagini digitali . . . . .	15
1.8 Valutazione sensoriale . . . . .	16
<b>2 Metodi di statistica multivariata . . . . .</b>	<b>19</b>
2.1 <b>Analisi delle componenti principali (PCA) . . . . .</b>	<b>19</b>
2.1.1 <i>Interpretazione dei risultati: analisi delle correlazioni . . . . .</i>	<i>22</i>
2.1.2 <i>Distanza euclidea multidimensionale tra loading . . . . .</i>	<i>23</i>
2.1.3 <i>Distanza euclidea multidimensionale tra correlation loading. . . . .</i>	<i>24</i>
2.1.4 <i>Mappe di correlazione. . . . .</i>	<i>27</i>
2.2 <b>Proiezione su strutture latenti (PLS) . . . . .</b>	<b>28</b>
2.2.1 <i>Pretrattamenti delle analisi spettroscopiche NIR . . . . .</i>	<i>30</i>
2.2.2 <i>Robustezza e interpretazione dei risultati . . . . .</i>	<i>31</i>
2.3 <b>Analisi delle correlazioni canoniche (CCA). . . . .</b>	<b>32</b>
2.3.1 <i>Forma ed interpretazione dei risultati . . . . .</i>	<i>34</i>
2.4 <b>Analisi d'immagine . . . . .</b>	<b>37</b>
2.4.1 <i>Analisi del colore medio . . . . .</i>	<i>39</i>
2.4.2 <i>Analisi della tessitura - decomposizione wavelet . . . . .</i>	<i>41</i>
2.5 <b>Metodiche di classificazione . . . . .</b>	<b>43</b>
2.5.1 <i>Analisi di sensitività e specificità . . . . .</i>	<i>45</i>
<b>3 Analisi esplorativa preliminare e studio delle correlazioni . . . . .</b>	<b>47</b>
3.1 <b>Dati di filiera . . . . .</b>	<b>47</b>
3.1.1 <i>Caratteristiche dell'animale vivo. . . . .</i>	<i>47</i>
3.1.2 <i>Caratteristiche della carcassa macellata. . . . .</i>	<i>49</i>
3.1.3 <i>Analisi cromatografica del lardo fresco . . . . .</i>	<i>51</i>
3.1.4 <i>Variabili di qualità del prodotto stagionato . . . . .</i>	<i>53</i>

3.1.5	<i>Caratteristiche dell'animale vivo e della carcassa macellata</i> . . . . .	54
3.1.6	<i>Caratteristiche dell'animale vivo ed analisi cromatografica del lardo fresco</i>	55
3.1.7	<i>Caratteristiche dell'animale vivo e qualità del prodotto stagionato</i> . . . . .	57
3.1.8	<i>Caratteristiche della carcassa macellata e qualità del prodotto stagionato</i>	58
<b>3.2</b>	<b>Analisi spettroscopica NIR</b> . . . . .	<b>60</b>
3.2.1	<i>Carcassa macellata</i> . . . . .	60
3.2.2	<i>Prodotto stagionato</i> . . . . .	63
<b>3.3</b>	<b>Analisi Instron</b> . . . . .	<b>66</b>
<b>3.4</b>	<b>Analisi colorimetriche</b> . . . . .	<b>68</b>
<b>3.5</b>	<b>Conclusioni</b> . . . . .	<b>70</b>
<b>4</b>	<b>Stima delle variabili chiave per il monitoraggio della qualità del prodotto</b> . . . . .	<b>73</b>
<b>4.1</b>	<b>Costruzione dei modelli PLS</b> . . . . .	<b>73</b>
<b>4.2</b>	<b>Qualità della carcassa macellata</b> . . . . .	<b>74</b>
4.2.1	<i>Analisi spettroscopiche NIR</i> . . . . .	75
<b>4.3</b>	<b>Qualità dello speck stagionato</b> . . . . .	<b>78</b>
4.3.1	<i>Analisi colorimetriche</i> . . . . .	79
4.3.2	<i>Analisi Instron</i> . . . . .	81
4.3.3	<i>Analisi spettroscopiche NIR</i> . . . . .	82
4.3.3.1	Parte magra . . . . .	82
4.3.3.2	Fetta intera macinata . . . . .	84
4.3.3.3	Strumento portatile . . . . .	86
4.3.4	<i>Analisi d'immagine</i> . . . . .	87
4.3.4.1	Analisi del colore medio . . . . .	87
4.3.4.2	Analisi della tessitura - decomposizione wavelet . . . . .	89
<b>4.4</b>	<b>Conclusioni</b> . . . . .	<b>91</b>
<b>5</b>	<b>Classificazione dei campioni rispetto al sesso, ai regimi alimentari e ai parametri sensoriali</b> . . . . .	<b>93</b>
<b>5.1</b>	<b>Sesso dei suini</b> . . . . .	<b>93</b>
<b>5.2</b>	<b>Regime alimentare tradizionale o arricchito</b> . . . . .	<b>95</b>
<b>5.3</b>	<b>Parametri sensoriali</b> . . . . .	<b>99</b>
5.3.1	<i>Giudizi sul colore</i> . . . . .	99
5.3.2	<i>Giudizi sulla consistenza</i> . . . . .	101
5.3.3	<i>Giudizi sul rapporto magro/grasso</i> . . . . .	103
5.3.3.1	Confronto delle valutazioni assegnate . . . . .	106
<b>5.4</b>	<b>Conclusioni</b> . . . . .	<b>107</b>
	<b>Conclusioni e sviluppi futuri</b> . . . . .	<b>109</b>



---

Appendice A Lista dei simboli . . . . .	111
Appendice B Lista degli acronimi . . . . .	113
Riferimenti bibliografici . . . . .	114



# Elenco delle figure

1.1	Esempio di costruzione di $n$ variabili binarie a partire da una variabile categoriale con $n$ distinte classi di appartenenza. . . . .	7
1.2	Immagini esemplificative delle diverse facce dei campioni e dell'assegnazione delle specifiche sigle; il campione di riferimento è il numero 25. (a) lato A; (b) lato B; (c) lato C; (d) lato D. . . . .	11
1.3	Esempi di spettri NIR, campione 20: (a) carcassa macellata, strumento Foss; (b) carcassa macellata, strumento Unity; (c) prodotto stagionato, strumento Foss; (d) prodotto stagionato, strumento Unity; (e) ogni faccia del prodotto stagionato, con distinzione di parte magra e grassa. . . . .	12
1.4	Curve sperimentali ottenute dall'analisi Istron, campione 20: singole ripetute allineate in linee tratteggiate, profilo medio in grassetto. . . . .	14
1.5	Equalizzazione dei singoli canali RGB: campione 20, fetta intera lato A. (a) immagine originale; (b) immagine equalizzata. . . . .	16
2.1	Decomposizione della matrice di dati originali $\mathbf{X}$ nelle matrici $\mathbf{T}$ e $\mathbf{P}$ di score e loading, a meno di una matrice di residui $\mathbf{E}$ . . . . .	20
2.2	Rappresentazione geometrica dell'analisi PCA; significato di loading $\mathbf{p}_i$ , equivalenti ai coseni direttori del nuovo asse principale, e degli score $\mathbf{t}_i$ , equivalenti alla coordinata spaziale del nuovo asse principale. . . . .	22
2.3	Rappresentazione dei risultati di un'analisi PCA, relativi alle prime due componenti principali: (a) diagramma dei loading ; (b) diagramma degli score. . . . .	23
2.4	Distanza multidimensionale tra i loading delle cinque variabili analizzate in Figura 2.3. . . . .	24
2.5	Esempio di rappresentazione dei correlation loading; le cinque variabili analizzate sono le medesime delle Figure 2.3 e 2.4. . . . .	25
2.6	Distanza multidimensionale tra i correlation loading delle cinque variabili analizzate. . . . .	26
2.7	Distanza multidimensionale tra i loading delle cinque variabili analizzate - scala cromatica adattata. . . . .	26
2.8	Esempio di una mappa di correlazione, relativa alle stesse cinque variabili. . . . .	27
2.9	Diagramma a blocchi del metodo CCA. . . . .	32
2.10	Esempio di rappresentazione classica dei risultati della CCA, relativi alle prime due funzioni canoniche: (a) canonical loading; (b) coefficienti di struttura. . . . .	36

2.11 Nuova rappresentazione proposta dei risultati della CCA: (a) funzione canonica 1; (b) funzione canonica 2. . . . .	36
2.12 Esempi di applicazioni di maschere, relative nello specifico al campione 20 lato A: (a) immagine originale e (b) relativo istogramma degli score; (c) parte magra e cotenna; (d) maschera della parte magra nel diagramma degli score; (e) parte grassa ed (f) relativa maschera nel diagramma degli score. .	38
2.13 Diagramma a blocchi della procedura di segmentazione. . . . .	40
2.14 Istogrammi dell'intensità luminosa nei tre canali RGB, campione 20 lato A.	41
2.15 Esempio di applicazione della trasformata wavelet. . . . .	42
2.16 Attribuzione delle classi del metodo PLSDA mediante distribuzioni di probabilità e relative soglie. . . . .	44
2.17 Caratterizzazione dei modelli di classificazione in base ai valori di sensibilità e specificità. . . . .	46
3.1 Mappa di correlazione relativa alle variabili della sottomatrice Animale vivo.	48
3.2 Mappa di correlazione relativa alle variabili della sottomatrice Carcassa macellata. . . . .	49
3.3 Mappa di correlazione relativa alle variabili della sottomatrice Cromatografia.	52
3.4 Mappa di correlazione relativa alle variabili della sottomatrice Prodotto stagionato. . . . .	53
3.5 Mappa di correlazione relativa alle variabili delle sottomatrici Animale vivo e Carcassa macellata. . . . .	55
3.6 Mappa di correlazione relativa alle variabili delle sottomatrici Animale vivo e Cromatografia. . . . .	56
3.7 Mappa di correlazione relativa alle variabili delle sottomatrici Animale vivo e Prodotto stagionato. . . . .	57
3.8 Mappa di correlazione relativa alle variabili delle sottomatrici Carcassa macellata e Prodotto stagionato. . . . .	59
3.9 Intervalli di lunghezze d'onda associati alla vibrazione dei più comuni gruppi funzionali organici. . . . .	61
3.10 Lunghezze d'onda correlate agli acidi grassi monoinsaturi, ottenute mediante CCA: (a) spettro medio Foss della carcassa macellata; (b) C16:1 (4.5%); (c) C18:1n9 (86.5%); (d) C20:1n9 (1.7%); (e) sommatoria MUFA. . . . .	62
3.11 Lunghezze d'onda correlate, ottenute mediante CCA: (a) spettro medio Foss dello speck stagionato; (b) attività dell'acqua; (c) umidità del magro; (d) sostanza secca. . . . .	64
3.12 Lunghezze d'onda correlate, ottenute mediante CCA: (a) NaCl; (b) ceneri.	65
3.13 Mappa di correlazione relativa alle sole variabili dell'Analisi Instron. . . . .	66

---

3.14	Mappa di correlazione relativa ai dataset Dati di filiera ed Analisi Instron, concatenati in un'unica matrice. . . . .	67
3.15	Mappa di correlazione relativa al dataset Colore. . . . .	68
3.16	Mappa di correlazione relativa ai dataset Dati di filiera e Colore. . . . .	69
4.1	Distribuzione degli errori relativi medi nella stima del pH a 45 minuti da spettro Foss della carcassa macellata, pretrattamento SNV + derivata seconda (ottimale). . . . .	77
5.1	Distribuzione degli errori di classificazione della variabile sesso, con metodo QDA calibrato sugli spettri NIR Foss della parte magra dello speck stagionato, pretrattamento (5). . . . .	94
5.2	Caratterizzazione dei modelli di classificazione in base ai valori di sensitività e specificità. . . . .	95
5.3	Distribuzione degli errori di classificazione dei regimi alimentari, con metodo LDA calibrato sui Dati di filiera. . . . .	96
5.4	Caratterizzazione dei modelli di classificazione in base ai valori di sensitività e specificità dei diversi regimi alimentari: (a) standard; (b) standard + polifenoli; (c) standard + PUFA. . . . .	97
5.5	Collocazione della parte magra e grassa di ogni campione nello spazio tridimensionale RGB. . . . .	100
5.6	Distribuzione dell'intensità luminosa della parte magra, relativa ai canali del rosso e del verde. . . . .	100
5.7	Distribuzione dell'intensità luminosa della parte grassa, relativa ai canali del verde e del blu. . . . .	101
5.8	Profilo medio delle analisi Instron e relativa area sottesa. . . . .	102
5.9	Distribuzione della consistenza della parte magra dello speck, ottenuta mediante analisi Instron. Si riporta la media delle misure in grassetto e l'intervallo $\pm\sigma$ in linea tratteggiata. . . . .	102
5.10	Distribuzione della consistenza della parte grassa dello speck, ottenuta mediante analisi Instron. . . . .	102
5.11	Applicazione di maschere per la selezione della: (a) parte magra e cotenna; (b) parte grassa. . . . .	104
5.12	Eliminazione della cotenna dalla parte magra. . . . .	104
5.13	Fitting per la determinazione della relazione analitica tra il rapporto magro/grasso ed il rispettivo giudizio numerico. . . . .	105
5.14	Stima del rapporto magro/grasso del campione n° 44: (a) immagine originale; (b) maschera sulla parte magra; (c) maschera sulla parte grassa. . . .	106

5.15 Scostamento di giudizio tra l'analisi d'immagine, oggettiva, e la valutazione sensoriale del giudice. . . . .	106
--	-----

# Elenco delle tabelle

1.1	Dettaglio di tutti i dataset disponibili. . . . .	5
1.2	Elenco delle variabili del dataset Dati di filiera, suddivise in sottomatrici caratteristiche delle varie fasi della filiera produttiva. . . . .	8
1.3	Classificazione comunitaria EUROP delle carcasse suine. . . . .	9
1.4	Elenco delle variabili aggiuntive, specifiche del dataset Analisi Instron. . . . .	14
1.5	Elenco delle variabili colorimetriche aggiuntive, specifiche del dataset Analisi colorimetriche. . . . .	14
1.6	Elenco delle caratteristiche analizzate durante la valutazione sensoriale e dei rispettivi range di attribuzione. . . . .	17
2.1	Esempio di rappresentazione tabellare dei risultati della CCA. . . . .	35
3.1	Elenco delle principali correlazioni individuate all'interno della sottomatrice Animale vivo. . . . .	48
3.2	Elenco delle principali correlazioni individuate all'interno della sottomatrice Carcassa macellata. . . . .	50
3.3	Elenco delle principali correlazioni individuate all'interno della sottomatrice Cromatografia. . . . .	52
3.4	Elenco delle principali correlazioni individuate all'interno della sottomatrice Prodotto stagionato. . . . .	54
3.5	Elenco delle principali correlazioni individuate tra la sottomatrice dell'Animale vivo e quella della Carcassa macellata. . . . .	55
3.6	Elenco delle principali correlazioni individuate tra la sottomatrice dell'Animale vivo e quella della Cromatografia. . . . .	56
3.7	Elenco delle principali correlazioni individuate tra la sottomatrice dell'Animale vivo e quella del Prodotto stagionato. . . . .	58
3.8	Elenco delle principali correlazioni individuate tra la sottomatrice della Carcassa macellata e quella del Prodotto stagionato. . . . .	59
3.9	Elenco delle principali correlazioni individuate tra i dataset Dati di filiera ed Analisi Instron. . . . .	67
3.10	Elenco delle principali correlazioni individuate tra i dataset Dati di filiera e Colore . . . . .	69
4.1	Elenco dei coefficienti di variazione degli indicatori di qualità analizzati. . . . .	74
4.2	Statistiche dei risultati nella stima del pH a 45 minuti con un numero di simulazioni pari a 1000. . . . .	76

4.3	Statistiche dei risultati nella stima del pH a 45 minuti con un numero di simulazioni pari a 10000. . . . .	76
4.4	Statistiche dei risultati nella stima del n° di iodio a partire dalle analisi NIR della carcassa. . . . .	78
4.5	Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato a partire dalle analisi colorimetriche sulla parte magra. . . . .	79
4.6	Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato a partire dalle analisi colorimetriche sulla parte grassa. . . . .	80
4.7	Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato a partire dalle analisi Instron sulla parte magra. . . . .	81
4.8	Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato a partire dalle analisi Instron sulla parte grassa. . . . .	81
4.9	Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato a partire dagli spettri NIR Foss sulla parte magra. . . . .	83
4.10	Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato a partire dagli spettri NIR Unity sulla parte magra. . . . .	83
4.11	Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato a partire dagli spettri NIR Foss sulla fetta intera macinata. . . . .	85
4.12	Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato a partire dagli spettri NIR Unity sulla fetta intera macinata. . . . .	85
4.13	Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato a partire dagli spettri NIR sul lato A - strumento portatile. . . . .	86
4.14	Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato a partire dalle analisi sul colore medio, lati A e B. . . . .	88
4.15	Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato a partire dalle analisi sul colore medio, lati C e D. . . . .	88
4.16	Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato ottenute mediante decomposizione wavelet - approccio multi-approssimazione, lati A e B. . . . .	90
4.17	Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato ottenute mediante decomposizione wavelet - approccio multi-dettaglio, lati A e B. . . . .	90
4.18	Statistiche dei risultati nella stima delle variabili NaCl e ceneri, ottenute mediante decomposizione wavelet dei lati esterni C e D. . . . .	91
4.19	Elenco delle prestazioni migliori ottenute nella stima delle variabili di qualità dello speck. . . . .	92



---

5.1	Errori medi ottenuti nella classificazione dei campioni in base al sesso dei suini; errori assoluti sui 6 campioni di convalida, mediati tra le 1000 simulazioni indipendenti. . . . .	94
5.2	Errori medi ottenuti nella classificazione dei campioni in base al regime alimentare dei suini. . . . .	96
5.3	Elenco della base di attribuzione dei giudizi numerici, riferiti a precisi rapporti magro/grasso. . . . .	105



# Introduzione

Lo speck è un salume stagionato dal sapore molto intenso tipico del Tirolo, una regione situata tra il nord Italia ed il sud dell'Austria. È ottenuto dalle cosce di maiale con una complessa lavorazione, in cui carni altamente selezionate sono sottoposte a fasi alternate di stagionatura ed affumicatura. Il processo di lavorazione inizia con una salatura delle carni ed un insaporimento con una tradizionale ricetta di aromi, tra cui pepe, ginepro, rosmarino ed alloro; dopo tre settimane di stagionatura, lo speck viene affumicato con legni poco resinosi ed a bassa temperatura (circa 20°C), in modo lento e graduale. Infine, dopo altre 22 settimane di stagionatura, esso è pronto per essere avviato alla commercializzazione e consumato. Negli ultimi anni questo prodotto ha acquisito un'importanza sempre più rilevante nel mercato internazionale, soprattutto dopo l'ottenimento dell'Indicazione Geografica Protetta (IGP) da parte dell'Unione Europea. Questa certificazione di qualità richiede il rispetto di requisiti molto stringenti in tutte le fasi della filiera produttiva, dall'allevamento alla lavorazione delle carni, al fine di assicurare un'eccellente qualità finale dello speck stagionato (*Regolamento Europeo 1107/96*). Quest'ultima è influenzata da numerose proprietà, tra cui la diminuzione del pH dopo la macellazione, il quantitativo e la composizione del grasso, e la presenza residua di acqua nella parte magra del prodotto stagionato. Una rapida diminuzione di pH, legata alla decomposizione anaerobica delle riserve di glicogeno, provoca la denaturazione delle proteine presenti nei tessuti, facilmente osservabile poiché le carni assumono una colorazione molto chiara e sono caratterizzate da una notevole perdita di liquidi (Andersen *et al.*, 1999). Questa degradazione può essere minimizzata trattando con particolare cura gli animali prima della macellazione e con un rapido raffreddamento delle carcasse (Maribo *et al.*, 1998); quindi, il monitoraggio del pH assume particolare rilevanza anche per l'ottimizzazione dei tunnel di raffreddamento. Il quantitativo di acidi grassi polinsaturi (*poly unsaturated fatty acids*, PUFA) influenza invece la conservabilità delle carni, poiché essi vengono rapidamente ossidati causando irrancidimento e perdita di proprietà cromatiche (Wood *et al.*, 2003). Infine, l'umidità e l'attività dell'acqua sono strettamente legati alla tessitura ed alla consistenza del prodotto stagionato, che rappresentano due dei criteri più importanti per ottenere la certificazione IGP. Il notevole gradiente di umidità a cui lo speck è sottoposto durante la stagionatura può infatti causare una disidratazione eccessiva in superficie ed una consistenza gommosa all'interno (Morales *et al.*, 2007); ciò provoca una maggiore difficoltà al taglio ed una sensazione sgradevole al palato.

Il controllo di qualità dei produttori richiede il rispetto ed il monitoraggio di questi parametri lungo l'intera filiera di lavorazione, dalla selezione delle carni alla stagionatura del prodotto. Inoltre, lo speck viene esaminato da ispettori, che ne giudicano il livello

in base a parametri sensoriali (sapore, colore, consistenza) e variabili chimiche (attività dell'acqua, contenuto di NaCl, ceneri, sostanza secca). Tuttavia, alcune di queste misure sono soggettive e poco ripetibili, perché dipendono dal personale giudizio dell'ispettore; altre, invece, richiedono analisi chimiche costose in termini di tempi e risorse, in alcuni casi distruttive sul campione, e generatrici di reflui da trattare. Emerge quindi l'esigenza di poter monitorare la qualità del prodotto mediante tecnologie rapide, accurate e non invasive, estendibili anche ad un controllo di qualità in linea. Una delle tecniche più promettenti a tal proposito è la spettroscopia al vicino infrarosso (*near infrared spectroscopy*, NIRS), la cui efficacia nelle applicazioni all'industria alimentare è riconosciuta e consolidata in letteratura (Ottavian *et al.*, 2012a,b; Prieto *et al.*, 2009). Inoltre, negli ultimi tempi sta acquisendo sempre maggiore interesse l'analisi d'immagine, poiché le fotocamere digitali rappresentano uno strumento di analisi molto diffuso ed economico, potenzialmente molto versatile ed interessante.

In questa Tesi è stata condotta un'analisi esplorativa molto approfondita al fine di evidenziare le variabili più correlate alla qualità in tutte le fasi della lavorazione, in modo da approfondire la conoscenza del processo e dei legami tra i diversi stadi. Sono state applicate metodiche di statistica multivariata (analisi delle componenti principali ed analisi delle correlazioni canoniche), necessarie per analizzare al meglio l'elevato numero di variabili a disposizione, altamente correlate tra loro. Quindi, utilizzando la proiezione su strutture latenti, si propone l'utilizzo della spettroscopia NIR e dell'analisi d'immagine per il monitoraggio e la stima della qualità dello speck lungo l'intera filiera di lavorazione. Vengono utilizzate ad ulteriore confronto anche analisi colorimetriche e analisi di tessitura Instron. Inoltre, viene analizzata l'efficacia di alcune modifiche di processo rispetto alla qualità dello speck ottenuto; tali variazioni riguardano l'arricchimento del regime alimentare tradizionale e la castrazione degli esemplari maschio, e sono entrambe finalizzate all'ottenimento di una qualità finale più elevata ed omogenea. Infine, rispondendo ad un'altra esigenza industriale legata alla necessità di una classificazione di qualità più oggettiva, vengono proposte delle tecniche di analisi multivariata d'immagine per selezionare in modo rapido ed efficace i campioni migliori, in termini di colore, consistenza (di entrambe le porzioni) e rapporto magro/grasso.

La Tesi è organizzata come segue. Nel Capitolo 1 vengono descritte tutte le tipologie di dati e di analisi a disposizione, mentre nel Capitolo 2 si introducono le metodiche di statistica multivariata utilizzate in questa Tesi. Segue quindi, nel Capitolo 3, un'analisi esplorativa dei dati, con uno studio approfondito di tutte le correlazioni più significative. Nel Capitolo 4 si stima la qualità della carcassa e dello speck stagionato a partire da analisi spettroscopiche NIR, analisi colorimetriche, analisi Instron ed immagini digitali. Infine, nel Capitolo 5, si utilizzano le immagini digitali per costruire una tecnica di classificazione rapida ed oggettiva, verificando altresì la possibilità di individuare gli effetti specifici di

alcune modifiche al processo.



# Capitolo 1

## Dati sulla lavorazione dello speck: struttura della filiera, misure e strumentazione

In questo Capitolo si illustra la struttura dei dati a disposizione, che comprendono un'ampia gamma di variabili sulla produzione di speck; il contenuto informativo è molto diversificato, con dettagli inerenti tutta la filiera produttiva e con specifici approfondimenti sulla carcassa macellata e sul prodotto stagionato da avviare alla commercializzazione.

Prima di procedere nell'analisi è quindi opportuno avere una piena conoscenza dei vari set di dati, delle modalità di acquisizione e di eventuali trattamenti preliminari, necessari per adattare la tipologia di dati all'analisi a cui devono essere sottoposti.

### 1.1 Dati disponibili

Si riporta in Tabella 1.1 un elenco completo dei dataset a disposizione. I dati raccolti si riferiscono a 36 campioni di speck stagionato ed alle rispettive carcasse da cui i tagli di carne sono stati ricavati; di questi, 30 costituiscono i campioni di calibrazione, utilizzati per tarare i parametri dei vari modelli, mentre 6 sono estratti in modo casuale per avere un gruppo di campioni di convalida, necessari per verificare l'accuratezza dei modelli su campioni non considerati in fase di calibrazione.

**Tabella 1.1** *Dettaglio di tutti i dataset disponibili.*

<b>Dataset</b>	<b>Dettagli</b>
Analisi del sangue sull'animale vivo	a 15 e 40 giorni dall'inizio della somministrazione del diverso mangime
Dati di filiera	caratteristiche dell'animale vivo caratteristiche della carcassa macellata analisi cromatografica del lardo fresco qualità dello speck stagionato
Analisi spettroscopiche NIR	carcassa macellata (Foss e Unity) parte magra dello speck stagionato (Foss e Unity) fetta intera dello speck stagionato (Foss e Unity) ogni faccia dello speck stagionato (portatile)

Analisi Instron	parte magra dello speck stagionato parte grassa dello speck stagionato
Analisi colorimetriche	parte magra dello speck stagionato parte grassa dello speck stagionato
Immagini ad alta risoluzione	ogni faccia dello speck stagionato
Valutazione sensoriale	giudizi e votazioni numeriche sulla qualità dello speck stagionato

Si discutono ora le principali caratteristiche di ciascun raggruppamento di dati, evidenziando le variabili di maggior interesse, gli indici di qualità ed i rispettivi limiti di specifica; vengono inoltre descritti gli strumenti utilizzati, la preparazione dei campioni e gli eventuali trattamenti preliminari sulle acquisizioni.

## 1.2 Analisi del sangue sull'animale vivo

Le analisi del sangue sugli esemplari vivi comprendono due distinti prelievi ematici, a 15 ed a 40 giorni dall'inizio della somministrazione dei mangimi sperimentali; il dataset è costituito da alcune caratteristiche dell'animale - sesso e tipologia di alimentazione - e da specifiche variabili ematiche determinate dopo il prelievo, al fine di verificare l'eventuale influenza del regime alimentare sulla composizione sanguigna.

Tuttavia, questa serie di dati è disponibile solo per un numero esiguo di campioni (13 su 36), e ciò limita le analisi statistiche a valle poiché, per avere consistenza, è necessario ridurre anche il numero di campioni di tutti gli altri set di dati, limitando l'analisi ai soli campioni per cui sono disponibili tutte le diverse tipologie di acquisizioni.

## 1.3 Dati di filiera

Come riportato in Tabella 1.1, il dataset è suddiviso in sottomatrici che raggruppano delle classi comuni di variabili lungo l'intera filiera produttiva, a partire dalle caratteristiche dell'animale vivo fino alle variabili di qualità del prodotto stagionato. A tal proposito, il dataset ha richiesto una fase di preparazione preliminare volta alla costruzione di alcune variabili categoriali, necessarie per garantire un corretto utilizzo del modello in fase di classificazione; una variabile di attribuzione di classe  $\mathbf{x}_i$  costituita da  $n$  distinti valori<sup>1</sup> è stata quindi suddivisa in  $n$  diverse variabili binarie 0-1, in cui si assegna il valore 1 ai campioni che appartengono alla specifica classe. Si chiarisce in Figura 1.1 la procedura

<sup>1</sup>Non di derivazione sperimentale, ma per assegnare semplicemente una classe di appartenenza.



utilizzata, con riferimento alle diverse giornate di macellazione dei suini.

Giornata di macellazione		1° giornata	2° giornata	3° giornata
1	→	1	0	0
3		0	0	1
2		0	1	0
2		0	1	0
3		0	0	1

**Figura 1.1** Esempio di costruzione di  $n$  variabili binarie a partire da una variabile categoriale con  $n$  distinte classi di appartenenza.

Ciò è necessario per attribuire lo stesso peso a tutte le diverse classi di appartenenza: infatti, se gli  $n$  valori numerici non fossero sostituiti con  $n$  colonne binarie, si forzerebbero delle correlazioni in realtà non presenti. Si riporta in Tabella 1.2 l'elenco completo delle variabili a disposizione.

Con riferimento alle variabili n° (26), (29) e (80), valgono rispettivamente le seguenti definizioni:

$$\text{resa (\%)} = 1 - \frac{\text{peso vivo} - \text{peso morto caldo}}{\text{peso morto caldo}} \quad (1.1)$$

$$\text{perdita (\%)} = \frac{\text{peso morto caldo} - \text{peso morto freddo}}{\text{peso morto caldo}} \quad (1.2)$$

EE SS = Estratto Etereo in % sulla Sostanza Secca (grasso in termini quantitativi)

La sottomatrice *Animale vivo* comprende le informazioni sul regime alimentare, la giornata di macellazione, l'azienda di nascita, il sesso ed il peso del suino. Ad eccezione delle variabili (24) e (25), tutte le altre sono variabili categoriali poiché individuano la classe di appartenenza, senza rappresentare una misura sperimentale diretta sul campione stesso<sup>2</sup>. Le variabili di maggior interesse sono la (1), (2) e (3), poiché si vuole analizzare l'eventuale incidenza del regime alimentare sugli indicatori di qualità del prodotto finale; i mangimi sperimentali sono arricchiti con un derivato delle olive, cioè (2) standard + polifenoli, o con semi di lino, cioè (3) standard + acidi grassi polinsaturi (PUFA; *poly-unsaturated fatty acids*).

Le variabili della *Carcassa macellata* rappresentano invece una serie di misure per quantificare gli spessori di grasso e di carne magra, mediante analisi in più punti<sup>3</sup>, sia nella parte interna che in quella esterna della carcassa. In base a questi, è possibile risalire alla

<sup>2</sup>In questo caso è da considerarsi una variabile categoriale anche il sesso (23) poiché, all'interno di essa, è stato attribuito il valore 0 agli esemplari maschi ed il valore 1 alle femmine.

<sup>3</sup>Le sigle 1T e 13T si riferiscono a diverse zone della carcassa.

**Tabella 1.2** *Elenco delle variabili del dataset Dati di filiera, suddivise in sottomatrici caratteristiche delle varie fasi della filiera produttiva.*

<b>Animale vivo</b>		<b>Carcassa macellata</b>		<b>Cromatografia</b>	
1	alimentazione standard	26	resa (%)	49	n° di iodio
2	standard + polifenoli	27	peso morto caldo	50	C10:0
3	standard + PUFA	28	peso morto freddo	51	C12:0
4	macello: 1° giornata	29	perdita (%)	52	C14:0
5	macello: 2° giornata	30	spessore di grasso	53	C14:1
6	macello: 3° giornata	31	spessore di carne	54	C16:0
7	azienda di nascita 1	32	carne magra (%)	55	C16:1
8	azienda di nascita 2	33	qualità media	56	C18:0
9	azienda di nascita 3	34	qualità alta	57	C18:1n9
10	azienda di nascita 4	35	qualità altissima	58	C18:2n6
11	azienda di nascita 5	36	lardo 1T totale	59	C18:3n6
12	madre 1	37	lardo 1T esterno	60	C18:3n3
13	madre 2	38	lardo 1T interno	61	C20:0
14	madre 3	39	carne 1T	62	C20:1n9
15	madre 4	40	lardo 13T totale	63	C20:2
16	madre 5	41	lardo 13T esterno	64	C20:3n6
17	madre 6	42	lardo 13T interno	65	C20:4n6
18	madre 7	43	carne 13T	66	C20:3n3
19	padre 1	44	lardo minimo totale	67	acidi saturi
20	padre 2	45	lardo minimo esterno	68	acidi monoinsaturi
21	padre 3	46	lardo minimo interno	69	acidi polinsaturi
22	padre 4	47	drip loss	70	acidi n3
23	sexso	48	pH a 45 minuti	71	acidi n6
24	incremento netto				
25	peso vivo				

<b>Prodotto stagionato</b>	
72	calo peso (%)
73	attività dell'acqua
74	pH
75	umidità del magro
76	NaCl
77	sostanza secca (%)
78	ceneri
79	proteina grezza
80	EE SS calcolato
81	acidità titolabile

percentuale di carne magra, variabile (32), ed attribuire una classe di qualità alla carcassa stessa, variabili (33), (34) e (35), secondo la classificazione comunitaria di Tabella 1.3 (Regolamento CEE 3220/84, aggiornato dal CEE 3513/94).

Vengono inoltre rilevati i pesi morti a caldo ed a freddo, e la relativa perdita in peso dovuta sia al gocciolamento del muscolo, (48) *drip loss*, che all'evaporazione dell'acqua contenu-

**Tabella 1.3** *Classificazione comunitaria EUROP delle carcasse suine.*

Carne magra (%)	Classe di qualità
> 55%	E (altissima)
50 ÷ 55	U (alta)
45 ÷ 50	R (media)
40 ÷ 45	O (bassa)
< 40	P (bassissima)

ta nei tessuti. Una variabile di primaria importanza è il pH a 45 minuti (49) perché è strettamente legato alla qualità del prodotto finale, in partiar modo al colore; è stabilita una soglia minima pari a 5.8, mentre valori molto bassi sono riconducibili ad uno stato di stress dell'animale in fase di macellazione: sotto stress, la carne che ne deriva è molto chiara, caratterizzata da una forte perdita di acqua e da un contenuto di NaCl più elevato. Un'altra variabile chiave è la perdita in peso (29), che può suggerire la necessità di una diversa gestione delle celle frigorifere o dei tunnel di raffreddamento (Gigieli *et al.*, 1989; Savell *et al.*, 2005)

Da ogni carcassa macellata è stato quindi prelevato un campione di lardo fresco, sottoposto alla misura del numero di iodio (49) e ad un'analisi cromatografica completa degli acidi grassi; questi ultimi, espressi sia per singolo valore che per classe di appartenenza, vanno a costituire la sottomatrice *Cromatografia*.

I singoli acidi grassi devono essere letti nel modo seguente:

$C_n:m \implies n = \text{numero di atomi di carbonio della catena};$   
 $m = \text{numero di insaturazioni presenti}.$

Un parametro molto importante è rappresentato dal numero di iodio che, attraverso la titolazione degli alogeni, dà una stima del grado di insaturazione degli acidi grassi presenti. Tale variabile è particolarmente importante perché, oltre a dare un'indicazione sulla qualità della dieta dell'esemplare vivo, influenza in modo significativo la qualità del prodotto finale. A tal proposito, estendendo l'applicazione dei prosciutti crudi di Parma e San Daniele, è stato fissato un limite di specificità pari a 70 che, sebbene non sia un riferimento di legge, è considerato un'ottimo indice tecnologico per garantire una buona stabilità del prodotto (Decreto del Ministero dell'Industria, del Commercio e dell'Artigianato del 18/12/1993). Infine, sono state raccolte nella sottomatrice *Prodotto stagionato* alcune variabili di qualità dello speck stagionato ed affumicato, pronto per la commercializzazione; tali indicatori riguardano il pH, l'umidità, il contenuto di sale, le ceneri e l'acidità. Tra queste dieci variabili di qualità, l'attività dell'acqua è senza dubbio il parametro più importante, utilizzato sia per i controlli di qualità interni del produttore, sia come indice di valore legale qualora richiesto in un disciplinare di produzione. Inoltre, può essere richiesta anche per la caratterizzazione dei requisiti igienici dei prodotti *ready to eat* (Regolamento CEE 2073/05).

## 1.4 Analisi spettroscopiche NIR

Le analisi spettroscopiche al vicino infrarosso (*near infrared, NIR*) rappresentano uno dei mezzi più efficaci per esaminare la struttura e la chimica di un materiale, poiché ogni gruppo funzionale è caratterizzato da un picco di assorbimento in un ristretto intervallo di lunghezze d'onda. Analizzando lo spettro NIR di una sostanza è quindi possibile risalire alla sua struttura chimica, verificando la presenza dei picchi caratteristici di uno o più gruppi funzionali in corrispondenza a specifiche lunghezze d'onda.

Il dataset in questione è costituito da parecchie analisi NIR, nello specifico:

1. analisi della carcassa macellata;
2. analisi dello speck stagionato;
3. analisi di ogni faccia dello speck stagionato.

Le acquisizioni 1. e 2. sono state eseguite con due strumenti diversi, il Foss NIRSystem 5000 e lo Unity SpectraStar 2500. Il primo è caratterizzato da un intervallo di analisi pari a 1100-2500 nm, passo 2 nm, *detector* PbSe, mentre il secondo ha un *range* di 680-2500 nm, passo 1 nm, e rilevatori InGaAs ( $\lambda < 1250$  nm) e Ge ( $\lambda > 1250$  nm).

Lo strumento portatile è invece un Grainit RXGrains, caratterizzato da un'accuratezza inferiore e da un intervallo di analisi compreso tra 300 e 1100 nm<sup>4</sup>, con passo di 2 nm, sensore *Monolithic Miniature Spectrometer 1* (MMS1) ed ottica Carl Zeiss. Nello specifico, le facce analizzate sono:

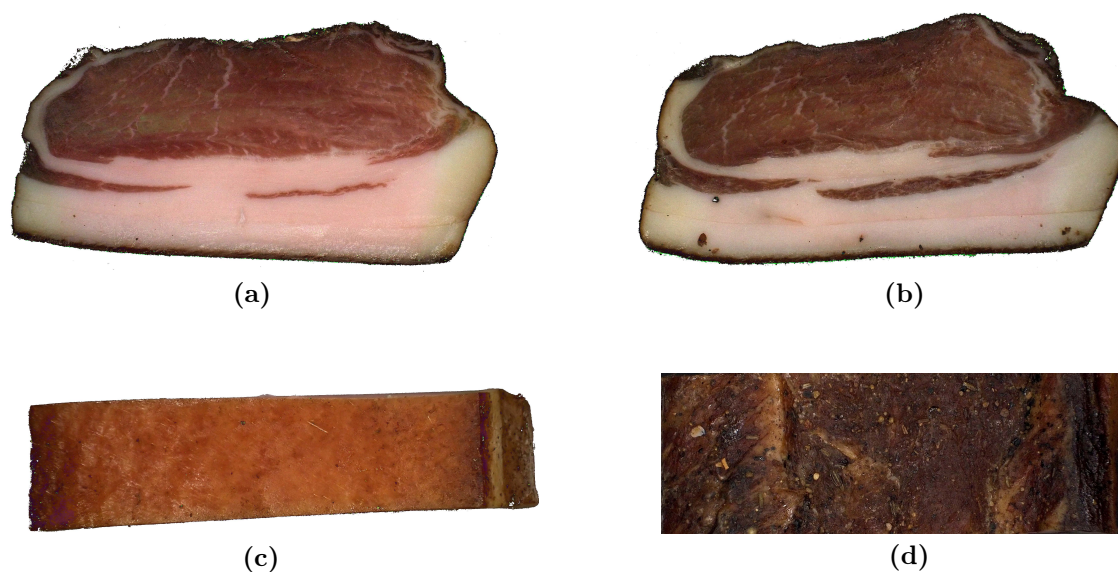
- lato A: una fetta laterale, comunemente esposta al taglio; è costituita da una parte magra e da una grassa, e costituisce una fonte di dati particolarmente significativa perché molte delle variabili di qualità dello speck stagionato sono relative a proprietà interne al prodotto, non analizzabili dai lati superficiali esterni;
- lato B: seconda fetta laterale, opposta alla faccia precedente;
- lato C: la parte inferiore dello speck è costituita dalla cotenna dell'animale, da cui si possono estrarre informazioni circa la salatura e l'affumicatura;
- lato D: l'ultimo lato è relativo alla faccia superiore dello speck, costituita dalla parte interna del taglio di carne della carcassa; in questa zona vengono cosparsi il sale e le spezie per conferire le caratteristiche organolettiche tipiche del prodotto.

Si riportano in Figura 1.2 alcune fotografie specifiche di ogni faccia, in modo da chiarirne l'assegnazione; si riportano invece in Figura 1.3 alcuni esempi di ogni analisi spettroscopica, relativi nello specifico al campione 20 (il primo campione).

Mentre per gli spettri sulla carcassa macellata è disponibile una singola acquisizione per entrambi gli strumenti, quelli sul prodotto stagionato sono invece acquisiti in due modalità, analizzando solamente la parte magra o macinando la fetta intera e sottoponendo all'analisi

---

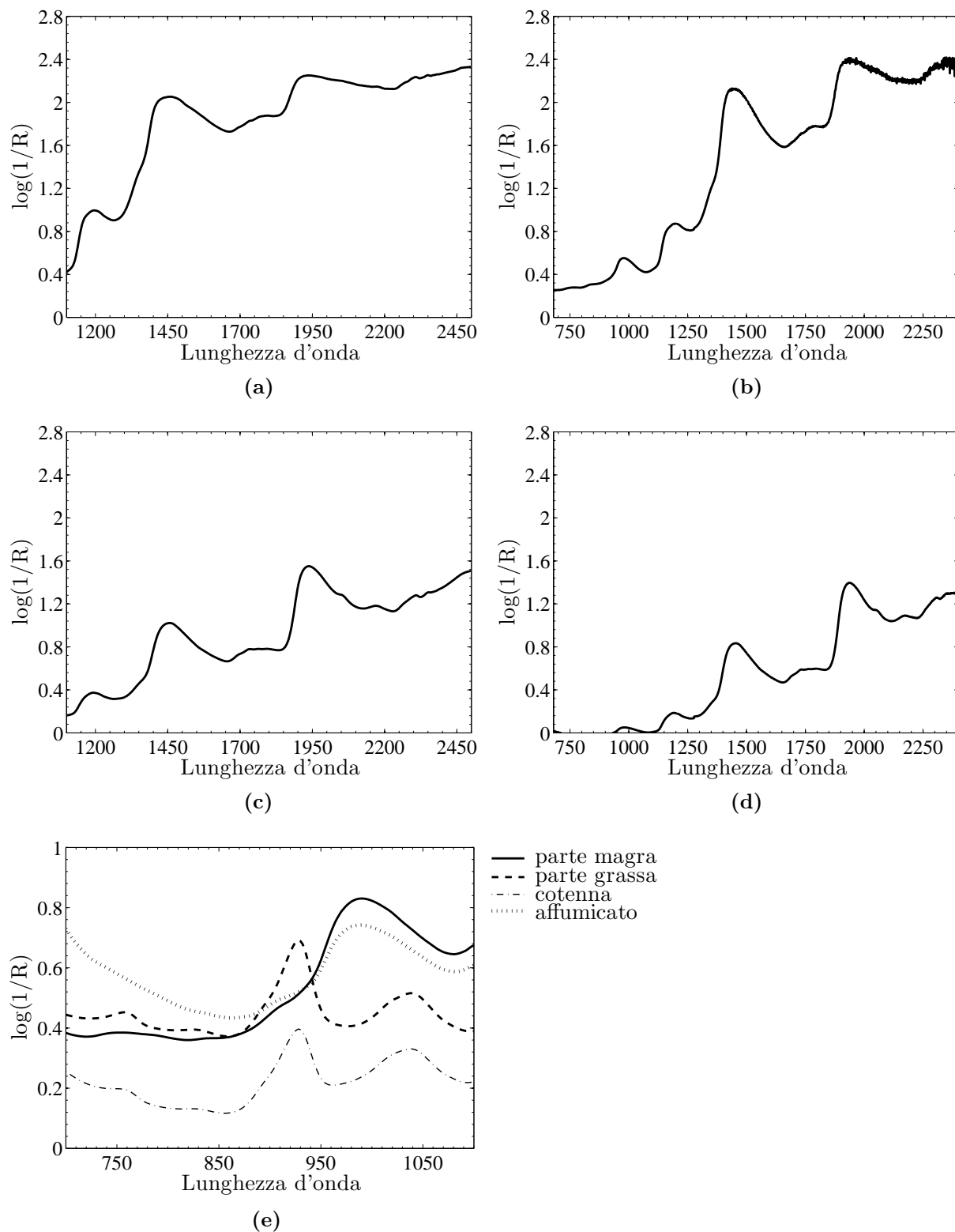
<sup>4</sup>Lo spettro considerato è stato poi ridotto all'intervallo 700÷1100 nm a causa di un problema alla lampada dello strumento, rilevato e corretto solo per la seconda parte delle acquisizioni.



**Figura 1.2** Immagini esemplificative delle diverse facce dei campioni e dell'assegnazione delle specifiche sigle; il campione di riferimento è il numero 25. (a) lato A; (b) lato B; (c) lato C; (d) lato D.

NIR un sottocampione. Tuttavia, quest'ultima analisi è stata ripetuta due volte poiché, nonostante la macinazione, il campione risultava comunque molto eterogeneo; le due serie di dati sono state quindi mediate, ma la totale casualità del sottocampione macinato scelto per l'analisi spettroscopica rende la trattazione statistica poco significativa ed i risultati poco affidabili. Viceversa, le acquisizioni sulla parte magra rappresentano una fonte di dati di particolare importanza e significatività, perché la specificità di uno spettro NIR consente di estrarre un gran numero di informazioni utili sia in fase di classificazione che di stima.

Tutti i campioni dello speck stagionato sono stati inoltre analizzati in ogni lato con l'ausilio di uno spettrometro NIR portatile, al fine di verificare se eventuali proprietà del prodotto finale possano essere predette a partire da specifiche analisi, relative ad una particolare faccia del campione. Ad esempio, potrebbe essere più indicato predire una particolare caratteristica del prodotto stagionato a partire dall'analisi della cotenna, piuttosto che dalla parte magra o da quella grassa.



**Figura 1.3** Esempi di spettri NIR, campione 20: (a) carcassa macellata, strumento Foss; (b) carcassa macellata, strumento Unity; (c) prodotto stagionato, strumento Foss; (d) prodotto stagionato, strumento Unity; (e) ogni faccia del prodotto stagionato, con distinzione di parte magra e grassa.

## 1.5 Analisi Instron

L’Instron è uno strumento per l’analisi della resistenza al taglio di un materiale, utile nel caso in esame per simulare la masticazione e quindi raccogliere informazioni circa la durezza e la compattezza del prodotto stagionato.

Da ogni campione è stata prelevata una fetta regolare di 3.0 cm di spessore, da cui sono state separate manualmente la parte magra e quella grassa; quindi è stato utilizzando una sonda cilindrica di 1.0 cm di diametro per prelevare un provino da sottoporre all’analisi Instron. Data l’eterogeneità di queste due zone, tale operazione non va ad intaccare la resistenza al taglio del prodotto, che giunge inalterato all’analisi. Sono stati estratti più provini cilindrici da entrambe parti, in modo da ottenere due distinte serie di analisi; ognuno di essi è stato sottoposto ad uno sforzo di taglio applicato mediante una configurazione monolama, registrando le variazioni sia nella forza che nella posizione della lama; da queste acquisizioni è quindi possibile ottenere un grafico forza / distanza della lama.

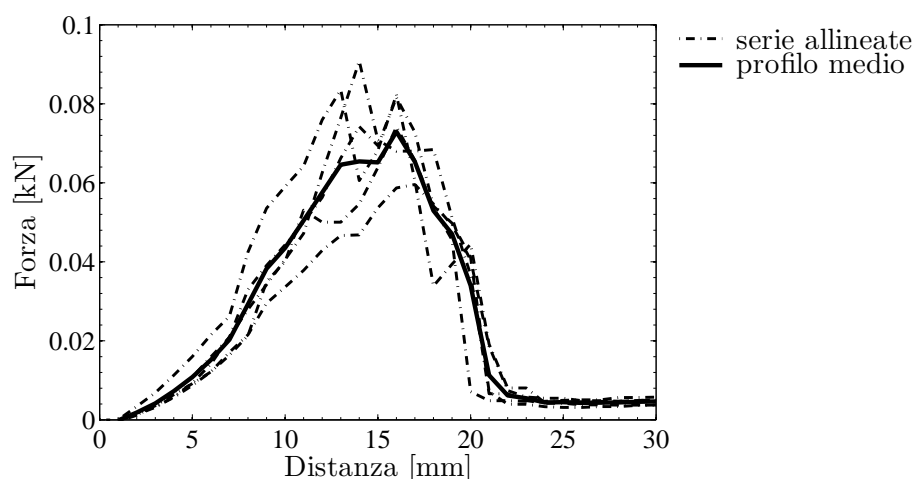
Il dataset ha poi richiesto un trattamento preliminare piuttosto articolato, necessario per allineare tutte le serie sperimentali ad un intervallo comune di analisi; per ogni campione sono infatti disponibili più misure ripetute, in numero variabile, ma la scansione della profondità di taglio non è fatta su valori coerenti fra ripetute diverse. Si è quindi imposto un intervallo di analisi pari a  $[0, 30]$  mm, suddiviso in unità pari ad 1.0 mm e, mediante un *down-sampling*, si sono approssimati tutti i punti sperimentali all’unità intera più prossima, rendendo coerenti i punti di ogni ripetuta.

Dopo aver riorganizzato e verificato ogni singola analisi, le ripetute complete relative al medesimo campione sono state mediate, ed il profilo medio risultante è stato utilizzato per calcolare altri indicatori sintetici per descrivere l’andamento della curva.

Si riporta in Figura 1.4 un esempio della procedura di media delle ripetute sperimentali allineate, mentre si definiscono in Tabella 1.4 le variabili aggiuntive, specifiche del dataset. Si è scelto di utilizzare come indicatori sintetici l’area sottesa dalla curva, che fisicamente rappresenta il lavoro di taglio compiuto, il valore massimo, il valore medio e la deviazione standard. Non è stata invece considerata la curtosità, potenzialmente utile per avere un’indicazione della forma della curva di analisi, poiché la distanza iniziale della lama dal campione non è costante e standardizzata, quindi le curve di campioni diversi non sono allineate dal punto di vista temporale e tale indicatore perde di significato.

## 1.6 Analisi colorimetriche

Una delle caratteristiche più importanti del prodotto stagionato è senza dubbio il colore, che a causa dell’impatto immediato sul consumatore può attirare il cliente all’acquisto o



**Figura 1.4** Curve sperimentali ottenute dall'analisi Instron, campione 20: singole ripetute allineate in linee tratteggiate, profilo medio in grassetto.

**Tabella 1.4** Elenco delle variabili aggiuntive, specifiche del dataset Analisi Instron.

Parte magra		Parte grassa	
82	area magro	86	area grasso
83	massimo magro	87	massimo grasso
84	media magro	88	media grasso
85	deviazione standard magro	89	deviazione standard grasso

meno; inoltre, anche da un punto di vista fisico, numerose proprietà sono legate al colore ed eventuali valori anomali delle stesse possono avere ripercussioni sull'aspetto dello speck. Ogni campione è stato quindi analizzato con l'ausilio di un colorimetro da banco, in modo da acquisire dettagli circa le principali variabili colorimetriche; le analisi sono state ripetute per più punti all'interno della fetta di prodotto stagionato, suddividendo le acquisizioni sulla parte magra da quelle sul grasso. Da ultimo, le serie di dati a disposizione sono state mediate, in modo da avere un'unica serie di variabili per ogni campione.

Si riportano in Tabella 1.5 le variabili aggiuntive, caratteristiche delle informazioni colorimetriche.

**Tabella 1.5** Elenco delle variabili colorimetriche aggiuntive, specifiche del dataset Analisi colorimetriche.

Parte magra		Parte grassa	
90	luminosità magro	95	luminosità grasso
91	indice del rosso magro	96	indice del rosso grasso
92	indice del giallo magro	97	indice del giallo grasso
93	croma magro	98	croma grasso
94	tinta magro	99	tinta grasso



Si riportano inoltre le relazioni tra le variabili aggiuntive di Tabella 1.5 (CIELAB, 1976).

$$C^* = \sqrt{a^{*2} + b^{*2}} \quad (1.3)$$

$$h = \arctan\left(\frac{b^*}{a^*}\right) \quad (1.4)$$

$a^*$  = indice del rosso

$b^*$  = indice del giallo

$C^*$  = croma

$h$  = tinta

Il principale attributo percettivo del colore è la tinta, con cui non si definisce un colore vero e proprio, ma una famiglia di colori, senza particolari distinzioni di luminosità e saturazione; in questo senso, rosso e rosa fanno parte della stessa famiglia. Tuttavia, non tutti i colori hanno una tinta: bianco, nero e varie gradazioni di grigio ne sono privi e sono per questo detti colori acromatici. L'occhio umano distingue tra 200 e 300 tinte diverse; tra queste, solo quattro non vengono percepite come mescolanze di altre tinte: si tratta del rosso, giallo, verde e blu, che sono dette tinte unarie.

Altri attributi del colore molto importanti sono:

- la luminosità: misura il grado di illuminazione ed è costituita da una parte cromatica ed una acromatica (il bianco per la luce, il grigio per altri oggetti);
- la pienezza: misura la quantità assoluta della parte cromatica;
- la saturazione: indica il rapporto tra la parte cromatica e quella acromatica;
- il croma: indica la pienezza di un'area in proporzione alla luminosità di un oggetto con illuminazione simile, che però appare bianco.

Rosso e rosa hanno la stessa tinta, ma la pienezza del primo è decisamente maggiore; per descrivere la saturazione si ricorre invece agli aggettivi “carico” o “pallido”.

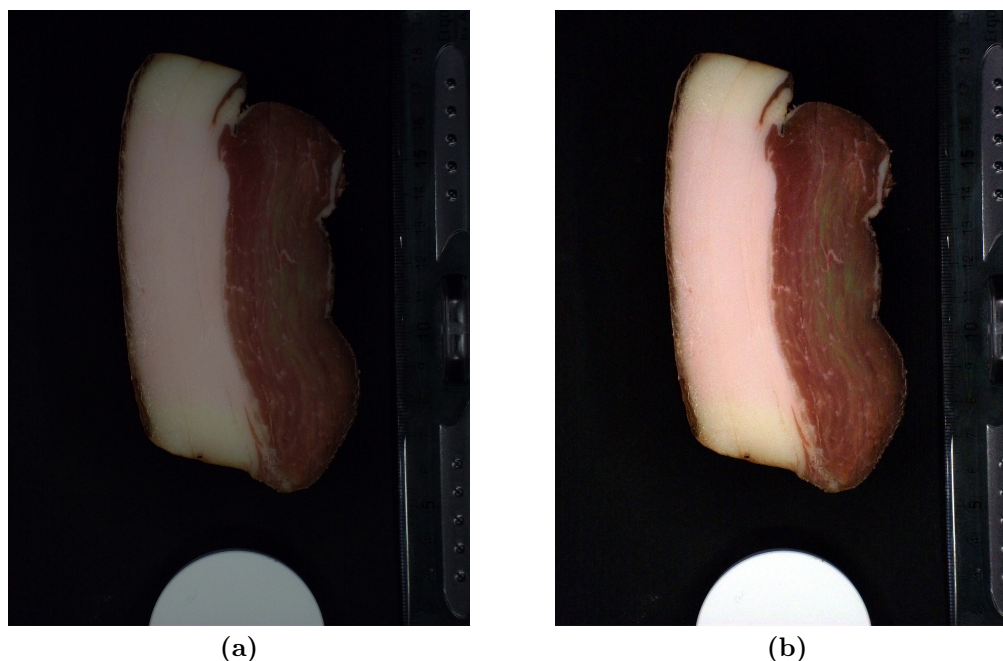
## 1.7 Immagini digitali

Per tutti i campioni sono inoltre disponibili delle fotografie ad alta risoluzione, utili per condurre ulteriori analisi sul colore, sulla frazione magro/grasso e sulla tessitura delle carni. La fotocamera utilizzata è la Kodak Easyshare M530, con 12 MP di risoluzione, zoom ottico 3X e tempo di esposizione di 1/60 di secondo; le immagini acquisite hanno dimensione  $4000 \times 3000$  pixel e coprono ogni faccia del campione: le due fette esterne, la cotenna e la parte affumicata.

Al fine di ottimizzare la qualità dell'immagine e consentire l'equalizzazione dei singoli canali RGB (rosso, verde e blu), i campioni sono stati fotografati su sfondo nero, con un riferimento di taratura per il bianco; è inoltre presente un righello per avere una stima

delle dimensioni.

Si riporta in Figura 1.5 un esempio del trattamento di equalizzazione dei singoli canali dell'immagine.



**Figura 1.5** Equalizzazione dei singoli canali RGB: campione 20, fetta intera lato A. (a) immagine originale; (b) immagine equalizzata.

Come si può notare, il trattamento di equalizzazione consente di ottimizzare la qualità dell'immagine, esaltando i colori e rendendo possibile l'individuazione di eventuali difetti, fluorescenze, macchie. Dopo questo pretrattamento, le immagini sono state segmentate eliminando il righello laterale ed il riferimento per il bianco, in modo da focalizzare l'analisi sulle proprietà della fetta di speck ed eliminare eventuali fonti di interferenza.

## 1.8 Valutazione sensoriale

L'ultima parte dei dati a disposizione è costituita da una serie di giudizi di qualità e variabili sensoriali, attribuiti da una giuria di esperti a tutti i campioni in esame; le caratteristiche analizzate comprendono il colore, la consistenza, la marezatura<sup>5</sup>, ed il rapporto tra la parte magra e quella grassa. Si riporta in Tabella 1.6 il dettaglio dei parametri considerati, specificando per ciascuno il *range* di attribuzione.

L'insieme di questi indicatori verrà analizzato nella fase di classificazione e stima, confrontando l'efficacia di un metodo oggettivo quale è il modello matematico calibrato su tutti i

<sup>5</sup>É una caratteristica delle fibre di carne, relativa al loro andamento sinuoso e poco contorto.

dati a disposizione, con il giudizio soggettivo seppur unificato dei giudici.

**Tabella 1.6** *Elenco delle caratteristiche analizzate durante la valutazione sensoriale e dei rispettivi range di attribuzione.*

<b>Caratteristica</b>	<b>Giudizio</b>
Colore della carne	1 = rosso pallido : 9 = rosso scuro
Difetti di colore del magro	[/] = assenza di difetto D = disomogeneo V = macchie verdi G = macchie grigie I = iridescenza
Colore del grasso	B = bianco BR = bianco rosato BG = bianco giallastro BV = bianco verde BGR = bianco grigio
Consistenza della carne	1 = flaccida : 9 = molto dura
Consistenza del grasso	1 = molle : 9 = compatto
Marezzatura	1 = assente : 9 = eccessiva
Rapporto magro/grasso	1.0 ÷ 2.5 = rapporto 5:1 2.6 ÷ 3.5 = rapporto 4:1 3.6 ÷ 4.5 = rapporto 3:1 4.6 ÷ 5.5 = rapporto 2:1 5.6 ÷ 6.5 = rapporto 1:1 6.6 ÷ 7.5 = rapporto 1:2 7.6 ÷ 8.5 = rapporto 1:3



# Capitolo 2

## Metodi di statistica multivariata

L'ampiezza dei dati a disposizione ed il gran numero di variabili fra loro correlate richiedono l'applicazione di tecniche statistiche multivariate, necessarie per ridurre la dimensione dei dati e riassumere le variabili originali in nuove variabili, tra loro ortogonali, che ne rappresentino la struttura di variabilità e correlazione.

Si discutono nel seguito le metodiche utilizzate per l'analisi statistica dei dati, indicandone pregi e svantaggi.

### 2.1 Analisi delle componenti principali (PCA)

L'analisi dei componenti principali (PCA, *Principal Component Analysis*) è una tecnica statistica che permette di riassumere il contenuto informativo di una matrice di dati  $\mathbf{X}$  selezionando le direzioni di massima variabilità del set di dati stesso; poiché le variabili originali, correlate tra loro, sono caratterizzate da direzioni di variabilità comuni, esse possono essere ridotte mediante trasformazioni lineari in un numero inferiore di variabili trasformate, ortogonali tra loro, chiamate componenti principali (Jackson, 1991). Inoltre, tali combinazioni lineari riassumono i dati originali minimizzando la perdita di informazione (Mardia *et al.*, 1979), in modo da condensare il contenuto informativo originale in un set di dati di dimensione inferiore, ma che spieghi la maggior parte della varianza sistematica. Si supponga di avere una matrice di dati originali  $\mathbf{X}$  di dimensione  $[N \times M]$ , in cui  $N$  è il numero dei campioni ed  $M$  quello delle variabili; da un punto di vista matematico, la procedura sopra descritta è equivalente a decomporre la matrice  $\mathbf{X}$  in una serie di  $R$  matrici  $\mathbf{C}_i$ ,  $i = 1, 2, \dots, R$ , in cui  $R$  è il rango di  $\mathbf{X}$ :

$$\mathbf{X} = \mathbf{C}_1 + \mathbf{C}_2 + \dots + \mathbf{C}_R \quad (2.1)$$

$$R = \text{rank}(\mathbf{X}) \quad . \quad (2.2)$$

Normalmente, tuttavia, un modello PCA è limitato alle prime  $D$  componenti principali (PC), lasciando le rimanenti quote marginali di varianza nella matrice dei residui  $\mathbf{E}$ . Il criterio di scelta del numero ottimale di componenti principali è basato sull'andamento della radice dell'errore quadratico medio (*Root Mean Square Error*, RMSE) in convalida incrociata, al variare del numero delle componenti principali considerate.

La convalida incrociata utilizzata in questa tesi segue un algoritmo *leave-one-out*, basato

$$RMSE = \sqrt{\frac{\sum_{n=1}^N (\mathbf{x}_n - \hat{\mathbf{x}}_n)^2}{N}} . \quad (2.3)$$

sulla costruzione di  $N$  distinti modelli, ottenuti eliminando di volta in volta un solo campione dal set di calibrazione, e proiettandolo sul modello costruito sugli altri. Il numero  $D$  di PC è scelto al fine di minimizzare tale parametro, che rappresenta lo scarto tra la matrice di dati originali e quella ottenuta dal modello PCA.

Inoltre, le sottomatrici  $\mathbf{C}_i$  possono essere scritte come prodotto vettoriale di due vettori  $\mathbf{t}_i$  e  $\mathbf{p}_i$ , definiti *score* e *loading*. L'equazione (2.1) è quindi riscrivibile come:

$$\mathbf{X} = \mathbf{t}_1 \mathbf{p}_1^T + \mathbf{t}_2 \mathbf{p}_2^T + \cdots + \mathbf{t}_D \mathbf{p}_D^T + \mathbf{E} \quad (2.4)$$

$$= \mathbf{TP}^T + \mathbf{E} \quad (2.5)$$

$$= \sum_{i=1}^D \mathbf{t}_i \mathbf{p}_i^T + \sum_{i=D+1}^R \mathbf{t}_i \mathbf{p}_i^T , \quad (2.6)$$

in cui la matrice dei residui  $\mathbf{E}$  è ottenuta come differenza tra la matrice  $\mathbf{X}$  dei dati originali e quella  $\hat{\mathbf{X}}$  ottenuta dal modello PCA, proiezione di  $\mathbf{X}$  nell'iper-spazio  $D$ -dimensionale delle componenti principali:

$$\mathbf{E} = \mathbf{X} - \mathbf{TP}^T \quad (2.7)$$

$$= \mathbf{X} - \hat{\mathbf{X}} . \quad (2.8)$$

Ciò può essere espresso anche dalla rappresentazione grafica di Figura 2.1.

$$\begin{aligned} \boxed{\mathbf{X}} &= \boxed{\mathbf{C}_1} + \boxed{\mathbf{C}_2} + \cdots + \boxed{\mathbf{C}_d} + \boxed{\mathbf{E}} \\ \\ \boxed{\mathbf{X}} &= \begin{matrix} \text{M} \\ \boxed{\mathbf{t}_1} \\ \text{N} \end{matrix} \begin{matrix} \text{1} & \text{M} \\ \boxed{\mathbf{p}_1^T} \\ \text{1} & \end{matrix} + \begin{matrix} \text{1} & \text{M} \\ \boxed{\mathbf{t}_2} \\ \text{N} & \end{matrix} \begin{matrix} \text{1} & \text{M} \\ \boxed{\mathbf{p}_2^T} \\ \text{1} & \end{matrix} + \cdots + \begin{matrix} \text{1} & \text{M} \\ \boxed{\mathbf{t}_D} \\ \text{N} & \end{matrix} \begin{matrix} \text{1} & \text{M} \\ \boxed{\mathbf{p}_D^T} \\ \text{1} & \end{matrix} + \begin{matrix} \text{M} \\ \boxed{\mathbf{E}} \\ \text{N} \end{matrix} \\ \\ &= \begin{matrix} \text{D} & \text{M} \\ \boxed{\mathbf{T}} & \boxed{\mathbf{P}^T} \\ \text{N} & \text{D} \end{matrix} + \begin{matrix} \text{M} \\ \boxed{\mathbf{E}} \\ \text{N} \end{matrix} \end{aligned}$$

**Figura 2.1** Decomposizione della matrice di dati originali  $\mathbf{X}$  nelle matrici  $\mathbf{T}$  e  $\mathbf{P}$  di score e loading, a meno di una matrice di residui  $\mathbf{E}$ .

I valori di *score* e *loading* possono essere calcolati a partire da una decomposizione ai valori singolari di una trasformata della matrice di covarianza  $\mathbf{S}$  di  $\mathbf{X}$ :

$$\mathbf{S} = \begin{bmatrix} s_{1,1}^2 & s_{1,2}^2 & \cdots & s_{1,m} & \cdots & s_{1,M} \\ s_{2,1} & s_{2,2}^2 & & & & \\ \vdots & & \ddots & & & \\ s_{m,1} & & & s_{m,m}^2 & & \\ \vdots & & & & \ddots & \\ s_{M,1} & \cdots & & & & s_{M,M}^2 \end{bmatrix}, \quad (2.9)$$

in cui  $s_{mm}^2$  è la varianza

$$s_{m,m}^2 = \frac{\sum_{n=1}^N (x_{n,m} - \bar{x}_m)^2}{N} \quad (2.10)$$

$$\bar{x}_m = \frac{\sum_{n=1}^N x_{n,m}}{N}, \quad (2.11)$$

mentre la covarianza  $s_{ij}$  è data da:

$$s_{i,m} = \frac{N \sum_{n=1}^N x_{i,n} x_{n,m} - \sum_{n=1}^N x_{i,n} \sum_{n=1}^N x_{n,m}}{[N(N-1)]} \quad (2.12)$$

Pre-moltiplicando e post-moltiplicando per una opportuna matrice ortonormale  $\mathbf{W}$ , è possibile decomporre  $\mathbf{S}$  in una matrice diagonale  $\mathbf{L}$ .

$$\mathbf{W}^T \mathbf{S} \mathbf{W} = \mathbf{L} \mathbf{U} \quad (2.13)$$

Gli elementi diagonali di  $\mathbf{L}$  sono le radici caratteristiche (cioé gli autovalori) di  $\mathbf{S}$ , mentre le colonne di  $\mathbf{U}$  vengono definite vettori caratteristici o autovettori. Gli autovalori possono essere ottenuti dalla soluzione della seguente equazione caratteristica o da una decomposizione ai valori singolari di  $\mathbf{S}$ :

$$|\mathbf{S} - \mathbf{I}| = 0, \quad (2.14)$$

in cui  $\mathbf{I}$  è la matrice identità. L'equazione (2.14) produce un polinomio in  $\mathbf{I}$ , da cui si ricavano gli autovalori  $l_d$ .

Gli autovettori vengono invece ottenuti risolvendo le equazioni:

$$[\mathbf{S} - \mathbf{I}] \mathbf{v}_i = 0 \quad (2.15)$$

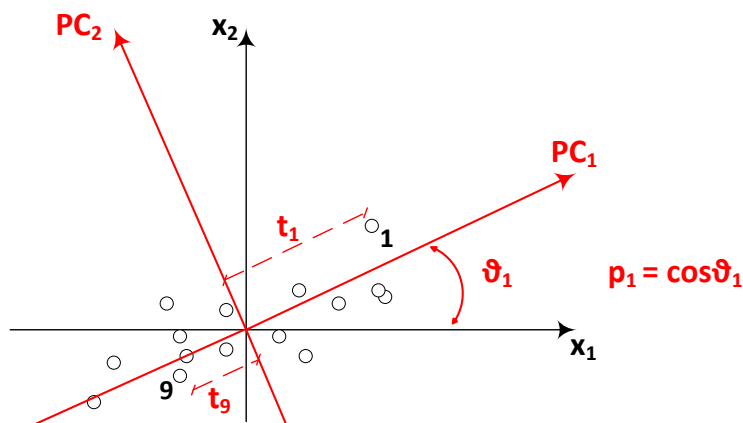
$$\mathbf{p}_i = \frac{\mathbf{v}_i}{\sqrt{\mathbf{v}_i^T \mathbf{v}_i}} \quad (2.16)$$

Ottenuti i valori dei *loading*, si ricavano gli *score* mediante l'equazione (2.17).

$$\mathbf{t}_i = \mathbf{p}_i^T \mathbf{x} \quad , \quad (2.17)$$

in cui  $\mathbf{x}$  è un vettore di campioni di variabili originali opportunamente (auto)scalato, cioè centrato sulla propria media e con varianza unitaria.

Da un punto di vista geometrico, la procedura appena descritta è equivalente ad una rotazione degli assi principali rispetto alle coordinate cartesiane originali  $x_1$  ed  $x_2$ ; i coseni direttori di questa rotazione sono dati dagli elementi degli autovettori  $\mathbf{p}_\alpha$ , mentre gli *score*  $\mathbf{t}_\alpha$  indicano la coordinata spaziale lungo questo nuovo asse principale. Si riporta in Figura 2.2 una rappresentazione grafica di quanto detto.



**Figura 2.2** Rappresentazione geometrica dell'analisi PCA; significato di *loading*  $\mathbf{p}_i$ , equivalenti ai coseni direttori del nuovo asse principale, e degli *score*  $\mathbf{t}_i$ , equivalenti alla coordinata spaziale del nuovo asse principale.

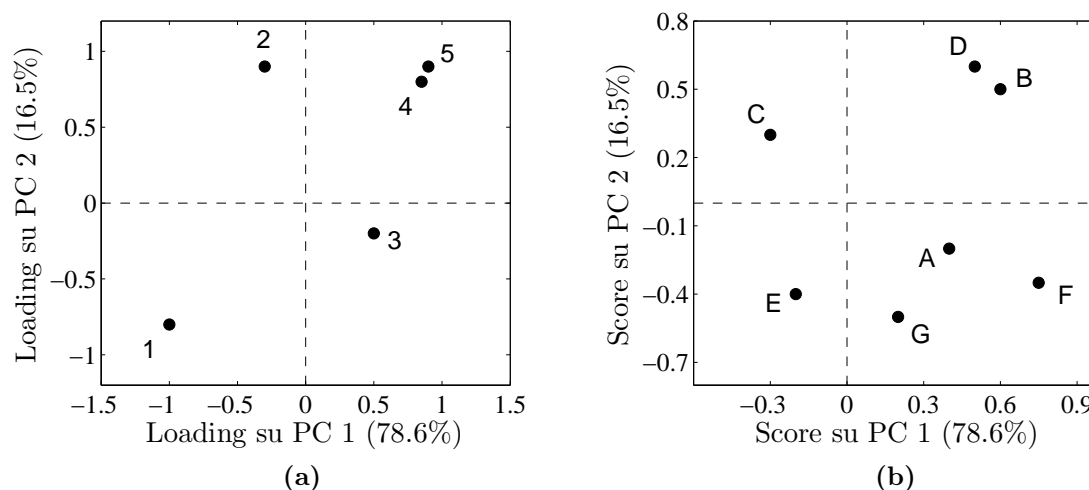
È tuttavia necessario sottolineare come, nella pratica, l'equazione (2.14) venga utilizzata solo per polinomi di grado  $\leq 3$ ; per ordini superiori, la risoluzione diretta viene abbandonata a favore di metodi iterativi (Geladi e Kowalski, 1986), più efficienti dal punto di vista computazionale.

### 2.1.1 Interpretazione dei risultati: analisi delle correlazioni

Oltre alla diminuzione del numero di variabili originali, l'analisi PCA consente anche di estrarre informazioni circa la correlazione tra le variabili e/o tra i campioni. Generalmente, ciò viene dedotto analizzando la distribuzione di *loading* e *score*, che nello spazio latente rappresentano rispettivamente le variabili ed i campioni. Nello specifico, la vicinanza di uno o più *loading* nello spazio delle componenti principali è caratteristica di una correlazione tra le variabili che rappresentano, mentre la vicinanza tra *score* esprime la correlazione



tra diversi campioni. Si riporta in Figura 2.3 un esempio di queste due rappresentazioni; si considerano 7 campioni (A,B,...,G), ciascuno dei quali è rappresentato da 5 variabili (1,2,...,5).



**Figura 2.3** Rappresentazione dei risultati di un'analisi PCA, relativi alle prime due componenti principali: (a) diagramma dei loading; (b) diagramma degli score.

In Figura 2.3a è possibile individuare una correlazione piuttosto significativa tra le variabili 4 e 5, i cui *loading* sono molto vicini tra loro; viceversa, la variabile 1 è collocata nel quadrante opposto rispetto ad esse, e ciò è rappresentativo di una anticorrelazione rispetto alla coppia di variabili precedenti. Dal grafico 2.3b, inoltre, si nota una significativa correlazione tra i campioni B e D, caratterizzati da una certa vicinanza nello spazio degli *score*. La procedura appena esposta è molto veloce ed efficace quando il numero delle componenti principali selezionate è limitato, poiché in questi casi le prime PC sono esplicative di quote sensibili di varianza spiegata e l'analisi grafica può essere limitata alle prime componenti principali, rappresentative delle maggiori quote di varianza spiegata. Tuttavia, quando il numero di PC è elevato e la percentuale di varianza spiegata è limitata anche per le prime componenti principali, queste rappresentazioni dovrebbero essere ripetute tra tutte le coppie di PC esplicative di quote di varianza sufficientemente elevate; ciò rende l'analisi grafica piuttosto lunga ed articolata, e fa emergere l'esigenza di un indicatore sintetico che possa riassumere tutti i singoli grafici in un unico dato.

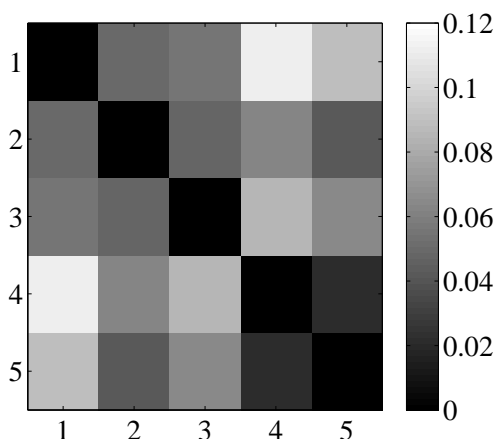
### 2.1.2 Distanza euclidea multidimensionale tra loading

Al fine di evitare il calcolo e l'interpretazione di numerosi *loading plot* tra coppie di componenti principali, è possibile di caratterizzare le correlazioni tra le variabili originali

calcolando la distanza euclidea multidimensionale tra *loading*,  $\mathbf{D}$  [ $D \times D$ ]:

$$\mathbf{D}_{i,j} = \sqrt{(\mathbf{p}_i - \mathbf{p}_j)(\mathbf{p}_i - \mathbf{p}_j)^T} \quad . \quad (2.18)$$

Tuttavia, per considerare la rappresentatività di ciascun *loading* rispetto alle variabili spiegate, è conveniente riscalare il valore di un *loading*  $p_{m,d}$  in base alla varianza spiegata della variabile  $m$  dalla componente principale  $i$ . Ciò mette al riparo da possibili interpretazioni errate, perché dare lo stesso peso a tutte le componenti principali può dare risultati impropri: infatti, se la componente principale  $j$  spiega una quota minima della varianza di  $m$ , non sarebbe corretto pesarlo alla pari di  $PC_i$ , che spiega invece una quota sensibile della varianza di  $m$ . Si riporta in Figura 2.4 un grafico esplicativo della distanza multidimensionale tra i *loading* di Figura 2.3a. La generalizzazione al caso multidimensionale è immediata.



**Figura 2.4** Distanza multidimensionale tra i *loading* delle cinque variabili analizzate in Figura 2.3.

Come si può notare, il calcolo della distanza euclidea multidimensionale conferma le correlazioni e le anticorrelazioni individuate analizzando Figura 2.3a: le correlazioni sono infatti rappresentate da minimi valori di distanza, evidenziati in nero, mentre le anticorrelazioni corrispondono all'estremo massimo dell'intervallo di distanze, evidenziato in bianco (per esempio, 4 vs 1).

### 2.1.3 Distanza euclidea multidimensionale tra correlation loading

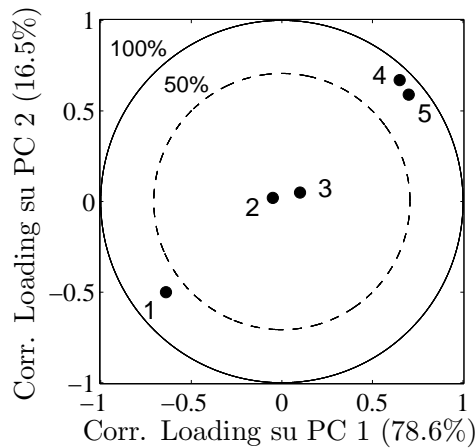
Per evidenziare ulteriormente le correlazioni tra variabili, è possibile calcolare anche i *correlation loading*  $\mathbf{R}$ . Questi parametri non solo esaltano le correlazioni, ma quantificano anche la significatività della varianza spiegata per ciascuna variabile, introducendo così un nuovo criterio discriminante. Vengono calcolati a partire dai *loading* standard ottenuti da un'analisi PCA, introducendo delle opportune modifiche tese ad evidenziare le correlazioni

presenti tra le variabili originali (Westad *et al.* (2003), Lorho *et al.* (2006)). Essi sono ottenuti mediante la relazione (2.19):

$$r_{m,d} = \frac{p_{m,d} \sqrt{\mathbf{t}_d^T \mathbf{t}_d}}{\sqrt{(\mathbf{x}_m - \bar{\mathbf{x}}_m)^T (\mathbf{x}_m - \bar{\mathbf{x}}_m)}} \quad , \quad (2.19)$$

in cui  $p_{m,d}$  è il *loading* standard relativo alla variabile  $m$  nel componente principale  $d$ ,  $\mathbf{t}_d$  è un vettore di *score* di  $PC_d$ , ed  $(\mathbf{x}_m - \bar{\mathbf{x}}_m)$  è la variabile  $\mathbf{x}_m$  centrata sulla sua media  $\bar{\mathbf{x}}_m$ .

Di norma i *correlation loading* vengono presentati per via grafica. L'interpretazione della vicinanza tra  $r_{m,d}$  relativi a variabili diverse è analoga a quella dei *loading* standard, mentre in questo caso vengono aggiunte anche due circonferenze per visualizzare le soglie di varianza spiegata pari al 50% e 100%, rappresentate rispettivamente da una linea tratteggiata ed una continua. Infatti, la somma delle coordinate dei *correlation loading* al quadrato esprime la varianza spiegata per quella variabile, e quindi quanto maggiore è la vicinanza di uno di essi alla circonferenza esterna, tanto più significativa è la varianza spiegata della variabile. Si riporta in Figura 2.5 un esempio della rappresentazione classica.

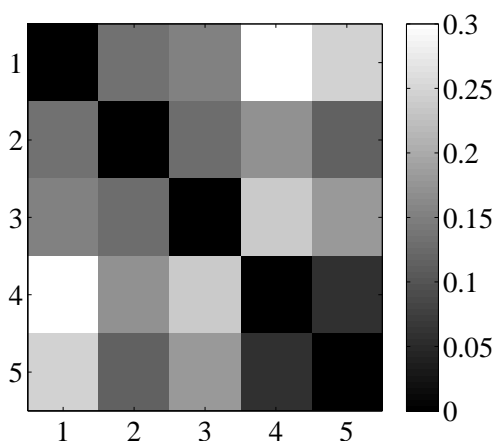


**Figura 2.5** Esempio di rappresentazione dei *correlation loading*; le cinque variabili analizzate sono le medesime delle Figure 2.3 e 2.4.

La Figura 2.5 evidenzia e conferma la correlazione tra le variabili 4 e 5, da considerarsi molto significative vista anche l'elevata varianza spiegata; viene confermata anche l'anticorrelazione con la variabile 1, mentre le variabili 2 e 3 risultano trascurabili nelle prime due componenti principali perché le rispettive varianze spiegate sono molto inferiori alle soglie di significatività minime.

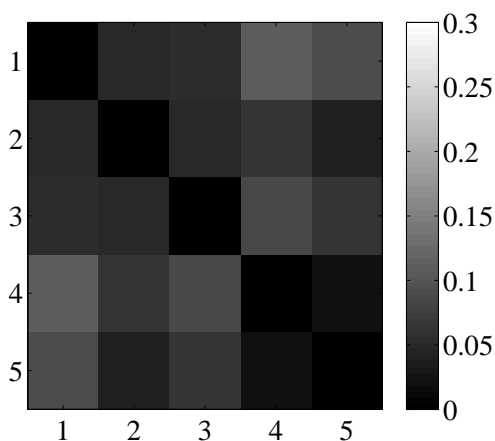
In piena analogia rispetto ai *loading* tradizionali, anche queste rappresentazioni risultano proficue solo quando il numero di componenti principali è limitato, poiché la ripetizione di più figure come la 2.5 fa perdere l'immediatezza dell'interpretazione e l'efficacia della lettura. Quindi, qualora siano presenti un numero non esiguo di componenti principali con quote di varianza spiegata simili, è conveniente il calcolo della distanza multidimensionale tra *correlation loading*, dipendente sempre dal numero  $d$  di PC scelti. Anche in

questo caso, si raccomanda lo *scaling* rispetto alla varianza spiegata della singola variabile nel componente principale considerato, al fine di rendere robusta l'analisi evitando interpretazioni errate. Tuttavia, è opportuno evidenziare la migliore efficacia dei *correlation loading* rispetto ai *loading* standard per quanto concerne l'analisi delle correlazioni, come riportato in Figura 2.6.



**Figura 2.6** Distanza multidimensionale tra i *correlation loading* delle cinque variabili analizzate.

Un confronto superficiale delle Figure 2.4 e 2.6 porterebbe ad affermare una ridondanza dei due diversi approcci, dal momento che l'impatto cromatico è molto simile. Tuttavia, l'utilizzo dei *correlation loading* amplifica notevolmente le distanze multidimensionali e quindi facilita l'individuazione delle variabili correlate, come evidente analizzando gli estremi della scala colorimetrica di Figura 2.6. Si riporta in Figura 2.7 un confronto equo delle due diverse metodiche, in cui sono stati uniformate le due diverse scale cromatiche. Figura 2.7 dimostra con notevole immediatezza la migliore efficacia dei *correlation loading* poiché, a parità di intervallo di distanze multidimensionali, l'interpretazione dei *loading* standard sarebbe del tutto indecifrabile.



**Figura 2.7** Distanza multidimensionale tra i *loading* delle cinque variabili analizzate - scala cromatica adattata.

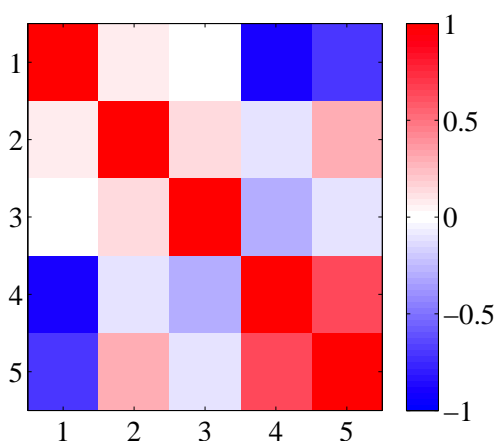
### 2.1.4 Mappe di correlazione

Le mappe di correlazione costituiscono una valida alternativa nello studio delle correlazioni tra le variabili originali della matrice  $\mathbf{X}$ . Queste mappe sono una rappresentazione grafica dei coefficienti di correlazione  $\rho_{ij}$ , che vengono calcolati a partire dalla matrice di covarianza  $\mathbf{S}$  (si faccia riferimento alle (2.9) e (2.12)):

$$\rho_{i,j} = \rho(m_i, m_j) = \frac{S_{ij}}{\sqrt{S_{ii}S_{jj}}} = \frac{S_{ij}}{\sigma_i\sigma_j} \quad , \quad (2.20)$$

in cui  $\sigma_m$  rappresenta la deviazione standard della variabile  $m$ . Si noti quindi come con variabili autoscalate, caratterizzate da deviazioni standard unitarie, le matrici di correlazione e covarianza siano equivalenti.

La rappresentazione dei risultati di Figura 2.8 è molto simile alla precedente Figura 2.4, ma in questo caso gli estremi dell'intervallo di variabilità sono fissati: il valore -1 è caratteristico di variabili anticorrelate, evidenziate in blu, mentre +1 è assegnato a variabili correlate, in rosso. In altre parole, i valori minori, nella fattispecie negativi, sono in questo caso rappresentativi di una anticorrelazione, contrariamente rispetto all'analisi della distanza multidimensionale tra *loading*.



**Figura 2.8** Esempio di una mappa di correlazione, relativa alle stesse cinque variabili.

Di norma, le mappe di correlazione sono da preferirsi all'analisi della distanza tra *loading* perché sono calcolate a partire dai dati originali tal quali, senza approssimazioni o perdita di informazioni dovuti all'analisi PCA. Tuttavia, è opportuno sottolineare che, qualora la varianza spiegata su  $\mathbf{X}$  sia molto elevata, i risultati dei due approcci sono simili e le relative interpretazioni pienamente confrontabili.

## 2.2 Proiezione su strutture latenti (PLS)

L'analisi delle componenti principali rappresenta una tecnica molto utile per diminuire il numero di variabili originali senza perdere contenuto informativo o per effettuare un'analisi delle correlazioni estesa a tutte le variabili in esame, ma in altri casi non rappresenta lo strumento statistico più efficace. Ad esempio, qualora si vogliano stimare una o più variabili di qualità, monitorare delle specifiche di processo o si vogliano invece classificare per categorie i campioni, una singola matrice di dati non è sufficiente. In questi casi è più conveniente lavorare con due matrici di dati, rappresentative dello spazio delle variabili predittrici  $\mathbf{X}$  [ $N \times M$ ] e della risposta  $\mathbf{Y}$  [ $N \times B$ ]; dato l'elevato numero di variabili e la correlazione tra di esse, una regressione multivariata su strutture latenti appare l'approccio più ragionevole per affrontare il problema. La selezione del numero di variabili latenti può avvenire utilizzando un'estensione dei metodi univariati o può seguire un approccio più generale, multivariato in senso stretto, considerando tutte le variabili in un unico modello. La prima tipologia di tecniche comprende la regressione lineare multipla (*Multiple Linear Regression*, MLR), che costruisce simultaneamente  $B$  modelli univariati ad altrettante variabili risposta e minimizza l'errore cumulato<sup>1</sup>, e la regressione su componenti principali (*Principal Component Regression*, PCR), in cui tuttavia le direzioni di massima variabilità di  $\mathbf{X}$  ed  $\mathbf{Y}$  vengono scelte in modo indipendente. Alternativamente, tra gli approcci multivariati si annoverano la *Canonical Correlation Regression* (CCR), tesa a massimizzare la correlazione tra  $\mathbf{X}$  ed  $\mathbf{Y}$ , la *Partial Least Squares regression* (PLS), che massimizza la covarianza, e la *Reduced Rank Regression* (RRR), una metodica intermedia (Burnham *et al.*, 1996). Tra tutte queste diverse metodiche, la PLS si è dimostrata la scelta più valida perché più robusta, con minori variazioni dei parametri del modello al variare dei campioni di convalida (Geladi e Kowalski, 1986).

La costruzione del modello può essere ottenuta per via iterativa definendo alcune relazioni esterne:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{F} = \sum \mathbf{t}_h \mathbf{p}_h^T + \mathbf{F} \quad (2.21)$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{G} = \sum \mathbf{u}_h \mathbf{q}_h^T + \mathbf{G} \quad , \quad (2.22)$$

in cui è necessario minimizzare gli errori  $\mathbf{F}$  ed  $\mathbf{G}$  per descrivere in maniera ottimale le matrici originali. Inoltre, sussiste una relazione interna per ciascuna variabile latente  $h$ , che dà il legame tra gli *score*  $\mathbf{t}$  ed  $\mathbf{u}$ ; il modello più semplice per tale relazione è dato da:

$$\mathbf{u}_h = \mathbf{b}_h \mathbf{t}_h \quad , \quad (2.23)$$

in cui  $\mathbf{b}_h = \mathbf{u}_h^T \mathbf{t}_h / \mathbf{t}_h^T \mathbf{t}_h$ . Tuttavia, questo modello non è ottimale perché le componenti principali sono selezionate in modo indipendente nelle due matrici originali, caratteristica

<sup>1</sup>Tutte le variabili risposta vengono trattate singolarmente, perdendo l'eventuale informazione derivante dalle loro correlazioni interne.

che rende molto debole la relazione interna<sup>2</sup>; una condivisione delle informazioni reciproche rende il modello molto più efficace, perché spesso sono sufficienti minime rotazioni delle componenti principali per agevolare notevolmente la regressione. Questo miglioramento della relazione interna avviene scambiando gli *score* delle due matrici nella selezione delle variabili latenti:

$$\mathbf{p}^T = \frac{\mathbf{t}^T \mathbf{X}}{\mathbf{t}^T \mathbf{t}} \quad \longrightarrow \quad \mathbf{p}^T = \frac{\mathbf{u}^T \mathbf{X}}{\mathbf{u}^T \mathbf{u}} \quad . \quad (2.24)$$

Tale modifica è particolarmente importante perché consente di determinare le direzioni di massima variabilità di  $\mathbf{X}$  che meglio spiegano  $\mathbf{Y}$ , eliminando i limiti dovuti all'indipendenza delle componenti principali. Tuttavia, viene persa l'ortogonalità tra gli *score*  $\mathbf{t}$ , caratteristica essenziale delle strutture latenti; per ripristinarla è necessario introdurre degli ulteriori parametri  $\mathbf{w}$ , detti pesi, che rappresentano i coefficienti delle combinazioni lineari delle variabili di  $\mathbf{X}$ . L'algoritmo di calcolo iterativo NIPALS (*Non-linear Iterative Partial Least Squares*) è (Geladi e Kowalski, 1986):

(1) scelta di un valore di tentativo,  $\mathbf{u} = \mathbf{u}_{start}$

$$(2) \quad \mathbf{w}^T = \frac{\mathbf{u}^T \mathbf{X}}{\mathbf{u}^T \mathbf{u}}$$

$$(3) \quad \mathbf{w}_{new}^T = \frac{\mathbf{w}_{old}^T}{\|\mathbf{w}_{old}^T\|}$$

$$(4) \quad \mathbf{t} = \frac{\mathbf{X} \mathbf{w}}{\mathbf{w}^T \mathbf{w}}$$

$$(5) \quad \mathbf{q}^T = \frac{\mathbf{t}^T \mathbf{Y}}{\mathbf{t}^T \mathbf{t}}$$

$$(6) \quad \mathbf{q}_{new}^T = \frac{\mathbf{q}_{old}^T}{\|\mathbf{q}_{old}^T\|}$$

$$(7) \quad \mathbf{u} = \frac{\mathbf{Y} \mathbf{q}}{\mathbf{q}^T \mathbf{q}}$$

(8) verifica della convergenza

$$(9) \quad \mathbf{p}^T = \frac{\mathbf{t}^T \mathbf{X}}{\mathbf{t}^T \mathbf{t}}$$

$$(10) \quad \mathbf{p}_{new}^T = \frac{\mathbf{p}_{old}^T}{\|\mathbf{p}_{old}^T\|}$$

$$(11) \quad \mathbf{t}_{new} = \mathbf{t}_{old} \|\mathbf{p}_{old}^T\|$$

$$(12) \quad \mathbf{w}_{new} = \mathbf{w}_{old} \|\mathbf{p}_{old}^T\|$$

$$(13) \quad \mathbf{b} = \frac{\mathbf{u}^T \mathbf{t}}{\mathbf{t}^T \mathbf{t}} \quad .$$

Il criterio di scelta del numero di variabili latenti è basato anche in questo caso sul *RMSE* in convalida incrociata, calcolato però sulle variabili risposta che si ha interesse a stimare.

<sup>2</sup>Problematica di cui è affetta anche la PCR.

La convalida avviene sempre con un algoritmo *leave-one-out*.

$$RMSE = \sqrt{\frac{\sum_{n=1}^N (\mathbf{y}_n - \hat{\mathbf{y}}_n)^2}{N}} \quad . \quad (2.25)$$

Infine, per identificare le variabili predittrici più influenti è stato utilizzato l'indice VIP (*variable importance in projection*; Andersen e Bro (2010)), definito come:

$$VIP_m = \sqrt{\frac{\sum_{h=1}^H w_{m,h}^2 SSY_h M}{SSY_{total} H}} \quad (2.26)$$

$$SSY = \sum_{n=1}^N (\mathbf{y}_n - \bar{\mathbf{y}})^2 \quad , \quad (2.27)$$

in cui  $w_{m,h}$  è il valore del peso del componente  $h$  relativo alla variabile  $m$ ,  $SSY_l$  è la somma dei quadrati delle varianze spiegate per l' $h$ -esimo componente,  $SSY_{total}$  è la somma dei quadrati della varianza spiegata della variabile dipendente, ed  $H$  è il numero totale di variabili latenti considerate. Normalmente, le variabili con indice  $VIP > 1$  sono considerate di grande importanza, mentre quelle con  $VIP < 1$  sono progressivamente meno influenti.

### 2.2.1 Pretrattamenti delle analisi spettroscopiche NIR

L'utilizzo dei modelli PLS ha trovato larga applicazione nella stima e predizione di alcune variabili di qualità o specifiche di processo a partire da analisi indirette, che non coinvolgessero cioè la determinazione analitica diretta, molto spesso dispendiosa in termini di tempo e risorse. Appartengono alla categoria di analisi indirette le analisi colorimetriche, spettroscopiche, di tessitura, etc...; tra queste, assumono particolare rilievo le analisi spettroscopiche al vicino infrarosso, caratterizzate da un contenuto informativo molto elevato. In letteratura questi metodi vengono utilizzati accoppiandoli a diversi pretrattamenti sulle acquisizioni NIR, al fine di migliorare la capacità e l'accuratezza di stima, predire e classificare. In questa Tesi si è scelto di implementare una procedura per tentativi per la scelta del pretrattamento più efficace fra i più comuni. Nello specifico, i pretrattamenti studiati sono stati:

- (1) nessun pretrattamento;
- (2) *standard normal variate* (SNV); è equivalente ad autoscalare la trasposta della matrice di acquisizione NIR (*auto-scaling* per righe, Barnes *et al.* (1989));
- (3) derivata prima;
- (4) derivata seconda;
- (5) SNV e derivata prima;



(6) SNV e derivata seconda.

Nel seguito, tutti i pretrattamenti verranno richiamati utilizzando la numerazione assegnata.

### 2.2.2 Robustezza e interpretazione dei risultati

Per la stima o predizione di alcune variabili di qualità è necessario suddividere i campioni in due gruppi, utilizzati rispettivamente per la calibrazione e la convalida del modello. Si creano quindi le matrici di calibrazione  $\mathbf{X}_{cal}$  [ $N_{cal} \times M$ ] e  $\mathbf{Y}_{cal}$  [ $N_{cal} \times B$ ], e quelle di convalida  $\mathbf{X}_{val}$  [ $(N - N_{cal}) \times M$ ] e  $\mathbf{Y}_{val}$  [ $(N - N_{cal}) \times B$ ]. I risultati vengono espressi in termini di errori relativi percentuali, coefficienti di determinazione in calibrazione,  $R_{cal}^2$ , e convalida,  $R_{val}^2$ , ed RPD (*ratio of prediction to deviation*), definiti rispettivamente da (Davies e Fearn, 2006):

$$\epsilon_b(\%) = \left( \frac{\mathbf{y}_{val,b} - \hat{\mathbf{y}}_{val,b}}{\mathbf{y}_{val,b}} \right) \times 100 \quad (2.28)$$

$$R_{cal}^2 = 1 - \frac{\left[ \sum_{n=1}^{N_{cal}} (\mathbf{y}_n - \hat{\mathbf{y}}_n)^2 \right]}{\left[ \sum_{n=1}^{N_{cal}} (\mathbf{y}_n - \bar{\mathbf{y}})^2 \right]} \quad R_{val}^2 = 1 - \frac{\left[ \sum_{n=1}^{N-N_{cal}} (\mathbf{y}_n - \hat{\mathbf{y}}_n)^2 \right]}{\left[ \sum_{n=1}^{N-N_{cal}} (\mathbf{y}_n - \bar{\mathbf{y}})^2 \right]} \quad (2.29)$$

$$RMSEP = \sqrt{\frac{\sum_{n=1}^{N-N_{cal}} (\mathbf{y}_n - \hat{\mathbf{y}}_n)^2}{N - N_{cal}}} \quad (2.30)$$

$$RPD = \frac{\sigma(\mathbf{y}_{val})}{RMSEP} \quad (2.31)$$

I coefficienti di determinazione costituiscono una proporzione tra la variabilità dei dati (a denominatore) e l'accuratezza del modello (a numeratore). Valori elevati, prossimi all'estremo superiore 1, indicano che il modello spiega la maggior parte della variabilità di  $\mathbf{Y}$ , mentre al diminuire di  $R^2$  l'accuratezza della stima peggiora. Inoltre, valori negativi di  $R_{val}^2$  indicano che il modello è mal calibrato o che il *fitting* della variabile non segue un andamento lineare. Tuttavia, Davies e Fearn (2006) hanno dimostrato la dipendenza di questo parametro dall'intervallo di variabilità di  $\mathbf{Y}$ , e quindi ritengono che l'utilizzo di  $R^2$  sia inappropriato con variabili caratterizzate da una variabilità molto limitata, quantificabile utilizzando il coefficiente di variazione  $CV$ :

$$CV_b(\%) = \left[ \frac{\sigma(\mathbf{y}_b)}{\bar{\mathbf{y}}_b} \right] \times 100 \quad (2.32)$$

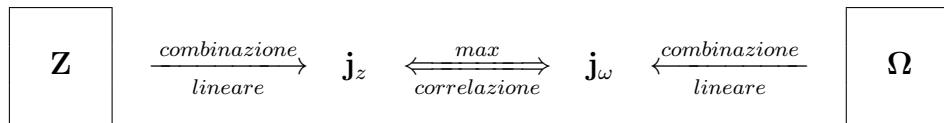
Viene quindi proposto l'utilizzo del RMSEP (*root mean square error of prediction*), che misura la variabilità della differenza tra i valori predetti e misurati di un set di campioni

di convalida. Tale parametro è indipendente dall'intervallo di variabilità di  $\mathbf{Y}$  e costituisce quindi un indicatore assoluto. Infine, si utilizza anche il fattore RPD (*ratio of prediction to deviation*), che misura l'incremento dell'accuratezza delle stime rispetto all'utilizzo dei valori medi  $\bar{\mathbf{Y}}$ .

### 2.3 Analisi delle correlazioni canoniche (CCA)

L'analisi delle correlazioni canoniche è una tecnica statistica che permette di individuare le correlazioni tra due matrici di dati, indicate come  $\mathbf{Z}$  [ $N \times \Phi$ ] e  $\mathbf{\Omega}$  [ $N \times \Delta$ ], caratterizzate dallo stesso numero  $N$  di campioni (righe), ma da un numero di variabili (colonne) non necessariamente uguale. La caratteristica peculiare di questa metodica risiede nella determinazione delle correlazioni “esterne” tra le variabili delle due matrici, senza alcun tipo di influenza da parte delle eventuali correlazioni “interne”, caratteristiche delle variabili di ciascuna matrice. Ciò è di fondamentale importanza perché in un'analisi PCA della matrice concatenata  $[\mathbf{Z}\mathbf{\Omega}]$  alcune correlazioni potrebbero essere trascurate perché di entità molto minore rispetto ad altre. L'analisi CCA consente invece di focalizzare lo studio su precise sottomatrici, apprezzando qualsiasi legame tra le variabili  $\mathbf{z}_\phi$  e  $\boldsymbol{\omega}_\delta$ , ed evitando interpretazioni parziali o errate dovute ad eventuali correlazioni molto significative tra  $\mathbf{z}_\phi$  e  $\mathbf{z}_{\phi+i}$  o tra  $\boldsymbol{\omega}_\delta$  e  $\boldsymbol{\omega}_{\delta+i}$ .

La tecnica si basa sulla costruzione di due iniziali indicatori sintetici  $\mathbf{j}_z$  ed  $\mathbf{j}_\omega$ , rispettivamente combinazione lineare delle variabili di  $\mathbf{Z}$  ed  $\mathbf{\Omega}$ ; successivamente, si determinano due basi  $\mathbf{A}$  e  $\mathbf{B}$ , costituite dai *canonical loading*, che moltiplicate per  $\mathbf{Z}$  e  $\mathbf{\Omega}$  massimizzano la correlazione tra  $\mathbf{j}_z$  ed  $\mathbf{j}_\omega$  (Nimon *et al.*, 2010; Sherry e Henson, 2005). Si riporta in Figura 2.9 una rappresentazione grafica della procedura appena descritta.



**Figura 2.9** Diagramma a blocchi del metodo CCA.

Il numero di coppie di indicatori sintetici, equivalente al numero di funzioni canoniche  $FC$ , è pari al rango minimo delle matrici di partenza:

$$FC = \min\{\text{rank}(\mathbf{Z}), \text{rank}(\mathbf{\Omega})\} \quad . \quad (2.33)$$

Il valore dei *canonical loading* è calcolato a partire da una decomposizione ai valori singolari di una trasformata della matrice di covarianza  $\mathbf{S}$  (Wilks, 2011).

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{z,\omega} & \mathbf{S}_{z,\omega} \\ \mathbf{S}_{\omega,z} & \mathbf{S}_{\omega,\omega} \end{pmatrix} \quad (2.34)$$

$$[\mathbf{S}_{z,z}]^{-1/2} [\mathbf{S}_{z,\omega}] [\mathbf{S}_{\omega,\omega}]^{-1/2} = [\mathbf{K}] [\mathbf{R}_c] [\mathbf{O}]^T \quad (2.35)$$

$$\mathbf{A} = [\mathbf{S}_{z,z}]^{-1/2} \mathbf{K} \quad (2.36)$$

$$\mathbf{B} = [\mathbf{S}_{\omega,\omega}]^{-1/2} \mathbf{O} \quad . \quad (2.37)$$

I coefficienti canonici di correlazione sono invece contenuti nella matrice diagonale  $\mathbf{R}_c$ ; essi quantificano la significatività della correlazione in ogni funzione canonica e permettono di capire se i risultati ottenuti siano o meno di rilievo.

$$\mathbf{r}_c = \text{diag}(\mathbf{R}_c) \quad . \quad (2.38)$$

Infine, le proiezioni  $\mathbf{\Lambda}$  e  $\mathbf{\Gamma}$  dei campioni originari nello spazio delle variabili canoniche sono ottenute dal prodotto delle variabili originali con i *canonical loading*.

$$\mathbf{\Lambda} = \mathbf{AZ} \quad (2.39)$$

$$\mathbf{\Gamma} = \mathbf{B}\mathbf{\Omega} \quad . \quad (2.40)$$

Tuttavia, come sarà meglio chiarito nel seguito, tali proiezioni non contengono alcuna informazione se interpretate singolarmente, ma il loro calcolo si rende necessario per ottenere altri coefficienti di maggiore significatività. Infatti, da una approfondita ricerca bibliografica sono emerse due diverse interpretazioni circa i risultati di un'analisi delle correlazioni canoniche; le si discutono entrambe nel seguito, evidenziandone i punti salienti.

- Courville e Thompson (2001) ritengono di fondamentale importanza il calcolo dei coefficienti di struttura (*structure coefficients*) da affiancare ai *canonical loading* per avere un'interpretazione robusta. Tali coefficienti sono ottenuti come correlazione tra le variabili originali  $\mathbf{Z}$  e le proiezioni  $\mathbf{\Lambda}$ , e chiariscono quale sia il contributo della generica variabile originale  $\mathbf{z}_\phi$  nel calcolo di ogni funzione canonica, e quindi del relativo indicatore sintetico tra cui viene massimizzata la correlazione per il calcolo dei *canonical loading*.

$$\mathbf{r}_s^z = \text{corr}(\mathbf{Z}, \mathbf{\Lambda}) \quad (2.41)$$

$$= [\mathbf{S}_{z,z}]^{1/2} \mathbf{K} \mathbf{R}_c \quad (2.42)$$

$$\mathbf{r}_s^\omega = \text{corr}(\mathbf{\Omega}, \mathbf{\Gamma}) \quad (2.43)$$

$$= [\mathbf{S}_{\omega,\omega}]^{1/2} \mathbf{O} \quad . \quad (2.44)$$

Assumono valori compresi tra 0 e 1, poiché il contributo di una singola variabile nella costruzione dell'indicatore sintetico può essere al massimo pari al 100%. La loro interpretazione è di fondamentale importanza perché si potrebbero ottenere dei risultati fuorvianti considerando solamente i *canonical loading*: se infatti il contributo della variabile  $\mathbf{z}_\phi$  nel calcolo della funzione canonica fosse molto limitato (coefficiente di struttura

trascurabile), l'entità del *canonical loading* perderebbe ogni significato. In altre parole, sarebbe inutile attribuire delle correlazioni basandosi su *canonical loading* elevati in valore assoluto, se ciò non è confermato da un effettivo contributo della variabile in questione al calcolo della funzione canonica. Inoltre, il quadrato dei coefficienti di struttura è un ottimo indicatore per stimare quanto la singola variabile partecipi alla costruzione delle diverse funzioni canoniche *FC*; se tali valori sono molto limitati in tutte le *FC*, la specifica variabile potrebbe essere eliminata dalla matrice di dati originali perché il suo contributo all'analisi è trascurabile.

$$(\mathbf{h}^z)^2 = \sum_{\theta=1}^{FC} (\mathbf{r}_{s,\theta}^z)^2 \quad (2.45)$$

in cui il pedice  $\theta$  si riferisce a ciascuna funzione canonica. La procedura è del tutto analoga per la matrice  $\mathbf{\Omega}$ .

- Rencher (1992) ritiene invece inutile e fuorviante il calcolo dei coefficienti di struttura, perché rappresentativi di un'analisi univariata, in pieno contrasto con l'approccio multivariato della CCA.

Entrambe queste interpretazioni sono state tenute in considerazione nell'analisi dei risultati, analizzando criticamente quale delle due fosse più efficace nell'analisi dei risultati ottenuti.

### 2.3.1 Forma ed interpretazione dei risultati

Analogamente a quanto riportato in letteratura, i risultati della CCA verranno presentati in due diverse forme, tabellare e grafica.

- *Forma tabellare*: si riportano tutti i valori numerici di *canonical loading*, coefficienti di struttura e coefficienti di struttura al quadrato. Sebbene il contenuto informativo sia molto preciso e dettagliato, la lettura è piuttosto complessa perché è necessario individuare i valori più rappresentativi di ogni matrice e confrontarli con gli analoghi indici dell'altra. Inoltre, è presente un set di risultati per ogni funzione canonica, ed in presenza di matrici con rango elevato il formato tabellare può risultare molto esteso. Infine, vista la scelta di non privilegiare una delle due filosofie interpretative, è necessario confrontare sia coefficienti di struttura che *canonical loading*, verificando volta per volta se siano in accordo o in contrasto. Si riporta in Tabella 2.1 un esempio di questo layout. Nell'esempio di Tabella 2.1 si riportano i risultati delle prime due funzioni canoniche relative all'analisi CCA su due matrici  $\mathbf{Z}$  ed  $\mathbf{\Omega}$ , costituite rispettivamente da quattro e cinque variabili. I risultati evidenziano una correlazione significativa tra le variabili  $\mathbf{z}_3 - \omega_5$  (prima funzione canonica), e tra le variabili  $\mathbf{z}_2 - \omega_2$ ,  $\mathbf{z}_4 - \omega_3$ ,  $\mathbf{z}_4 - \omega_4$  (seconda

**Tabella 2.1** Esempio di rappresentazione tabellare dei risultati della CCA.

	Funzione Canonica 1			Funzione Canonica 2			
Variabile	$\mathbf{A}_1$	$\mathbf{r}_s^z$	$(\mathbf{r}_s^z)^2$	$\mathbf{A}_2$	$\mathbf{r}_s^z$	$(\mathbf{r}_s^z)^2$	$(\mathbf{h}^z)^2$
$\mathbf{z}_1$	0,003	0,029	0,08	-0,020	-0,028	0,08	0,16
$\mathbf{z}_2$	-0,013	-0,294	8,62	0,379	0,699	48,80	57,42
$\mathbf{z}_3$	0,994	1,000	99,97	0,032	-0,010	0,01	99,98
$\mathbf{z}_4$	-0,020	-0,103	1,06	-0,781	-0,940	88,44	89,49

	Funzione Canonica 1			Funzione Canonica 2			
Variabile	$\mathbf{B}_1$	$\mathbf{r}_s^\omega$	$(\mathbf{r}_s^\omega)^2$	$\mathbf{B}_2$	$\mathbf{r}_s^\omega$	$(\mathbf{r}_s^\omega)^2$	$(\mathbf{h}^\omega)^2$
$\omega_1$	-0,005	-0,474	22,47	-0,040	-0,252	6,37	28,84
$\omega_2$	-0,026	-0,280	7,86	0,254	0,760	57,78	65,63
$\omega_3$	-0,183	0,811	65,73	0,074	-0,560	31,31	97,04
$\omega_4$	0,090	0,013	0,02	-0,860	-0,975	94,98	94,99
$\omega_5$	1,138	1,000	99,96	0,003	0,003	0,00	99,96

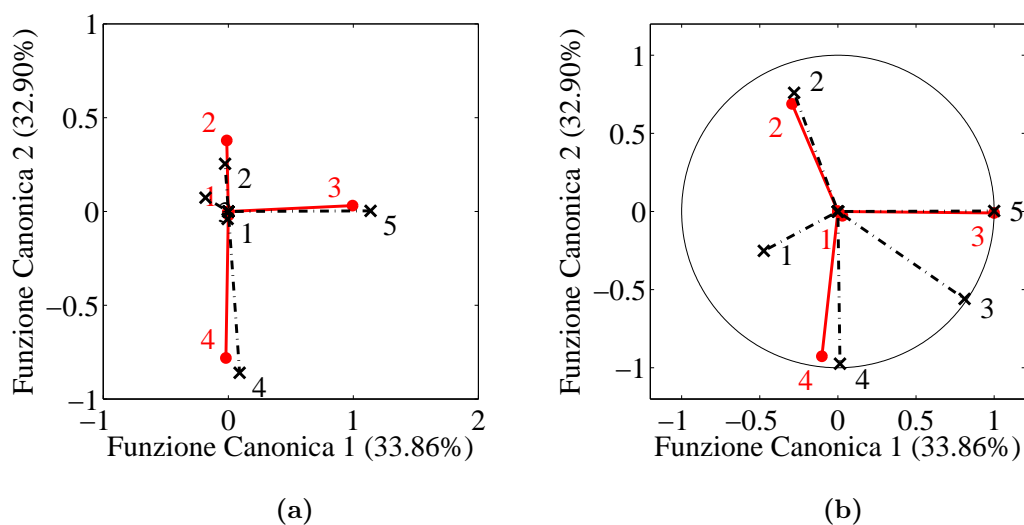
funzione canonica); inoltre, la colonna relativa ad  $(\mathbf{h}^z)^2$  mostra un contributo molto limitato da parte della variabile  $\mathbf{z}_1$  che, a giudicare dalle prime due funzioni canoniche, potrebbe essere eliminata dall'analisi. Nel complesso, in questa forma i risultati sono completi, ma la loro lettura è poco immediata.

- *Forma grafica*: in questo caso la comprensione dei risultati è molto più diretta, sebbene siano evidenziate solo le statistiche più significative. Generalmente si riportano i valori dei coefficienti di struttura delle prime due funzioni canoniche, ripetendo queste rappresentazioni per tutte le coppie di *canonical functions* esplicative di quote significative di varianza. La correlazione tra due variabili è individuabile applicando la regola del prodotto scalare a coppie di variabili di matrici diverse, rappresentate con segmenti di colore e tratto differente. L'attenzione deve quindi ricadere sulla lunghezza di entrambi i segmenti e sull'angolo compreso tra essi, ricordando che il prodotto scalare è massimizzato da lunghezze prossime ad uno<sup>3</sup> ed angoli quanto più acuti possibile.

Si riportano nel seguito due possibili rappresentazioni:

- a) la prima, Figura 2.10b, riporta i valori dei coefficienti di struttura delle prime due funzioni canoniche di Tabella 2.1; si tratta di una rappresentazione comune in letteratura (Cajo e Ter Braak, 1990; Graffelman, 2005), a cui è stato aggiunto per completezza e confronto un grafico dei *canonical loading*, Figura 2.10a. Per aumentare l'immediatezza di lettura, sono stati utilizzati due colori e tratti diversi: una linea rossa continua per le variabili di  $\mathbf{Z}$ , nera e tratteggiata per quelle di  $\mathbf{\Omega}$ .

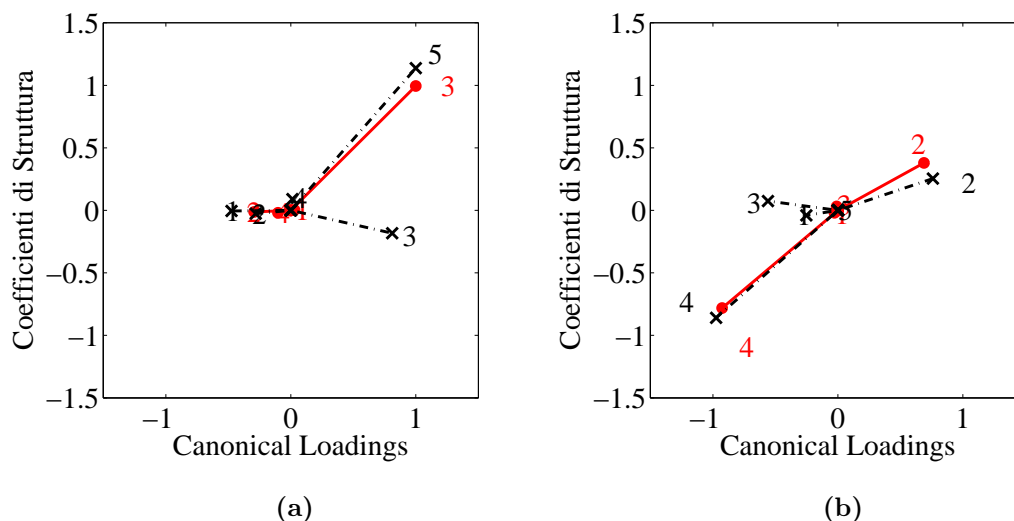
<sup>3</sup>Si ricordi a tal proposito che i coefficienti di struttura assumono valori compresi tra 0 e 1.



**Figura 2.10** Esempio di rappresentazione classica dei risultati della CCA, relativi alle prime due funzioni canoniche: (a) canonical loading; (b) coefficienti di struttura.

Come si può notare, le correlazioni individuate sono ovviamente analoghe a quelle di Tabella 2.1, ma la loro lettura è molto più facile ed immediata.

- b) l'altra opzione, rappresentata in Figura 2.11, non trova riscontro in letteratura e nasce dall'esigenza di evitare il confronto tra più funzioni canoniche, limitandosi ad esporre i risultati di una singola funzione (che di fatto è quanto voluto). In Figura 2.11 si riportano i valori dei *canonical loading* in ascissa e dei coefficienti di struttura in ordinata: una correlazione è tanto più significativa quanto maggiore è lo scostamento dall'origine degli assi.



**Figura 2.11** Nuova rappresentazione proposta dei risultati della CCA: (a) funzione canonica 1; (b) funzione canonica 2.

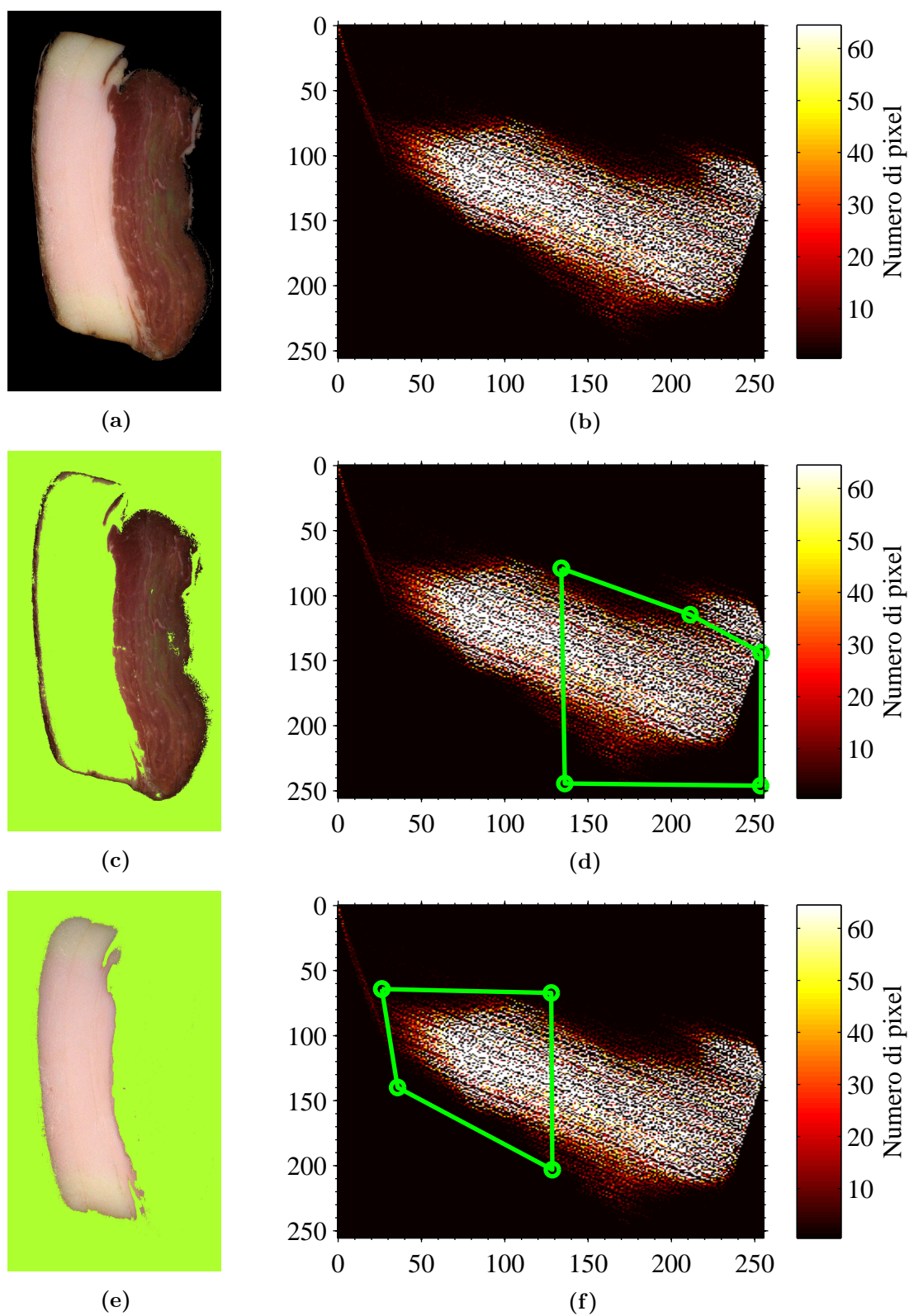
Questa tipologia di rappresentazione evidenzia i risultati specifici di una funzione canonica, evitando possibili errori di interpretazione e chiarendo l'eventuale presenza

di multicollinearità nei dati di origine, fenomeno che si manifesta con valori di *canonical loading* e coefficienti di struttura molto diversi tra loro. Inoltre, permette di individuare il contributo di ciascuna funzione canonica, caratteristica non disponibile nella rappresentazione di Figura 2.10; nello specifico, è ora possibile attribuire la correlazione tra  $\mathbf{z}_3 - \omega_5$  alla prima funzione canonica, mentre le correlazioni  $\mathbf{z}_2 - \omega_2$  e  $\mathbf{z}_4 - \omega_4$  sono individuate dalla seconda.

## 2.4 Analisi d'immagine

Il recente sviluppo di apparecchiature per l'acquisizione di immagini digitali ha consentito di sviluppare delle tecniche di analisi di immagine in grado di misurare o stimare specifiche proprietà chimico-fisiche di un prodotto, legate principalmente al suo colore o alle caratteristiche superficiali di rugosità e tessitura. L'applicazione di queste metodiche ai prodotti alimentari risulta di notevole interesse perché numerose proprietà organolettiche sono strettamente legate al colore (basti pensare alla freschezza di un ortaggio o all'ossidazione degradativa di una fetta di carne rancida). Nel caso in esame, l'utilizzo delle immagini digitali può costituire uno strumento molto interessante per caratterizzare la qualità dello speck, dal momento che numerosi indicatori sono legati al colore del prodotto stagionato. Inoltre, nell'immaginario comune è risaputo che la scelta del consumatore finale è fortemente influenzata dall'aspetto e dal colore della fetta di speck esposta, che può indurre una sensazione di freschezza e invito all'assaggio ovvero un rifiuto, legato a colori scuri, aspetto secco e colorazioni fluorescenti. Oltre a poter essere utilizzata per la stima e la classificazione dei campioni, l'analisi d'immagine è anche rapida, economica, e non invasiva.

Nel corso di questa Tesi sono state utilizzate immagini digitali ad alta definizione, nel seguito definite con  $\mathbf{J}$ , di dimensione  $[i \times j \times 3]$ , in cui  $i$  e  $j$  rappresentano il numero di *pixel* dei due lati e 3 è uguale al numero di canali dell'immagine RGB (rosso, verde e blu). Si riporta in Figura 2.12a un tipico esempio di immagine a disposizione. Le immagini sono state inizialmente equalizzate per ogni singolo canale e quindi segmentate in modo automatico per rimuovere il riferimento per il bianco ed il righello (come discusso alla sezione 1.7); quindi, per selezionare la fetta di speck, si è fatto ricorso all'analisi multivariata di immagine (MIA, *Multivariate Image Analysis*). L'analisi MIA è stata introdotta da Geladi e Esbensen (1989) con l'obiettivo primario di risolvere problemi di classificazione. Questo primo lavoro propone di riarrangiare la matrice tridimensionale  $\mathbf{J}$  in una matrice bidimensionale  $\mathbf{K}$ ,  $[jk \times 3]$ .  $\mathbf{K}$  è quindi decomposta mediante un'analisi PCA e genera due vettori di *score*, indicati generalmente con  $\mathbf{t}_1$  e  $\mathbf{t}_2$ . Nell'immagine originale, ogni *pixel* è associato ad una terna di scalari, corrispondenti rispettivamente alle intensità di rosso, verde e blu;



**Figura 2.12** Esempi di applicazioni di maschere, relative nello specifico al campione 20 lato A: (a) immagine originale e (b) relativo istogramma degli score; (c) parte magra e cotenna; (d) maschera della parte magra nel diagramma degli score; (e) parte grassa ed (f) relativa maschera nel diagramma degli score.



dopo l'utilizzo del modello PCA, un *pixel* è invece rappresentato da due soli scalari, corrispondenti ai valori degli *score*. La rappresentazione tipica è un diagramma bidimensionale  $t_1$  vs.  $t_2$ , in cui si riportano con intensità di colore diverso le numerosità dei *pixel* che stanno in un determinato intervallo di intensità di luce nell'immagine originale (Figura 2.12b). Inoltre, in un grafico di questo tipo simili intensità e rapporti di rosso, verde e blu sono rappresentati in un unico *cluster* (Geladi e Grahn, 1997). Selezionando quindi specifiche zone nello spazio degli *score* è possibile individuare uno o più di questi *cluster*, ciascuno associato ad una tipologia cromatica; tali zone, chiamate maschere (*masks*), consentono di identificare delle regioni dell'immagine originale caratterizzate da caratteristiche colorimetriche simili e quindi correlate (Liu e MacGregor, 2007). Ad esempio, si riporta in Figura 2.12d l'individuazione della maschera per la parte magra e la cotenna, mentre in Figura 2.12f quella relativa alla parte grassa. Figura 2.12 chiarisce come dall'istogramma degli *score* dell'immagine digitale (Figura 2.12b) sia possibile selezionare delle zone di interesse caratterizzate da simili indici cromatici, definendo delle maschere come quelle evidenziate in verde in Figura 2.12d e 2.12f.

Completato il pretrattamento, sono state implementate due diverse tecniche statistiche per la caratterizzazione delle immagini: l'analisi del colore medio analizza l'informazione legata al colore, mentre la trasformata *wavelet* descrive la tessitura e la rugosità superficiale del prodotto. La tessitura è definita come un descrittore delle variazioni locali di intensità tra *pixel* situati nelle immediate vicinanze gli uni degli altri (Russ, 1999); alternativamente, Bharati *et al.* (2004) la identificano con un attributo che rappresenta gli arrangiamenti spaziali delle intensità di luce dei singoli canali RGB.

#### 2.4.1 Analisi del colore medio

Si è scelto di analizzare il colore medio di entrambe le parti della fetta di speck, magra e grassa, che vengono trattate singolarmente vista la notevole differenza cromatica che le caratterizza. Dopo aver applicato la maschera di interesse (del magro o del grasso), è stato implementato un algoritmo di segmentazione automatica che consente di evidenziare la massima porzione di magro (o grasso) priva di interferenze esterne, costituite di fatto dal colore dello sfondo. L'algoritmo è basato su segmentazioni iterative su ogni lato dell'immagine, con un criterio di arresto tarato sulla percentuale di *pixel* selezionati. In Figura 2.13 si riporta un diagramma a blocchi di questa procedura, completato da un esempio fotografico per maggiore chiarezza. Come si può notare, ad ogni iterazione viene segmentato ciascun lato dell'immagine, eliminando quelle righe o quelle colonne che non contengono una percentuale sufficiente di *pixel* selezionati. La soglia minima  $L^*$  di *pixel* viene alzata progressivamente all'aumentare del numero di iterazioni, in modo da avere una segmenta-

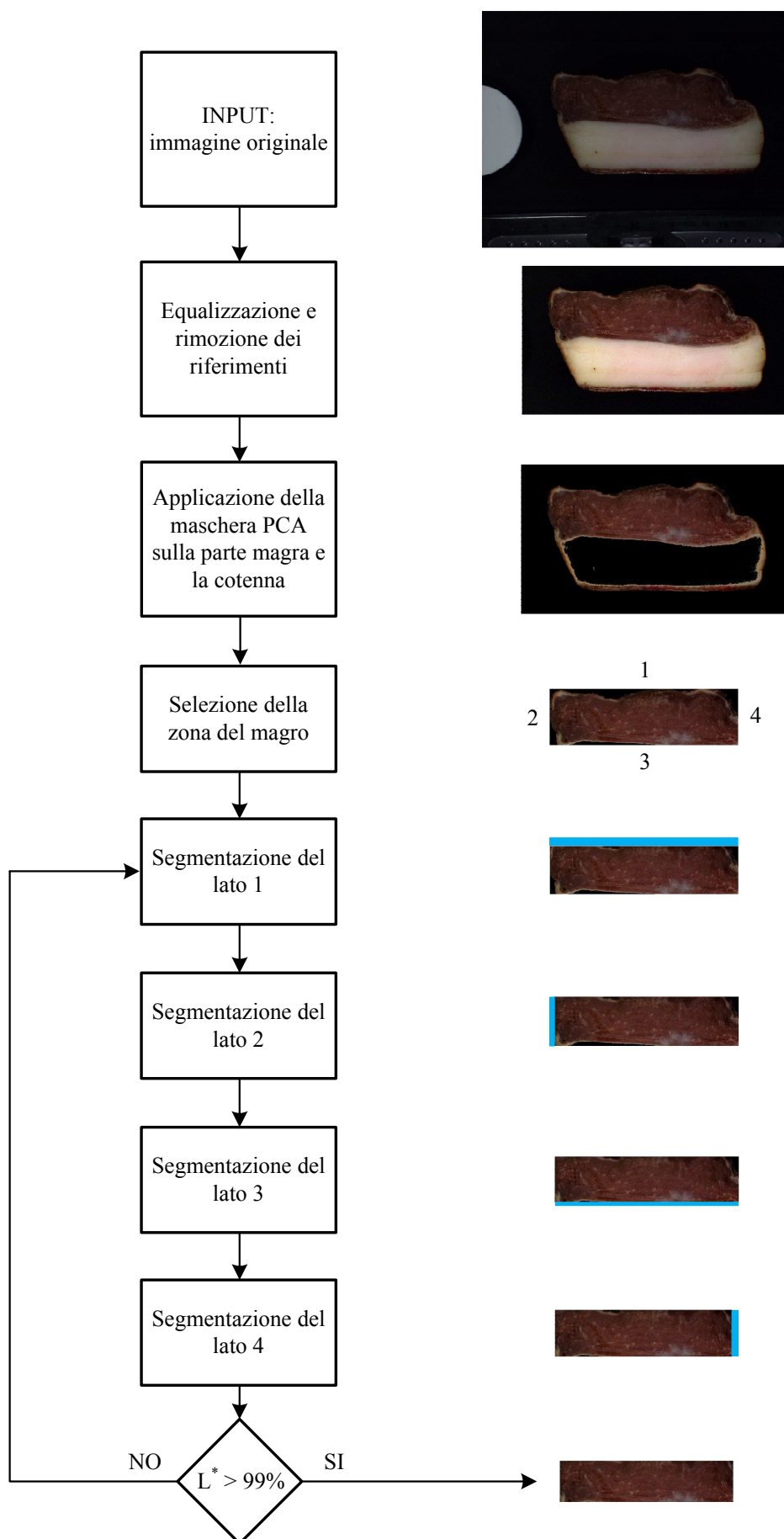
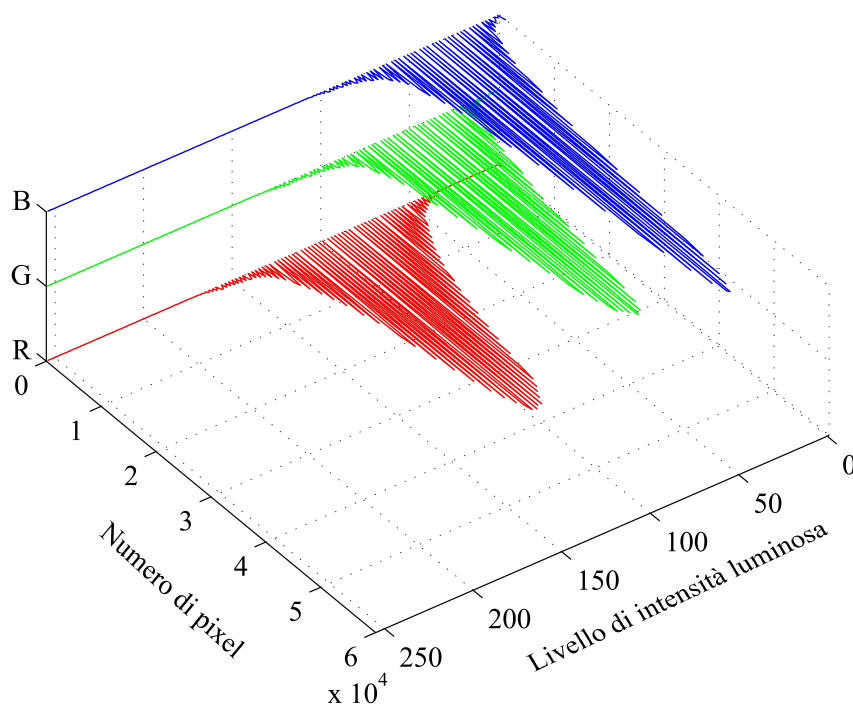


Figura 2.13 Diagramma a blocchi della procedura di segmentazione.

zione progressiva della fetta di speck. Ciò si rende necessario a causa della forma irregolare del prodotto, poiché un aumento troppo brusco della soglia può causare segmentazioni eccessive, che non massimizzano l'area considerata e causano anzi una perdita di informazione.

Ottenuta la porzione segmentata, sono state estratte le distribuzioni dell'intensità luminosa dei tre canali RGB, concatenandoli orizzontalmente in un vettore; la procedura è stata quindi ripetuta per tutti i campioni. Ad ulteriore chiarimento, si riporta in Figura 2.14 un esempio delle distribuzioni di intensità dei tre canali RGB. Gli istogrammi RGB di tutti i campioni (i.e., immagini) sono stati quindi collezionati in una matrice  $\mathbf{C}$  [ $N \times (255 * 3)$ ], utilizzata per predire le proprietà di interesse con modello PLS.



**Figura 2.14** Istogrammi dell'intensità luminosa nei tre canali RGB, campione 20 lato A.

### 2.4.2 Analisi della tessitura - decomposizione wavelet

L'analisi della tessitura tramite trasformata *wavelet* (Addison, 2002) è un metodo che utilizza la frequenza spaziale delle variazioni di intensità di luce nei *pixel* per convertire l'immagine originale in nuove forme, che ne sono un'approssimazione o ne esaltano i dettagli. In una trasformazione *wavelet* discreta e bidimensionale (DWT, *discrete wavelet transform*), un'immagine viene infatti processata mediante filtri passa alto e passa basso, che identificano rispettivamente gli elementi a bassa frequenza, chiamati approssimazioni,

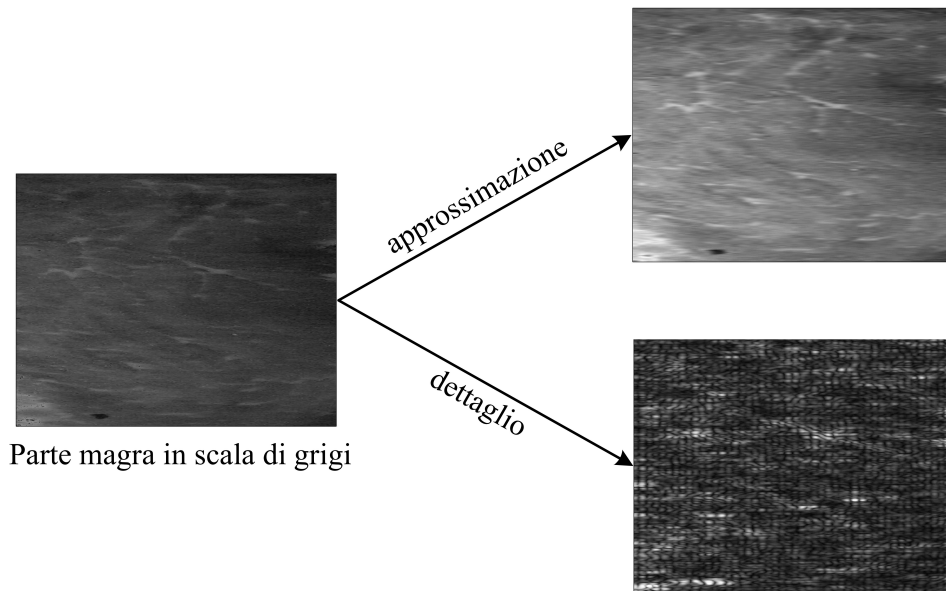
e quelli a frequenza elevata, definiti dettagli. L'operazione principale caratteristica di una DWT è la convoluzione del segnale originale con basi ortonormali costituite da versioni dilatate e traslate di una funzione *wavelet* assegnata (Addison, 2002). La convoluzione è definita da:

$$W(a, b) = \frac{1}{\sqrt{a}} \sum_{x=1}^{\text{n}^\circ \text{ pixel}} f(x) \psi\left(\frac{b-x}{a}\right) \quad , \quad (2.46)$$

in cui  $f(x)$  è il segnale originale,  $\psi$  è la funzione *wavelet* scelta, mentre  $a$  e  $b$  sono due parametri di dilatazione e traslazione, rispettivamente.

In questo studio, si è scelto analizzare le immagini in scala di grigi utilizzando le *wavelet* di tipo Coiflets 5; la scelta del numero di scale di risoluzione, chiamate anche livelli di decomposizione, è basata sulla correlazione bidimensionale tra l'immagine originale e quella ricostruita, fissando una soglia minima pari a 0.8 (Tomba *et al.*, 2010).

In questa Tesi sono stati condotti entrambi gli approcci, sia multiapprossimazione che multidettaglio, di cui si riportano due esempi in Figura 2.15. Come si può notare, il primo approccio approssima per scale successive l'immagine originale, che nella ricostruzione risulta caratterizzata da variazioni di intensità molto più ampie. Viceversa, l'approccio multidettaglio ne esalta i particolari, a cui corrispondono brusche variazioni di intensità luminosa, e ciò consente di evidenziare con notevole efficacia il grasso infiltrato tra le fibre.



**Figura 2.15** Esempio di applicazione della trasformata wavelet.

Al fine di caratterizzare i parametri legati alla tessitura dell'immagine ricostruita  $\mathbf{J}_{ric,s}$ , ad ogni livello  $s$  sono stati calcolati sei descrittori statistici, costituiti dalla media dell'intensità luminosa dei *pixel*, dalla deviazione standard, e da:

Nelle equazioni riportate sopra,  $P_s(l)$  indica la probabilità di trovare un *pixel* di coordinate

$$\text{asimmetria}|_s = \frac{M_{s,3}}{\sigma_s^3} \quad (2.47)$$

$$\text{curtosi}|_s = \frac{M_{s,4}}{\sigma_s^4} \quad (2.48)$$

$$\text{entropia}|_s = - \sum_{j=1}^{J-2} \sum_{k=1}^{K-2} P_s(l) \log_2[P_s(l)] \quad (2.49)$$

$$\text{energia}|_s = \|\mathbf{J}_{ric,s}\| \quad (2.50)$$

$(j,k)$  con intensità luminosa  $l$  alla scala di risoluzione  $s$ ,  $\|\cdot\|$  è la norma di Frobenius,  $M_{s,3}$  e  $M_{s,4}$  sono il terzo ed il quarto momento centrale di  $\mathbf{J}_{ric,s}$ , e  $\mu$  è l'intensità luminosa media in  $\mathbf{J}_{ric,s}$ . Si ricorda infine che un momento centrale di ordine  $h$ , relativo al livello  $s$ , è definito come:

$$\mathbf{j} = \text{vect}(\mathbf{J}) \quad (2.51)$$

$$M_{s,h} = \sum_{i=1}^I (\mathbf{j}_i - \bar{\mathbf{j}})^h P_s(\mathbf{j}_i) \quad (2.52)$$

in cui  $\mathbf{j}$  è la vettorizzazione dell'immagine  $\mathbf{J}$ .

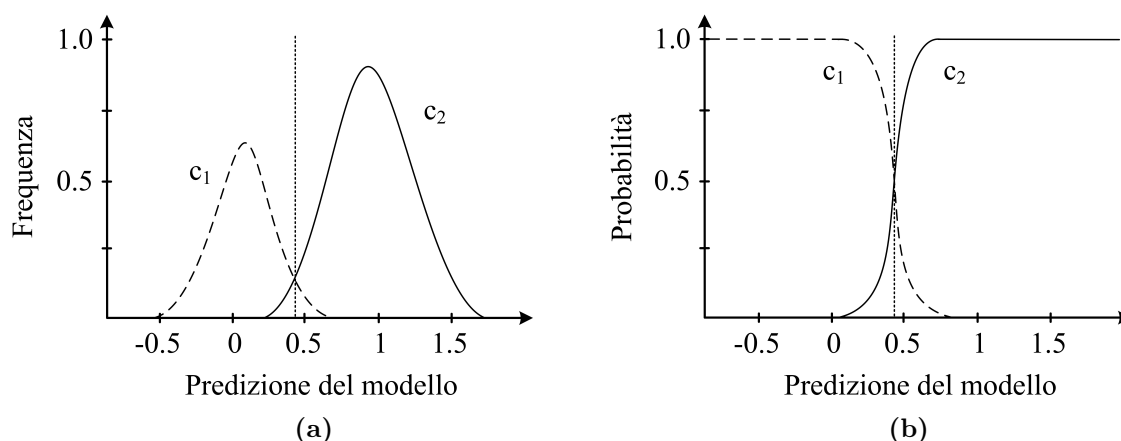
## 2.5 Metodiche di classificazione

Per classificare i campioni rispetto ad alcune delle variabili categoriali a disposizione sono state utilizzate diverse metodiche, in particolare:

- *k-nearest neighbour* (*k*NN): è uno degli algoritmi più utilizzati, che attribuisce l'appartenenza ad una classe basandosi sulle caratteristiche dei  $k$  campioni più vicini a quello considerato (Friedman *et al.*, 1977). In altre parole, un campione viene assegnato ad una classe  $c_1$  se questa è la più frequente tra i  $k$  campioni più vicini, con distanze di tipo euclideo. In questa Tesi il valore di  $k$  è stato imposto pari a  $(2C - 1)$ , in cui  $C$  è il numero di variabili categoriali che compongono una determinata classe. Ad esempio, per una variabile binaria quale è il sesso,  $k$  è uguale a 3, mentre con la terna dei regimi alimentari il valore di  $k$  sale a 5. Ciò assicura di basarsi sempre su un numero di campioni maggiore di uno, garantendo altresì la prevalenza netta di una classe tra i campioni considerati. Il pregio di questo algoritmo è la sua semplicità, anche se il calcolo delle distanze è computazionalmente oneroso, soprattutto con un numero elevato di campioni;
- analisi discriminante lineare LDA (*linear discriminant analysis*): è un metodo utilizzato per determinare una combinazione lineare di caratteristiche che consente di separare in due o più classi i campioni in esame (Friedman, 1988). In questo senso l'analisi LDA è molto simile alla PCA, ma mentre in quest'ultima le combinazioni lineari sono costruite basandosi sulla correlazione delle variabili, LDA cerca invece di modellare le differenze tra

le classi di dati. Inoltre, a differenza delle analisi della varianza ANOVA e di regressione, le variabili dipendenti sono costituite da variabili categoriali, rappresentate da interi positivi o da una stringa di testo;

- analisi discriminante quadratica QDA (*quadratic discriminant analysis*): rappresenta un'estensione del metodo LDA, in cui la suddivisione tra le diverse classi è ottenuta con una combinazione quadratica (Friedman, 1988). Inoltre, a differenza di LDA, in questo caso è necessario avere una distribuzione gaussiana dei campioni all'interno di ciascuna classe affinché il metodo sia efficiente;
- metodo PLSDA (*partial least squares discriminant analysis*): deriva dalla proiezione su strutture latenti (PLS), in cui però lo spazio della risposta è costituito da variabili categoriali: l'appartenenza ad una classe viene indicata con 1, la non appartenenza con 0. Queste ultime, tuttavia, non vengono predette con precisi valori interi, equivalenti alle singole classi, ma assumono valori intermedi. Devono per questo essere convertiti in variabili categoriali basandosi su delle soglie di distribuzione probabilistica, determinate durante la calibrazione del modello stesso (Pérez *et al.*, 2009). Per maggiore chiarezza, si riporta in Figura 2.16 un esempio di questa procedura, relativo ad una classificazione binaria.



**Figura 2.16** *Attribuzione delle classi del metodo PLSDA mediante distribuzioni di probabilità e relative soglie.*

In Figura 2.16a sono riportate le distribuzioni di probabilità delle due classi ottenute dalla calibrazione del modello PLSDA; come si può notare, dal punto di intersezione delle due curve di distribuzione si determina la soglia discriminante, necessaria per convertire le predizioni del modello stesso in variabili categoriali. Per tutti i valori predetti inferiori alla soglia, la probabilità di attribuzione della classe  $c_1$  è massima (Figura 2.16b); viceversa, oltre la soglia viene attribuita la classe  $c_2$ .

In questa Tesi sono stati utilizzati anche altri metodi più articolati, ad esempio:

- modificando la relazione interna del metodo PLSDA lineare, è possibile ottenere una nuova metodica di classificazione, caratterizzata da una relazione interna quadratica. Anche in questo caso le predizioni del modello vengono trasformate in variabili categoriali utilizzando le soglie di distribuzione probabilistica determinate in calibrazione, in modo analogo rispetto alla Figura 2.16;
- l'algoritmo SIMCA (*soft independent modelling of class analogy*) è la metodica di classificazione più articolata utilizzata in questa Tesi, poiché prevede di costruire un modello PCA dedicato per ciascuna classe presente (Wold e Sjostrom, 1977). In ognuno di questi  $C$  modelli viene calcolata la distanza ortogonale media dei campioni di calibrazione dall'iper-piano ottenuto, che costituisce la distanza critica di classificazione. Quindi, i campioni di convalida vengono proiettati in ciascun modello PCA e ne viene calcolata la distanza residua. Un campione è assegnato ad una data classe quando la sua distanza residua dal relativo modello è all'interno del limite di confidenza.

Poiché i problemi di collinearità tra le variabili originali impediscono l'utilizzo dei metodi  $k$ NN, LDA e QDA sui dati tal quali, questi sono stati applicati sugli *score* di un'analisi PCA, utilizzata per filtrare e rendere ortogonali le variabili. Il numero di componenti principali è stato scelto minimizzando gli errori di classificazione in convalida incrociata (Balabin *et al.*, 2010); tale criterio caratterizza anche gli algoritmi PLSDA, sia con relazione interna lineare che quadratica, anche se questi ultimi vengono applicati direttamente ai dati originali, senza la necessità di un'analisi PCA preliminare. Il metodo SIMCA, invece, è costruito in modo da tenere già in considerazione questo fattore, che non deve quindi essere specificato esternamente.

### 2.5.1 Analisi di sensitività e specificità

Oltre agli errori medi, l'efficacia di un metodo di classificazione viene caratterizzata calcolandone anche la sensitività e la specificità di attribuzione di ciascuna classe, definite nel seguito con riferimento ad una classificazione binaria (Fawcett, 2006):

$$\text{sensitività} = \frac{\text{positivi corretti PC}}{\text{positivi corretti PC} + \text{negativi errati NE}} \quad (2.53)$$

$$\text{specificità} = \frac{\text{negativi corretti NC}}{\text{negativi corretti NC} + \text{positivi errati PE}} \quad (2.54)$$

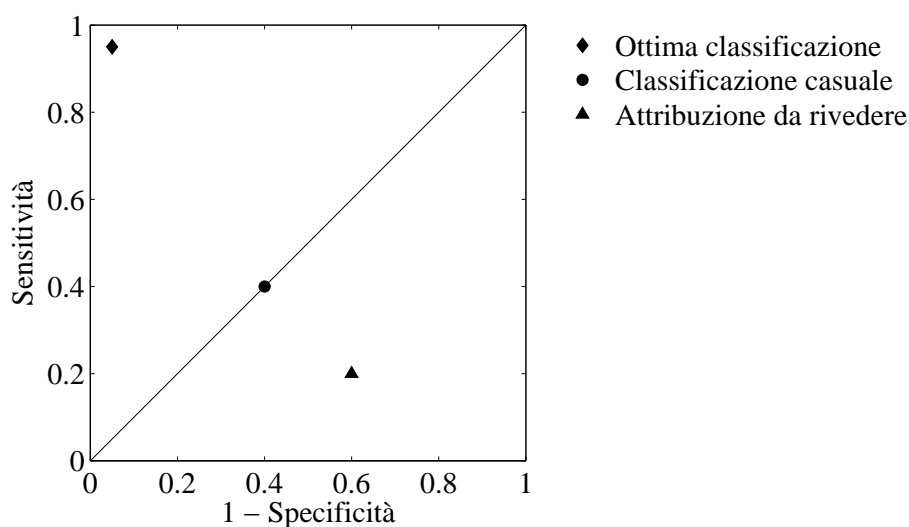
I risultati di un metodo di classificazione sono infatti esprimibili utilizzando una matrice "di confusione quadrata", costituita nello specifico da:

$$\text{matrice di confusione} = \begin{pmatrix} \text{PC} & \text{PE} \\ \text{NE} & \text{NC} \end{pmatrix}, \quad (2.55)$$

in cui i termini della diagonale principale indicano le classificazioni corrette, mentre la sommatoria degli elementi esterni è equivalente al numero di errori. I valori di sensitività e

specificità possono essere utilizzati per costruire un grafico come quello riportato in Figura 2.17, in cui le diverse prestazioni di classificazione sono identificabili in modo semplice e veloce.

L'interpretazione di Figura 5.2 è la seguente (Fawcett, 2006): le prestazioni ideali sono caratteristiche del punto  $(0,1)$ , poiché a queste coordinate corrispondono zero errori di classificazione; viceversa, il punto  $(1,0)$  rappresenta le peggiori prestazioni ottenibili, con nessuna classificazione corretta<sup>4</sup>.



**Figura 2.17** Caratterizzazione dei modelli di classificazione in base ai valori di sensibilità e specificità.

Le prestazioni di qualsiasi modello sono quindi comprese tra gli estremi opposti di questo quadrato. Inoltre, in tutti i punti della diagonale evidenziata in nero la classificazione è casuale, con probabilità di attribuzione equivalenti, mentre il triangolo inferiore è caratteristico di modelli che potrebbero essere informativi, ma in cui è necessario invertire il criterio di attribuzione delle classi. Nel complesso, quindi, un modello di classificazione è da considerarsi valido solo se le sue coordinate appartengono al triangolo superiore e sono quanto più vicine possibile all'estremo  $(0,1)$ .

<sup>4</sup>Si faccia riferimento alle definizioni (2.53) e (2.54).



# Capitolo 3

## Analisi esplorativa preliminare e studio delle correlazioni

Vista la grande varietà dei dati a disposizione, una prima parte dell'analisi è stata dedicata alla comprensione delle correlazioni tra le variabili presenti, individuandone legami e indipendenze, verificando altresì che la rilevanza statistica trovi riscontro sul significato fisico delle variabili stesse. L'approccio utilizzato si basa sull'utilizzo delle mappe di correlazione, utili ad esprimere in modo immediato la correlazione o anticorrelazione tra due o più variabili; per quanto concerne l'analisi delle acquisizioni spettroscopiche, invece, si è reso necessario l'utilizzo dell'analisi delle correlazioni canoniche (CCA).

Si discutono nel seguito le maggiori correlazioni individuate, suddivise per dataset di appartenenza; la trattazione statistica è sempre seguita da una interpretazione fisica dei risultati, in modo da procedere nei due approcci in modo parallelo, identificando consistenze ed anomalie.

### 3.1 Dati di filiera

Il set dei *Dati di filiera* è costituito da ben 81 variabili, che ne rendono articolata e dispersiva una trattazione unitaria; si è quindi scelto di analizzare dapprima le singole sottomatrici, estendendo poi lo studio agli eventuali legami tra coppie di sottomatrici diverse.

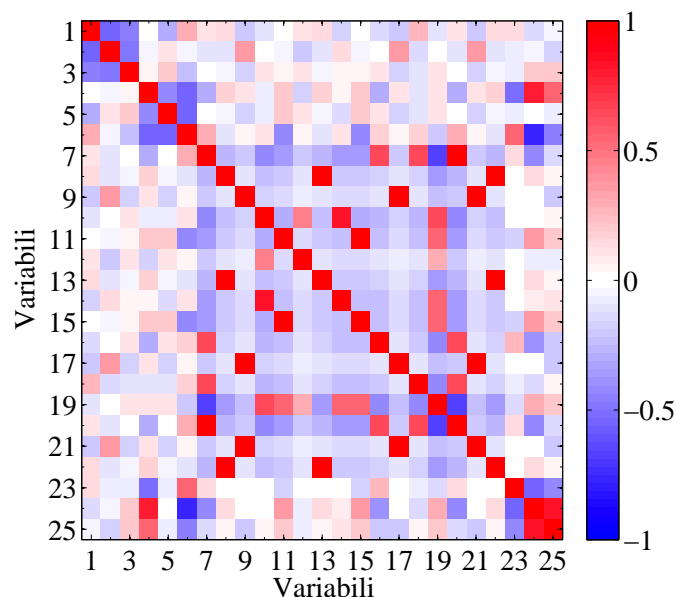
Si riporta nel seguito un elenco dettagliato delle maggiori correlazioni ed anticorrelazioni individuate, con le relative mappe di correlazione a supporto. Si evidenzia fin da ora che, a causa della scarsa significatività statistica dei risultati ottenuti, non verranno discusse per esteso le sezioni relative alle correlazioni tra le sottomatrici *Carcassa Macellata* e *Cromatografia* e tra *Cromatografia* e *Prodotto Stagionato*.

#### 3.1.1 Caratteristiche dell'animale vivo

Le variabili dell'animale vivo coinvolgono le informazioni sul regime alimentare, l'azienda di nascita, il sesso, il peso, e la giornata di macellazione del suino. Nell'analisi condotta sono state trascurate le correlazioni tra le variabili padre e madre perché difficilmente riproducibili e poco significative all'interno dell'ottica globale dello studio; inoltre, anche le anticorrelazioni tra le variabili categoriali non sono state considerate perché banali, dal

momento che l'attribuzione di una classe esclude di fatto la possibilità di appartenere a tutte le altre della medesima categoria.

Si riportano in Figura 3.1 ed in Tabella 3.1 i risultati ottenuti.



**Figura 3.1** Mappa di correlazione relativa alle variabili della sottomatrice *Animale vivo*.

**Tabella 3.1** Elenco delle principali correlazioni individuate all'interno della sottomatrice *Animale vivo*.

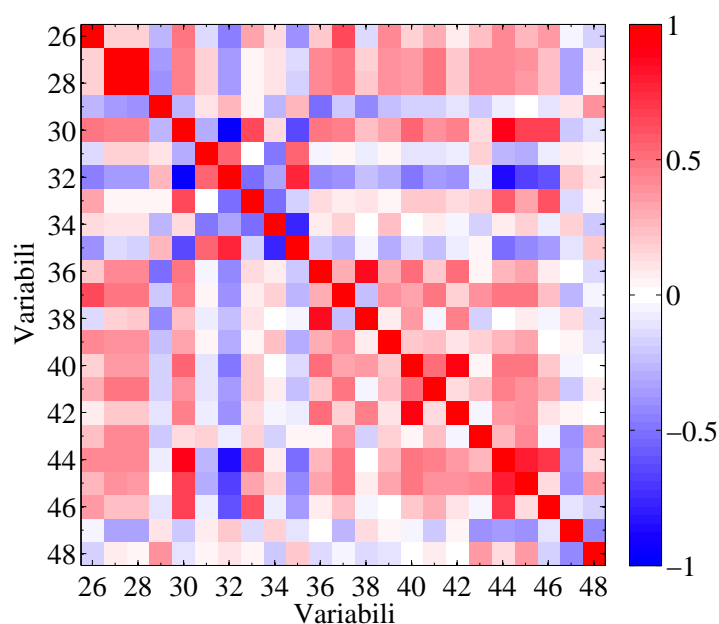
Variabile	Correlazioni	Anticorrelazioni
(4) macello: 1° giornata	(24) incremento netto (25) peso vivo	(23) sesso
(6) macello: 3° giornata	(23) sesso	(24) incremento netto (25) peso vivo
(23) sesso		(24) incremento netto (25) peso vivo
(24) incremento netto	(25) peso vivo	

La trattazione statistica è consistente con il significato fisico delle variabili perché nella prima giornata di macellazione sono stati abbattuti gli esemplari ingrassati più rapidamente, caratterizzati da un incremento netto superiore; viceversa, nell'ultima giornata sono stati macellati gli esemplari ingrassati più lentamente che normalmente hanno un peso vivo inferiore. Risulta verificata anche la correlazione relativa al sesso, poiché gli esemplari femmina ingrassano più lentamente e raggiungono pesi inferiori rispetto ai maschi. Si noti che in questo caso è da considerarsi una variabile categoriale anche il sesso poiché, all'interno

di essa, è stato attribuito il valore 0 agli esemplari maschi ed il valore 1 alle femmine; l'anticorrelazione deve quindi essere letta nel modo seguente: all'aumentare del sesso (gli esemplari femmina) incremento netto e peso vivo diminuiscono.

### 3.1.2 Caratteristiche della carcassa macellata

Si riporta in Figura 3.2 la mappa di correlazione delle variabili considerate.



**Figura 3.2** Mappa di correlazione relativa alle variabili della sottomatrice Carcassa macellata.

Come si può notare, rispetto al caso precedente le colorazioni sono molto più vivaci, indice sia di un numero maggiore di correlazioni sia di una significatività molto più accentuata; tuttavia, sono identificabili sin da ora dei quadrati rossi di variabili consecutive, legati presumibilmente alle diverse misure e quantificazioni degli spessori di lardo. Infatti, per i due diversi punti di analisi, sono disponibili le misure degli spessori di lardo esterno - variabili (37) e (41) - , interno - variabili (38) e (42) - , e le loro sommatorie - variabili (36) lardo totale 1T e (40) lardo totale 13T.

Si riporta in Tabella 3.2 il dettaglio delle correlazioni individuate. Anche in questo caso il significato fisico delle variabili sostiene e conferma l'analisi statistica per le ragioni indicate di seguito.

- La resa esprime la percentuale in peso della carcassa rimasta dopo il raffreddamento della stessa, ed è complementare alla perdita dovuta all'evaporazione dell'acqua contenuta nei tessuti ed al gocciolamento del muscolo. La presenza di grasso incide su questo parametro per due diverse ragioni:

**Tabella 3.2** *Elenco delle principali correlazioni individuate all'interno della sottomatrice Carcassa macellata.*

Variabile	Correlazioni	Anticorrelazioni
(26) resa (%)	(30) spessore di grasso (37) lardo 1T esterno	(32) carne magra (%) (35) qualità altissima
(27) peso morto caldo	(30) spessore di grasso (36) lardo 1T totale	(29) perdita (%) (32) carne magra (%)
(28) peso morto freddo	(37) lardo 1T esterno (41) lardo 13T esterno	(47) drip loss
(29) perdita (%)	(48) pH a 45 minuti	(36) lardo 1T totale (38) lardo 1T interno
(30) spessore di grasso	(33) qualità media (44) lardo minimo totale (45) lardo minimo esterno (46) lardo minimo interno	(32) carne magra (%) (35) qualità altissima
(31) spessore di carne	(32) carne magra (%) (35) qualità altissima	(34) qualità alta
(32) carne magra (%)	(35) qualità altissima	(30) spessore di grasso (36) lardo 1T totale (40) lardo 13T esterno (44) lardo minimo totale (45) lardo minimo esterno (46) lardo minimo interno
(33) classe di qualità media	(44) lardo minimo totale (45) lardo minimo esterno (46) lardo minimo interno	(32) carne magra (%)
(34) classe di qualità alta		(31) spessore di carne
(35) qualità altissima	(31) spessore di carne (32) carne magra (%)	(30) spessore di grasso (44) lardo minimo totale
(36) lardo 1T totale	(38) lardo 1T interno	
(44) lardo minimo totale	(46) lardo minimo interno	
(47) drip loss		(27) peso morto caldo (28) peso morto freddo (43) carne 13T (44) lardo minimo totale (45) lardo minimo esterno (48) pH a 45 minuti
(48) pH a 45 minuti	(29) perdita (%) (43) carne 13T (45) lardo minimo esterno	

$$\text{resa (\%)} = 1 - \frac{\text{peso vivo} - \text{peso morto caldo}}{\text{peso morto caldo}} \quad (3.1)$$

- a) minore è lo spessore di grasso, maggiori sono la quantità di carne e la superficie esposta: ciò causa un aumento della perdita in peso ed una diminuzione della resa;
- b) lo strato di grasso ha proprietà isolanti dal punto di vista termico, quindi limita la perdita in peso dovuta all'evaporazione dell'acqua. Si noti che subito dopo la macellazione, la carcassa ha una temperatura di circa  $30\div 35$  °C; l'evaporazione dell'acqua contenuta nei tessuti non è quindi un fenomeno trascurabile.

Tali aspetti giustificano quindi la correlazione riscontrata tra le variabili (26) resa % e (30), (37) - spessore di grasso e lardo 1T esterno; le due anticorrelazioni si spiegano invece in un minor quantitativo di grasso nelle carcasse di qualità altissima, caratterizzate invece da un'elevata percentuale di carne magra.

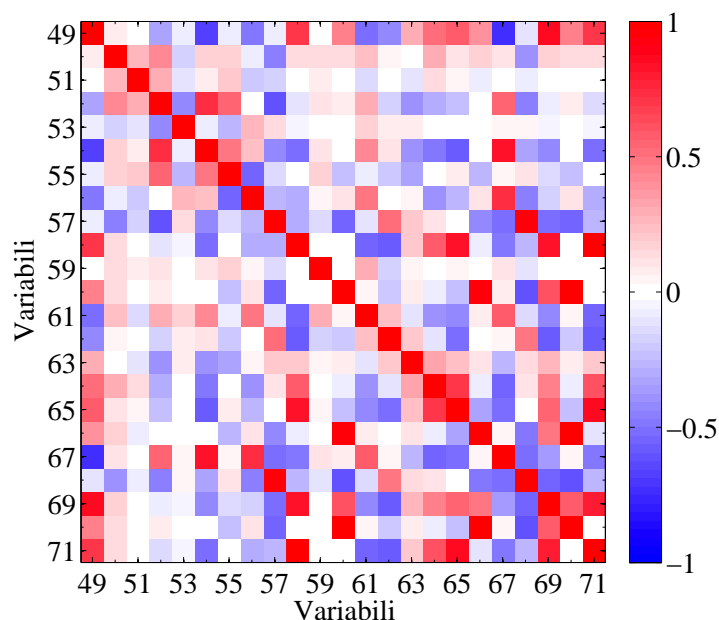
- Il peso morto freddo è ovviamente correlato all'equivalente a caldo, mentre è anticorrelato alla perdita ed al *drip loss*, che costituisce la diminuzione di peso dovuta al gocciolamento del muscolo; tale fenomeno è caratteristico della carne magra, e poiché gli strati di lardo non ne sono interessati, le rispettive variabili risultano giustamente anticorrelate.
- La qualità di una carcassa e quindi la sua classe di attribuzione è legata al suo quantitativo di grasso; una carcassa molto grassa ha una qualità media, una con discreti spessori di lardo una qualità alta, mentre una carcassa magra è classificata come ottima, di qualità altissima.
- Il *drip loss* è un parametro molto importante perché rappresenta un indicatore di qualità del taglio di carne macellato; il suo valore deve essere intermedio poiché un'eccessiva perdita di acqua determina una colorazione molto chiara ed un ridotto valore di pH, mentre un *drip loss* troppo limitato causa una colorazione di un rosso carico ed un pH elevato anche molte ore dopo la macellazione. Esso è anticorrelato al pH a 45 minuti - variabile (48) - perché una rapida diminuzione di pH facilita le perdite di acqua, mentre un pH elevato determina una maggiore ritenzione idrica: infatti, lontano dal punto isoelettrico, la massa proteica è in grado di legare un maggior numero di molecole d'acqua, per mezzo delle cariche superficiali; al punto isoelettrico, invece, la carica netta è nulla ed il *drip loss* molto elevato. Risulta invece anticorrelato:
  - a) al lardo - variabili (44) e (45) - per le proprietà isolanti di quest'ultimo;
  - b) ad entrambi i pesi morti - variabili (27) e (28) - perché ad un maggior peso della carcassa corrisponde un maggior quantitativo di grasso.

### 3.1.3 Analisi cromatografica del lardo fresco

Si riporta in Tabella 3.3 un breve sommario delle correlazioni individuate; la relativa mappa di correlazione è disponibile in Figura 3.3.

**Tabella 3.3** *Elenco delle principali correlazioni individuate all'interno della sottomatrice Cromatografia.*

Variabile	Correlazioni	Anticorrelazioni
(49) n° di iodio	(58) C18:2n6 (64) C20:3n6 (65) C20:4n6 (71) acidi n6 (69) acidi polinsaturi	(54) C16:0 (56) C18:0 (61) C20:0 (67) acidi saturi
singoli acidi grassi	raggruppamenti per classi di appartenenza	



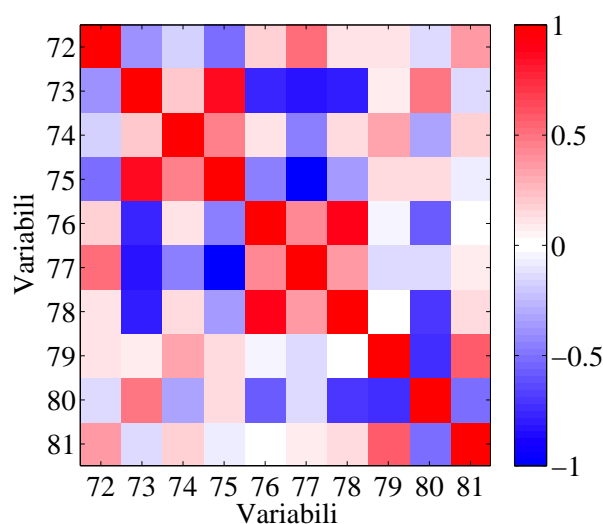
**Figura 3.3** *Mapa di correlazione relativa alle variabili della sottomatrice Cromatografia.*

La prima colonna di Figura 3.3 conferma il significato fisico del numero di iodio - variabile (49): infatti, è correlato agli acidi grassi polinsaturi, n3 ed n6 - variabili (58), (60), (64), (65), (69), (70), (71) - mentre è anticorrelato agli acidi grassi saturi - variabili (52), (54), (56), (61), (67). Ciò trova pieno riscontro nella sua quantificazione analitica, poiché rappresenta una stima del numero di insaturazioni degli acidi grassi presenti.

A titolo esplicativo, si riporta anche un esempio di correlazione tra alcuni singoli acidi e la loro categoria di appartenenza: gli acidi grassi n3 - variabile (70) - risultano molto correlati alle variabili (60) e (64), rappresentative appunto degli acidi C18:3n3 e C20:3n3.

### 3.1.4 Variabili di qualità del prodotto stagionato

La mappa di correlazione di Figura 3.4 indica delle correlazioni ed anticorrelazioni piuttosto significative tra le differenti variabili di qualità; nello specifico, si nota una stretta interdipendenza tra l'attività dell'acqua - variabile (73) - e l'umidità del magro - variabile (75) - e tra il contenuto di NaCl e le ceneri - variabili (76) e (78). Quest'ultima correlazione trova pieno riscontro nel significato fisico delle variabili, poiché una parte delle ceneri, assimilabili agli elettroliti, è proprio costituita dal sale; all'aumentare del contenuto di ceneri diminuisce invece l'attività dell'acqua libera, poiché le molecole di H<sub>2</sub>O solvatano i soluti per mezzo di ponti idrogeno.



**Figura 3.4** Mappa di correlazione relativa alle variabili della sottomatrice Prodotto stagionato.

Inoltre, tra le anticorrelazioni spicca anche quella tra l'EE SS calcolato - variabile (80) - e NaCl, ceneri e proteina grezza - variabili (76), (78) e (79); ciò trova conferma nella rappresentazione fisica delle variabili stesse, dal momento che queste hanno una solubilità molto limitata nella parte grassa, solubilizzandosi preferenzialmente nella parte acquosa della massa proteica.

Merita invece un ulteriore approfondimento l'anticorrelazione tra il grasso in termini quantitativi - variabile (80) - e l'acidità titolabile del magro stagionato - variabile (81): da un punto di vista teorico, infatti, ci si aspetterebbe una correlazione e non una anticorrelazione. Fisicamente, infatti, il grasso è presente sotto forma di fosfolipidi, esteri formati da una molecola di glicerolo e da tre acidi grassi; durante la stagionatura, però, una parte di questi esteri viene idrolizzata, provocando l'aumento degli acidi grassi liberi e quindi dell'acidità titolabile. Il processo di idrolisi è tuttavia legato all'attività enzimatica, che dipende da molti altri fattori; l'anticorrelazione individuata spinge quindi a ritenere che ci sia una buona attività enzimatica anche in presenza di limitati spessori di grasso.

Si riporta in Tabella 3.4 un elenco completo delle correlazioni individuate.

**Tabella 3.4** *Elenco delle principali correlazioni individuate all'interno della sottomatrice Prodotto stagionato.*

Variabile	Correlazioni	Anticorrelazioni
(72) calo peso (%)	(77) sostanza secca (%) (81) acidità titolabile	(73) attività dell'acqua (75) umidità del magro
(73) attività dell'acqua	(75) umidità del magro (80) EE SS calcolato	(76) NaCl (77) sostanza secca (%) (78) ceneri
(74) pH	(75) umidità del magro (79) proteina grezza	(77) sostanza secca (%) (80) EE SS calcolato
(75) umidità del magro		(77) sostanza secca (%)
(76) NaCl	(77) sostanza secca (%) (78) ceneri	(80) EE SS calcolato
(78) ceneri		(80) EE SS calcolato
(79) proteina grezza	(81) acidità titolabile	(80) EE SS calcolato
(80) EE SS calcolato		(81) acidità titolabile

### 3.1.5 Caratteristiche dell'animale vivo e della carcassa macellata

Dopo aver analizzato le singole sottomatrici di variabili, si individuano ora le eventuali correlazioni presenti tra coppie di sottomatrici differenti; tuttavia, l'interpretazione dei risultati diviene molto più complessa, rendendone la discussione prettamente qualitativa.

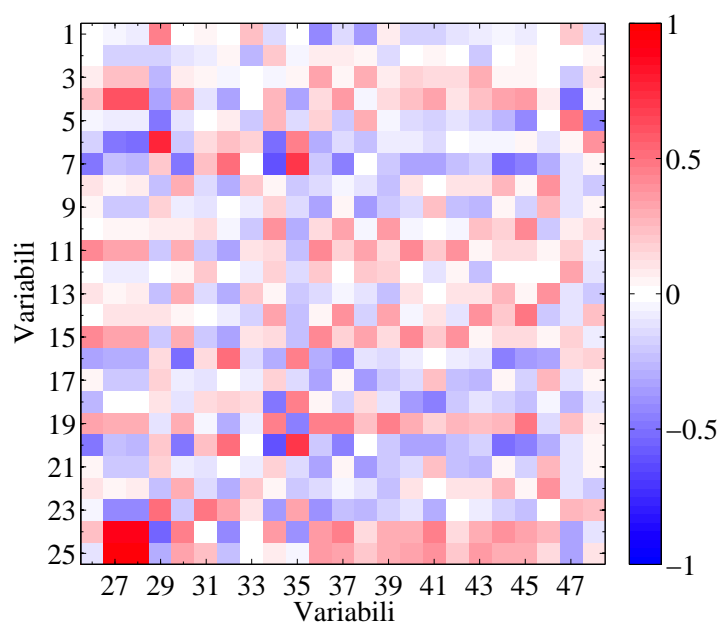
Concatenando le prime due sottomatrici, si è cercato di focalizzare eventuali legami tra le variabili dell'animale vivo e le caratteristiche della sua carcassa in fase di macellazione.

I risultati di questa analisi sono riportati in Figura 3.5 ed in Tabella 3.5.

Le correlazioni tra le due sottomatrici non sono particolarmente significative e, per poterle interpretare con significato fisico, è opportuno seguire una procedura indiretta, riconducendosi alle correlazioni delle singole sottomatrici già trattate nelle sezioni 3.1.1 e 3.1.2.

Ad esempio, la correlazione tra la qualità altissima e la terza giornata di macellazione - variabili (35) e (6) - deve essere letta ricordando che nell'ultima giornata sono stati macellati gli esemplari meno grassi; questi ultimi hanno anche una maggiore percentuale di carne magra, e quindi la correlazione con l'elevata qualità della carcassa è corretta. Nel complesso, comunque, non sono state individuate correlazioni degne di ulteriori approfondimenti.





**Figura 3.5** Mappa di correlazione relativa alle variabili delle sottomatrici Animale vivo e Carcassa macellata.

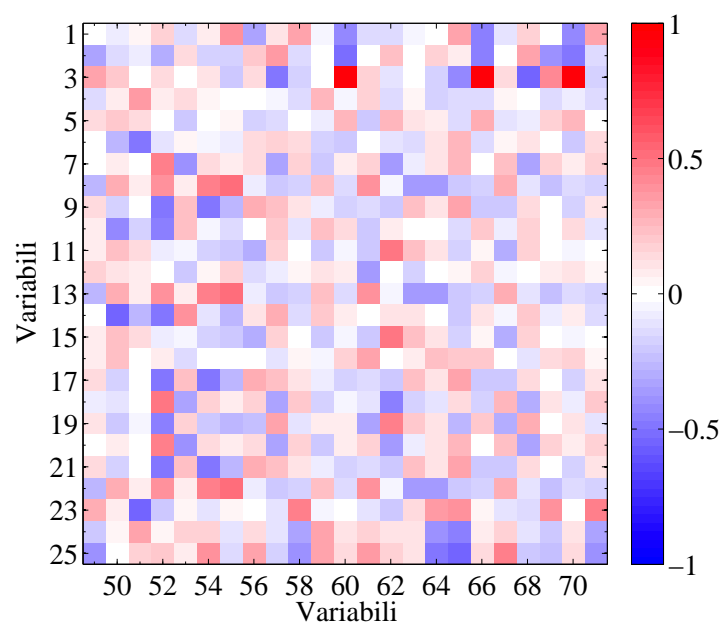
**Tabella 3.5** Elenco delle principali correlazioni individuate tra la sottomatrice dell'Animale vivo e quella della Carcassa macellata.

Variabile	Correlazioni	Anticorrelazioni
(27) peso morto caldo	(4) macello: 1° giornata	(6) macello: 3° giornata
(28) peso morto freddo	(24) incremento netto (25) peso vivo	(23) sesso
(29) perdita	(1) alimentazione standard (6) macello: 3° giornata (23) sesso	(3) standard + PUFA (5) macello: 2° giornata (24) incremento netto
(34) qualità alta		(6) macello: 3° giornata
(35) qualità altissima	(6) macello: 3° giornata	
(47) drip loss	(5) macello: 2° giornata	(4) macello: 1° giornata
(48) pH a 45 minuti	(6) macello: 3° giornata	(5) macello: 2° giornata

### 3.1.6 Caratteristiche dell'animale vivo ed analisi cromatografica del lardo fresco

Combinando le informazioni sull'animale vivo con le analisi cromatografiche del lardo fresco è possibile verificare la presenza di correlazioni tra le abitudini alimentari ed il sesso e la composizione chimica del grasso di copertura fresco.

Si riportano in Figura 3.6 ed in Tabella 3.6 i legami individuati.



**Figura 3.6** Mappa di correlazione relativa alle variabili delle sottomatrici Animale vivo e Cromatografia.

**Tabella 3.6** Elenco delle principali correlazioni individuate tra la sottomatrice dell'Animale vivo e quella della Cromatografia.

Variabile	Correlazioni	Anticorrelazioni
(60) C18:3n3 (66) C20:3n3 (69) acidi polinsaturi (70) acidi n3	(3) standard + PUFA	(1) alimentazione standard (2) standard + polifenoli
(57) C18:1n9 (68) acidi monoinsaturi	(2) standard + polifenoli	(3) standard + PUFA
(58) C18:2n6 (64) C20:3n6 (65) C20:4n6 (71) acidi n6	(23) sesso	(24) incremento netto (25) peso vivo

Il significato fisico delle variabili conferma anche in questo caso i risultati statistici, poiché il regime alimentare standard + PUFA - variabile (3) - è arricchito con semi di lino; una parte di questi acidi grassi polinsaturi è soggetta ad un assorbimento intestinale e ad un deposito negli strati adiposi, che hanno quindi un maggior contenuto di acidi polinsaturi ed n3 - variabili (60), (66), (69) e (70).

L'arricchimento della dieta con polifenoli - variabile (2) - sembra invece interagire con i fenomeni metabolici, in particolare con la sintesi degli acidi grassi, catalizzando la sintesi dell'acido oleico - variabile (57) - con un meccanismo di feedback positivo.

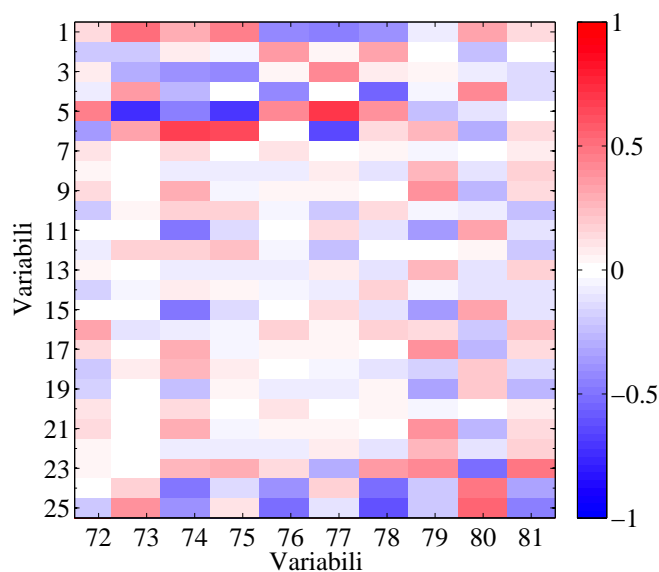
Infine, la correlazione tra il sesso - variabile (23) - e gli acidi grassi n6 - variabili (58), (64), (65) e (71) - si deve interpretare per via indiretta: gli esemplari femmina sono più leggeri, ed hanno quindi un minor quantitativo di grasso; minore è il peso, minore è la sintesi e quindi maggiore è il contenuto di acidi grassi n6.

### 3.1.7 Caratteristiche dell'animale vivo e qualità del prodotto stagionato

Come ultima analisi incrociata con la sottomatrice *Animale vivo*, è stata verificata la presenza di correlazioni tra le caratteristiche del suino e la qualità dello speck stagionato da esso prodotto. Si evidenziano le principali correlazioni in Figura 3.7 ed in Tabella 3.7.

L'attività dell'acqua e l'umidità del magro stagionato - variabili (73) e (75) - risultano correlate al sesso poiché gli esemplari femmina hanno un minor peso e quindi un minor quantitativo di grasso (idrofobo); è invece presente una maggior quantità di proteina grezza, in cui l'acqua trova maggiore affinità. Inoltre, la castrazione a cui tutti gli esemplari maschi sono sottoposti determina un maggior deposito di grasso, e ciò spiega ulteriormente tali legami.

La maggior parte delle rimanenti correlazioni è di difficile interpretazione perché le relazioni individuate riguardano variabili dell'animale vivo, come ad esempio i diversi regimi alimentari, ed alcune qualità dello speck stagionato; tra questi due estremi sono racchiusi i processi metabolici del suino, la stagionatura, l'affumicatura, e molti altri processi biologici non standardizzabili, con estrema variabilità intrinseca.



**Figura 3.7** Mappa di correlazione relativa alle variabili delle sottomatrici *Animale vivo* e *Prodotto stagionato*.

**Tabella 3.7** *Elenco delle principali correlazioni individuate tra la sottomatrice dell'Animale vivo e quella del Prodotto stagionato.*

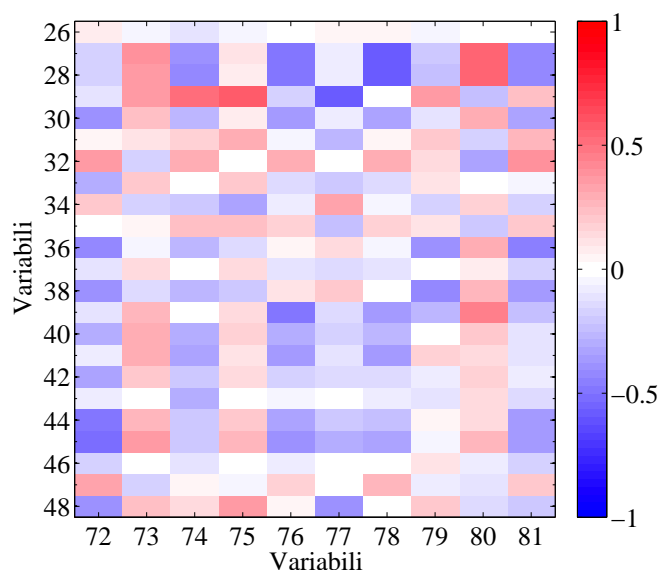
Variabile	Correlazioni	Anticorrelazioni
(72) calo peso (%)	(5) macello: 2° giornata	(6) macello: 3° giornata
(73) attività dell'acqua	(1) alimentazione standard (4) macello: 1° giornata (6) macello: 3° giornata (25) peso vivo	(5) macello: 2° giornata
(74) ph	(1) alimentazione standard (6) macello: 3° giornata (23) sesso	(3) standard + PUFA (4) macello: 1° giornata (5) macello: 2° giornata (24) incremento netto (25) peso vivo
(75) umidità del magro	(1) alimentazione standard (6) macello: 3° giornata (23) sesso	(3) standard + PUFA (5) macello: 2° giornata
(76) NaCl	(2) standard + polifenoli (5) macello: 2° giornata	(1) alimentazione standard (4) macello: 1° giornata (24) incremento netto (25) peso vivo
(77) sostanza secca (%)	(3) standard + PUFA (5) macello: 2° giornata	(1) alimentazione standard (6) macello: 3° giornata
(78) ceneri	(2) standard + polifenoli (5) macellazione (2) (23) sesso	(1) alimentazione standard (4) macello: 1° giornata (24) incremento netto (25) peso vivo
(79) proteina grezza	(23) sesso	
(80) EE SS calcolato	(1) alimentazione standard (4) macello: 1° giornata (24) incremento netto (25) peso vivo	(2) standard + polifenoli (6) macello: 3° giornata (23) sesso
(81) acidità titolabile	(23) sesso	(24) incremento netto (25) peso vivo

### 3.1.8 Caratteristiche della carcassa macellata e qualità del prodotto stagionato

Si è inoltre cercato di analizzare se le variabili di qualità dello speck stagionato siano correlate alle caratteristiche della carcassa da cui il taglio di speck deriva; tali correlazioni, se presenti, sarebbero di particolare interesse perché, nota la struttura e la percentuale di carne magra della carcassa, sarebbe possibile predire a priori la qualità del prodotto commercializzabile, a monte della fase di stagionatura e quindi ben prima rispetto al termine

della filiera produttiva.

I risultati ottenuti sono riportati in Figura 3.8 ed in Tabella 3.8.



**Figura 3.8** Mappa di correlazione relativa alle variabili delle sottomatrici Carcassa macellata e Prodotto stagionato.

**Tabella 3.8** Elenco delle principali correlazioni individuate tra la sottomatrice della Carcassa macellata e quella del Prodotto stagionato.

Variabile	Correlazioni	Anticorrelazioni
(27) peso morto caldo	(73) attività dell'acqua	(74) pH
(28) peso morto freddo	(80) EE SS calcolato	(76) NaCl
		(78) ceneri
		(80) acidità titolabile
(29) perdita (%)	(74) pH	(77) sostanza secca (%)
	(75) umidità del magro	

I risultati di Figura 3.8 sono rappresentativi di una limitata correlazione tra le variabili, proprio perché le successive fasi di stagionatura ed affumicatura a cui il taglio fresco è sottoposto modificano in maniera sensibile le sue proprietà, scorrelando di fatto le caratteristiche della carcassa dalla qualità del prodotto finito. In altre parole, la qualità dello stagionato è maggiormente correlata ad *altre* variabili che intervengono nella filiera produttiva a valle della macellazione, e che non sono quindi riconducibili alle proprietà della carcassa.

## 3.2 Analisi spettroscopica NIR

Esaurita l'analisi delle correlazioni tra le variabili del dataset *Dati di filiera*, si sposta ora la trattazione statistica al set di dati delle analisi spettroscopiche NIR. In questo caso, la sensibile correlazione tra tutte le lunghezze d'onda di uno spettro ha richiesto l'utilizzo di altri metodi statistici, diversi dalle mappe di correlazione, discussi nel dettaglio nel seguito. La continuità e la limitata variabilità di uno spettro NIR determinano infatti una forte correlazione tra tutte le lunghezze d'onda che lo costituiscono; in questo caso, una mappa di correlazione o un'analisi delle componenti principali sarebbero del tutto inefficaci perché la significativa correlazione tra tutte le variabili della matrice di acquisizione NIR maschererebbe qualsiasi altra correlazione. La concatenazione della matrice di acquisizione stessa con un'altra matrice di variabili, ad esempio quella di qualità del prodotto stagionato, è quindi da escludere a priori perché i risultati sarebbero privi di interesse pratico.

Per ovviare a questa problematica si è quindi scelto di implementare l'analisi delle correlazioni canoniche (CCA); nello specifico, le due matrici in input sono costituite dalla matrice di acquisizione NIR e da quella del singolo acido grasso dell'analisi cromatografica o della variabile di qualità del prodotto stagionato.

Si è scelto inoltre di implementare la procedura per via iterativa, individuando ad ogni step le tre lunghezze d'onda più correlate; dopo averle rimosse dallo spettro originale, l'analisi è stata ripetuta per cento iterazioni, selezionando così un totale di trecento lunghezze d'onda.

Si riportano nel seguito i risultati ottenuti utilizzando gli spettri NIR Foss, sia sulla carcassa macellata che sullo speck stagionato; i risultati sono tuttavia confermati anche utilizzando le acquisizioni Unity, previo adattamento del numero di lunghezze d'onda considerate<sup>1</sup>. Oltre a verificare la corrispondenza fisica tra il singolo acido grasso o la variabile di qualità in esame ed i rispettivi intervalli di lunghezze d'onda selezionati, caratteristici di specifici gruppi funzionali, si sono confrontati i risultati di questa analisi con quelli ottenuti dall'interpretazione delle mappe di correlazione, alle sezioni 3.1.3 e 3.1.4.

### 3.2.1 Carcassa macellata

Gli spettri NIR della carcassa macellata sono stati utilizzati per individuare eventuali correlazioni tra lo spettro stesso e le variabili dell'analisi cromatografica, che rappresentano la matrice di dati più vicina alla macellazione nell'iter della filiera produttiva.

Alla sezione 3.1.3 sono state individuate delle correlazioni particolarmente significative tra gli acidi grassi della stessa tipologia, siano essi saturi, monoinsaturi o polinsaturi, e la loro

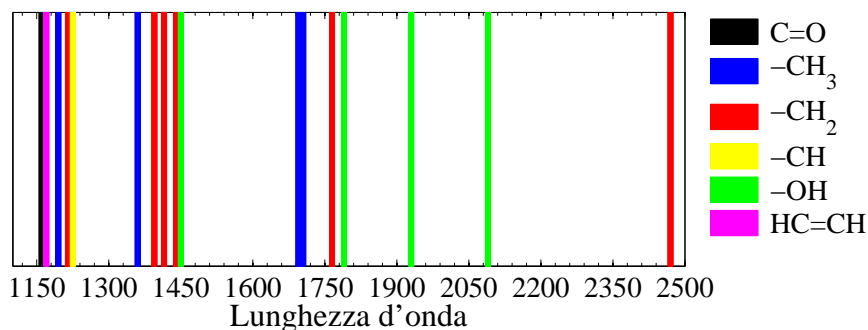
---

<sup>1</sup>I due strumenti hanno infatti un intervallo di acquisizione ed una risoluzione diversa.

variabile sommativa; si verifica ora se tali legami siano confermati anche dall'analisi delle correlazioni canoniche. Si riportano in Figura 3.10 lo spettro medio Foss della carcassa macellata e gli intervalli di lunghezze d'onda selezionati per gli acidi grassi monoinsaturi più rappresentativi, evidenziati in scala di grigi; le prime lunghezze d'onda selezionate sono colorate in nero, mentre la tinta si sbiadisce progressivamente al diminuire della correlazione.

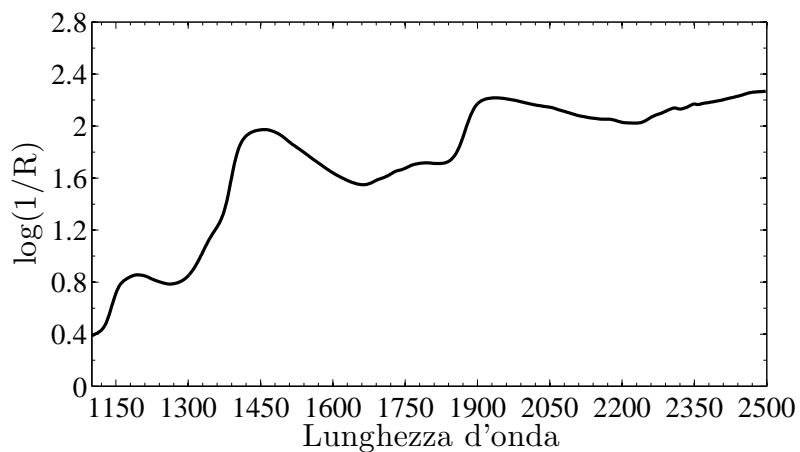
Per interpretare nel modo corretto i grafici di Figura 3.10 è necessario considerare anche la frazione ponderale di ogni acido grasso rispetto al totale di categoria, in modo da pesare le correlazioni dei singoli acidi in base ai rispettivi quantitativi; si noti infatti come le lunghezze d'onde correlate alla variabile sommativa, Figura 3.10 (e), siano molto simili a quelle dell'acido grasso C18:1n9, Figura (c), a cui corrisponde il contributo principale nella quantificazione degli acidi grassi monoinsaturi presenti. L'effetto degli altri due acidi MUFA (*mono unsaturated fatty acid*) è visibile nella banda di lunghezze d'onda più elevate, che sono selezionate con maggiore continuità nella variabile sommativa rispetto a quanto accade nel caso dell'acido oleico, C18:1n9.

Al fine di irrobustire ulteriormente i risultati ottenuti, è stata condotta un'approfondita ricerca bibliografica per convalidare gli intervalli di lunghezze d'onda correlati a ciascuna variabile. Tuttavia, le applicazioni dell'analisi NIR attualmente disponibili riguardano prevalentemente la sfera vegetale e, anche laddove l'analisi è estesa alle carni, bovine o suine, il grado di dettaglio è molto più superficiale rispetto alle necessità di confronto. Non sono infatti disponibili degli intervalli di lunghezze d'onda specifici di una variabile, ma vengono solamente identificate le bande caratteristiche di vari gruppi funzionali. A tal proposito, si riporta in Figura 3.9 una rappresentazione riassuntiva delle lunghezze d'onda caratteristiche dei più comuni gruppi funzionali; si noti come ad ognuno di essi corrispondano più intervalli, poiché le molecole possono essere eccitate su livelli vibrazionali diversi, tecnicamente detti *overtones* (Burns e Ciurczak, 2008; Ozaki *et al.*, 2007).

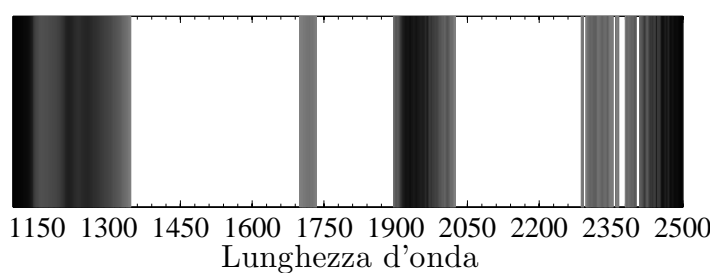


**Figura 3.9** Intervalli di lunghezze d'onda associati alla vibrazione dei più comuni gruppi funzionali organici.

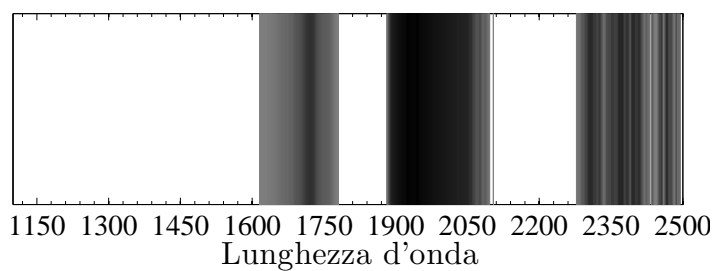
Dal confronto delle Figure 3.10 e 3.9 è possibile attribuire la banda più correlata, caratteri-



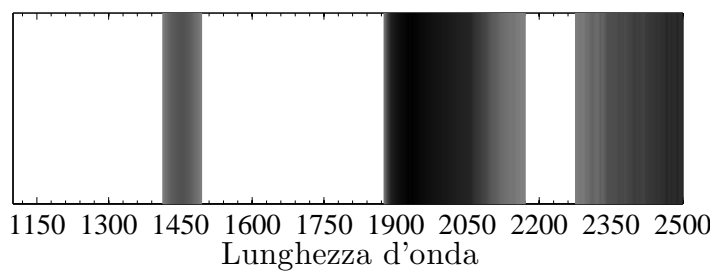
(a)



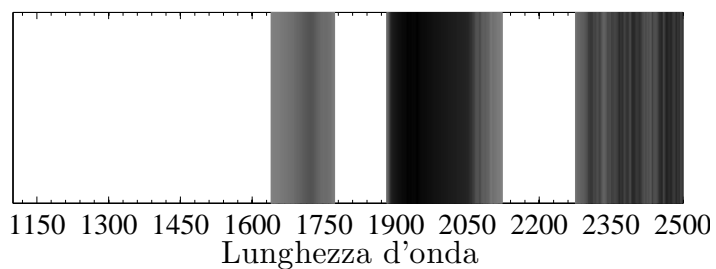
(b)



(c)



(d)



(e)

**Figura 3.10** Lunghezze d'onda correlate agli acidi grassi monoinsaturi, ottenute mediante CCA: (a) spettro medio Foss della carcassa macellata; (b) C16:1 (4.5%); (c) C18:1n9 (86.5%); (d) C20:1n9 (1.7%); (e) sommatoria MUFA.



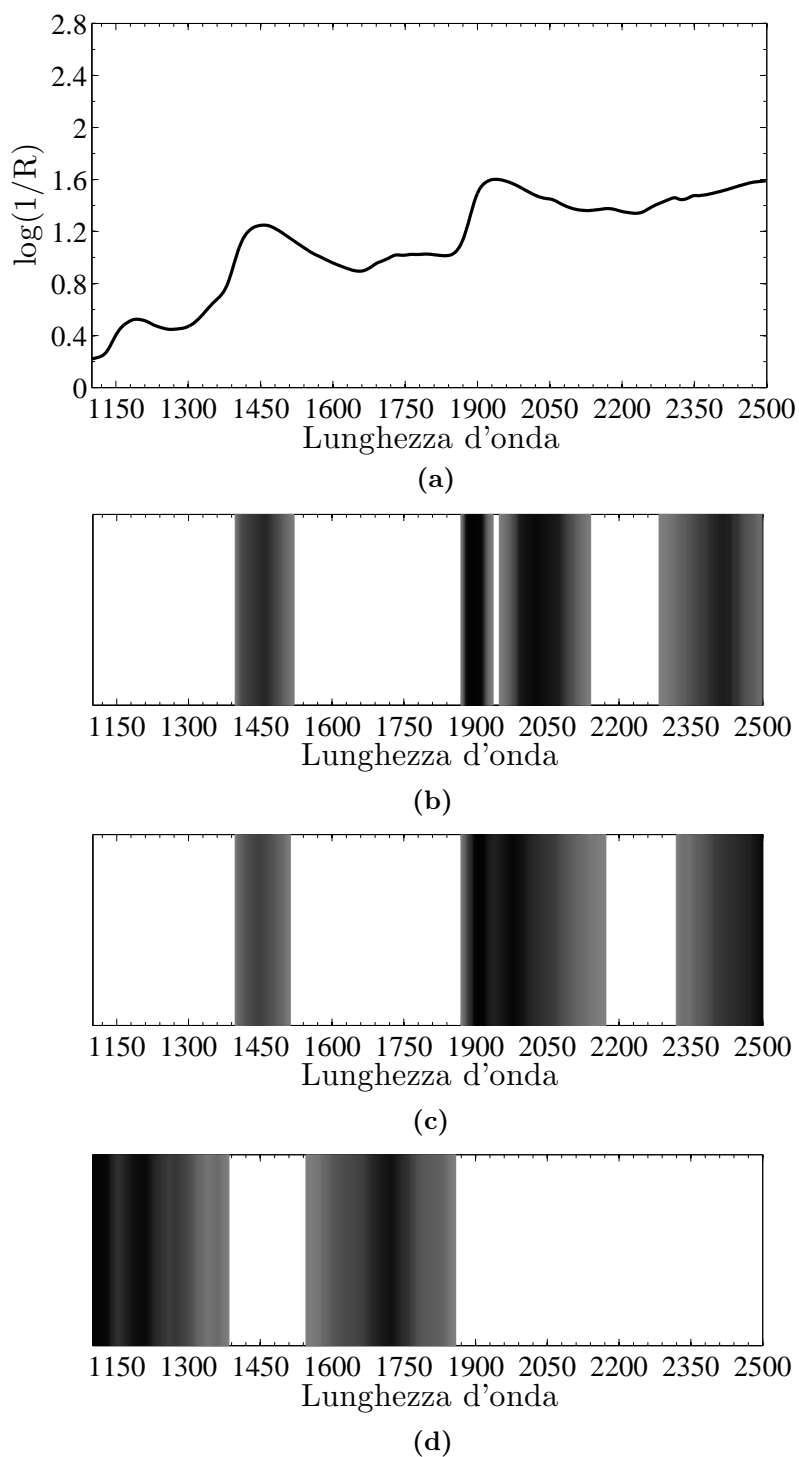
stica dell'intervallo 1900÷2100 nm, alla vibrazione del gruppo ossidrilico, presente in posizione terminale nella catena lineare di ogni acido grasso monoinsaturo; si riconoscono inoltre i gruppi metilici che costituiscono la catena dell'acido grasso, eccitati intorno a 1650 nm, ed il gruppo metilenico associato al doppio legame, caratterizzato da una banda prossima ai 1750 nm. La zona terminale dello spettro, a partire da circa 2000÷2100 nm, è invece attribuibile alla combinazione delle deformazioni e vibrazioni di più gruppi funzionali; non è quindi possibile associare un intervallo di lunghezze d'onda ad uno specifico gruppo funzionale, ma queste combinazioni di più gruppi sono ampiamente documentate in letteratura (Burns e Ciurczak, 2008).

I risultati ottenuti con gli acidi grassi monoinsaturi sono estendibili anche a tutte le altre tipologie di acidi grassi individuate dall'analisi cromatografica. Nel complesso, le lunghezze d'onda correlate alla variabile sommatoria sono pienamente confrontabili a quelle dell'acido grasso più rappresentativo all'interno della specifica categoria, mentre le lievi differenze sono spiegabili a partire dagli intervalli selezionati per gli acidi grassi marginali. Inoltre, utilizzando sempre la Figura 3.9 come riferimento, è possibile interpretare le lunghezze d'onda selezionate in base alla struttura chimica dei singoli acidi grassi.

### 3.2.2 Prodotto stagionato

Si ripete ora la stessa procedura della sezione precedente, applicandola però agli spettri NIR Foss sullo speck stagionato; in questo caso, sono state individuate le lunghezze d'onda più correlate alle variabili di qualità del prodotto stagionato, confrontando i legami ottenuti con quelli di Tabella 3.4. È opportuno premettere che l'interpretazione delle bande selezionate è molto più complessa rispetto al caso degli acidi grassi rilevati dalle analisi cromatografiche, poiché molte variabili di qualità non sono associabili a particolari gruppi funzionali né possono essere descritte con una struttura chimica precisa. Nella maggior parte dei casi, l'interpretazione fisica è quindi limitata ad un livello qualitativo, utile però come termine di confronto rispetto alle mappe di correlazione precedentemente descritte. In Tabella 3.4 è stata riportata una correlazione molto significativa tra l'attività dell'acqua e l'umidità del magro, variabili (73) e (75); entrambe risultano inoltre anticorrelate alla sostanza secca, variabile (77), come intuibile anche da un punto di vista fisico. L'applicazione dell'analisi delle correlazioni canoniche a questa terna di variabili porta ai risultati di Figura 3.11.

Come si può notare, la sovrapposizione degli intervalli selezionati per l'attività dell'acqua e l'umidità del magro è ottima, a conferma della correlazione tra queste due variabili; viceversa, le bande correlate alla sostanza secca sono complementari alle precedenti, ad ulteriore riprova dell'anticorrelazione individuata in Figura 3.4. Per le prime due variabili

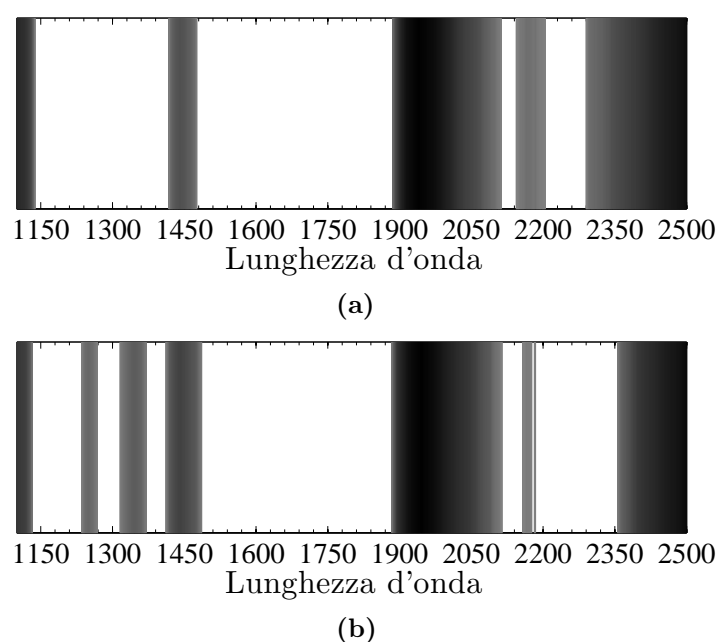


**Figura 3.11** Lunghezze d'onda correlate, ottenute mediante CCA: (a) spettro medio Foss dello speck stagionato; (b) attività dell'acqua; (c) umidità del magro; (d) sostanza secca.

è possibile avere anche un riscontro fisico rispetto alle lunghezze d'onda individuate, poiché entrambe sono essenzialmente legate all'acqua, e quindi al gruppo ossidrilico; le bande selezionate sono perfettamente sovrapponibili alle lunghezze d'onda di vibrazione del gruppo  $-OH$ , evidenziate in Figura 3.9: nello specifico ci si riferisce a circa 1410 nm (primo

*overtone*), all'intervallo tra 1900 e 2100 nm (*stretching* del legame O–H e combinazione delle vibrazioni dei gruppi ossidrile), ed alle lunghezze d'onda terminali, caratteristiche sempre della combinazione delle deformazioni di diversi gruppi funzionali (Burns e Ciurczak, 2008).

Si è scelto di estendere l'analisi delle correlazioni canoniche anche alle variabili (76) e (78), rappresentative del contenuto di NaCl e di ceneri; entrambe risultano correlate in modo significativo sia dalle mappe di correlazione che dall'interpretazione fisica, come discusso nel dettaglio alla sezione 3.1.4. Si riportano in Figura 3.12 gli intervalli di lunghezze d'onda selezionati dall'analisi CCA, relativi ad entrambe le variabili.



**Figura 3.12** Lunghezze d'onda correlate, ottenute mediante CCA: (a) NaCl; (b) ceneri.

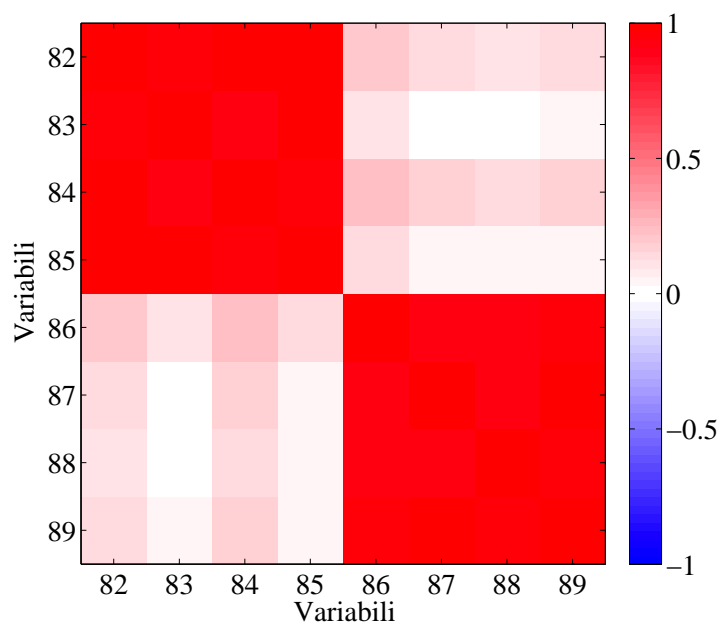
Anche in questo caso la corrispondenza degli intervalli selezionati è ottima, sia per quanto concerne le lunghezze d'onda selezionate, sia circa i livelli di significatività delle correlazioni tra singole  $\lambda$ ; infatti, le Figure 3.12 (a) e (b) sono pienamente confrontabili non solo per le bande evidenziate, ma anche per le scale cromatiche delle lunghezze d'onda evidenziate. In altre parole, la correlazione individuata è molto significativa, perché si manifesta sia nella sovrapposizione degli intervalli selezionati, sia nella piena confrontabilità delle lunghezze d'onda maggiormente correlate. In questo caso, tuttavia, non è possibile associare delle specifiche zone dello spettro alle variabili considerate, soprattutto a causa dello stato fisico in cui si trovano; i solidi inorganici vengono infatti eccitati in modo molto marginale dall'analisi NIR, che non possiede un contributo energetico tale da influenzare in modo significativo i legami chimici presenti, caratterizzati da un'estrema vicinanza e quindi energia di legame. In termini tecnici, è comune dire che i composti inorganici solidi ed i sali non hanno una *spectral fingerprint*, ovvero non sono identificabili da un'analisi spettroscopica

NIR (Burns e Ciurczak, 2008). Tuttavia, la presenza di variabili ioniche come ad esempio il cloruro di sodio si può misurare attraverso l'effetto che queste hanno nei solventi presenti, in questo caso l'acqua; come evidente, ciò trova pieno riscontro in Figura 3.12, in cui le lunghezze d'onda evidenziate sono pienamente associabili alle vibrazioni dei gruppi O–H.

### 3.3 Analisi Instron

Per lo studio delle correlazioni del set di dati *Analisi Instron* si è tornati all'interpretazione delle mappe di correlazione, che rappresentano uno strumento meno complesso per analizzare una tipologia di dati meno correlati rispetto agli spettri NIR.

Si presentano e discutono nel seguito due diverse mappe: la prima (Figura 3.13) è relativa alle sole variabili dell'analisi *Instron*, mentre la seconda (Figura 3.14) deriva dalla concatenazione dei due dataset *Dati di filiera* e *Analisi Instron*.

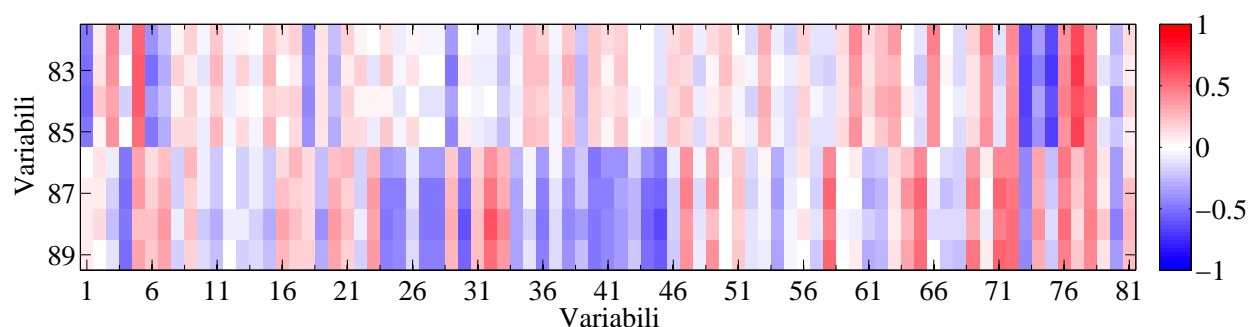


**Figura 3.13** Mappa di correlazione relativa alle sole variabili dell'Analisi Instron.

A primo impatto, è immediato verificare la forte correlazione tra tutti gli indicatori sintetici utilizzati per descrivere le curve di analisi Instron, sia per la parte magra che per quella grassa; ciò si nota non solo da due quadrati rossi di Figura 3.13, ma anche dal fatto che in Figura 3.14 le correlazioni con le altre variabili si presentano sottoforma di bande, indice che il legame è comune a tutte le variabili *Instron* della parte magra o grassa.

Si riporta in Tabella 3.9 il dettaglio completo delle correlazioni individuate.

Il significato fisico supporta anche in questo caso i risultati statistici poiché:



**Figura 3.14** Mappa di correlazione relativa ai dataset Dati di filiera ed Analisi Instron, concatenati in un'unica matrice.

**Tabella 3.9** Elenco delle principali correlazioni individuate tra i dataset Dati di filiera ed Analisi Instron.

Variabile	Correlazioni	Anticorrelazioni
gruppo Instron Magro	(3) standard + PUFA (5) macello: 2° giornata (60) C18:3n3 (63) C20:2 (66) C20:3n3 (70) acidi n3 (72) calo peso % (76) NaCl (77) sostanza secca % (78) ceneri	(1) alimentazione standard (6) macello: 3° giornata (29) perdita (73) attività dell'acqua (74) pH (75) umidità del magro
gruppo Instron Grasso	(5) macello: 2° giornata (32) carne magra (%) (35) qualità altissima (47) drip loss (49) n° di iodio (58) C18:2n6 (64) C20:3n6 (65) C20:4n6 (69) acidi polinsaturi (71) acidi n6 (72) calo peso % (74) pH (76) NaCl (78) ceneri	(4) macello: 1° giornata (24) incremento netto (25) peso vivo (27) peso morto caldo (28) peso morto freddo (30) spessore di grasso (36) lardo 1T totale (38) lardo 1T interno (40) lardo 13T totale (41) lardo 13T esterno (42) lardo 13T interno (44) lardo minimo totale (45) lardo minimo esterno (73) attività dell'acqua (80) EE SS calcolato

- l'area sottesa dalla curva di analisi è, come detto, esplicativa del lavoro compiuto; riferendosi alla parte magra, l'area risulta correlata agli acidi grassi n3 ed in generale alla sostanza secca presente nello stagionato, mentre è anticorrelata ad altre variabili inerenti umidità e pH. Ciò appare ragionevole poiché il diminuire del contenuto di acqua e

- parallelamente l'aumentare della sostanza secca sono indicativi di una carne stagionata più asciutta, dura e resistente, che si esplica appunto in un lavoro di taglio superiore;
- è invece più difficile arrivare a considerazioni analoghe anche per i dati relativi alla parte grassa, perché all'aumentare dello strato adiposo cambiano radicalmente struttura interna e proprietà. È comunque indicativo rilevare una diffusa anticorrelazione verso tutte le variabili caratteristiche di spessori di grasso o velocità di ingrassamento.

### 3.4 Analisi colorimetriche

In piena analogia rispetto alla sezione precedente, anche lo studio delle correlazioni delle *Analisi colorimetriche* è stato condotto utilizzando le mappe di correlazione; si riportano in Figura 3.15 la mappa relativa solamente al set di dati in questione, mentre in Figura 3.16 le correlazioni individuate dalla concatenazione di questo dataset con quello *Dati di filiera*.

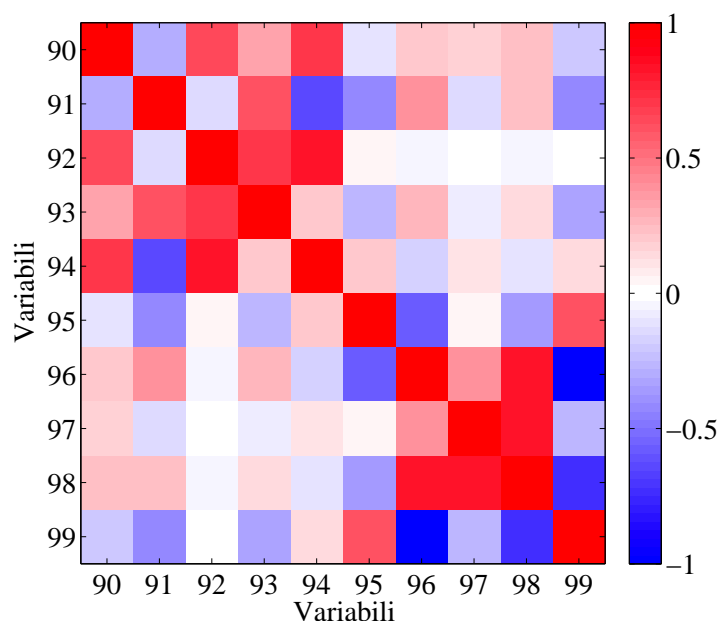
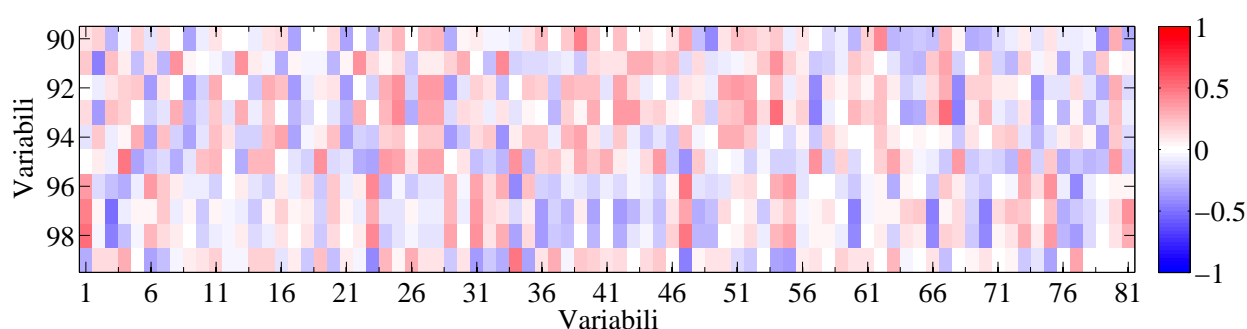


Figura 3.15 Mappa di correlazione relativa al dataset *Colore*.

In Figura 3.15 si notano dei legami piuttosto significativi tra i singoli indicatori colorimetrici, ma ciò appare per lo più riconducibile alla loro definizione matematica, riportata nelle Equazioni 1.3 e 1.4; infatti, dividendo la figura in quattro quadranti uguali, si nota una piena analogia strutturale tra i quadranti 1-3 e 2-4, relativi nello specifico alle correlazioni “miste” tra indicatori della parte magra e di quella grassa (1-3), ed alle correlazioni “interne” agli indicatori del magro o del grasso (2-4). Ciò fa presumere che le correlazioni individuate siano di puro effetto matematico, anche se è comunque possibile individuare delle differenze:

- il croma del magro - variabile (93) - risulta correlato, seppur debolmente, alla luminosità ed alla tinta del magro - variabili (90) e (95) - , mentre nelle analisi della parte grassa tali legami risultano opposti, rappresentativi di due anticorrelazioni;
- l'indice del giallo - variabile (92) per il magro e (97) per il grasso - risulta fortemente correlato alla luminosità ed alla tinta del magro - variabili (90) e (94) - , mentre debolmente anticorrelato agli stessi parametri della parte grassa - variabili (95) e (99).

In Figura 3.16 è invece immediato visualizzare l'assenza di correlazioni particolarmente significative tra i due dataset, come evidente dalle colorazioni molto pallide della mappa. Si riporta comunque in Tabella 3.10 un elenco delle correlazioni individuabili.



**Figura 3.16** Mappa di correlazione relativa ai dataset *Dati di filiera e Colore*.

**Tabella 3.10** Elenco delle principali correlazioni individuate tra i dataset *Dati di filiera e Colore*

Variabile	Correlazioni	Anticorrelazioni
(90) luminosità magro	(39) carne 1T (54) C16:0 (92) indice del giallo magro (94) tinta magro	(3) standard + PUFA (49) n° di iodio
(91) indice del rosso magro	(33) qualità media (54) C16:0 (93) croma magro (96) indice del rosso grasso	(2) standard + polifenoli (76) NaCl (94) tinta magro (95) luminosità G (99) tinta grasso
(92) indice del giallo magro	(43) carne 13T (93) croma magro (94) tinta magro	(46) lardo minimo interno
(93) croma magro	(54) C16:0 (67) acidi saturi	(2) standard + polifenoli (37) lardo 1T esterno (57) C18:1n9 (99) tinta grasso

(94) tinta magro	(32) carne magra (%) (39) carne 1T	(29) perdita (%) (30) spessore di grasso (33) qualità media (46) lardo minimo esterno
(95) luminosità grasso	(4) macello: 1° giornata (34) qualità alta (63) C20:2 (99) tinta grasso	(5) macello: 2° giornata (47) drip loss (96) indice del rosso grasso
(96) indice del rosso grasso	(1) alimentazione standard (6) macello: 3° giornata (23) sesso (47) drip loss (74) pH (98) cromia grasso	(24) incremento netto (99) tinta grasso
(97) indice del giallo grasso	(1) alimentazione standard (57) C18:1n9 (68) acidi monoinsaturi (72) calo peso % (98) cromia grasso	(36) lardo 1T totale (38) lardo 1T interno (42) lardo 13T interno (67) acidi saturi
(98) cromia grasso	(1) alimentazione standard (23) sesso (47) drip loss (57) C18:1n9 (68) acidi monoinsaturi	(36) lardo 1T totale (38) lardo 1T interno (42) lardo 13T interno (99) tinta grasso
(99) tinta grasso	(24) incremento netto (34) qualità alta	(6) macello: 3° giornata (74) pH

### 3.5 Conclusioni

Conclusa l'analisi esplorativa iniziale, si conoscono ora tutte le principali correlazioni tra le diverse variabili in esame; ciò è di estrema importanza per poter condurre in modo proficuo la successiva fase di stima e classificazione, perché si potranno sfruttare i legami individuati nel corso di questo capitolo per predire in modo accurato alcune variabili di interesse. Tra le correlazioni individuate si annoverano legami banali, come quello tra gli spessori di grasso ed il peso del suino, e relazioni meno scontate, quali ad esempio le interazioni tra il regime alimentare e le variabili cromatografiche o di qualità finale; alcune di queste, di difficile interpretazione, necessitano di uno studio più approfondito anche dal punto di vista metabolico, poiché coinvolgono processi biologici non generalizzabili a priori.

L'applicazione dell'analisi delle correlazioni canoniche ha consentito inoltre di individuare le lunghezze d'onda più correlate alle variabili cromatografiche o di qualità del prodotto stagionato, verificando quanto riportato in letteratura circa le bande di vibrazione dei prin-



cipali gruppi funzionali; a tal proposito, tuttavia, lo stato dell'arte piuttosto superficiale non ha consentito una conferma dettagliata delle lunghezze d'onda selezionate.

Infine, lo studio delle correlazioni ha consentito di assumere piena conoscenza dei *dataset* a disposizione, esplicitando quali siano i set di variabili più idonei per avviare la fase di predizione partendo già da buoni risultati.



# Capitolo 4

## Stima delle variabili chiave per il monitoraggio della qualità del prodotto

Il controllo della qualità delle carni lungo la filiera di lavorazione rappresenta uno strumento molto importante per monitorare l'evoluzione delle proprietà fisico-chimiche dello speck, utile ad adottare eventuali contromisure prima che il processo di stagionatura sia concluso. In questo Capitolo si estenderà l'applicazione della spettroscopia NIR al caso di studio sullo speck, verificando la possibilità di stimare le principali variabili di qualità lungo l'intera filiera produttiva. Inoltre, verranno utilizzate anche analisi colorimetriche, analisi di tessitura Instron, ed immagini digitali.

### 4.1 Costruzione dei modelli PLS

Una metodologia statistica appropriata per monitorare alcune variabili risposta, quali gli indicatori di qualità, a partire da altre variabili predittrici è la proiezione su strutture latenti, di cui sono discussi nel seguito i dettagli di implementazione.

Le analisi spettroscopiche a disposizione sono costituite da una matrice di acquisizione Foss, di dimensione  $[36 \times 700]$ , e da una matrice  $[36 \times 1721]$  relativa alle analisi Unity; a seconda della proprietà da stimare, gli spettri NIR utilizzati sono relativi alla carcassa macellata o allo speck stagionato. Le variabili di qualità sono invece estratte una per volta dal *dataset Dati di filiera*, e vanno a costituire il vettore  $\mathbf{y}$   $[36 \times 1]$ .

I 36 campioni a disposizione sono stati suddivisi in due gruppi, uno utilizzato per la calibrazione del modello e l'altro per la convalida; nello specifico, si è scelto di utilizzare 30 campioni in calibrazione,  $\mathbf{X}_{cal}$   $[30 \times 700]$  e  $\mathbf{y}_{cal}$   $[30 \times 1]$ , e 6 in convalida,  $\mathbf{X}_{val}$   $[6 \times 700]$  e  $\mathbf{y}_{val}$   $[6 \times 1]$ . Tuttavia, per condurre un'analisi il più generale possibile e migliorare la robustezza dei risultati, è stato implementato un algoritmo Monte Carlo in modo da rendere casuale la suddivisione dei campioni nei *dataset* di calibrazione e convalida. Il numero totale di combinazioni è però pari a  $36!/30! = 1.40 \times 10^9$ , che corrisponde ad un numero di simulazioni troppo elevato dal punto di vista computazionale. Si è quindi scelto di condurre due approcci, uno con 1000 e l'altro con 10000 simulazioni; dal confronto dei risultati ottenuti è possibile capire se l'aumento di un ordine di grandezza nel numero di simulazioni comporti sensibili variazioni o meno, in modo da individuare il giusto compromesso tra la robustezza

dei risultati ed i tempi di calcolo.

Il criterio di scelta del numero di variabili latenti è basato sulla minimizzazione del  $RMSE$  in convalida incrociata, utilizzando un algoritmo *leave-one-out* presentato alla sezione 2.2. I risultati ottenuti sono espressi in termini di errori relativi, coefficienti di determinazione  $R^2$  in calibrazione e convalida, ed RPD. Inoltre, si è scelto di scartare le iterazioni caratterizzate da valori negativi di  $R_{val}^2$ , in modo da evitare di includere modelli privi di significato. Infatti, si possono ottenere  $R_{val}^2$  negativi qualora i campioni di calibrazione coprano un insufficiente intervallo di variabilità, che causa un'estrapolazione all'esterno dello spazio del modello quando i campioni di convalida vengono predetti.

Questa procedura è stata implementata per tutte le tipologie di analisi a disposizione; gli specifici risultati vengono discussi nel seguito. Inoltre, al fine di valutare l'accuratezza delle stime anche rispetto all'effettiva variabilità dei singoli indicatori di qualità, si riportano in Tabella 4.1 i  $CV$  (cfr. paragrafo 2.2.2) di tutte le variabili analizzate in questo Capitolo. In via del tutto generale, ci si attende di ottenere stime tanto più accurate quanto minore è il  $CV$  della variabile in esame.

**Tabella 4.1** *Elenco dei coefficienti di variazione degli indicatori di qualità analizzati.*

Variabile	CV (%)
pH a 45 minuti	4.57
N° di iodio	3.03
Calo peso (%)	12.90
Attività dell'acqua	4.17
pH	0.85
Umidità	8.08
NaCl	23.99
Sostanza secca	6.57
Ceneri	17.13
Proteina grezza	3.14
EE SS	33.20
Acidità titolabile	24.49

## 4.2 Qualità della carcassa macellata

La prima possibilità di monitorare la qualità del taglio di carne in lavorazione è rappresentata dalle analisi sulla carcassa macellata; nello specifico, le variabili più importanti utilizzate nella prassi per caratterizzarla sono il pH a 45 minuti dalla macellazione ed il numero di iodio del lardo fresco. Si ricordi che il pH è strettamente legato alla perdita di liquidi per gocciolamento del muscolo (*drip loss*) e ciò influenza notevolmente la tessitura

delle carni e, a valle, la percezione organolettica del consumatore. Il numero di iodio è invece una misura del numero di insaturazioni degli acidi grassi presenti, ed è un parametro molto importante perché legato alle conservabilità delle carni: i doppi legami sono infatti facilmente degradabili per via ossidativa, e ciò compromette il colore, il sapore e le altre proprietà organolettiche dello speck.

Le uniche analisi condotte sulla carcassa macellata sono le acquisizioni NIR, poiché necessitano di una quantità di campione molto limitata, sono rapide e non invasive, tutte caratteristiche ottime per monitorare la qualità del taglio di carne senza intralciare la catena di lavorazione.

#### 4.2.1 Analisi spettroscopiche NIR

La spettroscopia NIR è caratterizzata da un contenuto informativo molto elevato sulla struttura e sulla chimica della sostanza analizzata, e frequenti sono i casi studio di letteratura in cui gli spettri NIR vengono utilizzati per stimare in modo accurato numerose proprietà chimiche di interesse (Collell *et al.*, 2011; Cozzolino *et al.*, 2005; Guillen e Cabo, 1998; Ottavian *et al.*, 2012a; Prieto *et al.*, 2009; Velasco e Becker, 1998). La stima della qualità dello speck da spettri NIR sarebbe di notevole rilevanza applicativa perché consentirebbe di evitare le analisi chimiche dirette, molto spesso dispendiose in termini di tempi e risorse, invasive, e caratterizzate dall'utilizzo di solventi chimici e reagenti potenzialmente rischiosi per gli operatori e l'ambiente.

Si riportano in Tabella 4.2 i risultati della stima del pH a 45 minuti a partire da spettri NIR della carcassa macellata, sia Foss che Unity. Il numero di iterazioni è pari a 1000, mentre i pretrattamenti utilizzati sono quelli elencati alla sezione 2.2.1; tutti i risultati sono relativi al *dataset* di convalida. Si evidenziano in grassetto le prestazioni ottimali, caratterizzate dai minimi errori medi relativi; vista la variabilità molto limitata di questa variabile, caratterizzata da un *CV* pari al 4.6%, è bene interpretare con cautela i coefficienti di determinazione.

Dall'analisi dei risultati di Tabella 4.2 si nota una limitata efficacia dei diversi pretrattamenti, come evidenziato dall'estrema omogeneità degli errori medi relativi, comunque molto buoni in tutte le stime. Utilizzando le acquisizioni Foss si evidenziano valori piuttosto modesti dei coefficienti di determinazione, dovuti sia a quote di varianza spiegata su  $y$  sempre inferiori al 25%, sia alla limitata variabilità del pH stesso; utilizzando gli spettri Unity, invece, si ottengono  $R_{cal}^2$  molto più elevati, prossimi all'unità, a riprova di una miglior *fitting* dei dati sperimentali in calibrazione. Nel complesso, l'analisi NIR ottimale ai fini della stima è costituita dallo spettro Foss con pretrattamento SNV + derivata seconda, sebbene altre acquisizioni e pretrattamenti si attestino su valori di errore molto

**Tabella 4.2** *Statistiche dei risultati nella stima del pH a 45 minuti con un numero di simulazioni pari a 1000.*

Spettro	Pretratt.	Varianza spiegata su X	Varianza spiegata su y	$\epsilon_{\text{med}}$ (%)	$R_{\text{cal}}^2$	$R_{\text{val}}^2$	RPD
Foss carcassa	(1)	93.08	6.16	3.446	0.062	0.145	1.2
	(2)	80.49	5.34	3.420	0.053	0.147	1.2
	(3)	62.65	7.92	3.262	0.079	0.190	1.2
	(4)	39.37	19.14	3.151	0.191	0.213	1.3
	(5)	56.64	9.13	3.278	0.091	0.191	1.2
	<b>(6)</b>	<b>32.71</b>	<b>23.11</b>	<b>3.138</b>	<b>0.231</b>	<b>0.225</b>	<b>1.3</b>
Unity carcassa	(1)	95.08	2.65	3.552	0.026	0.085	1.1
	(2)	78.03	5.39	3.539	0.054	0.131	1.2
	(3)	34.74	38.97	3.543	0.390	0.168	1.2
	<b>(4)</b>	<b>42.41</b>	<b>90.86</b>	<b>3.259</b>	<b>0.909</b>	<b>0.213</b>	<b>1.3</b>
	(5)	35.88	58.83	3.727	0.588	0.147	1.2
	(6)	22.60	87.39	3.402	0.874	0.209	1.3

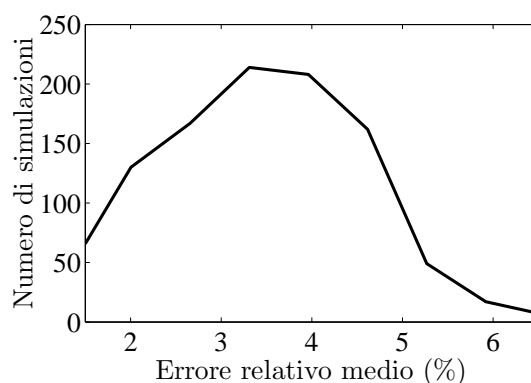
simili, evidenziati da RPD analoghi. Si confrontano ora i risultati di Tabella 4.2 con quelli ottenuti da un'analisi con un numero di simulazioni pari a 10000.

**Tabella 4.3** *Statistiche dei risultati nella stima del pH a 45 minuti con un numero di simulazioni pari a 10000.*

Spettro	Pretratt.	Varianza spiegata su X	Varianza spiegata su y	$\epsilon_{\text{med}}$ (%)	$R_{\text{cal}}^2$	$R_{\text{val}}^2$	RPD
Foss carcassa	(1)	93.09	6.32	3.438	0.063	0.147	1.2
	(2)	80.50	5.30	3.422	0.053	0.146	1.2
	(3)	62.73	7.93	3.324	0.079	0.183	1.2
	(4)	39.27	19.38	3.169	0.194	0.215	1.3
	(5)	56.56	9.11	3.301	0.091	0.195	1.2
	<b>(6)</b>	<b>32.78</b>	<b>23.09</b>	<b>3.153</b>	<b>0.231</b>	<b>0.226</b>	<b>1.3</b>
Unity carcassa	(1)	95.01	2.61	3.526	0.026	0.087	1.1
	(2)	78.04	5.44	3.535	0.054	0.133	1.2
	(3)	34.55	38.16	3.537	0.382	0.169	1.2
	<b>(4)</b>	<b>42.47</b>	<b>90.61</b>	<b>3.249</b>	<b>0.906</b>	<b>0.211</b>	<b>1.3</b>
	(5)	36.84	61.00	3.664	0.610	0.153	1.2
	(6)	22.77	87.71	3.394	0.877	0.200	1.2

Il confronto di Tabella 4.2 e 4.3 evidenzia la piena equivalenza dei risultati ottenuti, sia in termini di errori relativi commessi, sia circa la scelta dello spettro e del pretrattamento ottimale. Questa analogia viene confermata anche dai coefficienti di determinazione e dai valori di RPD, estremamente simili tra loro. Al fine di considerare anche il numero di

variabili latenti scelto, che in caso di difformità potrebbe avere un'influenza notevole sui risultati ottenuti, è stata analizzata anche la distribuzione delle variabili latenti selezionate in ciascun modello PLS, per entrambi i *range* di simulazioni. Tuttavia, anche questa analisi conferma l'equivalenza dei due approcci, poiché le distribuzioni delle variabili latenti selezionate sono pressochè identiche, riscalate di un ordine di grandezza. Inoltre, è possibile notare come il pretrattamento non incida sul numero di variabili latenti selezionate, e ciò dimostra che le migliori prestazioni dei modelli con pretrattamento non sono dovute ad un maggior numero di *LV* considerate, ma dipendono dall'efficacia del pretrattamento stesso. Dall'analisi delle statistiche sui risultati e del numero di variabili latenti selezionate è quindi possibile giustificare la scelta di un numero di simulazioni pari a 1000, caratterizzato da risultati equivalenti all'approccio superiore. Tuttavia, ad ulteriore completezza, si riporta in Figura 4.1 anche la distribuzione degli errori relativi medi del pretrattamento ottimale, con un numero di simulazioni pari a 1000.



**Figura 4.1** *Distribuzione degli errori relativi medi nella stima del pH a 45 minuti da spettro Foss della carcassa macellata, pretrattamento SNV + derivata seconda (ottimale).*

L'andamento degli errori relativi medi è prossimo ad una dispersione gaussiana intorno al valore medio, quindi l'efficacia di un numero di simulazioni pari a 1000 è confermata anche sotto questo punto di vista.

Stabilito il numero di simulazioni ottimale, si riportano in Tabella 4.4 le statistiche sulle stime del numero di iodio, ottenute utilizzando sempre gli spettri NIR della carcassa macellata. Tra i risultati di Tabella 4.4 si evidenziano i coefficienti di determinazione scadenti sia in calibrazione che in convalida, mentre i valori di RPD denotano un minimo vantaggio nella stima rispetto all'utilizzo dei valori medi del set di calibrazione. Inoltre, le quote di varianza spiegata su  $y$  dei modelli ottimali sono estremamente limitate, ma un loro aumento provoca tuttavia un peggioramento nelle stime, caratterizzate da errori relativi medi superiori. Questi ultimi sono sì molto limitati, ma tali valori sono imputabili esclusivamente al ridotto *CV*, pari al 3.0%, piuttosto che ad una effettiva efficacia predittiva dei modelli. Tutte queste caratteristiche indicano un'assenza di correlazione tra il numero di iodio e gli

**Tabella 4.4** Statistiche dei risultati nella stima del n° di iodio a partire dalle analisi NIR della carcassa.

Spettro	Pretratt.	Varianza spiegata su X	Varianza spiegata su y	$\epsilon_{\text{med}}$ (%)	$R^2_{\text{cal}}$	$R^2_{\text{val}}$	RPD
Foss carcassa	<b>(1)</b>	<b>89.14</b>	<b>0.47</b>	<b>1.897</b>	<b>0.005</b>	<b>0.035</b>	<b>1.1</b>
	(2)	69.43	1.50	1.934	0.015	0.033	1.1
	(3)	61.80	12.50	2.227	0.125	0.082	1.1
	(4)	40.60	37.99	2.160	0.380	0.124	1.2
	(5)	53.41	16.76	2.220	0.168	0.088	1.2
	(6)	27.51	35.41	2.042	0.354	0.101	1.2
Unity carcassa	<b>(1)</b>	<b>93.19</b>	<b>0.62</b>	<b>2.085</b>	<b>0.006</b>	<b>0.033</b>	<b>1.1</b>
	(2)	76.25	12.55	2.416	0.125	0.089	1.2
	(3)	28.70	45.75	2.671	0.458	0.133	1.2
	(4)	10.36	38.65	2.569	0.387	0.112	1.2
	(5)	33.71	67.71	2.597	0.677	0.164	1.2
	(6)	13.06	62.21	2.425	0.622	0.124	1.2

spettri NIR, probabilmente dovuta a soglie di rumore sperimentale molto elevate, tali da assimilare questa variabile a rumore bianco. Ciò può trovare riscontro nella complessità del metodo analitico diretto (*Regolamento CEE 2568/91*), che richiede l'utilizzo di particolari solventi, reagenti, e necessita di tempi morti e di reazione piuttosto lunghi (mezz'ora o più, a seconda della quantità analizzata). Le numerose fasi richieste possono aver determinato una sovrapposizione dei rumori sperimentali, che nella stima finale del numero di iodio costituiscono la quota preponderante della misura analitica. Nel complesso, quindi, tale variabile non risulta stimabile a partire dagli spettri NIR; tuttavia, vista la notevole efficacia della spettroscopia NIR nella rilevazione dei doppi legami organici, di cui il numero di iodio è una misura, ci si propone di rivedere in futuro la metodologia analitica diretta, al fine di individuare ed escludere eventuali fonti di rumore ed ottimizzare così la misura.

### 4.3 Qualità dello speck stagionato

Oltre ai parametri più significativi in fase di macellazione, si è scelto di stimare anche tutte le variabili di qualità del prodotto stagionato da avviare alla commercializzazione. Tali indicatori esprimono infatti la qualità dello speck in termini di tessitura delle carni, consistenza, proprietà organolettiche, e sono di estremo interesse sia per il controllo interno dei produttori, sia per l'ottenimento della certificazione IGP, a garanzia del consumatore ed a tutela del prodotto. In questo caso le tipologie di analisi condotte sono molto più diversificate, in particolare sono disponibili:



- analisi colorimetriche, specifiche della parte magra e grassa;
- analisi Instron sulla tessitura delle carni, anch'esse suddivise in parte magra e grassa;
- acquisizioni NIR (Foss e Unity), relative alla parte magra o ottenute su un sottocampione della fetta intera macinata;
- analisi NIR eseguite con lo strumento portatile, specifiche di ogni lato dello speck;
- immagini digitali di ogni lato dello speck.

Si discutono nel seguito le stime ottenute a partire da ogni tipologia di analisi, evidenziando le proprietà caratterizzabili ed i relativi margini di accuratezza.

### 4.3.1 Analisi colorimetriche

Si riportano in Tabella 4.5 i risultati ottenuti dalle stime delle dieci variabili di qualità dello speck stagionato a partire dalle analisi colorimetriche sulla parte magra; Tabella 4.6 riporta invece i risultati complementari, ottenuti a partire dalle analisi sulla porzione di grasso della fetta di speck.

**Tabella 4.5** *Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato a partire dalle analisi colorimetriche sulla parte magra.*

Variabile	Varianza spiegata su X	Varianza spiegata su y	$\epsilon_{\text{med}}$ (%)	$R_{\text{cal}}^2$	$R_{\text{val}}^2$	RPD
Calo peso %	67.75	10.69	13.065	0.107	0.076	1.1
Attività dell'acqua	71.68	3.58	3.470	0.036	0.057	1.1
pH	70.33	3.07	0.591	0.031	0.062	1.1
Umidità	70.48	3.92	5.868	0.039	0.110	1.2
NaCl	86.10	8.98	19.884	0.090	0.133	1.2
Sostanza secca	70.95	3.99	4.614	0.040	0.106	1.2
Ceneri	85.60	8.16	13.113	0.082	0.132	1.2
Proteina grezza	68.16	5.65	2.464	0.057	0.126	1.2
EE SS	69.44	3.24	31.069	0.032	0.074	1.1
Acidità titolabile	64.66	2.68	16.996	0.027	0.070	1.1

I modelli che utilizzano le analisi colorimetriche sul magro sono tutti caratterizzati da quote di varianza spiegata su  $y$  estremamente limitate, e ciò indica una dipendenza poco significativa delle variabili di qualità dal colore della parte magra. Sebbene alcuni errori relativi di Tabella 4.5 siano molto bassi, ciò è da ricondursi prevalentemente ai  $CV$  molto limitati di alcune variabili piuttosto che ad una effettiva efficacia nella stima da parte dei modelli. Viceversa, i risultati di Tabella 4.6 indicano la possibilità di poter utilizzare le analisi colorimetriche sulla parte grassa per la stima di alcuni indicatori di qualità: attività dell'acqua, umidità e sostanza secca sono tutte caratterizzate da valori di  $R_{\text{cal}}^2$  molto buoni, sostenuti anche in convalida da RPD sensibilmente superiori all'unità. Valori simili

**Tabella 4.6** Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato a partire dalle analisi colorimetriche sulla parte grassa.

Variabile	Varianza spiegata su X	Varianza spiegata su y	$\epsilon_{\text{med}}$ (%)	$R_{\text{cal}}^2$	$R_{\text{val}}^2$	RPD
Calo peso %	89.08	9.10	9.175	0.091	0.132	1.2
<b>Attività dell'acqua</b>	<b>99.94</b>	<b>78.97</b>	<b>2.317</b>	<b>0.790</b>	<b>0.470</b>	<b>1.7</b>
pH	99.18	34.99	0.554	0.350	0.308	1.4
<b>Umidità</b>	<b>99.92</b>	<b>68.23</b>	<b>5.044</b>	<b>0.682</b>	<b>0.394</b>	<b>1.5</b>
NaCl	99.28	39.37	20.038	0.394	0.235	1.3
<b>Sostanza secca</b>	<b>99.92</b>	<b>68.10</b>	<b>4.143</b>	<b>0.681</b>	<b>0.400</b>	<b>1.5</b>
Ceneri	99.60	35.30	13.429	0.353	0.271	1.3
Proteina grezza	99.64	54.12	2.020	0.541	0.319	1.4
EE SS	88.91	9.08	34.923	0.091	0.096	1.2
Acidità titolabile	86.19	0.45	19.085	0.004	0.020	1.1

caratterizzano anche pH e proteina grezza, sebbene in questo caso i livelli di accuratezza raggiunti siano leggermente inferiori. Il confronto con i risultati di Tabella 4.5 conferma le statistiche appena discusse, poiché tutte le variabili evidenziate in grassetto sono caratterizzate da prestazioni migliori in termini di errori relativi medi, escludendo quindi che l'aumento della varianza spiegata su  $y$  sia dovuto ad un *overfitting* delle variabili stesse. Un modello statistico si definisce infatti in *overfitting* quando l'aumento di varianza spiegata non è legato ad una migliore spiegazione della varianza sistematica, quanto piuttosto alla descrizione del rumore sperimentale di misura; ciò peggiora le prestazioni in predizione, poiché minime fluttuazioni nei dati possono avere effetti estremamente rilevanti sui risultati.

Nel complesso, le analisi colorimetriche rappresentano uno strumento rapido, non invasivo ed abbastanza accurato per la stima di cinque delle dieci variabili di qualità in esame; i valori di RPD prossimi ad 1.5 indicano infatti stime con caratteristiche non propriamente ottimali, ma comunque sufficienti a cogliere un andamento medio della proprietà. È tuttavia necessario prestare particolare attenzione alla scelta del punto di misura, poiché la presenza di grasso infiltrato o di fibre di muscolo magro può alterare notevolmente le acquisizioni, non potendole più attribuire alla sola parte magra o grassa, rispettivamente. Un'eventuale applicazione in linea richiederebbe quindi una supervisione sulle acquisizioni colorimetriche da processare, eseguita o da parte di un operatore o, meglio, con l'implementazione di una verifica automatica dei dati, che sono caratterizzati da valori molto diversi di luminosità ed indice del rosso a seconda che il campione analizzato sia relativo alla parte magra o grassa.

### 4.3.2 Analisi Instron

Si riportano in Tabella 4.7 e 4.8 i risultati ottenuti dalle stime degli indicatori di qualità dello speck stagionato a partire dalle analisi Instron, relative rispettivamente alla parte magra e grassa.

**Tabella 4.7** *Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato a partire dalle analisi Instron sulla parte magra.*

Variabile	Varianza spiegata su X	Varianza spiegata su y	$\epsilon_{\text{med}}$ (%)	$R^2_{\text{cal}}$	$R^2_{\text{val}}$	RPD
Calo peso %	32.20	11.67	10.837	0.117	0.123	1.2
<b>Attività dell'acqua</b>	<b>33.60</b>	<b>25.57</b>	<b>2.852</b>	<b>0.256</b>	<b>0.319</b>	<b>1.4</b>
pH	29.71	7.58	0.674	0.076	0.090	1.2
Umidità	31.49	18.69	6.303	0.187	0.189	1.2
NaCl	33.98	15.06	17.679	0.151	0.218	1.3
Sostanza secca	31.34	17.86	5.106	0.179	0.181	1.2
Ceneri	35.08	33.90	12.166	0.339	0.274	1.3
Proteina grezza	25.15	4.81	2.390	0.048	0.070	1.1
EE SS	33.86	18.76	33.587	0.188	0.192	1.2
Acidità titolabile	40.08	27.66	21.755	0.277	0.184	1.2

**Tabella 4.8** *Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato a partire dalle analisi Instron sulla parte grassa.*

Variabile	Varianza spiegata su X	Varianza spiegata su y	$\epsilon_{\text{med}}$ (%)	$R^2_{\text{cal}}$	$R^2_{\text{val}}$	RPD
Calo peso %	85.14	45.89	12.402	0.459	0.259	1.3
Attività dell'acqua	60.13	28.96	3.134	0.290	0.248	1.3
pH	72.45	24.68	0.659	0.247	0.178	1.2
Umidità	61.43	24.38	6.144	0.244	0.160	1.2
NaCl	38.38	6.06	19.405	0.061	0.085	1.1
Sostanza secca	60.08	23.41	4.969	0.234	0.151	1.2
Ceneri	45.49	22.32	13.024	0.223	0.221	1.3
Proteina grezza	47.69	12.30	2.666	0.123	0.104	1.2
EE SS	46.63	11.19	34.277	0.112	0.172	1.2
Acidità titolabile	78.28	32.81	20.766	0.328	0.142	1.2

Come si può notare dalle colonne relative all'RPD, il contenuto informativo delle analisi Instron non risulta particolarmente legato alla qualità del prodotto stagionato, sia per quanto riguarda le analisi sul magro che sul grasso. L'unica variabile stimabile con una discreta accuratezza è l'attività dell'acqua, valutata con un errore relativo medio sufficientemente limitato a partire dalle analisi sulla parte magra. Le scarse prestazioni in predizione

sono tuttavia giustificabili considerando l'estrema eterogeneità interna dello speck, poiché la presenza di setti o nervature nel sottocampione analizzato va ad influenzare in modo significativo il profilo di sforzo ottenuto. Inoltre, gli stessi processi di stagionatura ed affumicatura a cui lo speck è sottoposto non lo condizionano in modo uniforme, aumentando la variabilità delle proprietà all'interno del prodotto. La tessitura delle carni rimane senza dubbio un indice di estrema rilevanza per la caratterizzazione della qualità dello speck, ma l'analisi Instron è molto sensibile alla natura del sottocampione, estratto nella prassi in modo casuale. Considerando inoltre le tempistiche di estrazione del sottocampione e l'invasività della misura, l'analisi Instron è da ritenersi poco idonea ad un utilizzo per la caratterizzazione rapida della qualità del prodotto.

### 4.3.3 *Analisi spettroscopiche NIR*

Visti i numerosi casi studio di letteratura che riportano l'efficace utilizzo della spettroscopia al vicino infrarosso per la stima di diverse proprietà chimiche, si è scelto di utilizzare queste analisi seguendo molteplici procedure, al fine di evidenziare la migliore tipologia di sottocampione e di strumento:

1. si è estratto un campione dalla parte magra di ogni fetta di speck, sottoponendolo all'analisi NIR con gli strumenti Foss e Unity;
2. è stata macinata un'intera fetta di speck di ogni campione, analizzando solamente un sottocampione scelto casualmente; gli strumenti utilizzati sono sempre il Foss e lo Unity;
3. si è scelto di utilizzare anche uno strumento portatile, in modo da verificare non solo le caratteristiche specifiche di ogni lato dello speck, ma anche la possibilità di caratterizzare la qualità del prodotto senza la necessità di uno strumento fisso da banco, seppur più accurato.

Si discutono nel seguito queste tre tipologie di analisi NIR, in modo da evidenziarne gli specifici pregi e svantaggi; inoltre, nonostante siano stati applicati tutti i pretrattamenti indicati alla sezione 2.2.1, al fine di non appesantire la lettura vengono riportati solamente i risultati ottenuti con il pretrattamento ottimale.

#### 4.3.3.1 Parte magra

Si riportano in Tabella 4.9 e 4.10 le statistiche sui risultati ottenuti stimando le dieci variabili di qualità dello speck a partire dagli spettri NIR sulla parte magra; nello specifico, Tabella 4.9 si riferisce agli spettri Foss, mentre Tabella 4.10 a quelli Unity.

Tabella 4.9 evidenzia in modo immediato il notevole contenuto informativo della spettroscopia NIR, che consente di stimare in modo accurato la quasi totalità degli indicatori di

**Tabella 4.9** *Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato a partire dagli spettri NIR Foss sulla parte magra.*

Variabile	Pretratt.	Varianza spiegata su X	Varianza spiegata su y	$\epsilon_{\text{med}}$ (%)	$R^2_{\text{cal}}$	$R^2_{\text{val}}$	RPD
Calo peso %	(3)	49.04	9.79	7.990	0.098	0.200	1.2
Attività dell'acqua	(2)	<b>99.92</b>	<b>99.86</b>	<b>0.391</b>	<b>0.999</b>	<b>0.974</b>	<b>9.1</b>
pH	(4)	<b>86.01</b>	<b>97.26</b>	<b>0.468</b>	<b>0.973</b>	<b>0.443</b>	<b>1.6</b>
Umidità	(6)	<b>76.73</b>	<b>99.68</b>	<b>0.999</b>	<b>0.997</b>	<b>0.961</b>	<b>6.9</b>
NaCl	(5)	<b>89.41</b>	<b>86.16</b>	<b>8.229</b>	<b>0.862</b>	<b>0.743</b>	<b>2.7</b>
Sostanza secca	(5)	<b>94.49</b>	<b>99.66</b>	<b>0.695</b>	<b>0.997</b>	<b>0.969</b>	<b>8.0</b>
Ceneri	(3)	<b>98.00</b>	<b>99.89</b>	<b>2.107</b>	<b>0.999</b>	<b>0.965</b>	<b>7.5</b>
Proteina grezza	(6)	<b>87.00</b>	<b>99.96</b>	<b>1.405</b>	<b>1.000</b>	<b>0.602</b>	<b>2.0</b>
EE SS	(6)	<b>86.69</b>	<b>99.96</b>	<b>11.374</b>	<b>1.000</b>	<b>0.745</b>	<b>2.8</b>
Acidità titolabile	(3)	77.94	49.85	17.236	0.499	0.369	1.5

**Tabella 4.10** *Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato a partire dagli spettri NIR Unity sulla parte magra.*

Variabile	Pretratt.	Varianza spiegata su X	Varianza spiegata su y	$\epsilon_{\text{med}}$ (%)	$R^2_{\text{cal}}$	$R^2_{\text{val}}$	RPD
Calo peso %	(1)	95.40	7.14	7.843	0.071	0.162	1.2
Attività dell'acqua	(2)	<b>99.75</b>	<b>99.89</b>	<b>0.494</b>	<b>0.999</b>	<b>0.961</b>	<b>7.6</b>
pH	(6)	<b>18.82</b>	<b>49.71</b>	<b>0.495</b>	<b>0.497</b>	<b>0.388</b>	<b>1.5</b>
Umidità	(1)	<b>99.98</b>	<b>99.43</b>	<b>1.099</b>	<b>0.994</b>	<b>0.947</b>	<b>6.2</b>
NaCl	(3)	<b>71.17</b>	<b>94.71</b>	<b>10.345</b>	<b>0.947</b>	<b>0.632</b>	<b>2.1</b>
Sostanza secca	(1)	<b>99.99</b>	<b>99.93</b>	<b>0.666</b>	<b>0.999</b>	<b>0.973</b>	<b>8.5</b>
Ceneri	(1)	<b>100.00</b>	<b>99.85</b>	<b>3.271</b>	<b>0.998</b>	<b>0.908</b>	<b>5.0</b>
Proteina grezza	(3)	<b>74.86</b>	<b>96.95</b>	<b>1.609</b>	<b>0.970</b>	<b>0.503</b>	<b>1.7</b>
EE SS	(1)	<b>99.97</b>	<b>92.44</b>	<b>16.431</b>	<b>0.924</b>	<b>0.552</b>	<b>1.9</b>
Acidità titolabile	(5)	44.65	62.84	17.836	0.628	0.315	1.4

qualità, ad eccezione di due sole variabili, calo peso % ed acidità titolabile. Le varianze spiegate su  $y$  sono prossime al 100%, mentre i coefficienti di determinazione sono ottimi sia in calibrazione che in convalida. Inoltre, l'indice RPD evidenzia un notevole vantaggio nella stime rispetto all'utilizzo dei valori medi, con valori sensibilmente superiori ad 1: è opportuno sottolineare come per quattro variabili (attività dell'acqua, umidità, sostanza secca e ceneri) l'RPD assuma addirittura valori superiori a 6, fatto di notevole significatività statistica dal momento che valori pari a 3 possono già essere considerati ottimi ai fini della stima (Fearn, 2003). Gli eccellenti risultati in predizione sono confermati anche in Tabella 4.10, che riporta valori molto simili di varianze spiegate e coefficienti di determinazione. Inoltre, sebbene nel complesso siano da preferirsi le analisi Foss, è opportuno evidenziare

l'ottima stima della sostanza secca utilizzando gli spettri Unity: l'errore relativo medio è inferiore al caso precedente, con un valore di RPD pari a 8.5.

Nell'analisi degli errori relativi medi di entrambe le Tabelle 4.9 e 4.10 è necessario considerare anche gli andamenti dei coefficienti di variazione di Tabella 4.1, poiché i modelli ottimali caratterizzati da errori superiori al 10% sono specifici delle variabili con la maggiore variabilità, rispettivamente NaCl ed EE SS. Errori relativi e statistiche sui risultati non sono quindi discordanti tra loro: la calibrazione ed il *fitting* sono molto buoni, ma la notevole variabilità delle proprietà determina errori relativi leggermente più elevati. A tal proposito è opportuno tener conto anche di un altro fattore: nella calibrazione dei modelli PLS, si è scelto di non vincolare la presenza nel *dataset* di calibrazione dei campioni caratterizzati dagli estremi della proprietà in esame. Ciò assicura la massima generalità e robustezza dei risultati, ma in questo modo può accadere che i valori estremi della proprietà siano associati a campioni di convalida, generando un'estrapolazione della stima al di fuori dello spazio del modello e quindi un peggioramento delle statistiche del singolo modello.

#### 4.3.3.2 Fetta intera macinata

Verificata la notevole efficacia degli spettri NIR sulla parte magra, si analizza ora se l'inclusione della parte grassa porti o meno dei vantaggi in termini di accuratezza delle stime ottenute. Prima di procedere nell'analisi dei risultati, è tuttavia opportuno fare una precisazione: sebbene venga macinata l'intera fetta di speck, l'estrazione casuale del sottocampione non assicura l'effettiva presenza di grasso nel sottocampione analizzato. I risultati sono quindi relativi ad un sottocampione macinato, ma il relativo rapporto magro/grasso non rientra nelle specifiche soggette a controllo da parte dell'operatore.

Anche in questo caso si riportano in due tabelle diverse i risultati degli spettri Foss e Unity: i primi in Tabella 4.11, i secondi in Tabella 4.12.

Il confronto dei risultati degli spettri Foss e Unity evidenzia in questo caso delle differenze abbastanza significative, anche se come detto non è possibile imputare queste variazioni ad uno dei due strumenti poiché potrebbero essere semplicemente dovute alla diversa composizione dei sottocampioni analizzati. Innanzitutto, si evidenzia il peggioramento generale delle stime rispetto alle acquisizioni sulla parte magra, con l'unica eccezione del pH che si attesta su margini di accuratezza leggermente migliori. Anche i valori dell'indice RPD risultano sensibilmente inferiori rispetto alle stime sulla parte magra, sebbene si stabilizzino sempre su livelli ottimali, prossimi a 3. In questo caso, tuttavia, si notano delle differenze nelle stime a seconda dello strumento utilizzato nell'analisi spettroscopica: ad esempio, la proteina grezza risulta poco stimabile a partire dagli spettri Foss, mentre predetta con

**Tabella 4.11** *Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato a partire dagli spettri NIR Foss sulla fetta intera macinata.*

Variabile	Pretratt.	Varianza spiegata su X	Varianza spiegata su y	$\epsilon_{\text{med}}$ (%)	$R_{\text{cal}}^2$	$R_{\text{val}}^2$	RPD
Calo peso %	(1)	81.75	19.66	7.840	0.197	0.279	1.3
<b>Attività dell'acqua</b>	<b>(3)</b>	<b>96.03</b>	<b>97.72</b>	<b>0.918</b>	<b>0.977</b>	<b>0.873</b>	<b>3.8</b>
<b>pH</b>	<b>(2)</b>	<b>82.17</b>	<b>27.72</b>	<b>0.444</b>	<b>0.277</b>	<b>0.442</b>	<b>1.6</b>
<b>Umidità</b>	<b>(5)</b>	<b>93.48</b>	<b>96.53</b>	<b>2.301</b>	<b>0.965</b>	<b>0.798</b>	<b>3.1</b>
<b>NaCl</b>	<b>(6)</b>	<b>95.75</b>	<b>92.30</b>	<b>8.291</b>	<b>0.923</b>	<b>0.727</b>	<b>2.6</b>
<b>Sostanza secca</b>	<b>(5)</b>	<b>93.20</b>	<b>96.23</b>	<b>1.886</b>	<b>0.962</b>	<b>0.796</b>	<b>3.1</b>
<b>Ceneri</b>	<b>(2)</b>	<b>99.97</b>	<b>99.46</b>	<b>4.531</b>	<b>0.995</b>	<b>0.815</b>	<b>3.3</b>
Proteina grezza	(4)	44.54	22.16	2.186	0.222	0.251	1.3
EE SS	(3)	90.39	67.61	23.048	0.676	0.414	1.5
Acidità titolabile	(3)	77.40	57.24	17.236	0.572	0.349	1.5

**Tabella 4.12** *Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato a partire dagli spettri NIR Unity sulla fetta intera macinata.*

Variabile	Pretratt.	Varianza spiegata su X	Varianza spiegata su y	$\epsilon_{\text{med}}$ (%)	$R_{\text{cal}}^2$	$R_{\text{val}}^2$	RPD
Calo peso %	(1)	83.62	17.38	7.509	0.174	0.267	1.3
<b>Attività dell'acqua</b>	<b>(2)</b>	<b>99.72</b>	<b>97.78</b>	<b>1.434</b>	<b>0.978</b>	<b>0.737</b>	<b>2.5</b>
<b>pH</b>	<b>(3)</b>	<b>53.51</b>	<b>38.39</b>	<b>0.471</b>	<b>0.384</b>	<b>0.381</b>	<b>1.5</b>
<b>Umidità</b>	<b>(3)</b>	<b>87.75</b>	<b>99.56</b>	<b>2.466</b>	<b>0.996</b>	<b>0.785</b>	<b>2.8</b>
<b>NaCl</b>	<b>(1)</b>	<b>99.95</b>	<b>99.50</b>	<b>7.566</b>	<b>0.995</b>	<b>0.763</b>	<b>2.7</b>
<b>Sostanza secca</b>	<b>(3)</b>	<b>88.88</b>	<b>99.70</b>	<b>1.915</b>	<b>0.997</b>	<b>0.791</b>	<b>2.9</b>
<b>Ceneri</b>	<b>(1)</b>	<b>99.97</b>	<b>99.73</b>	<b>5.702</b>	<b>0.997</b>	<b>0.734</b>	<b>2.7</b>
<b>Proteina grezza</b>	<b>(3)</b>	<b>74.86</b>	<b>96.95</b>	<b>1.609</b>	<b>0.970</b>	<b>0.503</b>	<b>1.7</b>
EE SS	(1)	99.48	90.73	24.861	0.907	0.385	1.5
<b>Acidità titolabile</b>	<b>(3)</b>	<b>77.99</b>	<b>93.06</b>	<b>15.429</b>	<b>0.931</b>	<b>0.436</b>	<b>1.6</b>

accuratezza utilizzando le acquisizioni Unity. Inoltre, fatto ben più rilevante, risulta stimabile anche l'acidità titolabile, predetta con una buona accuratezza utilizzando gli spettri Unity, pretrattati con derivata prima.

Nel complesso, le analisi spettroscopiche sulla fetta intera macinata risultano meno informative rispetto alle acquisizioni sulla sola parte magra; tuttavia, qualora fossero disponibili solamente le analisi sullo speck macinato, sarebbe comunque possibile ottenere delle stime molto buone, sebbene meno accurate rispetto alle analisi NIR ottimali sul magro.

4.3.3.3 Strumento portatile

L'utilizzo di uno strumento NIR portatile per la stima indiretta di alcune proprietà chimiche trova limitato riscontro in letteratura poiché la minore accuratezza delle acquisizioni è nota a priori. Tuttavia, visto il notevole contenuto informativo degli spettri NIR, anche un livello di accuratezza inferiore potrebbe essere sufficiente per ottenere ottime stime, con il notevole vantaggio di non dover utilizzare uno strumento da banco, molto più costoso, e con la possibilità di ottenere una misura direttamente sul campo. Ad esempio, per il controllo di qualità interno di un piccolo produttore risulterebbe molto più indicato lo strumento portatile, più economico, versatile, e facile da usare. Inoltre, caratteristica molto importante, lo strumento portatile può essere spostato su ogni lato dello speck, consentendo di ottenere acquisizioni dedicate; ciò può essere molto utile per evidenziare eventuali correlazioni o caratteristiche peculiari, specifiche di un lato dello speck.

Tuttavia, nel caso in esame l'intervallo analizzato dallo strumento è molto limitato, e ciò limita l'efficacia delle stime ottenute. I risultati presentati nel seguito devono quindi essere considerati un punto di partenza per ulteriori sviluppi futuri, e servono come primo test di efficacia per verificare se la via delle analisi con lo strumento portatile sia percorribile o meno.

Si riportano in Tabella 4.13 le statistiche dei risultati ottenuti stimando le variabili di qualità a partire dagli spettri NIR sul lato A, acquisiti con lo strumento portatile. Come si può notare, l'accuratezza delle stime è decisamente inferiore rispetto agli strumenti da banco Foss e Unity, ma pienamente confrontabile ed anzi superiore rispetto alle analisi colorimetriche ed Instron. Nello specifico, la stima più accurata è quella relativa all'attività dell'acqua, caratterizzata da un errore relativo del 2.5% e da un discreto valore di RPD.

**Tabella 4.13** *Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato a partire dagli spettri NIR sul lato A - strumento portatile.*

Variabile	Pretratt.	Varianza spiegata su X	Varianza spiegata su y	$\epsilon_{\text{med}}$ (%)	$R^2_{\text{cal}}$	$R^2_{\text{val}}$	RPD
Calo peso %	(6)	36.53	15.45	10.111	0.155	0.170	1.2
<b>Attività dell'acqua</b>	<b>(5)</b>	<b>93.39</b>	<b>74.08</b>	<b>2.517</b>	<b>0.741</b>	<b>0.434</b>	<b>1.6</b>
pH	(3)	47.93	11.61	0.551	0.116	0.183	1.2
Umidità	(2)	94.46	41.09	5.727	0.411	0.297	1.4
NaCl	(6)	69.23	64.70	15.881	0.647	0.309	1.4
Sostanza secca	(2)	93.90	39.85	4.598	0.399	0.294	1.4
Ceneri	(5)	79.59	47.44	12.254	0.474	0.232	1.3
Proteina grezza	(6)	42.69	21.78	2.436	0.218	0.194	1.2
EE SS	(2)	89.46	36.47	33.449	0.365	0.290	1.3
Acidità titolabile	(4)	46.24	25.37	21.441	0.254	0.141	1.2



Poichè tale andamento è caratteristico anche dei risultati ottenuti utilizzando le acquisizioni sugli altri tre lati, si è scelto di non riportare le ulteriori stime e le relative statistiche, considerate ridondanti e poco significative. Questo andamento può essere spiegato in un'assenza di informazioni peculiari (specifiche di un lato) all'interno dell'intervallo analizzato, tuttavia troppo limitato per trarre conclusioni generalizzate.

Nel complesso, queste analisi con lo strumento portatile non sono risultate indicate per la caratterizzazione quantitativa di una proprietà, ma potrebbero essere utili per avere un'indicazione qualitativa sull'andamento del processo di stagionatura dello speck. Si rimanda tuttavia a sviluppi futuri la verifica dell'efficacia di queste analisi, raccomandando l'utilizzo di un intervallo spettrale di acquisizione più ampio, in modo da poter avere un quadro completo sulla validità di questo strumento ai fini della stima. Anche la specificità di alcune proprietà su determinati lati dello speck è lasciata a futuri sviluppi.

#### 4.3.4 *Analisi d'immagine*

L'utilizzo delle immagini digitali ad alta definizione per la stima delle proprietà chimiche dello speck presenta numerose caratteristiche di interesse, tra cui:

- l'ampia diffusione ed il costo molto limitato delle fotocamere digitali;
- la possibilità di condurre una stima a partire da un'immagine della cotenna o del lato affumicato (lati C e D di Figura 1.2), senza nemmeno tagliare il prodotto; ciò rappresenterebbe un notevole vantaggio perché per i campioni sottoposti al controllo di qualità non sarebbe preclusa la vendita;
- la possibilità di analizzare la tessitura delle carni, confrontando i risultati con quelli ottenuti dallo strumento Instron, dedicato a queste analisi ma distruttivo sul campione.

Al fine di verificare se questi potenziali vantaggi trovino un concreto riscontro sulle stime ottenute, sono state utilizzate due diverse metodologie: la prima si basa sul colore medio della carne, mentre la seconda sulla tessitura. Si presentano nel seguito i risultati ottenuti da ciascuna di esse, confrontandoli con l'accuratezza raggiunta utilizzando la spettroscopia NIR.

##### 4.3.4.1 Analisi del colore medio

Si riportano in Tabella 4.14 le statistiche sulle stime degli indicatori di qualità dello speck stagionato ottenute utilizzando i livelli di intensità medi dei singoli canali RGB, presentati alla sezione 2.4.1. Sebbene le analisi sui lati A e B (Figura 1.2) siano state condotte sia sulla parte magra che su quella grassa, si riportano per brevità solamente i risultati migliori, specificando a che porzione si riferiscano. Si è invece scelto di riportare in una

Tabella distinta, la 4.15, i risultati ottenuti a partire dalle analisi sui lati C e D, poiché delle stime accurate su queste facce dello speck consentirebbero di non tagliare il prodotto per caratterizzarne la qualità, con notevoli vantaggi dal punto di vista industriale.

I risultati di Tabella 4.14 evidenziano la migliore efficacia delle analisi sulla parte magra, che consentono di ottenere una migliore accuratezza per nove delle dieci variabili consi-

**Tabella 4.14** *Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato a partire dalle analisi sul colore medio, lati A e B.*

Variabile	Lato e Parte	Varianza spiegata su X	Varianza spiegata su y	$\epsilon_{\text{med}}$ (%)	$R_{\text{cal}}^2$	$R_{\text{val}}^2$	RPD
Calo peso %	A magro	26.31	78.02	5.813	0.780	0.250	1.3
Attività dell'acqua	B magro	14.51	57.90	2.474	0.579	0.362	1.4
pH	B magro	9.84	57.03	0.555	0.570	0.245	1.3
Umidità	B magro	23.39	85.84	4.667	0.858	0.380	1.5
NaCl	B magro	13.19	53.30	16.048	0.533	0.277	1.3
Sostanza secca	B magro	23.23	84.48	3.840	0.845	0.361	1.5
Ceneri	A magro	11.25	66.34	11.056	0.663	0.194	1.2
Proteina grezza	B magro	20.77	83.51	2.195	0.835	0.232	1.3
EE SS	B grasso	19.63	41.23	25.899	0.412	0.222	1.3
Acidità titolabile	A magro	11.00	54.86	15.504	0.549	0.189	1.2

**Tabella 4.15** *Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato a partire dalle analisi sul colore medio, lati C e D.*

Variabile	Lato	Varianza spiegata su X	Varianza spiegata su y	$\epsilon_{\text{med}}$ (%)	$R_{\text{cal}}^2$	$R_{\text{val}}^2$	RPD
Calo peso %	D	23.87	23.04	7.888	0.230	0.151	1.2
<b>Attività dell'acqua</b>	<b>D</b>	<b>35.16</b>	<b>68.08</b>	<b>2.532</b>	<b>0.681</b>	<b>0.427</b>	<b>1.6</b>
pH	D	34.15	54.97	0.579	0.550	0.268	1.3
Umidità	D	30.70	49.63	5.511	0.496	0.332	1.4
NaCl	D	34.61	69.71	15.445	0.697	0.374	1.5
Sostanza secca	C	63.98	79.52	4.417	0.795	0.373	1.5
Ceneri	D	36.12	63.13	10.254	0.631	0.365	1.5
Proteina grezza	D	27.69	56.58	2.221	0.566	0.234	1.3
EE SS	C	28.44	24.68	28.499	0.247	0.130	1.2
Acidità titolabile	D	36.28	61.16	17.947	0.612	0.207	1.3

derate. Le stime raggiunte non sono tuttavia molto significative in termini di  $R^2$  e RPD, poiché consentono di ottenere un riscontro piuttosto qualitativo sull'andamento delle variabili di qualità, descrivendone per lo più un profilo medio. Tabella 4.15 riporta invece risultati piuttosto sorprendenti, poiché molte delle variabili di qualità vengono stimate in modo più accurato a partire dai lati esterni, relativi nello specifico alla parte affumicata

(C) ed alla cotenna (D). Tuttavia, nonostante alcuni singoli valori di RPD siano buoni, questo miglioramento non consente di raggiungere i livelli di accuratezza caratteristici delle analisi NIR, di gran lunga più informative. A tal proposito, è stata utilizzata anche una seconda metodica di analisi d'immagine basata sulla maschera di covarianza (Yu e MacGregor, 2003), finalizzata alla descrizione delle proprietà locali, non mediate sull'intera fetta di speck. Tuttavia, i risultati ottenuti sono molto simili a quelli di Tabella 4.14 e 4.15, e quindi la complicazione del metodo di analisi non trova giustificazione. Si rimanda tuttavia a sviluppi futuri un miglioramento delle prestazioni di stima utilizzando l'informazione legata al colore, sfruttando anche strumenti di acquisizione più sofisticati e con un miglior bilanciamento del bianco.

#### 4.3.4.2 Analisi della tessitura - decomposizione wavelet

Le immagini digitali possono essere utilizzate anche per analizzare la tessitura delle carni, che come discusso nella sezione relativa alle analisi Instron rappresenta un parametro molto rilevante nella percezione sensoriale del consumatore. Vista le notevoli differenze tra il tessuto magro e quello adiposo, si è scelto di analizzare separatamente le due diverse porzioni di ciascuna fetta di speck. La tecnica utilizzata si basa sulla decomposizione *wavelet*, applicata a diverse scale di risoluzione con due diversi approcci, multi-approssimazione e multi-dettaglio. Anche in questo caso, si riportano solamente le stime caratterizzate dai risultati migliori, specificando il lato e la parte di speck a cui l'immagine digitale fa riferimento. Nello specifico, Tabella 4.16 riporta i risultati dell'approccio multi-approssimazione, mentre Tabella 4.17 è relativa alle stime dell'approccio multi-dettaglio; in entrambi i casi, i lati di riferimento sono A e B.

Dal confronto dei risultati ottenuti emergono delle prestazioni poco significative nella stima della qualità finale del prodotto, ad eccezione delle variabili umidità e sostanza secca. Queste ultime vengono infatti predette con una buona accuratezza, soprattutto utilizzando l'approccio multi-dettaglio, caratterizzato da un valore di RPD prossimo a 2; approssimando invece l'immagine originale a più scale di decomposizione si ottengono delle stime leggermente peggiori, con un errore relativo medio superiore di circa lo 0.5%. Per tutte le altre variabili di qualità le stime sono modeste e non utilizzabili ai fini di una predizione quantitativa degli indicatori; la migliore efficienza di uno dei due approcci è variabile a seconda della singola variabile in esame. È interessante notare come le due variabili caratterizzate da buone stime siano anche le più legate da un punto di vista fisico alla tessitura delle carni, confermando di fatto l'obiettivo del metodo. Infatti, l'umidità e la sostanza secca influenzano in modo significativo la consistenza e la sensazione al palato dello speck, determinando una sensazione gommosa, morbida o stopposa a seconda dei

**Tabella 4.16** *Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato ottenute mediante decomposizione wavelet - approccio multi-approssimazione, lati A e B.*

Variabile	Lato e Parte	Varianza spiegata su X	Varianza spiegata su y	$\epsilon_{\text{med}}$ (%)	$R^2_{\text{cal}}$	$R^2_{\text{val}}$	RPD
Calo peso %	A grasso	43.27	15.58	7.605	0.156	0.243	1.3
Attività dell'acqua	B magro	90.89	40.41	3.174	0.404	0.283	1.3
pH	B magro	47.58	18.29	0.525	0.183	0.306	1.4
<b>Umidità</b>	<b>B magro</b>	<b>92.63</b>	<b>57.84</b>	<b>4.787</b>	<b>0.578</b>	<b>0.468</b>	<b>1.6</b>
NaCl	B magro	68.31	13.10	19.253	0.131	0.198	1.2
<b>Sostanza secca</b>	<b>B magro</b>	<b>92.69</b>	<b>57.37</b>	<b>3.794</b>	<b>0.574</b>	<b>0.449</b>	<b>1.6</b>
Ceneri	B magro	81.51	28.68	12.849	0.287	0.206	1.3
Proteina grezza	A grasso	34.88	10.03	2.371	0.100	0.155	1.2
EE SS	A grasso	51.06	21.71	25.791	0.217	0.360	1.5
Acidità titolabile	A grasso	51.00	27.85	17.258	0.279	0.312	1.4

**Tabella 4.17** *Statistiche dei risultati nella stima delle variabili di qualità dello speck stagionato ottenute mediante decomposizione wavelet - approccio multi-dettaglio, lati A e B.*

Variabile	Lato e Parte	Varianza spiegata su X	Varianza spiegata su y	$\epsilon_{\text{med}}$ (%)	$R^2_{\text{cal}}$	$R^2_{\text{val}}$	RPD
Calo peso %	A magro	80.53	19.82	6.917	0.198	0.290	1.4
Attività dell'acqua	A magro	99.71	58.95	2.762	0.590	0.413	1.5
pH	B magro	71.28	11.68	0.544	0.117	0.301	1.3
<b>Umidità</b>	<b>A magro</b>	<b>99.85</b>	<b>70.23</b>	<b>4.220</b>	<b>0.702</b>	<b>0.504</b>	<b>1.8</b>
NaCl	A magro	83.09	45.81	19.160	0.458	0.264	1.3
<b>Sostanza secca</b>	<b>A magro</b>	<b>99.80</b>	<b>68.73</b>	<b>3.426</b>	<b>0.687</b>	<b>0.521</b>	<b>1.8</b>
Ceneri	A magro	98.50	36.95	13.624	0.369	0.215	1.3
Proteina grezza	B grasso	93.52	17.81	2.319	0.178	0.183	1.2
EE SS	A grasso	59.83	23.91	26.689	0.239	0.287	1.4
Acidità titolabile	B magro	91.84	14.41	17.648	0.144	0.227	1.3

valori assunti. Visto il rilevante impatto di questi indicatori nella scelta del consumatore, un loro controllo, seppur qualitativo, con tecniche poco costose e non invasive potrebbe trovare applicazione anche nel mondo industriale.

Applicando la stessa tecnica alle fotografie dei lati esterni C e D si ottengono risultati poco significativi, caratterizzati da valori di RPD molto prossimi all'unità e da coefficienti di determinazione scadenti. Le uniche variabili per cui si riscontrano delle buone statistiche, superiori alla media, sono il contenuto di NaCl e le ceneri, stimate con una sufficiente ac-

curatezza con entrambi gli approcci<sup>1</sup>. Si riportano in Tabella 4.18 questi specifici risultati, ritenuti interessanti visto che consentono di non tagliare il prodotto per caratterizzarne la qualità. Sebbene gli errori relativi medi siano molto superiori rispetto ai risultati ottenuti con la spettroscopia NIR, i risultati di Tabella 4.18 indicano la possibilità di utilizzare una fotografia della cotenna (D) per avere una stima qualitativa del contenuto di NaCl e ceneri. Questi risultati assumono maggiore significatività ricordando i coefficienti di variazione di Tabella 4.1, che dimostrano come queste due variabili siano tra quelle caratterizzate da una maggior variabilità.

**Tabella 4.18** *Statistiche dei risultati nella stima delle variabili NaCl e ceneri, ottenute mediante decomposizione wavelet dei lati esterni C e D.*

Variabile	Lato	Varianza spiegata su X	Varianza spiegata su y	$\epsilon_{\text{med}}$ (%)	$R_{\text{cal}}^2$	$R_{\text{val}}^2$	RPD
Multi-approssimazione							
NaCl	C	60.67	23.99	17.140	0.240	0.278	1.3
	D	53.86	22.78	14.866	0.228	0.313	1.4
Ceneri	C	62.97	30.27	11.695	0.303	0.315	1.4
	D	72.94	33.09	12.102	0.331	0.306	1.4
Multi-dettaglio							
NaCl	C	87.37	38.65	16.786	0.387	0.332	1.4
	D	68.38	27.06	14.657	0.271	0.336	1.4
Ceneri	C	92.79	43.05	11.057	0.430	0.340	1.4
	D	86.39	40.61	10.030	0.406	0.400	1.5

## 4.4 Conclusioni

Nel corso di questo Capitolo sono stati utilizzati diversi *dataset* sperimentali per verificare la possibilità di monitorare la qualità dello speck lungo l'intera filiera di lavorazione, evitando le analisi chimiche dirette e risparmiando così tempi e risorse. Si riassumono in Tabella 4.19 i migliori risultati ottenuti, specificando le relative statistiche e la tipologia di analisi utilizzate.

Come si può notare, sette delle dodici variabili di qualità vengono stimate in modo ottimo, come dimostrano gli errori relativi molto limitati ed i significativi valori dei coefficienti di

<sup>1</sup>Si ricordi tuttavia che NaCl e ceneri rappresentano di fatto una sola variabile, dal momento che il cloruro di sodio costituisce la quota preponderante delle ceneri.

**Tabella 4.19** *Elenco delle prestazioni migliori ottenute nella stima delle variabili di qualità dello speck.*

Variabile	Analisi	$\epsilon_{\text{med}}$ (%)	$R^2_{\text{cal}}$	$R^2_{\text{val}}$	RPD
pH a 45 minuti	NIR Foss della carcassa	3.138	0.231	0.225	1.3
n° di iodio	— non stimabile —				
Calo peso %	Colore medio lato A - maschera sulla parte magra	5.813	0.780	0.250	1.3
Attività dell'acqua	NIR Foss parte magra	0.391	0.999	0.974	9.1
pH	NIR Foss parte magra	0.468	0.973	0.443	1.6
Umidità	NIR Foss parte magra	0.999	0.997	0.961	6.9
NaCl	NIR Foss parte magra	8.229	0.862	0.743	2.7
Sostanza secca	NIR Unity parte magra	0.666	0.999	0.973	8.5
Ceneri	NIR Foss parte magra	2.107	0.999	0.965	7.5
Proteina grezza	NIR Foss parte magra	1.405	1.000	0.602	2.0
EE SS	NIR Foss parte magra	11.374	1.000	0.745	2.8
Acidità titolabile	NIR Unity intero macinato	15.429	0.931	0.436	1.6

determinazione ed RPD. Il numero di iodio è l'unica variabile non stimabile, mentre pH a 45 minuti, calo peso %, pH (del magro stagionato) ed acidità titolabile vengono predette in modo discreto, con valori di RPD che indicano un *fitting* centrato prevalentemente sull'andamento medio della proprietà. La tipologia di analisi più informativa ai fini della stima è costituita dalla spettroscopia NIR, con una netta prevalenza delle analisi Foss sulla parte magra dello speck stagionato. Le altre analisi utilizzate sono gli spettri Unity, sia sulla parte magra del prodotto stagionato che sulla fetta intera macinata, gli spettri Foss sulla carcassa macellata e le immagini ad alta risoluzione, relative nello specifico al lato A. Le tipologie di analisi rimanenti (analisi colorimetriche, Instron, NIR portatile) hanno consentito di ottenere stime poco significative, di carattere qualitativo; un discorso analogo vale anche per le immagini ad alta risoluzione, sebbene in alcuni casi possano risultare utili per avere una stima qualitativa, rapida ed economica di specifici indicatori di qualità.

Riassumendo:

- è stata verificata l'ottima efficacia della spettroscopia NIR per il monitoraggio della qualità dello speck stagionato in modo rapido, accurato e ripetibile;
- l'utilizzo delle immagini digitali per la stima della qualità appare ancora prematuro perché i margini di accuratezza raggiunti sono decisamente inferiori; si lascia a sviluppi futuri il miglioramento delle prestazioni, introducendo anche l'utilizzo di apparecchiature più sofisticate.

La robustezza dei risultati ottenuti è inoltre garantita dall'approccio Monte Carlo, basato su mille simulazioni indipendenti. Tutte queste caratteristiche si sposano al meglio con le necessità industriali del controllo qualità, a cui potrebbero essere trasferite con successo tutte le tecniche sviluppate.

# Capitolo 5

## Classificazione dei campioni rispetto al sesso, ai regimi alimentari e ai parametri sensoriali

Nel corso di questo Capitolo viene esaminata la possibilità di classificare i 36 campioni a disposizione in base al sesso dei suini, al regime alimentare a cui sono stati sottoposti durante l'allevamento, ed ai parametri sensoriali attribuiti da una giuria di esperti.

La classificazione del sesso e dell'alimentazione si basa sulle analisi spettroscopiche NIR e sui dati raccolti lungo l'intera filiera produttiva; l'obiettivo è di verificare se i diversi regimi alimentari inducano delle differenze sensibili nella qualità dello speck ottenuto, studiando altresì la possibilità di ottenere un prodotto finale di qualità omogenea ed ottimale. Inoltre, si vuole costruire una metodologia oggettiva di analisi dei parametri sensoriali, che consenta di esprimere un giudizio obiettivo e ripetibile sulla qualità dello speck.

### 5.1 Sesso dei suini

La classificazione dello speck rispetto al sesso dei suini è di notevole interesse poiché si vuole verificare se l'operazione di castrazione a cui sono sottoposti gli esemplari maschio sia sufficiente per ottenere una qualità finale omogenea. Infatti, viste le notevoli differenze di attività ormonale tra gli esemplari maschio e femmina, l'equilibratura ottenuta con la castrazione potrebbe non essere sufficiente, causando la persistenza di proprietà peculiari e quindi disomogeneità qualitative. La classificazione è stata condotta a partire da diversi *dataset*  $\mathbf{X}$ , costituiti dai dati raccolti lungo la filiera di lavorazione o dalle analisi NIR sulla carcassa, sulla parte magra dello speck stagionato e su un sottocampione della fetta intera macinata. Lo spazio della risposta  $\mathbf{Y}$  è invece costituito dalla variabile binaria relativa al sesso. Inoltre, vengono confrontate le prestazioni di diverse metodiche di classificazione, introdotte al paragrafo 2.5.

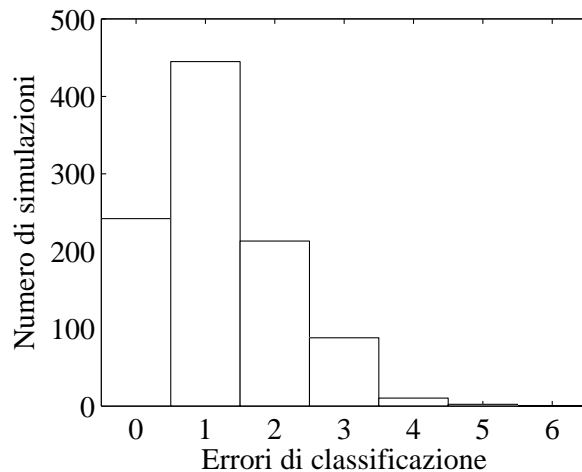
Si riportano in Tabella 5.1 le migliori prestazioni ottenute, in termini di errori medi assoluti di classificazione. Anche in questo caso è stato seguito un approccio Monte Carlo, caratterizzato da un numero di simulazioni pari a 1000, con 30 campioni in calibrazione e 6 in convalida; inoltre, il numero di componenti principali è stato scelto seguendo l'andamento degli errori di classificazione commessi, e minimizzandone il valore (Balabin *et al.*,

2010). Analogamente al Capitolo precedente, i pretrattamenti utilizzati sono quelli elencati al paragrafo 2.2.1.

Si noti come gli errori di classificazione vengano presentati con una cifra decimale poiché ottenuti dalla media delle 1000 simulazioni indipendenti; tuttavia, ad ulteriore chiarezza, si riporta in Figura 5.1 un istogramma relativo alla distribuzione degli errori di classificazione della metodica e del *dataset* ottimali.

**Tabella 5.1** Errori medi ottenuti nella classificazione dei campioni in base al sesso dei suini; errori assoluti sui 6 campioni di convalida, mediati tra le 1000 simulazioni indipendenti.

Dataset	Pretratt.	$k$ NN	LDA	QDA	PLSDA
Dati di filiera	-	2.6	1.9	2.0	3.9
NIR Foss carcassa	(3)	2.1	1.5	1.7	4.3
NIR Unity carcassa	(3)	2.4	1.6	1.4	4.2
NIR Foss speck magro	(5)	2.0	1.6	<b>1.2</b>	4.4
NIR Unity speck magro	(5)	2.3	1.8	1.3	4.3
NIR Foss speck intero	(5)	2.1	2.1	1.6	4.0
NIR Unity speck intero	(6)	1.8	1.9	1.3	4.3

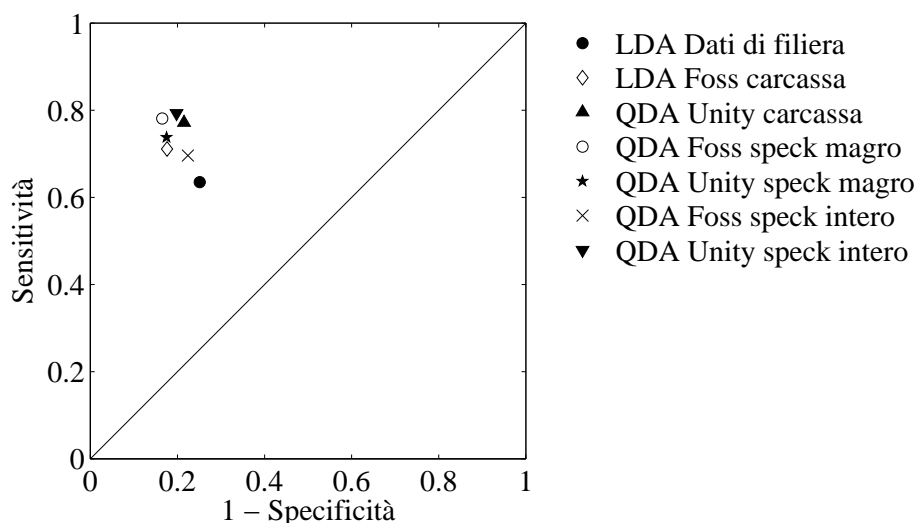


**Figura 5.1** Distribuzione degli errori di classificazione della variabile sesso, con metodo QDA calibrato sugli spettri NIR Foss della parte magra dello speck stagionato, pretrattamento (5).

Come si può notare dall'analisi di Tabella 5.1, i risultati di classificazione dei diversi modelli sono poco accurati utilizzando tutti i *dataset* a disposizione, con un margine di errore mai inferiore al 20%. Nel complesso, la metodica PLSDA presenta le prestazioni peggiori, seguita dall'algoritmo  $k$ NN; i metodi di classificazione lineare (LDA) e quadratico (QDA) sono invece caratterizzati da prestazioni leggermente migliori, seppur ancora insufficienti per un'efficiente suddivisione dei campioni. Tuttavia, oltre agli errori medi, l'efficacia di



un metodo di classificazione viene caratterizzata calcolandone anche la sensibilità e la specificità di attribuzione di ciascuna classe, rappresentate in modo efficace in Figura 5.2.



**Figura 5.2** Caratterizzazione dei modelli di classificazione in base ai valori di sensibilità e specificità.

Nel caso in esame, tutti i modelli ottimali di ogni *dataset* sono rappresentati da punti piuttosto distanti dall'estremo ideale, seppur scostati rispetto alla diagonale. La Figura 5.2 conferma inoltre i risultati di Tabella 5.1, poiché i modelli più prossimi al punto (1,0) sono anche quelli caratterizzati da un numero di errori medi inferiori.

L'insieme di questi risultati evidenzia la difficoltà di classificare in maniera precisa il sesso dei campioni, sia a partire dai *Dati di filiera* sia utilizzando le analisi spettroscopiche NIR. Queste conclusioni sono confermate anche utilizzando metodi più complessi, quali SIMCA e PLSDA con relazione interna quadratica, i cui risultati sono analoghi a quanto già esposto e non vengono quindi riportati per brevità.

Le considerazioni finali confermano inoltre quanto ci si attendeva da un punto di vista chimico-fisico, poiché si ritiene che il sesso dell'animale vivo non influenzi in modo significativo la qualità finale dello speck, soprattutto a causa della castrazione degli esemplari maschio. Tuttavia, è possibile che permangano delle differenze nella composizione e nella struttura della parte grassa, ma poiché non sono disponibili analisi dedicate su questa parte dello speck non è possibile condurre una classificazione mirata sul grasso, cui si rimanda in caso di successivi sviluppi futuri.

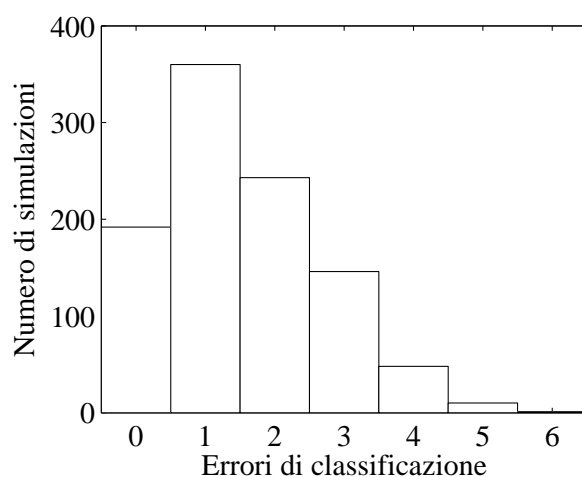
## 5.2 Regime alimentare tradizionale o arricchito

La classificazione della tipologia di alimentazione è stata condotta utilizzando i dati di filiera e le analisi NIR come variabili predittrici  $\mathbf{X}$ , e le tre variabili categoriali del regime

alimentare come spazio della risposta  $\mathbf{Y}$ . Si riportano in Tabella 5.2 i risultati di un'analisi analoga rispetto alla sezione precedente, condotta però sulle variabili relative al regime alimentare a cui i suini sono stati sottoposti. Si noti che in questo caso la classificazione è ternaria, poiché l'alimentazione può essere tradizionale o può essere arricchita con polifenoli o PUFA. Questo aumento del numero di classi non modifica la rappresentazione tabulare degli errori medi di classificazione, ma complica notevolmente la rappresentazione grafica di sensitività e specificità, poiché questi indici caratterizzano ogni singola classe e sarebbe quindi necessario utilizzare una figura tridimensionale. Per garantire maggiore chiarezza, si è scelto di riportare in Figura 5.4 tre grafici bidimensionali relativi a ciascuna coppia di classi; anche in questo caso viene rappresentata in Figura 5.3 la distribuzione degli errori di classificazione del metodo e del *dataset* ottimali.

**Tabella 5.2** Errori medi ottenuti nella classificazione dei campioni in base al regime alimentare dei suini.

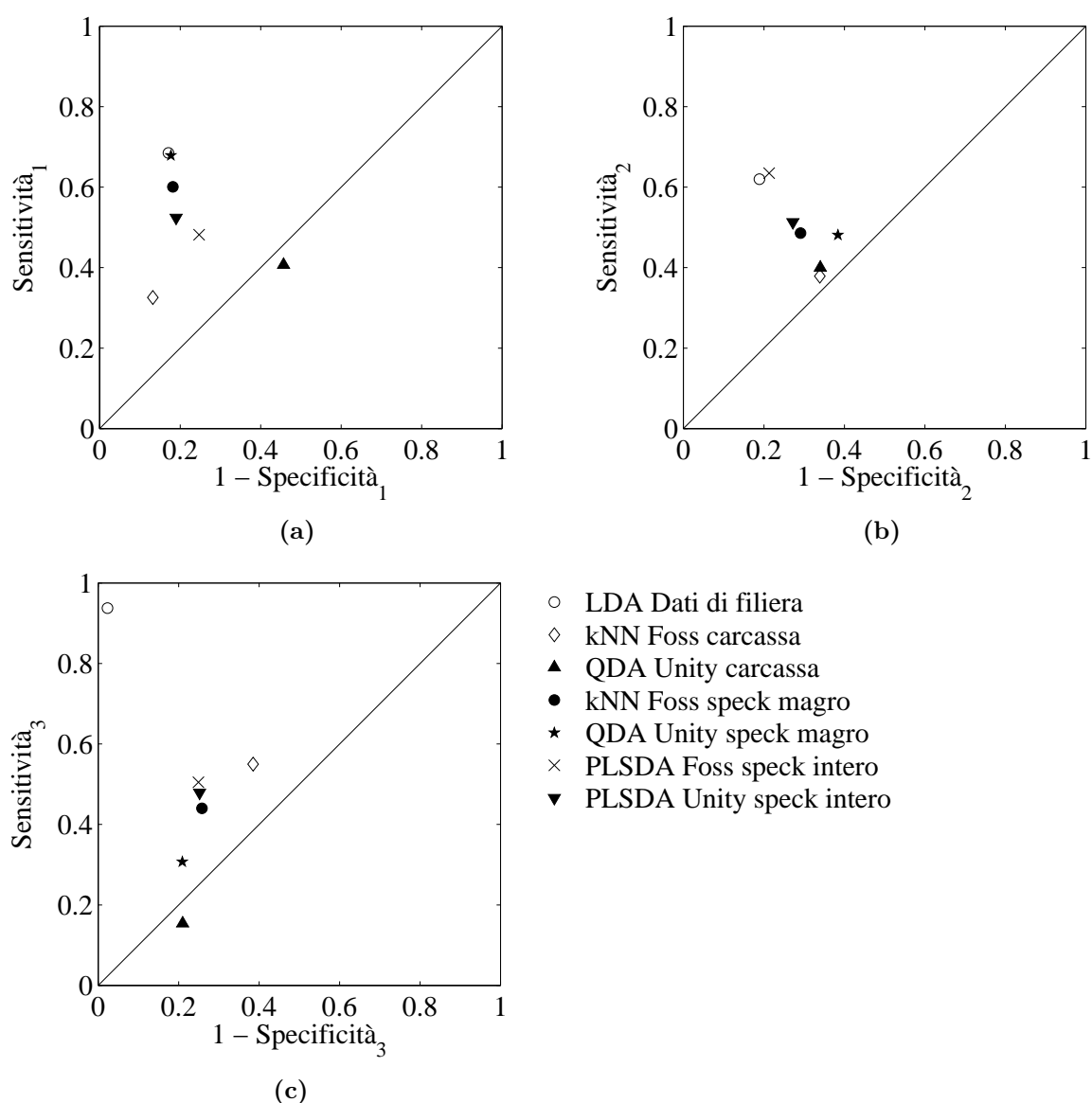
Dataset	Pretratt.	kNN	LDA	QDA	PLSDA
Dati di filiera	-	3.4	<b>1.5</b>	4.0	2.2
NIR Foss carcassa	(4)	3.6	3.6	3.6	4.0
NIR Unity carcassa	(4)	4.4	4.7	4.1	4.8
NIR Foss speck magro	(1)	3.1	3.8	3.4	3.4
NIR Unity speck magro	(4)	3.6	3.6	3.1	3.3
NIR Foss speck intero	(4)	4.2	3.5	4.1	2.9
NIR Unity speck intero	(3)	3.8	3.5	4.1	3.1



**Figura 5.3** Distribuzione degli errori di classificazione dei regimi alimentari, con metodo LDA calibrato sui Dati di filiera.

Ad eccezione dei *Dati di filiera*, i risultati ottenuti evidenziano una scarsa capacità di classificazione a partire dalle analisi spettroscopiche NIR, che conducono ad una percentuale di errori sempre superiore al 50%. Le prestazioni migliorano utilizzando i dati raccolti

lungo la filiera di lavorazione, a causa di una maggiore incidenza dei regimi alimentari nelle analisi cromatografiche del lardo fresco. Tuttavia, come evidente in Figura 5.3, ciò non limita la distribuzione degli errori di classificazione, che per alcune simulazioni sono addirittura 5 su 6. Per analizzare nel dettaglio le prestazioni dei diversi modelli, si riportano in Figura 5.4 le analisi sulla sensibilità e sulla specificità di attribuzione di ogni singola classe. L'interpretazione di questi andamenti è molto interessante perché consente di cogliere i punti di forza e di debolezza di ogni algoritmo: ad esempio, in Figura 5.4c si



**Figura 5.4** Caratterizzazione dei modelli di classificazione in base ai valori di sensibilità e specificità dei diversi regimi alimentari: (a) standard; (b) standard + polifenoli; (c) standard + PUFA.

evidenzia una classificazione pressochè perfetta dei campioni alimentati con l'arricchimento di PUFA a partire dai *Dati di filiera*. Questa ottima classificazione influenza positivamente i risultati complessivi del metodo LDA, che è in grado di classificare le altre due classi in

modo buono, ancorchè non ottimo. Inoltre, confrontando le Figure 5.4a e 5.4c si notano delle caratteristiche anomale del metodo PLSDA applicato alle analisi Unity dello speck macinato: infatti, le coordinate delle classi 1 e 3 appartengono al triangolo inferiore della Figura, e ciò suggerisce di invertire l'attribuzione delle due classi stesse per migliorare le prestazioni complessive del modello. Entrambe queste considerazioni possono essere discusse e giustificate richiamando il significato fisico delle variabili ed i risultati di Tabella 5.2, poiché:

- il regime alimentare arricchito con PUFA ha un contenuto superiore di semi di lino, molto ricchi di acidi grassi polisostituiti ed n3; questi ultimi, ingeriti in maggiore quantità, si depositano nei tessuti dei suini e sono quindi presenti in frazione più rilevante anche nel lardo di copertura fresco. Quindi, l'ottima classificazione di questo regime alimentare, specifica del modello calibrato sui *Dati di filiera*, è dovuta alle correlazioni molto significative con gli acidi grassi polisostituiti rilevati dall'analisi cromatografica, che costituiscono un contenuto informativo assente da tutti gli altri *dataset* utilizzati. Si noti che la significativa correlazione tra questa categoria di acidi grassi ed il regime alimentare arricchito con PUFA era già stata evidenziata nel corso dell'analisi esplorativa iniziale, in particolare in Tabella 3.6;
- sebbene i criteri di attribuzione delle classi 1 e 3 del metodo PLSDA (simbolo ▲) non siano ottimali, l'analisi di Figura 5.4 evidenzia come le coordinate di tutte le classi siano molto prossime alla diagonale, caratteristica di classificazioni casuali. Anche invertendo le classi 1 e 3, la robustezza del modello rimarrebbe comunque molto limitata, e quindi le discrete prestazioni riportate in Tabella 5.2 sono da attribuirsi prevalentemente al caso piuttosto che ad una effettiva capacità di classificazione. A titolo di confronto, si consideri l'algoritmo  $k$ NN applicato agli spettri Foss dello speck magro (rappresentato in Figura 5.4 con un pallino ●): tale modello è caratterizzato dagli stessi errori di classificazione medi, ma le sue coordinate sono molto più scostate dalla diagonale rispetto al caso precedente. Ciò conferma da un lato la casualità di classificazione del metodo PLSDA, mentre dall'altro chiarisce la necessità di non limitarsi all'interpretazione dei soli errori medi, ma di approfondire l'analisi del modello calcolando anche i valori di sensibilità e specificità.

Nel complesso, la classificazione dei campioni rispetto ai diversi regimi alimentari risulta molto poco accurata, con errori medi decisamente elevati. Tuttavia, anche queste modeste prestazioni trovano riscontro nei risultati attesi basandosi sul significato fisico delle variabili, poiché le variazioni di composizione dei tre regimi alimentari sono considerate troppo limitate per poter causare differenze macroscopiche dal punto di vista chimico ed organolettico. In questo senso, quindi, è giustificabile un'incapacità di classificazione, poiché la

modifica dei parametri nutrizionali causa effetti nel comportamento dell'animale vivo<sup>1</sup> e nella composizione del lardo di copertura fresco, ma non è tale da migliorare o alterare i requisiti di qualità dello speck stagionato.

### 5.3 Parametri sensoriali

Nel corso di questa sezione si presentano delle tecniche oggettive per la classificazione dei campioni di speck sulla base dei parametri sensoriali di Tabella 1.6, legati principalmente al colore, alla consistenza della carne ed al rapporto magro/grasso. Poichè i dati sensoriali a disposizione sono affetti da una notevole soggettività di giudizio, che si attesta in media sul 25% ma che per alcuni campioni arriva al 50%, si è scelto di slegare la trattazione dai giudizi degli esperti, che non vengono utilizzati nemmeno come termine di confronto. La classificazione è basata unicamente sulle immagini digitali ad alta risoluzione e sulle analisi di tessitura Instron, che costituiscono le tipologie di analisi ottimali per ottenere un giudizio oggettivo.

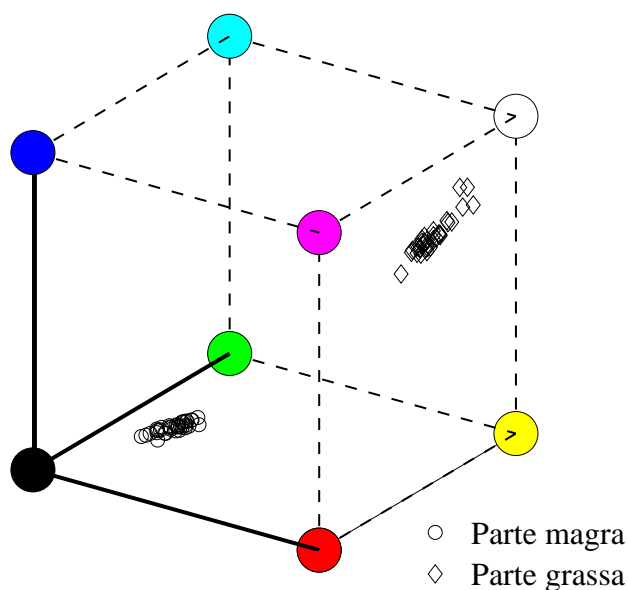
#### 5.3.1 Giudizi sul colore

Lo strumento più rapido e comune per classificare in modo rigoroso il colore dello speck è costituito da un'immagine digitale RGB di una sua fetta. Le immagini a disposizione sono state sottoposte ai pretrattamenti di equalizzazione e rimozione del riferimento per il bianco; quindi, grazie a maschere dedicate, vengono selezionate la parte magra e quella grassa, segmentate ulteriormente per ottenere un rettangolo di area massima (si rimanda a tal proposito alla procedura di Figura 2.13). Si calcola il colore medio di entrambe le parti di ogni campione, e lo si proietta nello spazio tridimensionale RGB, come rappresentato in Figura 5.5.

I tre assi principali corrispondono alle intensità dei tre canali RGB, rispettivamente del rosso, del verde e del blu; gli altri vertici del cubo sono costituiti dalle combinazioni di coppie o terne di colori primari, che costituiscono i colori secondari, il bianco ed il nero. Nel complesso, ogni possibile colorazione trova una collocazione precisa in un questo grafico tridimensionale, e può quindi essere analizzata in modo oggettivo. Analizzando la distribuzione della parte magra e grassa di ogni campione è immediato notare una collocazione molto diversa, prossima al nero/rosso per il magro ed al bianco per il grasso. Inoltre, si nota una certa variabilità interna tra i campioni delle medesima regione, e ci si propone di analizzare proprio questi scostamenti locali per caratterizzare il colore della fetta di speck.

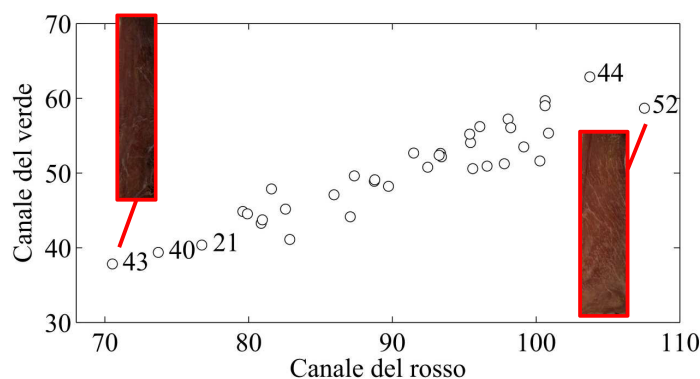
---

<sup>1</sup>È stato infatti registrato un aumento di aggressività negli esemplari alimentati con un arricchimento di PUFA, tale da indurre una riduzione dell'arricchimento stesso.



**Figura 5.5** Collocazione della parte magra e grassa di ogni campione nello spazio tridimensionale RGB.

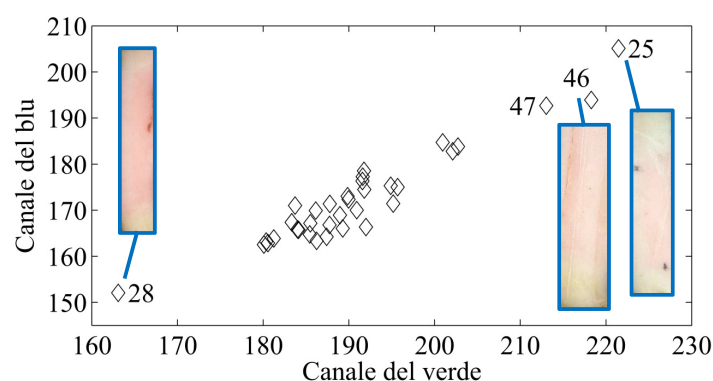
Per quanto concerne la parte magra, la massima variabilità si concentra nei canali del rosso e del verde; quindi, al fine di semplificare la visualizzazione, si è scelto di limitarsi ad un grafico bidimensionale su questi due soli canali. Come evidente in Figura 5.6, la distribuzione delle intensità luminose del rosso e del verde è piuttosto ampia, e consente di classificare in modo accurato il colore complessivo della parte magra. Infatti, i campioni più scuri, come il n° 43, sono caratterizzati da valori inferiori sia per il canale del rosso che del verde; viceversa, i campioni più luminosi come il n° 52 si collocano all'estremo opposto, a valori di intensità luminosa molto più elevati.



**Figura 5.6** Distribuzione dell'intensità luminosa della parte magra, relativa ai canali del rosso e del verde.

La procedura appena descritta viene applicata in modo analogo anche per la parte grassa, con l'unica modifica relativa alla scelta dei canali da analizzare: in questo caso, infatti, la massima variabilità si concentra nelle intensità luminose del verde e del blu, come riportato

in Figura 5.7.



**Figura 5.7** Distribuzione dell'intensità luminosa della parte grassa, relativa ai canali del verde e del blu.

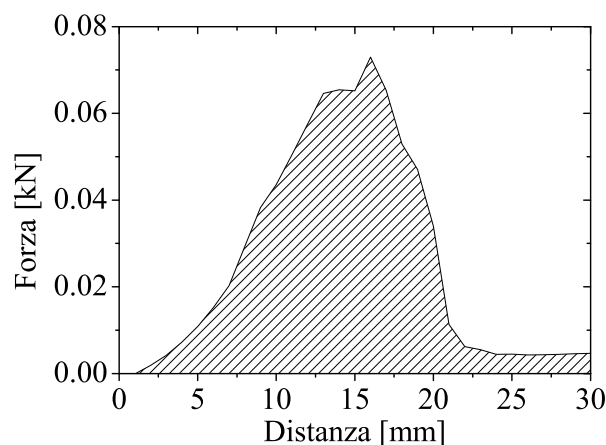
La classificazione risulta molto buona anche per la parte grassa: i campioni con le intensità luminose più elevate, n° 25 e 46, sono caratterizzati da colorazioni molto pallide, prossime al bianco; all'estremo opposto, invece, il grasso appare molto più rosato, con tinte decisamente più cariche.

Nel complesso, quindi, è sufficiente una semplice fotocamera digitale ed alcune tecniche di segmentazione per riuscire a classificare in modo rapido ed efficace il colore medio di entrambe le porzioni della fetta di speck. A parità di fotocamera e di condizioni di illuminazione del campione, infatti, il metodo risulta oggettivo e non affetto da considerazioni appartenenti alla sfera di discrezionalità del giudice.

### 5.3.2 Giudizi sulla consistenza

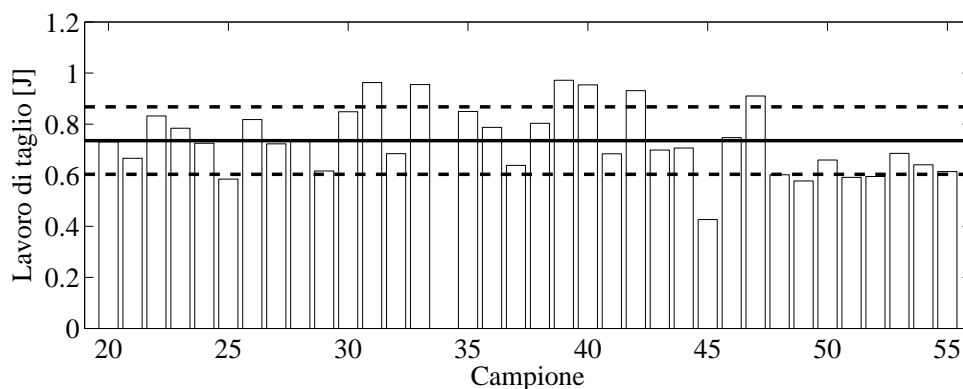
Per caratterizzare la consistenza di entrambe le regioni della fetta di speck sono state utilizzate le analisi di tessitura Instron, che rappresentano uno strumento dedicato per la stima di questa proprietà. Al fine di ottenere un indice sintetico per ciascun campione, è stata calcolata l'area sottesa alla curva di Figura 5.8, rappresentativa del lavoro di taglio. L'interpretazione del significato fisico dell'area evidenziata è immediata, poiché quanto maggiore è il lavoro di taglio tanto più elevata è la consistenza del prodotto.

Si ricorda che sono disponibili due distinte serie di analisi Instron, specifiche della parte magra e di quella grassa; dopo averle sottoposte ai pretrattamenti discussi alla sezione 1.5, è stata calcolata l'area sottesa a ciascuna curva di analisi, caratteristica di ogni campione. Si riportano i risultati ottenuti nei grafici a barre di Figura 5.9 e 5.10, relativi rispettivamente al magro ed al grasso. Si noti che per alcuni campioni le analisi sono state interrotte prima del punto finale, a causa di un'interferenza locale di nervature, frammenti di cartilagine o altri setti estranei rispetto al tessuto da analizzare; al fine di non alterare i risultati,

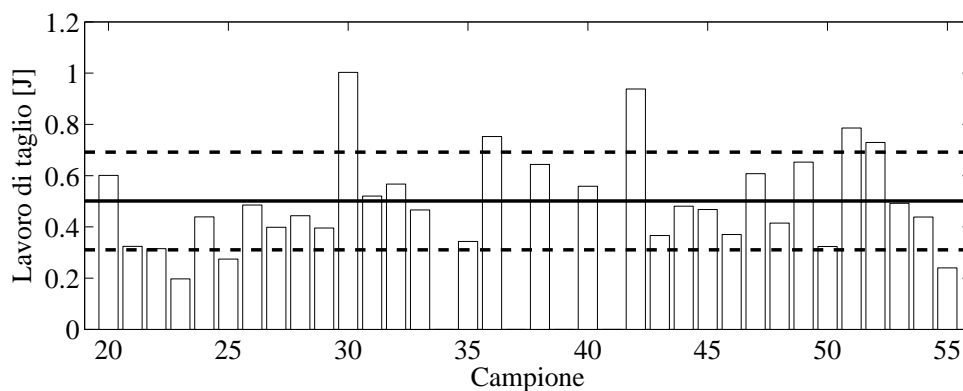


**Figura 5.8** Profilo medio delle analisi Instron e relativa area sottesa.

questi campioni sono stati rimossi, e non vengono quindi rappresentati nei due grafici a seguire. Si riportano inoltre la media delle misure, con un tratto continuo, ed in linea tratteggiata l'intervallo  $\pm\sigma$ , in cui  $\sigma$  rappresenta la deviazione standard tra i campioni.



**Figura 5.9** Distribuzione della consistenza della parte magra dello speck, ottenuta mediante analisi Instron. Si riporta la media delle misure in grassetto e l'intervallo  $\pm\sigma$  in linea tratteggiata.



**Figura 5.10** Distribuzione della consistenza della parte grassa dello speck, ottenuta mediante analisi Instron.



L'analisi delle Figure 5.9 e 5.10 evidenzia una maggiore variabilità nella consistenza della parte grassa, caratterizzata da un valore di  $\sigma$  più elevato e quindi da una distribuzione dei lavori di taglio molto più ampia. I sottocampioni della parte magra risultano invece più omogenei, con scostamenti limitati dal valore medio.

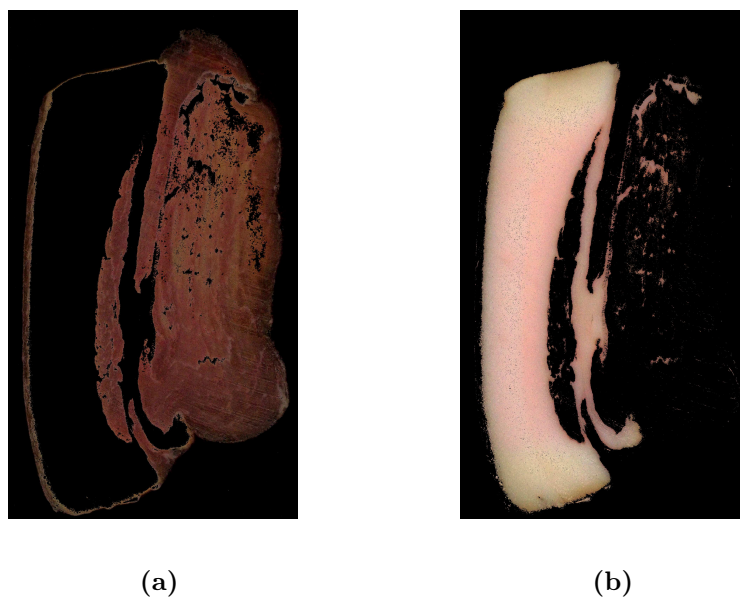
Nel complesso, l'utilizzo congiunto di queste due Figure consente di caratterizzare in modo rapido ed efficace la consistenza dello speck, sia per quanto concerne la parte magra che per quella grassa. Tuttavia, l'unico limite di questa tecnica risiede nella notevole specificità dell'analisi Instron: vengono infatti misurate proprietà estremamente localizzate, che possono variare notevolmente all'interno del prodotto. Anche la standardizzazione del punto di campionamento e di misura non garantisce una maggiore riproducibilità, perché i processi di stagionatura ed affumicatura condizionano comunque lo speck in modo diverso da campione a campione. D'altro canto, però, sono stati analizzati più sottocampioni da ogni speck intero, e le singole analisi sono state mediate per ottenere un andamento medio, in modo da evitare una rappresentatività troppo modesta. La validità di questa procedura è ancora da verificare; tuttavia, in caso di esito positivo, sarebbe sufficiente assegnare ad ogni giudizio di consistenza uno specifico lavoro di taglio per ottenere una classificazione rapida ed oggettiva, basata su principi analitici riproducibili.

### 5.3.3 Giudizi sul rapporto magro/grasso

La determinazione del rapporto magro/grasso di una fetta di speck è un parametro normalmente determinato mediante ispezione visuale, ma molto spesso la colorazione simile della parte magra e della cotenna esterna può indurre errori di valutazione. Inoltre, la presenza di grasso infiltrato può non essere rilevata da un'analisi rapida sul colore d'insieme, mentre viene facilmente colta da una tecnologia automatica di analisi d'immagine. Per stimare anche questo parametro in modo oggettivo è stato quindi utilizzato il *dataset* di immagini digitali, usando delle maschere specifiche per isolare le singole zone di interesse. Si riporta in Figura 5.11 un esempio di applicazione di due diverse maschere, una sulla parte magra e la cotenna, l'altra sulla parte grassa.

In Figura 5.11a è evidente la medesima colorazione del muscolo magro e della cotenna esterna, che deve tuttavia essere rimossa per ottenere un calcolo corretto del rapporto magro/grasso. Si riporta in Figura 5.12 l'immagine ottenuta dopo aver applicato un algoritmo di segmentazione automatica, basato sulla percentuale di *pixel* con intensità luminosa diversa da zero.

Come evidente, viene rimossa quasi interamente la cotenna, ad eccezione di due minime zone molto prossime alla parte magra; queste ultime, tuttavia, non vengono eliminate poiché tra di esse è presente una sensibile porzione di tessuto magro, che deve essere quantificata



**Figura 5.11** Applicazione di maschere per la selezione della: (a) parte magra e cotenna; (b) parte grassa.



**Figura 5.12** Eliminazione della cotenna dalla parte magra.

ai fini del calcolo del rapporto magro/grasso.

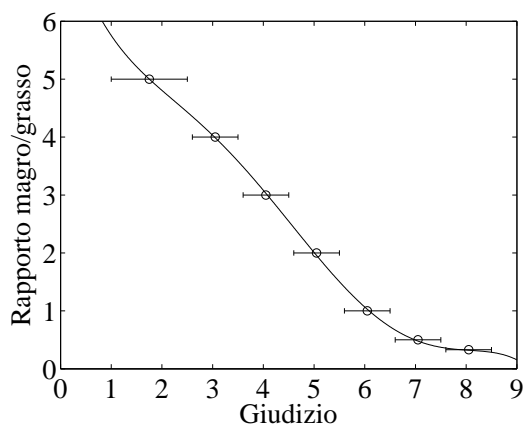
A questo punto, ottenute le due regioni di interesse, Figura 5.12 e 5.11b, è sufficiente calcolarne l'area per stimare in modo oggettivo il rapporto magro/grasso. Rimane tuttavia da gestire un'ultima questione, poiché la classificazione dei campioni non è basata sull'effettivo rapporto delle aree, ma su un intervallo di giudizio riferito a dei rapporti predeterminati, come riportato in Tabella 5.3.

Si è scelto di rappresentare i valori medi di ogni intervallo rispetto ai relativi rapporti magro/grasso; quindi, come riportato in Figura 5.13, è stata utilizzata un polinomio di quinto grado per fittare l'andamento ottenuto e ricavare un'espressione analitica della loro relazione. Si noti che sebbene l'utilizzo di un'equazione cubica sia sufficiente per ottenere

**Tabella 5.3** *Elenco della base di attribuzione dei giudizi numerici, riferiti a precisi rapporti magro/grasso.*

Magro/grasso	Intervallo di giudizio
5:1	1.0 ÷ 2.5
4:1	2.6 ÷ 3.5
3:1	3.6 ÷ 4.5
2:1	4.6 ÷ 5.5
1:1	5.6 ÷ 6.5
1:2	6.6 ÷ 7.5
1:3	7.6 ÷ 8.5

un coefficiente di determinazione  $R^2$  molto elevato, pari al 99.97%, ciò non assicura un andamento monotono della relazione. Per polinomi di ordine inferiore a cinque, infatti, si evidenzia un punto di minimo assoluto di coordinate prossime a (8,0.4), che determina una non biunivocità della relazione per rapporti magro/grasso inferiori a 0.4.

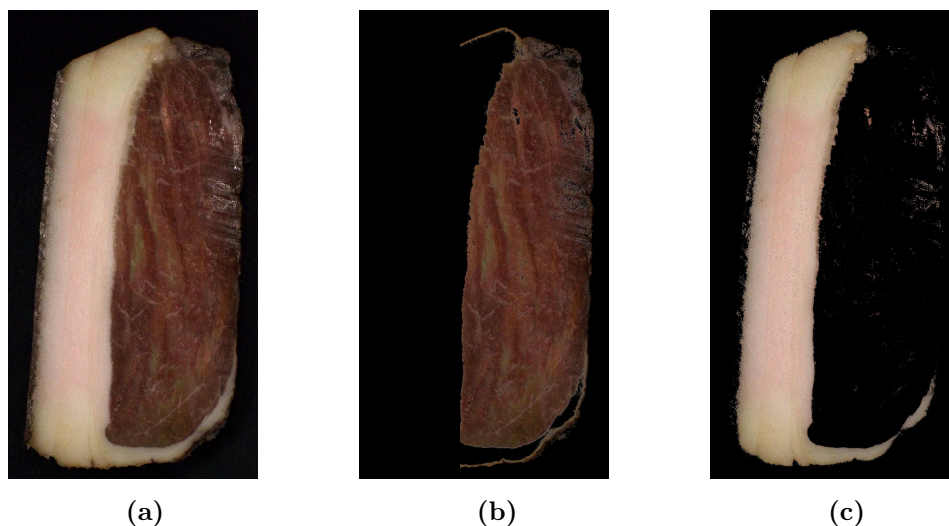


**Figura 5.13** *Fitting per la determinazione della relazione analitica tra il rapporto magro/grasso ed il rispettivo giudizio numerico.*

Con questa procedura si è ora in grado di quantificare in modo rapido ed accurato il rapporto magro/grasso. Molto spesso la stima di questo parametro è considerata banale perché evidente a livello macroscopico; tuttavia, analizzando le due diverse maschere di Figura 5.12 le possibili fonti di errore appaiono molto chiare, poiché il contributo della cotenna e del grasso infiltrato può essere facilmente incluso nella frazione della parte magra. La disponibilità di un metodo rapido ed oggettivo può quindi rappresentare uno strumento di supporto estremamente utile, vista anche l'estrema diffusione dello strumento di acquisizione richiesto. A titolo di esempio, si riporta nel Paragrafo successivo un confronto tra le tecnica proposta e la valutazione visiva del giudice.

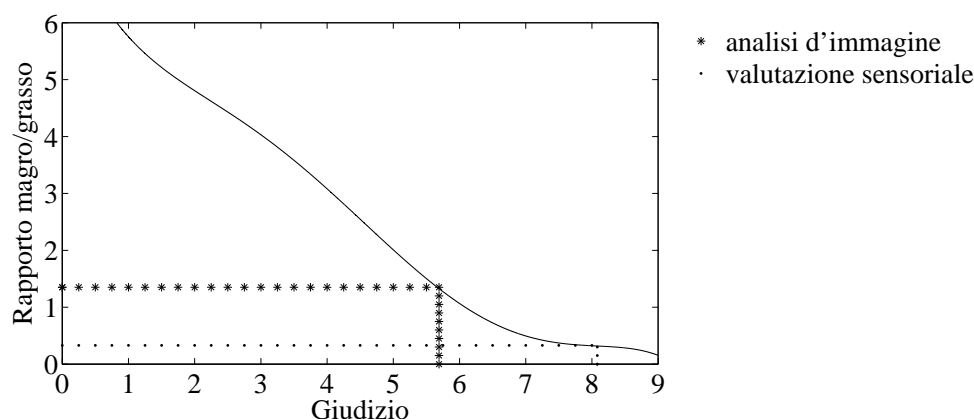
### 5.3.3.1 Confronto delle valutazioni assegnate

Si consideri a titolo di esempio il campione numero 44, di cui si riporta un'immagine digitale in Figura 5.14a; le maschere specifiche sulla parte magra e quella grassa sono invece rappresentate in Figura 5.14b e 5.14c.



**Figura 5.14** Stima del rapporto magro/grasso del campione n° 44: (a) immagine originale; (b) maschera sulla parte magra; (c) maschera sulla parte grassa.

Calcolando le aree di ciascuna regione si ottiene un rapporto magro/grasso pari a 1.35, che sostituito nella relazione analitica rappresentata in Figura 5.15 determina un giudizio finale pari a 5.69. La valutazione data dal giudice è invece di 8.08, e ciò indica una significativa sovrastima della frazione di grasso.



**Figura 5.15** Scostamento di giudizio tra l'analisi d'immagine, oggettiva, e la valutazione sensoriale del giudice.

Risulta quindi evidente la migliore accuratezza della tecnica di analisi d'immagine proposta, basata su criteri oggettivi e ripetibili che consentono di evitare banali errori di

valutazione, dovuti principalmente alla sovrastima di una delle due regioni della fetta di speck.

## 5.4 Conclusioni

I risultati di questo Capitolo evidenziano l'impossibilità di condurre una classificazione accurata sia rispetto al sesso che al regime alimentare, a partire da tutti i *dataset* a disposizione. Ciò rappresenta un risultato di rilievo poiché:

- conferma l'efficacia della castrazione degli esemplari maschio, dal momento che l'equilibratura dell'attività ormonale consente di ottenere uno speck stagionato con proprietà organolettiche indifferenti rispetto al sesso;
- conferma le attese sull'effetto dell'arricchimento alimentare, che determina delle variazioni nelle proprietà chimico-fisiche molto limitate e poco distinguibili.

L'utilizzo di tecniche oggettive per la caratterizzazione dei parametri sensoriali consente inoltre di ottenere un valido strumento di supporto per la classificazione dei campioni. Queste tecniche si dimostrano rapide, accurate e soprattutto riproducibili, evitando qualsiasi criterio soggettivo per l'attribuzione di un giudizio di qualità.



# Conclusioni e sviluppi futuri

In questa Tesi è stata valutata la possibilità di monitorare la qualità dello speck lungo i principali stadi della filiera di lavorazione, per garantire il rispetto dei requisiti richiesti dalla certificazione IGP in ogni fase del processo produttivo. È stata condotta un'analisi esplorativa molto dettagliata, al fine di approfondire la conoscenza del processo ed evidenziare le correlazioni più significative tra le proprietà del suino vivo, della carcassa macellata e dello speck stagionato. Alcune di queste relazioni hanno confermato legami già noti, altre invece necessitano di ulteriori approfondimenti perché sono strettamente connesse ai processi metabolici, scarsamente modellati fino ad ora. I risultati di questa analisi correlativa sono stati molto utili nella successiva fase di stima, in cui si è cercato di predire gli indicatori di qualità dello speck evitando le analisi chimiche dirette, a favore di altre acquisizioni più rapide ed economiche quali le analisi spettroscopiche NIR e le immagini digitali. In particolare, gli spettri NIR si sono dimostrati particolarmente efficaci nella stima di numerose variabili di qualità dello speck (attività dell'acqua, umidità, ceneri, NaCl, proteina), poiché caratterizzati da un'ottima accuratezza, ripetibilità e rapidità di analisi. Necessita invece di ulteriori approfondimenti l'utilizzo di uno strumento NIR portatile, per il quale si raccomanda l'uso di un intervallo spettrale più esteso. Si rimanda a sviluppi futuri anche il giudizio sulle immagini digitali, che in questa Tesi si sono dimostrate idonee solo per stime di carattere qualitativo. L'impiego di strumenti di acquisizione più sofisticati, dotati soprattutto di un miglior bilanciamento del bianco, potrebbe infatti influenzare in modo significativo l'accuratezza delle stime ottenute.

I risultati di classificazione denotano invece una limitata tracciabilità del sesso e dei diversi regimi alimentari, a riprova dell'efficacia della castrazione sugli esemplari maschio e della modesta influenza dei mangimi arricchiti, caratterizzati da variazioni troppo limitate per produrre effetti tangibili. La valutazione dei parametri sensoriali a partire dalle immagini digitali si è dimostrata rapida ed efficace, soprattutto per quel che concerne il colore dello speck ed il rapporto magro/grasso. Si lascia invece a sviluppi futuri l'applicazione e la convalida delle analisi Instron per il giudizio della consistenza del prodotto, poiché è necessario valutare se la media di più analisi locali possa costituire una buona descrizione della proprietà all'interno dello speck.

Nel complesso, le tecniche di statistica multivariata utilizzate in questa Tesi hanno consentito di sviluppare degli strumenti di supporto molto utili per la caratterizzazione e la classificazione della qualità dello speck, rispondendo ad una precisa esigenza del mondo industriale. In particolare, le analisi NIR garantiscono una stima accurata di otto su dieci variabili qualità dello speck stagionato, mentre le immagini digitali costituiscono un mezzo rapido ed ampiamente diffuso per classificare in modo oggettivo la qualità del prodotto.





# Appendice A

## Lista dei simboli

<b>A, B</b>	matrici dei <i>canonical loading</i>
<b>C</b>	matrice di decomposizione di <b>X</b> , di rango 1
<b>D</b>	matrice della distanza multidimensionale tra <i>loading</i>
<b>E</b>	matrice dei residui della PCA
<b>F, G</b>	matrici dei residui della PLS
<b>I</b>	matrice identità
<b>P, Q</b>	matrici di <i>loading</i>
<b>R<sub>c</sub></b>	matrice dei coefficienti canonici di correlazione
<b>S</b>	matrice di covarianza
<b>T, U</b>	matrici di <i>score</i>
<b>X</b>	matrice di dati originali
<b><math>\hat{X}</math></b>	matrice delle proiezioni dei dati originali sul modello PCA
<b>Y</b>	matrice delle variabili risposta
<b>Z, <math>\Omega</math></b>	matrici utilizzate nell'analisi delle correlazioni canoniche
<b><math>\Lambda, \Gamma</math></b>	matrici delle proiezioni dei campioni nello spazio delle variabili canoniche
<b><math>\mathbf{b}_h</math></b>	vettore della relazione interna, specifico di ogni variabile latente $h$
<b><math>\mathbf{j}_z, \mathbf{j}_\omega</math></b>	indicatori sintetici dell'analisi CCA sulle matrici <b>Z</b> e <b><math>\Omega</math></b>
<b><math>(\mathbf{h}^z)^2</math></b>	sommatoria delle varianze spiegate di ogni variabile $\mathbf{z}_\phi$ in tutte le FC
<b>p, q</b>	vettori di <i>loading</i>
<b>r<sub>c</sub></b>	vettore dei coefficienti canonici di correlazione
<b><math>\mathbf{r}_s^z, \mathbf{r}_s^\omega</math></b>	vettori dei coefficienti di struttura di <b>Z</b> e <b><math>\Omega</math></b>
<b>t, u</b>	vettori di <i>score</i>
<b>w</b>	vettore di <i>weights</i>
<b><math>\bar{x}_m</math></b>	vettore medio della variabile $\mathbf{x}_m$
<b><math>\mathbf{y}_b</math></b>	vettore della $b$ -esima variabile risposta
<b><math>\hat{\mathbf{y}}_b</math></b>	vettore delle predizioni del modello PLS della variabile $\mathbf{y}_b$
<b>r</b>	rango della matrice <b>X</b>

$\rho$	coefficiente di correlazione
$\sigma$	deviazione standard
$B$	numero di variabili risposta
$D$	numero delle componenti principali scelte nell'analisi PCA
$H$	numero delle variabili latenti scelte nell'analisi PLS
$N$	numero di campioni della matrice $\mathbf{X}$
$M$	numero di variabili della matrice $\mathbf{X}$
$R^2$	coefficiente di determinazione
$\Phi$	numero di variabili della matrice $\mathbf{Z}$
$\Delta$	numero di variabili della matrice $\mathbf{\Omega}$

# Appendice B

## Lista degli acronimi

<i>CCA</i>	analisi delle correlazioni canoniche
<i>CV</i>	coefficiente di variazione
<i>FC</i>	funzione canonica
<i>LV</i>	variabile latente
<i>MIA</i>	analisi d'immagine multivariata
<i>NIR</i>	spettroscopia al vicino infrarosso
<i>PC</i>	componente principale
<i>PCA</i>	analisi delle componenti principali
<i>PCR</i>	regressione sulle componenti principali
<i>PLS</i>	proiezione su strutture latenti
<i>RMSE</i>	<i>Root Mean Square Error</i> radice quadrata dell'errore quadratico medio
<i>RMSEP</i>	<i>Root Mean Square Error of Prediction</i>
<i>RPD</i>	<i>Ratio of Prediction to Deviation</i>
<i>SNV</i>	<i>Standard Normal Variate</i>
<i>VIP</i>	<i>Variable Importance in Predictions</i>



# Riferimenti bibliografici

- Addison, P.S. (2002). *The illustrated wavelet transform handbook*. London (UK): Institute of Physics Publishing.
- Andersen, C.M., Bro, R. (2010). Variable selection in regression - a tutorial. *J. Chemom.*, **24**, 728–737.
- Andersen, J.R., Borggaard, C., Rasmussen, A.J., Houmoller, L.P. (1999). Optical measurements of pH in meat. *Meat Sci.*, **53**, 135–141.
- Balabin, R.M., Safieva, R.Z., Lomakina, E.I. (2010). Gasoline classification using near infrared (NIR) spectroscopy data: comparison of multivariate techniques. *Anal. Chim. Acta*, **671**, 27–35.
- Barnes, R.J., Dhanoa, M.S., Lister, S.J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *App. Spectrosc.*, **43**, 727–777.
- Bharati, M.H., Liu, J.J., MacGregor, J.F. (2004). Image texture analysis: methods and comparisons. *Chemom. Intell. Lab. Syst.*, **72**, 57–71.
- Burnham, A.J., Viveros, R., MacGregor, J.F. (1996). Frameworks for latent variable multivariate regression. *J. Chemom.*, **10**, 31–45.
- Burns, D.A., Ciurczak, E.W. (2008). *Handbook of near-infrared analysis, 3rd edition*. Vol. 35. Boca Raton, Florida (U.S.A.): CRC Press, Practical spectroscopy series.
- Cajo, J., Ter Braak, F. (1990). Interpreting canonical correlation analysis through biplots of structure correlations and weights. *Psychometrika*, **55**, 3, 519–531.
- CIELAB, Standard (1976). Method of measuring and specifying color rendering properties of light sources. *Publication CIE n° 13-2 (T.C. 3.2)*,
- Collell, C., Gou, P., Arnau, J., Comaposada, J. (2011). Non-destructive estimation of moisture, water activity and NaCl at ham surface during resting and drying using NIR spectroscopy. *Food Chem.*, **129**, 601–607.
- Courville, T., Thompson, B. (2001). Use of structure coefficients in published multiple regression articles:  $\beta$  is not enough. *Educ. Psychol. Meas.*, **61**, 229–248.
- Cozzolino, D., Murray, I., Chree, A., Scaife, J.R. (2005). Multivariate determination of free fatty acids and moisture in fish oils by partial least-squares regression and near-infrared spectroscopy. *Food Sci. Techn.*, **38**, 821–828.
- Davies, A.M.C., Fearn, T. (2006). Back to basics: calibration statistics. *Spectroscopy Europe*, **18**, 31–32.
- Decreto del Ministero dell'Industria, del Commercio e dell'Artigianato del 18.12.1993 per i prosciutti crudi di Parma e San Daniele (1993).
- Fawcett, T. (2006). An introduction to ROC analysis. *Patter Recogn. Lett.*, **27**, 861–874.
- Fearn, T. (2003). Assessing calibrations: SEP, RPD, RER and R2. *NIR News*, **13**, 12–14.

- Friedman, J.H. (1988). Regularized discriminant analysis. *J. Am. Stat. Assoc.*, **84**, 165–175.
- Friedman, J.H., Bentely, J., Finkel, R. (1977). An algorithm for finding best matches in logarithmic expected time. *ACM T. Math. Software*, **3**, 209–226.
- Geladi, P., Esbensen, K. (1989). Can image analysis provide information useful in chemistry? *J. Chemom.*, **3**, 419–429.
- Geladi, P., Grahn, H. (1997). *Multivariate image analysis*. New York (U.S.A.): John Wiley & Sons.
- Geladi, P., Kowalski, B.R. (1986). Partial least-squares regression: a tutorial. *Anal. Chim. Acta*, **185**, 1–17.
- Gigieli, A., Butler, F., Hudson, B. (1989). Alternative methods of pig chilling. *Meat Sci.*, **26**, 67–83.
- Graffelman, J. (2005). Enriched biplots for canonical correlation analysis. *J. Appl. Stat.*, **32**, 2, 173–188.
- Guillen, M.D., Cabo, N. (1998). Relationships between the composition of edible oils and lard and the ratio of the absorbance of specific bands of their Fourier transform infrared spectra. Role of some bands of the fingerprint region. *J. Agric. Food Chem.*, **46**, 1788–1793.
- Jackson, J.E. (1991). *A user's guide to principal components*. New York (U.S.A.): John Wiley & Sons.
- Liu, J.J., MacGregor, J.F. (2007). On the extraction of spectral and spatial information from images. *Chemom. Intell. Lab. Syst.*, **85**, 119–130.
- Mardia, K.V., Kent, J.T., Bibby, J.M. (1979). *Multivariate analysis*. Waltham, Massachusetts (U.S.A.): Academic Press.
- Maribo, H., Olsen, E.V., Barton-Gade, P., Moller, A.J., Karlsson, A. (1998). Effect of early post-mortem cooling on temperature, pH fall and meat quality in pigs. *Meat Sci.*, **50**, 115–129.
- Morales, R., Serra, X., Guerrero, L., Gou, P. (2007). Softness in dry-cured porcine *biceps femoris* muscles in relation to meat quality characteristics and processing conditions. *Meat Sci.*, **77**, 662–669.
- Nimon, K., Henson, R.K., Gates, M.S. (2010). Revisiting interpretation of Canonical Correlation Analysis: a tutorial and demonstration of Canonical Commonality Analysis. *Multivar. Behav. Res.*, **45**, 702–724.
- Ottavian, M., Facco, P., Barolo, M., Berzaghi, P., Segato, S., Novelli, E., Balzan, S. (2012a). Near-infrared spectroscopy to assist authentication and labeling of Asiago d'allevo cheese. *J. Food Eng.*, **113**, 289–298.
- Ottavian, M., Facco, P., Fasolato, L., Novelli, E., Mirisola, M., Perini, M., Barolo, M. (2012b). Use of near-infrared spectroscopy for fast fraud detection in seafoods: application to the authentication of wild European sea bass (*Dicentrarchus labrax*). *J. Agr. Food Chem.*, **60**, 639–648.

- Ozaki, Y., McClure, W.F., Christy, A.A. (2007). *Near-infrared spectroscopy in food science and technology*. New York (U.S.A.): John Wiley & Sons.
- Pérez, N.F., Ferré, J., Boqué, R. (2009). Calculation of the reliability of classification in discriminant partial least-squares binary classification. *Chemom. Intell. Lab. Syst.*, **95**, 121–128.
- Prieto, N., Roehe, R., Lavin, P., Batten, G., Andrés, S. (2009). Application of near infrared reflectance spectroscopy to predict meat and meat products quality: a review. *Meat Sci.*, **83**, 175–186.
- Regolamento CEE 2073/05, sui criteri microbiologici applicabili ai prodotti alimentari (2005). *GU L 338 del 22.12.2005*.
- Regolamento CEE 2568/91, relativo alle caratteristiche degli oli d'oliva e degli oli di sansa d'oliva nonché ai metodi ad essi attinenti (1991). *GU 248 del 5.9.1991*, Allegato XVI.
- Regolamento CEE 3220/84, che determina la tabella comunitaria di classificazione delle carcasse di suino (1984). *GU L 301 del 20.11.1984*, aggiornato dal Regolamento CEE 3513/94.
- Rencher, A.C. (1992). Interpretation of canonical discriminant functions, canonical variates, and principal components. *The American Statistician*, **46**, 217–225.
- Russ, J.C. (1999). *The image processing handbook, 3rd edition*. Florida (U.S.A.): CRC Press.
- Savell, J.W., Mueller, S.L., Baird, B.E. (2005). The chilling of carcasses. *Meat Sci.*, **70**, 449–459.
- Sherry, A., Henson, R.K. (2005). Conducting and interpreting Canonical Correlation Analysis in personality research: a user-friendly primer. *J. Pers Assess.*, **84**, 1, 37–48.
- Tomba, E., Facco, P., Roso, M., Modesti, M., Bezzo, F., Barolo, M. (2010). Artificial vision system for the automatic measurement of inter-fiber pore characteristics and fiber diameter distribution in nanofiber assemblies. *Ind. Eng. Chem. Res.*, **49**, 2957–2968.
- Velasco, L., Becker, H.C. (1998). Estimating the fatty acid composition of the oil in intact-seed rapeseed (*Brassica napus* L.) by near-infrared reflectance spectroscopy. *Euphytica*, **101**, 221–230.
- Wilks, D.S. (2011). *Statistical Methods in the Atmospheric Sci.s, 3rd edition*. Vol. 100. Waltham, Massachusetts (U.S.A.): Academic Press. Cap. 13.
- Wold, S., Sjostrom, M. (1977). *Chemometrics: theory and application*. Vol. 52. New York (U.S.A.): American Chemical Society. Cap. 12.
- Wood, J.D., Richardson, R.I., Nute, G.R., Fisher, A.V., Campo, M.M., Kasapidou, E., Sheard, P.R., Enser, M. (2003). Effects of fatty acids on meat quality: a review. *Meat Sci.*, **66**, 21–32.
- Yu, H., MacGregor, J.F. (2003). Multivariate image analysis and regression for prediction of coating content and distribution in the production of snack foods. *Chemom. Intell. Lab. Syst.*, **67**, 125–144.





# Ringraziamenti

A conclusione di questi sette mesi di lavoro desidero ringraziare tutte le persone che mi hanno supportato e dato fiducia nello sviluppo di questa Tesi, ad iniziare dal Prof. Barolo e da tutti i ragazzi di CAPE-Lab, Matteo e Pierantonio in primis. È doveroso ricordare anche il prezioso contributo del Prof. Novelli, del Dipartimento di Medicina Animale, e del laboratorio NIRS di Agripolis.

Infine, dopo oltre cento pagine di considerazioni tecniche e settoriali, è bene che anche una persona taciturna e riservata come me dedichi un po' di spazio per qualche ringraziamento più personale e sentito, soprattutto perché fino ad ora credo di averlo fatto di rado. Grazie quindi alla mia famiglia, che mi ha sempre sostenuto e dato la libertà di sbagliare scegliendo da solo: tenermi a bada non è sempre stato facile, anzi (vero Mamma?), ma molte delle mie soddisfazioni sono nate e cresciute su tutte quelle possibilità che non mi avete mai negato.

Grazie anche ai miei amici, quelli dei viaggi e della pizza prima del posticipo, dei tanti calici alzati ma anche delle pacche sulla spalla quando non tutto girava per il verso giusto. La vostra presenza è stata un appoggio su cui ho sempre potuto contare.

*Last but not least*, dovrei ringraziare anche chi ha sempre supportato la mia testardaggine, mi ha raggiunto a Padova in settimana per regalarmi un sorriso, e mi è sempre stata accanto, a modo suo, ogni qualvolta ne avessi bisogno. Tuttavia, ho scritto *dovrei* perché nel nostro rapporto non c'è mai stato bisogno di tante parole, quindi anche queste rimarranno tra noi due..!