



SAPIENZA
UNIVERSITÀ DI ROMA

FACOLTÀ DI INGEGNERIA CIVILE ED INDUSTRIALE

Dipartimento di Ingegneria Chimica, Materiali e Ambiente

**TECNICHE INNOVATIVE ED APPLICAZIONI
SENSORISTICHE PER LA CARATTERIZZAZIONE DI
MATERIE PRIME**

Dottorato di Ricerca in Ingegneria Elettrica, dei Materiali e delle Nanotecnologie

Ingegneria dei Materiali e delle Materie Prime

Candidato:

Ing. Riccardo Gasbarrone

Supervisore:

Prof. Giuseppe Bonifazi

CICLO XXXII

Anno Accademico 2018-2019

"Il primo dovere di un uomo è la continua ricerca e il perseguimento della verità"

Marco Tullio Cicerone

SOMMARIO

La presente tesi di dottorato è il risultato di uno studio interdisciplinare, che ha avuto quale tema centrale di ricerca la messa a punto di sistemi innovativi e “*sensor-based*” per la caratterizzazione di materie prime, nonché per il controllo qualità di prodotti. Particolare attenzione è stata rivolta all’uso di procedure innovative e allo sviluppo di appropriati algoritmi di statistica multivariata, *machine learning* e *computer vision* a complemento dell’uso di tecniche di spettroscopia classica ed iperspettrale.

Le metodiche analitiche impiegate hanno riguardato l’uso della spettroscopia, quale tecnologia non distruttiva potenzialmente impiegabile nella selezione *on-line* (i.e. *sorting*). L’analisi di immagine iper-spetttrale (*HyperSpectral Imaging*: HSI) e la spettroscopia puntuale vengono confrontate tra loro e con tecniche di analisi tradizionali. L’uso di tecnologie iper-spetttrali negli intervalli dello spettro elettromagnetico che vanno dal visibile (Vis) all’infrarosso ad onde corte (SWIR), 400-2500 nm, in combinazione con altre tecniche d’indagine (i.e. analisi chimiche e analisi in micro-fluorescenza ai raggi X) consentono l’implementazione di una metodologia di caratterizzazione rapida, non distruttiva e non invasiva. Tali tecniche offrono la possibilità di individuare le procedure applicabili più consone per il controllo qualità e/o per la tipologia di trattamento da utilizzare per le materie prime ed i prodotti presi in esame. L’attività di ricerca presentata in questa tesi dimostra come sia possibile ottenere importanti informazioni circa la caratterizzazione di materie prime secondarie e prodotti agroalimentari, mediante tecniche spettroscopiche supportate da indagini chemiometriche.

Dai casi applicativi analizzati e da una revisione sistematica della letteratura, si è visto come le tecniche basate sull’HSI siano estremamente versatili e scalabili per applicazioni di controllo qualità e classificazione *on-line* (i.e. come tecnologia di sensing per macchine selezionatrici), sia per materie prime che per prodotti agro-alimentari. L’HSI può essere considerato uno strumento valido per caratterizzare e monitorare differenti tipologie di materiali. Si è visto, inoltre, come la spettroscopia Vis-SWIR, presenta numerosi vantaggi rispetto all’uso di tecniche analitiche tradizionali (i.e. costi inferiori, dimensioni ridotte della strumentazione, rapidità di analisi, possibilità di effettuare analisi non distruttive e non invasive anche in situ e possibilità di ottenimento di risultati multi-parametrici), soprattutto per quanto riguarda la valutazione di parametri qualitativi di prodotti agro-alimentari.

Le prospettive di sviluppo di questa attività di ricerca risiedono nell’ottimizzazione delle informazioni ottenibili dall’applicazione di tecniche spettroscopiche, nonché di correlare i risultati delle analisi distruttive e non distruttive con i parametri caratteristici del materiale analizzato.

INDICE

| | | |
|----------|---|-----------|
| 1 | Introduzione | 11 |
| 1.1 | <i>Contesto generale.....</i> | 11 |
| 1.2 | <i>Struttura della tesi.....</i> | 12 |
| 1.3 | <i>L'importanza delle materie prime e del controllo qualità nel settore del riciclo.....</i> | 12 |
| 1.4 | <i>L'importanza del controllo qualità nel settore agro-alimentare.....</i> | 15 |
| 1.5 | <i>Stato dell'arte delle tecniche di spettrografia ad immagine e spettrofotometria puntuale</i> | 17 |
| 2 | Teoria | 21 |
| 2.1 | <i>Cenni di spettroscopia.....</i> | 21 |
| 2.1.1 | <i>Spettrofotometria nel vicino infrarosso.....</i> | 22 |
| 2.1.2 | <i>Spettrofotometria XRF</i> | 25 |
| 2.2 | <i>HyperSpectral imaging.....</i> | 27 |
| 2.2.1 | <i>Modalità di acquisizione delle immagini iperspettrali</i> | 29 |
| 2.2.2 | <i>Modalità di analisi delle immagini iperspettrali.....</i> | 30 |
| 2.3 | <i>Tecniche di statistica multivariata e metodi chemiometrici.....</i> | 33 |
| 2.3.1 | <i>Metodi di pre-elaborazione di dati</i> | 34 |
| 2.3.2 | <i>Tecniche di analisi esplorativa e semplificazione dei dati.....</i> | 42 |
| 2.3.3 | <i>Tecniche di regressione per analisi quantitative</i> | 46 |
| 2.3.4 | <i>Tecniche di classificazione supervisionate per analisi qualitative</i> | 49 |
| 2.3.5 | <i>Calibrazione, convalida incrociata e validazione di un modello.....</i> | 53 |
| 2.3.6 | <i>Tecniche di selezione delle variabili significative.....</i> | 55 |
| 3 | Strumenti utilizzati | 57 |
| 3.1 | <i>Sistema iperspettrale SWIR.....</i> | 57 |
| 3.2 | <i>Sistema iperspettrale Vis-NIR.....</i> | 60 |
| 3.3 | <i>Spettrofotoradiometro portatile VIS-SWIR FieldsSpec® 4.....</i> | 64 |
| 3.4 | <i>Spettrometro portatile NIR MicroNIR™.....</i> | 67 |
| 3.5 | <i>Spettrometro Micro-XRF M4 TORNADO™</i> | 69 |
| 4 | Casi di studio..... | 71 |
| 4.1 | <i>Materie prime secondarie.....</i> | 72 |
| 4.1.1 | <i>Caratterizzazione di un flusso di RAEE risultante da comminuzione e separazione magnetica per densità</i> | 72 |
| 4.1.2 | <i>Identificazione a cascata di materiali provenienti da monitor a schermo piatto giunti a fine vita mediante HyperSpectral Imaging</i> | 80 |
| 4.1.3 | <i>Identificazione a cascata di materiali provenienti da rifiuti di imballaggi mediante HyperSpectral Imaging</i> | 92 |

Indice

| | | |
|----------|--|------------|
| 4.2 | <i>Prodotti agro-alimentari</i> | 107 |
| 4.2.1 | Valutazione qualitativa della piccantezza di campioni di peperoncino essiccato mediante spettrofotometria puntuale | 107 |
| 4.2.2 | Analisi qualitativa di pasta surgelata mediante spettrofotometria puntuale | 116 |
| 4.2.3 | Classificazione mediante <i>HyperSpectral Imaging</i> di nocciole buone e relativi scarti | 124 |
| 4.2.4 | Classificazione mediante <i>HyperSpectral Imaging</i> di pistacchi buoni e relativi scarti | 138 |
| 4.2.5 | Studio sulla germinabilità di semi mediante <i>HyperSpectral Imaging</i> nell'infrarosso ad onde corte | 151 |
| 4.2.6 | Valutazione del processo di maturazione dell'oliva itrana mediante l'utilizzo di un sistema iperspettrale da laboratorio ed uno spettroradiometro portatile..... | 161 |
| 4.2.7 | Identificazione di olive attaccate da mosca olearia mediante l'utilizzo di un sistema iperspettrale da laboratorio ed uno spettroradiometro portatile | 173 |
| 4.2.8 | Valutazione del processo di maturazione del kiwi giallo mediante l'utilizzo di spettrofotometri NIR portatili..... | 189 |
| 4.2.9 | Valutazione di parametri qualitativi del grano tenero mediante spettrofotometria NIR: prima e dopo la macinazione..... | 199 |
| 5 | Considerazioni finali e prospettive future | 212 |
| | Riferimenti bibliografici | 214 |
| 6 | Allegati | 229 |
| | <i>Allegato I - MATLAB script - Sisuchema_SWIR_import.m</i> | 229 |
| | <i>Allegato II - MATLAB script - fieldspec_import2.m</i> | 230 |
| | <i>Allegato III - MATLAB script - micronir_import.m</i> | 232 |
| | <i>Allegato IV - MATLAB script - importxrfmap.m</i> | 233 |
| | <i>Allegato V - MATLAB script - create_datasetXRF.m</i> | 237 |
| | <i>Allegato VI - Case study – Analisi micro-XRF di sottoprodotti MDS provenienti da RAEE</i> | 238 |
| | <i>Allegato VII- Case study - Riconoscimento della tipologia di polimero dei frammenti di Light Guide Panel provenienti da monitor a schermo piatto giunti a fine vita</i> | 242 |
| | <i>Allegato VIII- Case study – Packaging</i> | 245 |
| | <i>Allegato IX- Case study - Peperoncino</i> | 248 |
| | <i>Allegato X- Case study – Pasta surgelata</i> | 253 |
| | <i>Allegato XI- Case study – Nocciole</i> | 259 |
| | <i>Allegato XII- Case study – Pistacchi</i> | 272 |
| | <i>Allegato XIII- Case study - Sementi</i> | 284 |
| | <i>Allegato XIV- Case study – Olive: invaiatura</i> | 286 |
| | <i>Allegato XV- Case study – Olive: mosca olearia</i> | 289 |
| | <i>Allegato XVI- Case study – Kiwi: maturazione</i> | 293 |
| | <i>Allegato XVII- Case study – Grano e farine</i> | 296 |

Indice

| | |
|--|------------|
| Lista delle pubblicazioni | 302 |
| Ringraziamenti | 304 |

Elenco delle Abbreviazioni

| | | | |
|---------|---|--------|--|
| AEE | <i>Apparecchiature Elettriche ed Elettroniche</i> | OLS | <i>Ordinary Least Squares</i> |
| ANN | <i>Artificial Neural Network</i> | PC | <i>Policarbonato</i> |
| ASCI | <i>American Standard Code for Information Interchange</i> | PCB | <i>Printed circuits board (circuiti stampati)</i> |
| BLU | <i>Backlight Unit (unità di retroilluminazione)</i> | PE | <i>Polietilene</i> |
| C&D | <i>Rifiuti da Costruzione e Demolizione</i> | PLS | <i>Partial Least Squares</i> |
| CCD | <i>Charge-Coupled Device</i> | PLS-DA | <i>Partial Least Squares – Discriminant Analysis</i> |
| DM | <i>Dry Matter</i> | PMMA | <i>Polimetilmetacrilato</i> |
| DoE | <i>Design of Experiment</i> | PQN | <i>Probabilistic Quotient Normalization</i> |
| EMSC | <i>Extended Multiplicative Scatter Correction</i> | QDA | <i>Quadratic Discriminant Analysis</i> |
| EoL | <i>End of Life</i> | RAEE | <i>Rifiuti di Apparecchiature Elettriche ed Elettroniche</i> |
| FOW | <i>Field Of View</i> | RDA | <i>Regularized Discriminant Analysis</i> |
| GLS-W | <i>Generalized Least Squares - Weighting</i> | RGB | <i>Red Green Blue</i> |
| HSI | <i>HyperSpectral Imaging</i> | RMSE | <i>Root Mean Square Error of Calibration</i> |
| iPLS | <i>interval Partial Least Squares</i> | RMSEC | <i>Root Mean Square Error of Calibration</i> |
| iPLS-DA | <i>interval Partial Least Squares-Discriminant Analysis</i> | RMSECV | <i>Root Mean Square Error of Cross Validation</i> |
| IR | <i>InfraRed (Infrarosso)</i> | RMSEP | <i>Root Mean Square Error of Prediction</i> |
| K/S | <i>Kennard/Stone</i> | ROI | <i>Region of Interest</i> |
| k-NN | <i>k-Nearest Neighbours</i> | SC | <i>Splice Correction</i> |
| LCD | <i>Liquid Cristal Display</i> | S-G | <i>Savitzky-Golay</i> |
| LDA | <i>Linear Discriminant Analysis</i> | SIMCA | <i>Soft Independent Modelling of Class Analogy</i> |
| LGP | <i>Light Guide Panel</i> | SNV | <i>Standard Normal Variate</i> |
| LV | <i>Latent Variable (variabile latente)</i> | SR | <i>Selectivity Ratio</i> |
| MC | <i>Mean Center</i> | SSC | <i>Soluble Solid Content (contenuto di solidi solubili)</i> |
| MDS | <i>Magnetic Density Separation</i> | SVM | <i>Support Vector Machine</i> |
| MI | <i>Maturity Index (indice di maturazione)</i> | SWIR | <i>Short Wave InfraRed (Infrarosso ad onde corte)</i> |
| MIR | <i>Mid-InraRed (Infrarosso medio)</i> | TSSC | <i>Total Soluble Solid Content (contenuto di solidi solubili totali)</i> |
| MLR | <i>Multiple Linear Regression</i> | UV | <i>Ultraviolet (ultravioletto)</i> |
| MPLS | <i>Modified Partial Least Squares</i> | VIP | <i>Variable Importance for Projection</i> |
| MSC | <i>Multiplicative Scatter Correction</i> | Vis | <i>Visible (visibile)</i> |
| NIPALS | <i>Nonlinear Iterative Partial Least Squares</i> | WEEE | <i>Waste Electrical and Electronic Equipment</i> |
| NIR | <i>Near InfraRed (Infrarosso vicino)</i> | XRF | <i>X-Ray Fluorescence</i> |
| N-W | <i>Norris-Williams</i> | | |

1.1 CONTESTO GENERALE

La presente tesi di dottorato è il risultato di uno studio interdisciplinare, che ha avuto quale tema centrale di ricerca la messa a punto di sistemi innovativi e “*sensor-based*” per la caratterizzazione di materie prime, nonché per il controllo qualità di prodotti. Particolare attenzione è stata rivolta all’utilizzo di procedure innovative e allo sviluppo di appropriati algoritmi di statistica multivariata, *machine learning* e *computer vision* a complemento dell’utilizzo di tecniche di spettroscopia classica e iperspettrale.

Le metodiche analitiche impiegate hanno riguardato prevalentemente l’uso della spettroscopia, quale tecnologia non distruttiva potenzialmente impiegabile nella selezione on-line (i.e. come tecnologia di *sorting*). L’analisi di immagine iper-spetttrale (*HyperSpectral Imaging*: HSI) e la spettroscopia puntuale sono state confrontate tra loro e con tecniche di analisi tradizionali. L’utilizzo di tecnologie iper-spetttrali negli intervalli spettrali Vis-NIR (*Visible – Near InfraRed*: 400-1000 nm), NIR (*Near InfraRed*: 1000-1700 nm), SWIR (*Short Wave InfraRed*: 1000-2500 nm) e Vis-SWIR (*Visible – Short Wave InfraRed*: 400-2500 nm), in combinazione con altre tecniche d’indagine (come ad esempio: analisi chimiche e analisi in micro-fluorescenza ai raggi X) consentono l’implementazione di una metodologia di caratterizzazione rapida, non distruttiva e non invasiva. I dati ottenuti sono stati sottoposti ad analisi chemiometrica con l’obiettivo di sviluppare ed ottimizzare la capacità discriminante delle funzioni di classificazione testate durante la sperimentazione e/o la capacità predittiva delle risposte spettrali.

Tali tecniche consentono l’individuazione delle procedure applicabili più consone per il controllo qualità e/o trattamento da utilizzare per le materie prime ed i prodotti presi in esame.

In particolare, le problematiche analizzate nella presente tesi, riguardano principalmente il settore del riciclo e dell’industria delle materie prime secondarie e quello dell’industria agro-alimentare.

Per quanto riguarda il campo delle materie prime secondarie, lo studio si è incentrato sul riconoscimento di diverse tipologie di materiali all’interno di un flusso di rifiuti eterogeneo (i.e. Rifiuti di Apparecchiature Elettriche ed Elettroniche, RAEE) e parzialmente omogeneo (i.e. *packaging* proveniente da raccolta differenziata della plastica), con le seguenti finalità: i) controllo qualità step by step in un processo di riciclo (classificazione di più materiali negli *output* analizzati); ii) recupero di metalli (i.e. identificazione di Cu in schede a circuito stampato e identificazione di circuiti stampati in un flusso di RAEE); iii) selezione/riciclo (i.e. impostazione di modelli a cascata per la separazione/selezione di diversi materiali) e riconoscimento di contaminanti (i.e. carta e cartone da imballi in plastica).

Mentre, per quanto concerne il settore agro-industriale e quello dell'industria alimentare, lo studio è stato focalizzato sui seguenti punti: i) valutazione della maturazione/invaiaitura di alcune tipologie di frutti (i.e. kiwi ed olive); ii) riconoscimento di prodotti finiti (i.e. frutta secca) da contaminanti; iii) valutazione di parametri qualitativi e/o percettivo-gustativi di diversi prodotti agro-alimentari (i.e. kiwi, farine, grano, olive e peperoncino); iv) valutazione e riconoscimento di prodotti alterati (i.e. olive attaccate da mosca olearia e/o danneggiate) e v) valutazione della granulometria di prodotti (i.e. peperoncino e farine).

1.2 STRUTTURA DELLA TESI

La presente tesi è organizzata nel seguente modo:

Capitolo 2 – Teoria; in questo capitolo viene fornita una panoramica generale sulle tecniche spettroscopiche utilizzate. In particolare, vengono fornite informazioni riguardanti la spettrofotometria nel vicino infrarosso convenzionale ed HSI, e spettrofotometria XRF. Inoltre, in questo capitolo vengono presentate le tecniche di statistica multivariata ed i metodi chemiometrici utilizzati nella presente tesi.

Capitolo 3 – Strumenti utilizzati; in questo capitolo vengono descritte le funzionalità, le configurazioni e le specifiche tecniche degli strumenti utilizzati, nonché le procedure di calibrazione da effettuare prima dell'acquisizione dei dati e la successiva gestione dei dati acquisiti.

Capitolo 4 – Casi di studio; questo capitolo è il corpo centrale della presente tesi. Ivi sono riportati i casi di studio che hanno interessato l'attività di ricerca durante il percorso di dottorato. In particolare, sono descritte diverse logiche e procedure, basate su metodiche di statistica multivariata, *machine learning* e *computer vision* a complemento dell'utilizzo di tecniche di spettroscopia classica ed iperspettrale, finalizzate alla caratterizzazione e alla soluzione di problematiche riguardanti il controllo qualità di materie prime e prodotti.

Capitolo 5 – Considerazioni finali e prospettive future; nel capitolo finale sono riportate alcune considerazioni sulle tecniche utilizzate nell'ambito del progetto di ricerca di dottorato e sono discusse ulteriori prospettive e scenari di ricerca futuri.

1.3 L'IMPORTANZA DELLE MATERIE PRIME E DEL CONTROLLO QUALITÀ NEL SETTORE DEL RICICLO

Sin dall'antichità, ogni scoperta dell'uomo ha portato ad una serie d'innovazioni ed applicazioni che hanno fornito un notevole benessere fino al momento in cui sono state

superate da altre tecnologie. Nel corso della storia, i progressi non si limitarono esclusivamente alla tecnologia militare, ma si estesero a strumenti agricoli, alla conservazione e alla preparazione di cibi, ad applicazioni terapeutiche e cosmetiche, nonché a molti altri aspetti della vita quotidiana. Con il passare del tempo, è migliorata anche la nostra capacità nella produzione di nuovi materiali, oltre alla capacità di caratterizzare, prevedere e sfruttare le proprietà fisiche e chimiche dei materiali. Così, è diventato possibile sviluppare nuove applicazioni che migliorano la prestazione tecnica, la durata e l'affidabilità dei prodotti, nonché la riduzione degli oneri ambientali. Nell'era dell'informazione, gli sviluppi nell'ingegneria e nella scienza dei materiali continuano a far crescere e a sostenere lo sviluppo di nuove tecnologie. Le proprietà uniche di minerali non combustibili, prodotti minerali, metalli e leghe contribuiscono alla fornitura di cibo, riparo, infrastrutture, trasporti, comunicazioni, sanità ed applicazioni di difesa. Il telefono cellulare è un esempio familiare, che illustra la dipendenza da una tecnologia d'importanza globale basata proprio sulle materie prime. Molti prodotti dipendono da minerali essenziali di tipo non combustibile e da altri prodotti minerali, la cui fornitura è a volte soggetta a interruzioni o a limitazioni. Nel breve e medio termine (poco meno di un decennio), l'equilibrio tra domanda e offerta risulta spesso fragile e i prezzi di queste materie prime potranno destabilizzarsi. Nel lungo periodo (più di circa dieci anni), la disponibilità di minerali non combustibili e di prodotti minerali dipende fortemente dagli investimenti in idee e in tecnologia. Investimenti insufficienti o politiche sbagliate nel settore delle materie prime, potranno portare a conseguenti restrizioni sulla disponibilità di alcuni materiali in futuro. In linee generali, la situazione nei paesi membri dell'Unione Europea è questa: la domanda di risorse è in continua crescita e la concorrenza per accaparrarsi alcune risorse è sempre più sfrenata. Inoltre, l'estrazione di nuove risorse non è più sostenibile da un punto di vista ambientale e queste non sono più "abbondanti, disponibili, ed accessibili a basso costo". Da questo contesto, nasce una tendenza sempre più sentita al riciclo e al recupero di materie prime a seconda vita (*Bonifazi et al., 2018c*).

Nel 2015 la Commissione europea ha adottato un piano d'azione per contribuire ad accelerare la transizione dell'Europa verso un'economia circolare, lasciando il vecchio modello di crescita lineare e promuovendo una crescita economica sostenibile. Il modello di crescita lineare (*Figura 1.3.1a*), che ha caratterizzato lo sviluppo delle economie a partire dalla rivoluzione industriale, è basato sul principio del "prendi, produci, usa e getta". Nei sistemi di economia circolare (*Figura 1.3.1b*), invece, i prodotti riescono a mantenere il loro valore aggiunto il più a lungo possibile e vi è assenza di rifiuti. Una volta che un prodotto raggiunge la fine del proprio ciclo di vita, diventando così un rifiuto, le risorse rimangono all'interno del sistema economico, in modo tale da poter essere riutilizzate più volte a fini produttivi e creare così nuovo valore, alimentando così la ciclicità delle risorse (*EUROPEAN COMMISSION, 2019*), in perfetta armonia con i principi di sviluppo sostenibile.

Seguendo questa logica, la possibilità di re-immissione nel mercato delle cosiddette materie prime seconde, ricavate da processi di riciclo e recupero, è ammissibile qualora esse abbiano caratteristiche compatibili e comparabili con quelle dei materiali originali

Introduzione

e/o naturali. La conoscenza delle caratteristiche del rifiuto e dei materiali che lo compongono è di fondamentale importanza. Infatti, da un'ottima caratterizzazione del materiale, si può perseguire l'ottimizzazione dell'intero sistema del processo di riciclo. L'obiettivo della caratterizzazione è, di fatto, proprio quello di indagare le caratteristiche fisiche, chimiche e morfologiche del prodotto in esame, al fine di scegliere la strada più adatta ai fini del recupero. In tale ottica, l'utilizzo di sistemi di controllo per la qualità dei prodotti riciclati è anch'essa di fondamentale importanza. Infatti, spesso, i materiali di recupero sono considerati qualitativamente inferiori e una certificazione accurata di tali prodotti gli conferirebbe un'accezione positiva, denotandoli come affidabili. In questo contesto, una strada perseguibile potrebbe essere quella di adottare attrezzature e procedure tecnologicamente avanzate che possano garantire la qualità del prodotto durante l'intero processo di riciclo. Monitorando ciascuna fase del processo, si può garantire la qualità della materia prima seconda ottenuta.



a



b

Figura 1.3.1. Confronto tra economia lineare (a) ed economia circolare (b).

1.4 L'IMPORTANZA DEL CONTROLLO QUALITÀ NEL SETTORE AGRO-ALIMENTARE

Il termine qualità viene definito come “l'insieme delle proprietà e delle caratteristiche di un prodotto o servizio che si riferiscono alla sua capacità di soddisfare bisogni espliciti o impliciti” (*UNI EN ISO 8402:1995*) o anche come “il grado in cui un insieme di caratteristiche intrinseche soddisfa i requisiti” (*UNI EN ISO 9000:2005*). Ad ogni modo, tale concetto non può essere definito in maniera univoca per i prodotti agro-alimentari; in quanto può essere inteso in maniera differente a seconda dei differenti soggetti che operano all'interno della filiera produttiva. Ciascuno di questi soggetti, infatti, si basa sui propri criteri di qualità ed accettabilità del prodotto, i quali sono spesso soggettivi e basati sulle proprie aspettative. Dunque, qualora si dovesse definire la qualità di un prodotto in maniera adeguata è necessario specificare il sistema di riferimento nel quale deve essere calato tale concetto.

L'industria agro-alimentare italiana è particolarmente vocata alla qualità dei prodotti, soprattutto per quanto concerne gli attributi qualitativi e quantitativi, ma anche per quanto riguarda l'individuazione dei target di mercato. Particolare attenzione viene data alla fase di trasferimento dei frutti e/o prodotti agro-alimentari generici dai luoghi di produzione/conservazione ai mercati di consumo; il focus principale è quello di limitare l'insorgenza di una rapida e generalizzata senescenza del prodotto, quale causa principale della riduzione della *shelf-life* (*Testoni & Lovati, 2010*). La *shelf-life* è quel lasso di tempo in cui il prodotto preserva le proprie caratteristiche qualitative ed organolettiche nelle normali condizioni di conservazione o di utilizzo.

Negli ultimi anni, in questo settore, la ricerca si è focalizzata sullo sviluppo di nuove tecnologie e tecniche innovative non-distruttive e non-invasive per determinare gli attributi qualitativi sia interni che esterni di prodotti agro-alimentari. Tali tecniche, pensate per essere speditive e di facile utilizzo, sono applicabili anche nei processi di controllo *on-line*, *at-line* ed *in-line* (*Figura 1.4.1*), nonché in sistemi di classificazione e/o separazione di prodotti (*Rantanen & Khinast, 2015*).

Nel settore agroalimentare, gli standard riferibili alla definizione, la misura, l'ottenimento e la garanzia dei livelli qualitativi presentano diversi aspetti particolari (*Cavicchi, 2002*). In genere, la qualità di un prodotto agro-alimentare può essere descritta da un punto di vista igienico-sanitario (i.e. presenza di tossine o parassiti, tracce di fitofarmaci e contaminazione biologica), alimentare (i.e. valori nutrizionali), organolettico-sensoriale (i.e. consistenza, aroma e sapore) ed estetico (i.e. forma, dimensione, colore e presenza di difetti). Per la valutazione di tali aspetti qualitativi ci si può avvalere di differenti strumenti. Ad esempio, i parametri estetici possono essere valutati con criteri di tipo visivo ed in genere sono valutati in maniera non-distruttiva. Ad ogni modo, tale criterio non tiene conto degli aspetti qualitativi riferiti alle proprietà organolettiche, nutrizionali ed igienico-sanitarie, valutati in genere con tecniche analitiche. Per quanto riguarda i frutti, i parametri analitici che vengono analizzati comunemente con tecniche distruttive, sono: la durezza della polpa (stimata con il

penetrometro), il contenuto di solidi solubili o zuccheri (valutata con il metodo del rifrattometro), l'acidità (valutata mediante titolazione) ed il pH (misurato con il pHmetro). Ad ogni modo, la misura di tali parametri è spesso onerosa in termini di tempo e di numerosità campionaria. Inoltre, è strettamente dipendente dalla modalità di campionamento che varia a seconda del luogo di campionamento (ad esempio, se ci si trova ad effettuare il campionamento in pieno campo, nel luogo di conservazione o nella struttura di lavorazione e/o trasformazione).

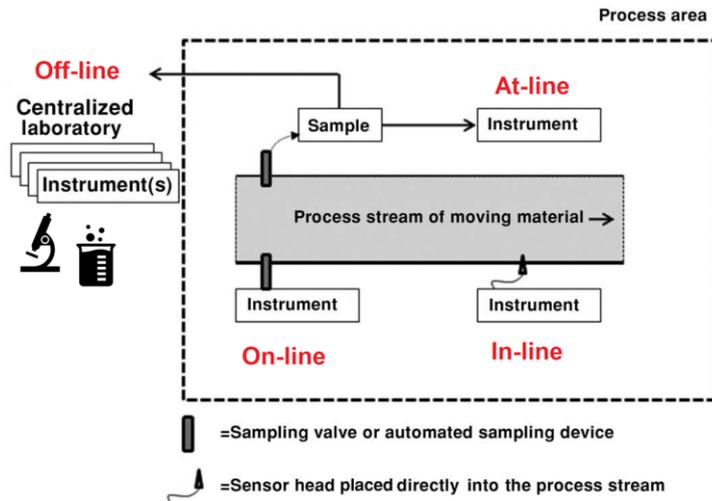


Figura 1.4.1. Rappresentazione schematica delle varie modalità di analisi di processo. Off-line: il campione viene portato in un laboratorio esterno all'area di processo. In-line: la sonda viene posta direttamente a contatto con il flusso di materiale. On-line: consiste in una fase di campionamento automatico e di analisi; di solito il campione viene reintrodotta nel flusso di materiale. At-line: il campione prelevato dalla linea viene analizzato nell'area di processo; di solito il campione prelevato non viene reintrodotta nel flusso di materiale (Rantanen & Khinast, 2015).

Ad esempio, nel caso specifico dei frutti, si ha un'estrema variabilità dello stato di maturazione degli stessi anche su uno stesso albero, nonché tra alberi differenti situati nello stesso impianto (Smith *et al.*, 1994). La determinazione degli standard qualitativi per i prodotti da immettere nel mercato richiede pertanto un'attenta opera di campionamento. In questo contesto, inoltre, un limite operativo e metodologico certamente non trascurabile è quello imposto dall'onerosità economica e dal tempo di analisi delle metodologie di tipo distruttivo (Costa *et al.*, 2001).

Da quanto visto, l'utilizzo di tecniche non-distruttive per l'analisi di qualità di frutti permetterebbe di affrontare il problema del campionamento nell'ambito produttivo e distributivo, sia da un punto prettamente pratico che da un punto di vista economico. Inoltre, l'utilizzo di tecniche non-invasive e non-distruttive consentirebbe la valutazione di attributi interni mediante misurazione esterna (Ortiz & Torregrosa, 2014; Giovanelli *et al.*, 2014). In aggiunta, con tali tecniche è possibile effettuare più misurazioni durante un arco di tempo, seguendone l'evoluzione in termini di maturazione e/o senescenza del prodotto (Costa *et al.*, 2009). Considerata l'importanza di mantenere elevate le caratteristiche

organolettiche e nutrizionali nel tempo e necessità del mantenimento della qualità globale di un prodotto, negli ultimi anni sono state effettuate numerose ricerche riguardanti lo sviluppo di sensori in grado di fornire informazioni circa “lo stato” del prodotto, grazie all’individuazione di correlazioni ottenibili con i parametri qualitativi (i.e. contenuto di solidi solubili, durezza della polpa, ecc.).

Tra questi metodi di analisi non-distruttivi emerge la spettroscopia Vis (*Visible*) e NIR (*Near InfraRed*), quale tecnica analitica basata sull’assorbimento di radiazioni elettromagnetiche.

1.5 STATO DELL’ARTE DELLE TECNICHE DI SPETTROGRAFIA AD IMMAGINE E SPETTROFOTOMETRIA PUNTUALE

I sistemi di *computer vision* sono diventati degli strumenti di utilizzo comune nell’ambito del controllo di processi manifatturieri e del monitoraggio della qualità di prodotti nell’industria delle materie prime e nel settore agri-alimentare (*Amigo et al., 2013*).

Negli ultimi anni, con il continuo progresso scientifico e tecnologico, nonché con la miniaturizzazione delle strumentazioni, l’*HyperSpectral Imaging* (HSI) è divenuta una delle tecniche più promettenti per il controllo qualità di prodotti dell’industria agro-alimentare e, più in generale, dell’industria delle materie prime. In questo panorama, le tecnologie basate sull’HSI nel vicino infrarosso stanno godendo di un ottimo successo (*Hollestein et al., 2015*). Infatti, l’HSI rappresenta una buona soluzione per differenti *task*, quali: caratterizzazione, classificazione e controllo di qualità di materiali e bio-materiali.

L’HSI rappresenta una tecnologia emergente che coniuga ed integra le tecniche di *image processing* a quelle della spettroscopia tradizionale, al fine di ottenere sia informazioni spaziali che spettrali. I suoi principali punti di forza risiedono nel fatto che, tale tecnica è non invasiva e non distruttiva, in più è rapida rispetto a numerose altre tecniche analitiche. Utilizzando la tecnica HSI, lo spettro di ciascun pixel del campione analizzato può essere utilizzato come “*fingerprint*”, ovvero come una vera e propria impronta digitale del materiale, al fine di caratterizzare la composizione chimico-fisica del pixel e visualizzarne la posizione spaziale (*Sun, 2010*).

Grazie a queste peculiarità, l’HSI ha avuto successo in differenti settori; in letteratura sono riportate differenti applicazioni nei seguenti campi (*Tabella 1.5.1*): i.e. beni culturali, industria agro-alimentare, industria chimico-farmaceutica, medicina ed applicazioni cliniche, per applicazioni in biologia e, più in generale, in chimica analitica.

Proprio negli ultimi anni, l’HSI sta emergendo anche nell’industria delle materie prime e nel settore del riciclo. In particolare, in letteratura sono riportate alcune applicazioni dell’HSI per: caratterizzazione e riconoscimento di polimeri vergini (*Leitner, et al., 2003; Gosselin et al., 2011*) e di riciclo (*Serranti et al., 2010; Serranti & Bonifazi, 2010; Palmieri et al.,*

2014a; Bonifazi 2015a), caratterizzazione estetica di rocce ornamentali (Bonifazi et al., 2005), caratterizzazione e riconoscimento di diversi materiali presenti in un flusso di Rifiuti da Costruzione e Demolizione -C&D- (Palmieri et al., 2014b), riconoscimento di materiali contenenti asbesto (Bonifazi et al., 2018a), caratterizzazione di prodotti di compostaggio (Serranti et al., 2009) e biomasse (Bonifazi & Serranti, 2010), caratterizzazione di “car fluff” (che consiste in una miscela di materiali metallici, ferrosi e non, di plastica, gomma, vetro, fibre tessili, carta, vernici ecc.) proveniente da veicoli giunti a fine vita (Bonifazi & Serranti, 2006a), caratterizzazione di “bottom ash” (ceneri pesanti) provenienti da inceneritori di rifiuti solidi urbani (Bonifazi & Serranti, 2007), identificazione di vetroceramica in rifiuti vetrosi (Bonifazi & Serranti, 2006b).

Tabella 1.5.1. Alcuni campi di applicazione della spettrografia ad immagine investigati da differenti autori.

| Campi di applicazione | Autori |
|---|---|
| Controllo qualità di prodotti agroalimentari ed applicazioni in campo agrario | Gowen et al., 2007; Del Fiore et al., 2010; Sun, 2010; Kumaravelu & Gopal, 2015; Tsuchikawa & Kobori, 2015; Kamruzzaman et al., 2016; Zhang et al., 2016; Serranti et al., 2018a; Serranti et al., 2018b. |
| Biologia | Pisani et al., 2013; Gowen et al., 2015. |
| Scienze forensi | Edelman et al., 2012; Fernández de la Ossa, Amigo, & García-Ruiz, 2014. |
| Farmaceutica | Roggo et al., 2007; Gowen et al., 2008; Rocha et al., 2010. |
| Medicina ed applicazioni cliniche | Liu et al., 2007; Jolivot et al., 2011; Vasefi et al., 2016. |
| Beni culturali | Kubik, 2007; Agresti et al., 2013; Capobianco et al., 2015; Daniel et al., 2016; Capobianco et al., 2017. |
| Monitoraggio ambientale (proximal e remote sensing) | Moroni et al., 2013; Stuart et al., 2019. |

Nell’ambito dell’industria agro-alimentare, la sempre crescente competitività di mercato determina la necessità di un continuo miglioramento dei sistemi di monitoraggio dei processi produttivi, prediligendo l’utilizzo di metodi rapidi, affidabili e non distruttivi. In questo contesto, le applicazioni di biofotonica (ovvero quell’area di ricerca multidisciplinare, basata sulle discipline tecnico-ingegneristiche che studiano l’interazione tra la luce ed i materiali biologici applicate alle scienze della terra, alla biologia ed alla medicina), dato l’elevato tasso di innovatività (il quale è basato sui comparti *high-tech*: ottica, elettronica, informatica ed algoritmica) e le esigenze richieste della produzione di sistemi di monitoraggio o controllo altamente informativi, multi-parametrici, speditivi, non distruttivi, per applicazioni in campo (*on-the-go*) o in linea di processo (*on-line*), sono in forte crescita (Jürgens et al., 2013).

La spettroscopia NIR è una tecnica analitica molto efficace per determinare in tempi ridotti ed in maniera simultanea numerose caratteristiche dei prodotti agro-alimentari. Le analisi effettuabili tramite spettroscopia NIR possono essere distinte in: i) quantitative, finalizzate alla determinazione di un particolare analita presente nei campioni studiati (i.e. messa a punto di modelli di calibrazione) e ii) qualitative, finalizzate alla classificazione dei campioni presi in esame (i.e. messa a punto di modelli di classificazione). Per determinare la composizione chimica di prodotti agro-alimentari tramite spettroscopia NIR è necessaria una fase di calibrazione, la quale consiste nell'individuazione di campioni rappresentativi e uno studio statistico riferito ai dati di composizione chimica e gli spettri acquisiti dei campioni.

In tempi recenti, sono stati effettuati numerosi studi riferiti allo sviluppo di nuove tecniche di analisi non distruttive e non-invasive per la determinazione di parametri qualitativi di prodotti agro-alimentari. In particolare, la spettroscopia Vis-NIR è stata utilizzata con successo nel campo orto-frutticolo per determinare parametri, quali: l'SSC (contenuto di solidi solubili), il contenuto di acidità, l'indice di maturazione ma anche la consistenza ed altre proprietà fisiologiche (*dos Santos et al., 2013; Misal & Deshmukh, 2016*). Inoltre, tali tecniche possono essere applicate per la valutazione del contenuto di grassi in prodotti di carne e pesce, e per la valutazione di differenti parametri qualitativi di cereali e/o prodotti cerealicoli. In letteratura, ci sono numerosissimi esempi applicativi della spettroscopia Vis-NIR in riflettanza per i prodotti agroalimentari. In [Tabella 1.5.2](#) ne sono riportati alcuni.

La grandezza di tali tecniche risiede soprattutto nella scalabilità e nella versatilità. In quanto, tali metodiche sono utilizzabili sia per valutazioni speditive in campo (*on-the-go*), grazie all'utilizzo di strumentazione portatile, o in linea di processo (*on-line*) per fini di classificazione o separazione (i.e. utilizzabile come tecnologia di *sensing* per macchine selezionatrici).

Introduzione

Tabella 1.5.2. Alcune applicazioni della spettroscopia Vis-NIR in riflettanza per analisi di prodotti agro-alimentari (dos Santos et al., 2013).

| | Specie / varietà analizzata | Caratteristiche analizzate | Metodi di regressione / classificazione multivariata | Referenze |
|---------------------------------------|--|---|---|---|
| Frutta e verdura | Mandarino | Durezza, SSC, acidità, pH, Indice di maturazione, Indice di colore, ecc. | MPLS | <i>Sun et al., 2009</i> |
| | Albicocca (su albero) | Durezza, SSC, acidità, colore, ecc. | PLS | <i>Christen et al., 2012</i> |
| | Pomodoro | Durezza, SSC, acidità, colore, ecc. | PLS | <i>Camps et al., 2012</i> |
| | Arancia | Durezza, SSC, acidità, Indice di maturazione, Indice di colore, ecc. | PCR, PLS | <i>Cayuela & Weiland, 2010a; Liu et al., 2012</i> |
| | Mela | Durezza, SSC, acidità, contenuto di clorofilla, polifenoli, ecc. | PLS | <i>Yang, et al., 2011</i> |
| | Pera | SSC | PLS, MLR | <i>Liu et al., 2010</i> |
| | Prugna | Durezza, SSC, classificazione secondo la varietà. | PLS-DA | <i>Pérez-Marín et al., 2010</i> |
| | Mango | SSC | PLS | <i>Saranwong et al., 2003</i> |
| | Uva | SSC, acidità, pH, ecc. | PLS | <i>Guidetti et al., 2010</i> |
| | Olive | DM, contenuto in olio, ecc. | PLS | <i>Cayuela et al., 2010b</i> |
| | Kiwi | DM | PLS | <i>Lue et al., 2010</i> |
| Cereali e prodotti cerealicoli | Riso | Contenuto di proteine, contenuto di amilosio, ecc. | PLS | <i>Li & Shaw, 1997</i> |
| | Insilato di mais | DM | MPLS | <i>Berzaghi et al., 2005</i> |
| | Farina di grano (in polvere ed in pasta) | Contenuto di umidità, Contenuto di proteine, contenuto in ceneri, Contenuto di grassi e carboidrati | PLS | <i>Kumagai et al., 2004</i> |
| Prodotti di carne e pesce | Pesce spada | Classificazione secondo lo stato di conservazione | PLS-DA | <i>Fasolato et al., 2012</i> |
| | Carne suina | Classificazione secondo differenti razze (da muscolo massetere) | ANN | <i>del Moral et al., 2009</i> |
| | Tonno | Contenuto di grassi | PLS | <i>Shimamoto et al., 2003</i> |

2.1 CENNI DI SPETTROSCOPIA

La spettroscopia è lo studio dell'interazione tra la materia e la radiazione elettromagnetica.

Storicamente, la spettroscopia nasce nel XVII secolo con gli studi effettuati sulla dispersione della luce bianca per mezzo di un prisma. *Isaac Newton* fu il primo ad utilizzare la parola "spettro" per descrivere la combinazione di colori che genera la luce bianca (*Bursej, 2017*). Successivamente, con la formulazione della teoria ondulatoria della luce di *Christiaan Huygens* nel 1678, il termine spettro venne riferito all'intensità della luce in funzione della lunghezza d'onda (λ) o della frequenza (ν). Durante i primi anni del XVII secolo, *William Hershel* studiò la distribuzione spettrale della radiazione solare e scoprì la radiazione infrarossa. Pochi anni più tardi, *Joseph von Fraunhofer* inventò il reticolo di diffrazione per misurare esattamente la lunghezza d'onda della luce solare, gettando le basi per la costruzione degli spettroscopi moderni. Ad oggi, il termine spettro può riferirsi a un flusso o un'intensità di radiazione elettromagnetica o particelle in funzione della loro energia, lunghezza d'onda, frequenza o massa.

Lo strumento che permette la misura di uno spettro è quindi detto spettrometro, spettrografo o spettrofotometro. In particolare, la spettrofotometria si occupa della misura dell'intensità luminosa dal vicino ultravioletto al vicino infrarosso.

Lo spettro elettromagnetico comprende l'insieme di tutte le possibili frequenze delle radiazioni elettromagnetiche. Pur essendo continuo, lo spettro è convenzionalmente suddiviso in vari intervalli o bande di frequenza, come riportato in [Figura 2.1.1](#).

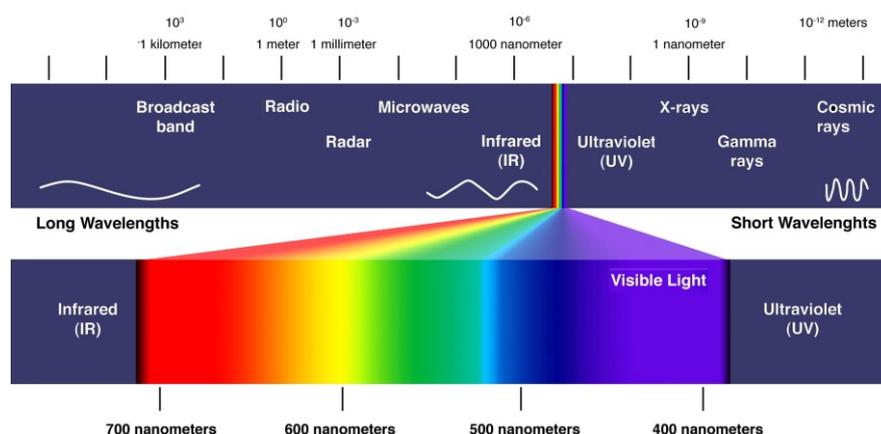


Figura 2.1.1. Raffigurazione convenzionale dello spettro elettromagnetico.

Le diverse tipologie di radiazioni possono differire in termini di lunghezza d'onda o di frequenza, ma fisicamente sono identiche. I.e. quando la luce interagisce con un materiale, possono aver luogo differenti fenomeni fisici, come ad esempio: la riflessione, la trasmissione, l'assorbimento o la dispersione (*scattering*) della luce e la fluorescenza. Tutti questi fenomeni possono aver luogo quando la radiazione incidente induce dei cambiamenti nei livelli energetici della struttura atomica del materiale, i quali possono essere di natura elettronica, vibrazionale o nucleare. Le differenze di energia tra i due livelli energetici possono coprire un ampio intervallo dello spettro elettromagnetico dalle onde radio ai raggi gamma.

2.1.1 Spettrofotometria nel vicino infrarosso

La spettroscopia nel vicino infrarosso o NIRS (*Near Infrared Spectroscopy*) è una tecnica che consiste nella misura non invasiva e non distruttiva della superficie di un campione. Il principio su cui è basata la tecnica della spettrofotometria nel vicino infrarosso è l'interazione tra la radiazione elettromagnetica riflessa o trasmessa e la struttura molecolare della superficie del materiale, che genera transizioni vibrazionali. Più nel dettaglio, la regione dello spettro elettromagnetico, che convenzionalmente corrisponde al NIR è compresa tra 780 e i 2500 nm, ovvero $12,500 - 4,000 \text{ cm}^{-1}$ ([Blanco & Villaroya, 2002](#)).

Lo spettrometro rileva la radiazione riflessa e di conseguenza le bande di assorbimento. Queste bande di assorbimento generano uno spettro specifico; se una radiazione nel vicino infrarosso interagisce con un materiale attivo nel campo del NIR, si potrà misurare lo spettro caratteristico dello stesso. Questo spettro caratteristico rappresenterebbe una vera e propria "impronta" del materiale stesso ([Schropp et al., 2014](#)).

Le differenze tra le firme spettrali caratteristiche possono permettere l'identificazione di una grande varietà di materiali, tra questi: plastiche, minerali e componenti organiche. Ad ogni modo, questa tecnica presenta delle limitazioni. I.e. nel caso in cui si abbia a che fare con materiali neri o non attivi nell'intervallo dell'infrarosso vicino, oppure quando la maggior parte delle radiazioni vengono assorbite o riflesse dal materiale, la quantità di radiazioni trasmesse o riflesse potrebbero risultare troppo basse per consentire una analisi rappresentativa del campione analizzato.

2.1.1.1 Il principio fisico della spettrofotometria nel vicino infrarosso

La luce può essere vista come un flusso di corpuscoli (i cosiddetti fotoni) o come un'onda elettromagnetica. L'interazione tra fotoni e molecole può essere descritta dall'equazione di *Bohr* e più in generale dalle leggi dell'ottica quantica. Mentre, l'interazione tra un'onda elettromagnetica ed un mezzo può essere descritta dalle equazioni di *Maxwell* e dalle teorie dell'elettromagnetismo.

Quando una superficie di un campione è sollecitata da una radiazione, l'energia assorbita si manifesta con la vibrazione molecolare, dove la frequenza della vibrazione molecolare è uguale alla frequenza della radiazione IR. Il processo di assorbimento dipende dalla composizione del campione e proprio per questo ne può fornire informazioni (*Schropp et al., 2014*).

Un qualunque corpo, se sottoposto ad irraggiamento termico o luminoso, possiede una determinata proprietà di riflessione, assorbimento e trasmissione sia del calore radiativo che della luce. La riflettanza (ρ) è un parametro percentuale adimensionale che rappresenta il potere riflettente di un corpo sottoposto a irraggiamento. L'intero raggio radiante che giunge sul corpo con la sua energia si frammenta rispettando il principio di conservazione dell'energia: una parte di energia è riflessa, una parte viene trasmessa, mentre una parte è assorbita (*Bruni, 1964*). Perciò si avrà che la somma dei parametri riflettanza (ρ), trasmittanza (τ) e assorbanza (α) è uguale ad 1:

$$\alpha + \rho + \tau = 1 \quad (2.1.1)$$

L'energia associata alla radiazione elettromagnetica è trasmessa in pacchetti indivisibili chiamati quanti, relativi ad ogni singolo fotone (*Schropp et al., 2014*). La dimensione E di un quanto di energia dipende dalla frequenza ν della radiazione secondo la formula:

$$E = h\nu \quad (2.1.2)$$

Dove h è la costante di *Planck* ($h = 6.626 \times 10^{-34} \text{ J s}$) e ν è la frequenza della radiazione.

La frequenza della radiazione è associata allo stato energetico finale ed iniziale, dalla relazione:

$$\Delta E \equiv E_2 \pm E_1 = h\nu \quad (2.1.3)$$

Dove E_2 e E_1 rappresentano rispettivamente lo stato energetico finale ed iniziale.

L'energia associata è convertita in movimenti di *stretching* (stiramento del legame chimico: vibrazione del legame lungo il piano, simmetrico o asimmetrico, in conseguenza al quale varia aritmicamente la distanza interatomica) e *bending* (deformazione dell'angolo di legame: movimento a forbice che cambia l'angolo formato tra i legami). Tipicamente, una vibrazione fondamentale non dà luogo ad attività nel vicino infrarosso, perché le modalità di vibrazione sembrano non influenzare lo spettro del NIR. Nel NIR, le bande di assorbimento sono generate da interazioni e/o combinazioni di vibrazioni fondamentali, prevalentemente dovute ai legami molecolari: O-H, C-H, S-H e N-H (*Blanco & Villaroya, 2002; Cen & He, 2007*). Sovrapposizioni e combinazioni di differenti tipologie di bande d'assorbimento indicano determinate caratteristiche all'interno dello spettro NIR.

2.1.1.2 Gli spettrometri nel vicino infrarosso

Uno spettrometro è un dispositivo che permette la misura dello spettro di radiazione elettromagnetica, o le proprietà della luce in funzione della sua lunghezza d'onda. La maggior parte degli spettrometri utilizza il principio di interferenza per scomporre la radiazione luminosa nelle sue lunghezze d'onda e misurarne l'intensità su un fotodiode.

Uno spettrofotometro NIR è comunemente costituito da una fonte di luce (di solito una luce alogena con bulbo a tungsteno), un monocromatore, un rivelatore (*detector*) e da componenti ottici (i.e. lenti, collimatori, separatore di luce o *beam splitter*, sfere integranti e fibre ottiche). In genere, come sorgente luminosa vengono utilizzate lampade che, nonostante abbiano una emissione nel visibile, possono essere considerate una fonte di radiazione NIR economica e disponibile. Il campione trasmette, riflette o riemette una parte di tale radiazione; la parte di radiazione trasmessa dal campione viene quindi raccolta ed inviata al monocromatore che seleziona le aree spettrali di interesse ([Glasstone, 1946](#)). Nei sistemi moderni, viene utilizzato il reticolo di diffrazione come monocromatore. Il reticolo di diffrazione consiste in una superficie incisa con numerose linee parallele che riflette la radiazione infrarossa; queste linee inducono la divisione della radiazione incidente per diffrazione in singole lunghezze d'onda. In genere, la selezione delle lunghezze d'onda da inviare al rivelatore viene effettuata ruotando il reticolo e quindi variando l'angolo di incidenza della radiazione proveniente dalla sorgente. La radiazione, una volta resa monocromatica, passa quindi al rivelatore che la analizza e fornisce un dato analitico. Per convenzione gli spettrofotometri vengono classificati in base al tipo di monocromatore, ovvero, il dispositivo che permette di scomporre la luce nelle sue componenti "cromatiche" e quindi di selezionare la lunghezza d'onda della radiazione emessa dalla sorgente ([Schropp et al., 2014](#)).

Gli spettrometri utilizzati per alcuni casi applicativi, riportati nella presente tesi, coprono diversi range spettrali i.e. lo spettrofotoradiometro ASD FieldSpec 4[®] Standard-Res ([ASD Inc. 2015a](#)) lavora nel Vis-SWIR (350-2500 nm), mentre lo spettrometro JDSU MicroNIR[™] ([JDS Uniphase Corporation, 2013](#)) opera nel NIR (1000-1700 nm). Tali dispositivi sono descritti dettagliatamente nel capitolo [3. Strumenti utilizzati](#).

Le modalità di presentazione del campione allo strumento possono essere di diversa tipologia. Più nel dettaglio, l'acquisizione degli spettri può avvenire attraverso quattro diverse modalità di misurazione:

- I) Trasmittanza: la luce incidente si propaga attraverso il campione e viene misurata quando lo attraversa, nel punto opposto alla fonte di luce ([Figura 2.1.2a](#));
- II) Riflettanza: la fonte di luce ed il detector sono posizionati dalla stessa parte. Per evitare problemi di riflessione, i rivelatori sono posizionati a 45° rispetto al piano del campione ([Figura 2.1.2b](#));
- III) Transflettanza (assorbimento in riflessione): è una combinazione dei due modi citati sopra. Il campione viene posizionato come nel caso della riflettanza, ma il riflettore è posizionato al lato opposto della fonte di luce, dalla parte più distante dal campione ([Figura 2.1.2c](#));

IV) Interattanza: è anch'essa una combinazione delle prime due modalità, ma riguarda soltanto i campioni solidi. In questa modalità il campione viene illuminato da un fascio di fibre ottiche a diretto contatto con la superficie del campione. I fasci della fibra sono posizionati sullo stesso lato della fonte di illuminazione (Figura 2.1.2d). Il contatto diretto tra la fibra ed il campione elimina l'effetto del riflesso della superficie e massimizza la profondità di penetrazione (Alander et al., 2013).

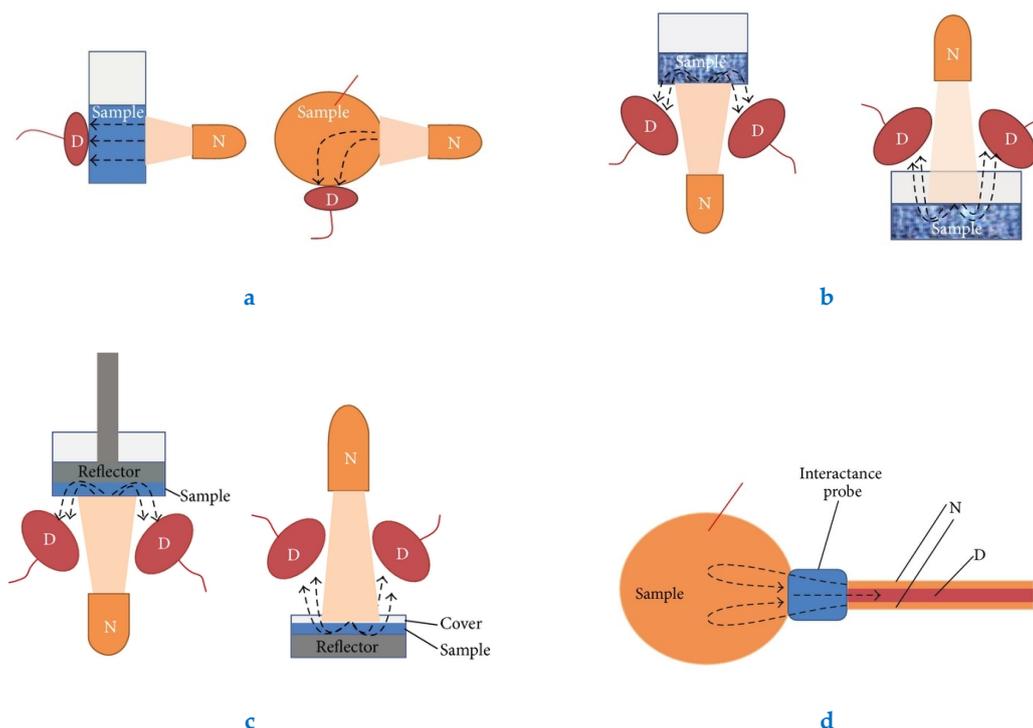


Figura 2.1.2. Modalità di presentazione del campione nella Spettroscopia NIR: a) trasmittanza, b) riflettanza, c) trasflettanza e d) interattanza. N: fonte di luce; d: detector.

2.1.2 Spettrofotometria XRF

L'analisi a fluorescenza a raggi X, o analisi XRF (*X-Ray Fluorescence*), si basa sul processo di eccitamento e conseguente rilassamento di un elettrone di valenza all'interno del guscio elettronico di un atomo. In fisica atomica, la fluorescenza è definita generalmente come un processo comprendente due fasi: nella prima fase un atomo passa ad uno stato eccitato grazie all'acquisto di energia dall'esterno, nella seconda fase l'elettrone torna al suo stato iniziale cedendo spontaneamente energia sotto forma di fotoni. A differenza della fluorescenza ottica, l'eccitamento dell'elettrone si verifica ad energie nettamente più elevate e la radiazione di fluorescenza emessa rientra nel range dei raggi X (Bruker Corporation, 2013).

Tramite analisi XRF si è in grado di rilevare gli elementi principali (Si, Ti, Al, Fe, Mn, Mg, Ca, Na, K e P) e gli elementi in traccia (i.e. > 1 ppm: Ba, Ce, Co, Cr, Cu, Ga, La, Nb, Ni, Rb, Sc, Sr, Rh, U, V, Y, Zr e Zn) che compongono il campione in esame. Tuttavia, l'analisi XRF presenta alcune limitazioni. I.e. la maggior parte degli strumenti in commercio non consentono di misurare con precisione ed accuratezza l'abbondanza di elementi con Z (numero atomico) < 11. Inoltre, l'analisi XRF non può distinguere le variazioni isotopiche di un elemento o distinguere ioni dello stesso elemento in diversi stati di valenza (*Bruker Corporation, 2009*).

2.1.2.1 Il principio fisico dell'analisi XRF

La ionizzazione degli atomi di un campione oggetto di analisi può avere luogo a seguito dell'esposizione ai raggi X con lunghezza d'onda corta o a raggi gamma. La ionizzazione consiste nella espulsione di uno o più elettroni dall'atomo (*Wirth, 2019*). L'espulsione dell'elettrone si può verificare se l'atomo è esposto ad una radiazione che ha energia maggiore del potenziale di ionizzazione. I raggi X e i raggi gamma, grazie alla loro energia, possono espellere gli elettroni dagli orbitali interni dell'atomo. Le principali transizioni energetiche che possono verificarsi vengono denominate in base alle loro modalità. La transizione $L \rightarrow K$ viene tradizionalmente chiamata K_α , una transizione $M \rightarrow K$ è detta K_β , mentre una transizione $M \rightarrow L$ viene chiamata L_α e così via (vedere *Figura 2.1.3*). La struttura elettronica dell'atomo diventa instabile se l'elettrone viene rimosso secondo una delle modalità suddette (*Bruker Corporation, 2009*). Quando questo accade, gli elettroni in orbitali superiori "cadono" nell'orbitale inferiore per riempire il vuoto lasciato precedentemente. L'energia rilasciata sotto forma di un fotone, durante la caduta, è pari alla differenza di energia dei due orbitali coinvolti; in questo modo, il campione emette radiazioni. l'assorbimento di radiazione di energia specifica si manifesta nella ri-emissione di radiazioni a diversa energia (generalmente inferiore); questo è il cosiddetto fenomeno di fluorescenza (*Figura 2.1.3*). La lunghezza d'onda della radiazione fluorescente può essere calcolata con la cosiddetta legge di *Planck*:

$$\lambda = \frac{hc}{E} \quad (2.1.4)$$

Dove h è la costante di *Planck*, E è la dimensione d'energia di un quanto, c è la velocità della luce.

2.1.2.2 Gli spettrometri XRF

La radiazione può essere analizzata sia alla luce della selezione dell'energia dei fotoni o separando le lunghezze d'onda della radiazione. L'intensità di ogni radiazione caratteristica è direttamente correlata alla quantità di ciascun elemento nel materiale (*Wirth, 2019*). Il fascio di raggi X incidente è tipicamente prodotto dal rodio. Quando il

fascio di raggi X primario raggiunge il campione, la struttura atomica del materiale passa ad uno stato eccitato, emettendo raggi X lungo uno spettro di lunghezze d'onda atomiche caratteristiche, che sono specifiche per ogni atomo. Quindi, l'uso di uno spettrometro XRF consente la separazione di un complesso spettro a raggi X emesso in lunghezze d'onda caratteristiche per ciascun elemento presente. Per misurare l'intensità del fascio emesso, vengono utilizzati vari tipi di rivelatori (contatori proporzionali a flusso di gas e scintillazioni). L'intensità dell'energia misurata da questi rivelatori è proporzionale alla quantità di un particolare elemento nel campione.

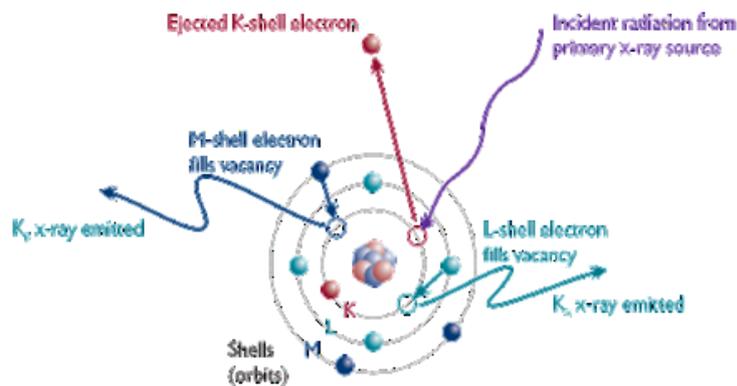


Figura 2.1.3. Principio fisico dell'analisi in fluorescenza a raggi X.

Le caratteristiche tecniche dello spettrometro XRF utilizzato nel caso di studio “4.1.1 Caratterizzazione di un flusso di RAEE risultante da comminazione e separazione magnetica per densità” presentato in questa tesi è riportato nel capitolo “3.5. Spettrometro Micro-XRF M4 TORNADO”.

2.2 HYPER SPECTRAL IMAGING

Le tecniche di monitoraggio e controllo basate sull'analisi d'immagine hanno ottenuto un grande successo nel campo della ricerca, in quanto sono in grado di caratterizzare morfologia e morfometria di campioni eterogenei. Un sistema di *imaging* classico si basa su immagini RGB (*Red-Green-Blue*). Una immagine RGB, che è formata da un certo numero di elementi chiamati pixel, è definita dalla combinazione di tre colori che miscelati tra di loro forniscono quasi tutto lo spettro dei colori del visibile (*Ornberg et al., 1999; Hecht, 2002*): rosso ($622 \text{ nm} < \lambda < 780 \text{ nm}$), verde ($492 \text{ nm} < \lambda < 577 \text{ nm}$) e blu ($455 \text{ nm} < \lambda < 492 \text{ nm}$). Ad ogni modo, per alcune applicazioni, sono necessarie informazioni che vanno oltre il campo del visibile.

Laddove è necessario analizzare anche l'aspetto chimico-fisico di un campione si può ricorrere all'utilizzo di sistemi di imaging iperspettrale (*Hans & Geladi, 2007*).

L'*HyperSpectral Imaging* (HSI) combina l'analisi di immagine digitale con la spettroscopia convenzionale, consentendo l'acquisizione simultanea di informazioni spettrali e spaziali (Mehta et al., 2018). Mentre un sistema di imaging RGB restituisce i valori di intensità dei tre canali R, G e B per ciascun pixel di un'immagine, i sistemi HSI consentono l'acquisizione di più immagini, attraverso la campionatura di più canali a differenti lunghezze d'onda (Figura 2.2.1).

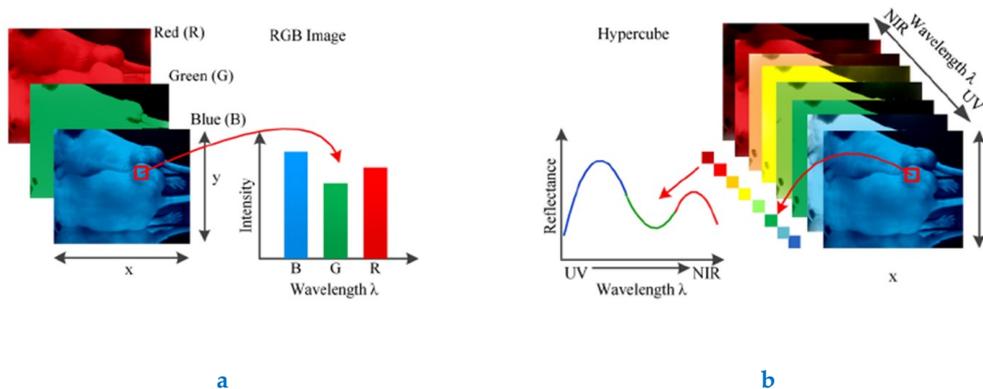


Figura 2.2.1. Rappresentazione di: a) un'immagine RGB e b) un'immagine iperspettrale (Mehta et al., 2018)

A seconda del numero di canali, si possono distinguere: immagini multispettrali, composte da un numero limitato di bande abbastanza larghe (in genere 3-10) ed immagini iperspettrali, composte da centinaia di bande per ciascun pixel (Gowen et al., 2007; Wu & Sun, 2013). Un'immagine iperspettrale è detta anche iper-cubo (Burgers & Gowen, 2011), data la sua natura tridimensionale ed è definita da due dimensioni spaziali (x, y) ed una dimensione spettrale (λ).

Nell'ambito delle tecniche di *chemical imaging*, il termine *HyperSpectral Imaging* comprende differenti metodologie di acquisizione ed elaborazione di dati spettrali risolti spazialmente: i.e. spettroscopia di *scattering Raman*, spettroscopia infrarossa a trasformata di Fourier, spettroscopia di fluorescenza ed infrarossa (Gowen et al., 2015).

Tipicamente, i sistemi HSI più diffusi lavorano nelle regioni che vanno dal visibile fino all'infrarosso ad onde corte (da 400 a 2500 nm). I costi di un sistema HSI sono di gran lunga superiori rispetto ai sistemi *imaging* RGB. Ma il costo è in genere compensato dalle grandi potenzialità che i sistemi HSI possono offrire (Polder et al., 2002; Noordam et al., 2005; Masoud et al., 2011). Un vantaggio delle tecniche basate su sistemi HSI rispetto alle tecniche di spettroscopia convenzionale (i.e. spettroscopia NIR) è quello di essere maggiormente sensibile alle variazioni spaziali delle concentrazioni di un analita (o composto analizzato), quando si analizzano matrici disomogenee. In altre parole, con le tecniche di spettroscopia convenzionale si può misurare un solo spettro su una superficie, nel caso in cui il contributo di un determinato analita sia trascurabile; mentre con tecniche di imaging iperspettrale, non solo si può rilevare la presenza dell'analita ma lo si può anche localizzare spazialmente.

D'altro canto, l'applicazione di tecniche HSI comporta l'utilizzo di computer con elevata capacità di calcolo e di archiviazione; un'immagine iperspettrale può arrivare a pesare centinaia di megabyte (Burgers & Gowen, 2011). Inoltre, l'acquisizione di un'immagine iperspettrale, richiede molto più tempo all'acquisizione di un'immagine digitale con una normale fotocamera. In genere, per specifiche applicazioni vengono selezionate solamente le bande maggiormente informative, velocizzando così il sistema e rendendolo efficiente per il monitoraggio *on-line* o *at-line* (Wu & Sun, 2013). In questo contesto, la chemiometria gioca un ruolo di fondamentale importanza per risolvere problemi concernenti la selezione delle bande spettrali più significative e per lo sviluppo di sistemi efficienti per la visualizzazione dei dati acquisiti.

2.2.1 Modalità di acquisizione delle immagini iperspettrali

In base alla modalità di acquisizione delle immagini iperspettrali, I sistemi HSI possono essere suddivisi in 3 categorie differenti (Burgers & Gowen, 2011): *whiskerbroom*, *pushbroom* e *staredown* (Figura 2.2.2).

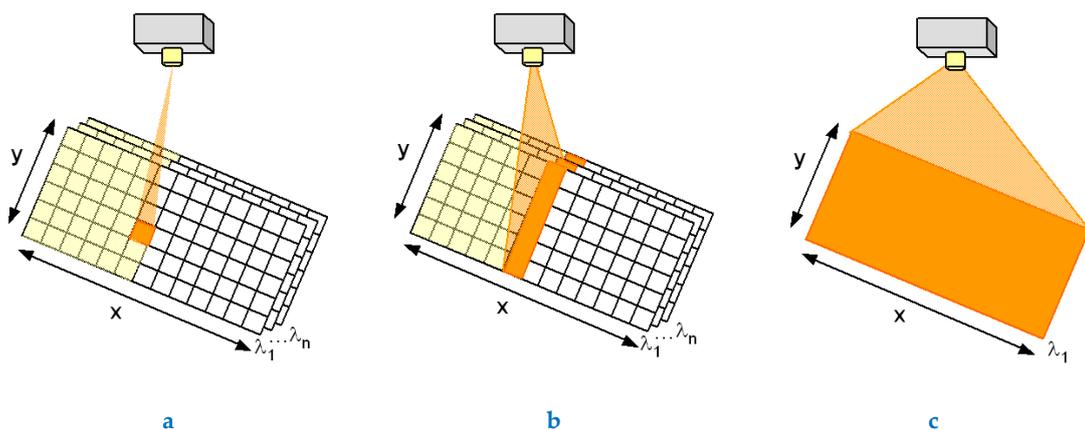


Figura 2.2.2. Le tre configurazioni di un sistema HSI (Burgers & Gowen, 2011): *whiskerbroom* (a), *pushbroom* (b), *staredown* (c).

Nella configurazione *whiskerbroom*, detta anche configurazione *point mapping*, il campione viene posto su uno stage e viene spostato in posizioni predefinite lungo x ed y, acquisendo uno singolo spettro (i.e. da λ_1 a λ_n) in corrispondenza di ciascuna posizione $\{x_i, y_j\}$. Invece, la configurazione *pushbroom*, detta anche *line imaging*, consente di acquisire l'ipercubo di dati in maniera più veloce rispetto alla configurazione *whiskerbroom*.

La configurazione *pushbroom* consente di acquisire tutti i dati spettrali (i.e. da λ_1 a λ_n) di tutti i pixel lungo y, ad una determinata posizione x_i . I dati 2D (dimensione spaziale y e dimensione spettrale λ) acquisiti per ciascuna linea, spostando il campione lungo x, vengono concatenati insieme per formare l'ipercubo.

Mentre, nella configurazione *staretdown*, l'ipercubo viene generato concatenando le immagini 2D (dimensione spaziale x ed y) acquisite per ciascuna lunghezza d'onda λ_i , in direzione λ . Nella configurazione *staretdown*, a differenza delle altre due configurazioni, la *Field of View* viene mantenuta fissa e non sono previste parti in movimento.

2.2.2 Modalità di analisi delle immagini iperspettrali

Le immagini iperspettrali contengono tipicamente molti spettri e sono molto informative. Si pensi che un'immagine iperspettrale può arrivare a "pesare" *gigabytes*. Lo scopo principale dell'analisi d'immagine iperspettrale è di solito quella di ridurre le dimensioni mantenendo al contempo informazioni sufficienti per classificare e/o quantificare dal punto di vista chimico-fisico le aree di un campione. L'analisi di dati iperspettrali si basa su tecniche che combinano il *signal processing*, l'*imaging processing* e la spettroscopia. Le fasi principali dell'analisi di immagine iperspettrale sono schematizzate in [Figura 2.2.3](#).

A partire dai dati grezzi dell'ipercubo è necessaria una fase di pre-elaborazione dati per rimuovere variazioni non desiderate dalle informazioni spettrali e per preparare i dati per ulteriori elaborazioni. I.e. I *detector* utilizzati nei sistemi HSI (in particolare i rilevatori InGaAs utilizzati nei sistemi NIR-HSI) sono inclini ad avere problemi di pixel morti o *spikes*. Questi artefatti devono essere individuati ed esclusi dalle successive analisi ([Burger, 2009](#); [Burger & Gowen, 2011](#)). Dopo aver individuato e corretto tali fonti di variazione nei dati o in alternativa a valle di tale procedura, potrebbe essere necessario rimuovere lo sfondo dell'immagine (il *background*, ovvero la porzione del campo visivo dell'immagine che non contiene il campione).

Una volta che i dati dell'ipercubo sono stati pre-elaborati, la fase successiva di analisi comunemente seguita è l'analisi esplorativa. Le bande spettrali nel NIR (e in una certa misura anche quelle nel MIR, ovvero nel *Mid InfraRed*) tendono ad essere più ampie ed altamente correlate, risulta quindi vantaggioso utilizzare le informazioni contenute in tutte le lunghezze d'onda simultaneamente. Proprio per questo, si utilizza un'approccio multivariato o "chemiometrico". L'analisi in componenti principali (PCA) è un approccio chemiometrico di base utile per l'analisi esplorativa dei dati HSI, il quale consente di esplorare tutte le possibili fonti di variabilità all'interno di un ipercubo. L'analisi esplorativa è uno strumento qualitativo. Di solito è necessaria un'analisi quantitativa, i.e. per classificare pixel al fine di identificare la presenza di un eventuale contaminante o di stimarne la concentrazione sulla superficie di un campione. Per quantificare o classificare i dati di un ipercubo è necessario sviluppare un modello robusto ed affidabile.

Per calibrare un modello, si possono utilizzare delle determinate porzioni dell'immagini, i.e. *Region of Interest* (ROI), quindi dei sottoinsiemi di spettri. Un esempio di questo processo è schematizzato in [Figura 2.2.4a](#). Nell'esempio vengono selezionate delle ROI etichettate con le variabili fittizie 0 ed 1 che identificano due classi. In alternativa possono essere utilizzati algoritmi che scelgono gli spettri all'interno dell'intero dominio

spaziale in maniera casuale, i.e. utilizzando l' algoritmo di Kennard e Stone (*Kennard & Stone, 1969*). In questo caso, sarà necessario far attenzione ad aver rimosso eventuali *outliers*.

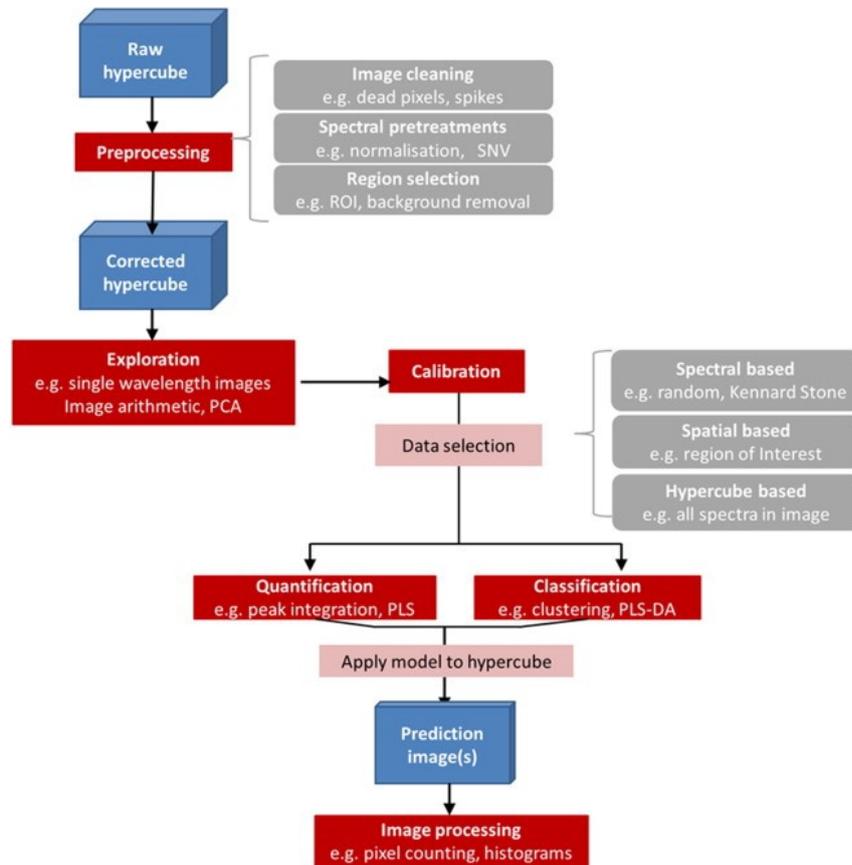


Figura 2.2.3. Diagramma di flusso che mostra i tipici passaggi richiesti nell'analisi d'immagine iperspettrale (Gowen et al., 2015).

Esistono numerosi algoritmi di classificazione *pixel-based* o *object-based*: metodi non supervisionati (i.e. *K-neighbors neighbors*, k-NN) e metodi supervisionati (i.e. *Partial Least Squares – Discriminant Analysis*, PLS-DA). Qualsiasi modello sviluppato a partire da dati spettrali può essere applicato a ciascun pixel presente in un ipercubo, risultando in una mappa di predizione (Figura 2.2.4b).

Un approccio simile può essere seguito per effettuare una modellazione quantitativa (i.e. *Partial Least Squares*, PLS).

Nel capitolo successivo verranno descritte con maggior dettaglio le tecniche chemiometriche utilizzate nei casi di studio presentati in questa tesi.

Teoria

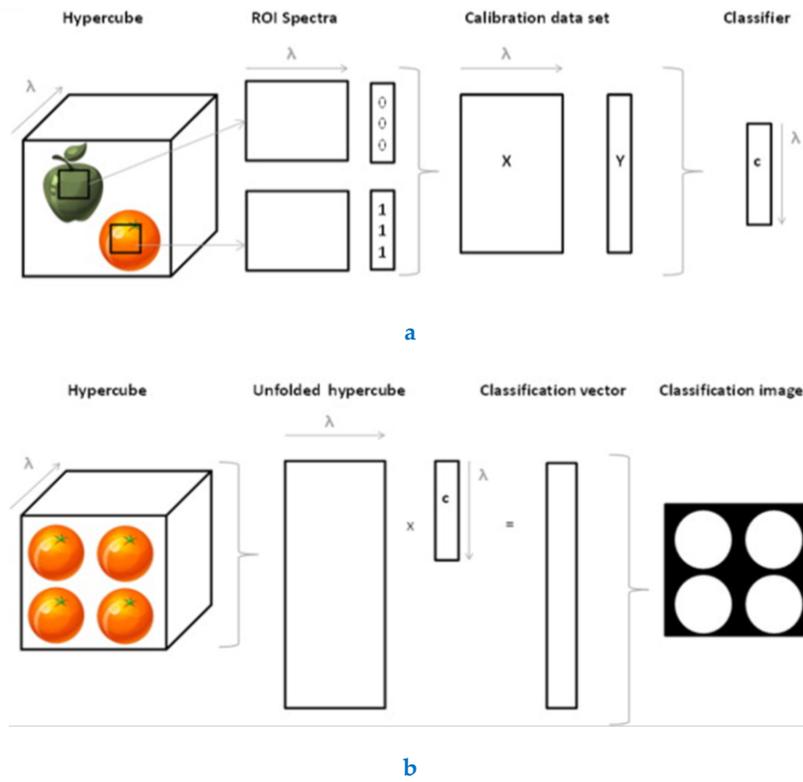


Figura 2.2.4. Rappresentazione schematica dello sviluppo di un modello di classificazione a parti da un ipercubo di dati (a) e applicazione del modello di classificazione ad un nuovo ipercubo (b); (Gowen et al., 2015).

2.3 TECNICHE DI STATISTICA MULTIVARIATA E METODI CHEMIOMETRICI

Nel 2002, l'*International Chemiometrics Society* ha definito la chemiometria come quella scienza che mette in relazione, tramite metodi matematici o statistici, le misure effettuate su un sistema o processo fisico con lo stato del sistema stesso.

Tuttavia, la definizione di chemiometria è un argomento, tutt'oggi, ancora di discussione e non è ancora stata trovata una definizione univoca, nonostante esistano differenti riviste internazionali e numerose società scientifiche dedicate alla chemiometria.

Nel 1974, *Svante Wold* coniò la parola chemiometria al fine di descrivere quella disciplina che è dedicata all'estrazione di importanti informazioni a partire da esperimenti chimici. Successivamente, egli cercò di ridefinire la parola chemiometria, come quell'insieme di processi e metodologie dedite all'estrazione e alla rappresentazione delle informazioni contenute in dati chimici misurati (*World, 1995*).

Secondo *Massart et al.*, invece, la chemiometria è quella disciplina che usa la matematica, la statistica e la logica per: i) definire o selezionare procedure sperimentali ottimali, ii) analizzare e massimizzare il contenuto informativo presente all'interno di un set di dati iii) ottenere un certo grado di "conoscenza" del sistema (*Massart et al., 1998*).

L'utilizzo di metodi chemiometrici implica analisi di dati multivariati, in cui diversi campioni correlati vengono analizzati contemporaneamente. D'altro canto, per statistica multivariata, invece, s'intende quella parte della statistica in cui l'oggetto dell'analisi è per sua natura formato da almeno due componenti.

Quando si ha a che fare con dati complessi è necessario, se non fondamentale, adottare un approccio multivariato. Un sistema complesso è, per natura, descritto da molteplici variabili. Quindi, per affrontare un problema complesso si ha la necessità di analizzare più variabili. Infatti, utilizzando e combinando più variabili, si possono ottenere maggiori informazioni sul sistema che si sta analizzando.

In generale, i dati da analizzare con un approccio multivariato sono disposti in tabelle o matrici. Un esempio è quello di una matrice, le cui righe rappresentano i campioni e nella quale ciascuna colonna rappresenta una variabile che caratterizza il sistema complesso. Adottando, quindi, un approccio chemiometrico per l'analisi dei dati, vengono considerate simultaneamente tutte quante le variabili che entrano in gioco e di conseguenza, l'informazione che si ottiene è rappresentativa del sistema nella sua totalità.

Negli anni, sono state messe a punto numerose tecniche di analisi multivariata, le quali possono essere suddivise in cinque filoni principali:

- i) Tecniche di *Design of Experiment* (DoE), le quali hanno come obiettivo principale quello dell'ottimizzazione delle strategie di analisi al fine di massimizzare il contenuto informativo dei dati analizzati e di minimizzare le risorse necessarie per effettuare l'analisi dei dati.

- ii) Tecniche di analisi esplorativa, come ad esempio la *Principal Component Analysis* (PCA), le quali hanno come obiettivo quello di ottenere una conoscenza approfondita del sistema studiato al fine di definire in maniera strategica le analisi successive.
- iii) Tecniche di regressione, come ad esempio la *Partial Least Squares* (PLS), le quali consistono nel mettere a punto un modello matematico in grado di stimare quantitativamente un determinato set di variabili (variabili risposta) utilizzando un altro set di dati (descrittori).
- iv) Tecniche di classificazione, come ad esempio la *Partial Least Squares – Discriminant Analysis* (PLS-DA), le quali consistono nel mettere a punto un modello matematico in grado di riconoscere le classi di appartenenza dei campioni studiati, a partire da un dato set di variabili sperimentali (descrittori).

In questo paragrafo è riportato un focus sulle tecniche di statistica multivariata e i metodi chemiometrici utilizzati nei casi di studio (vedere capitolo *IV. Casi di studio*).

2.3.1 Metodi di pre-elaborazione di dati

La pre-elaborazione dei dati, o in inglese *pre-processing*, è una tecnica di *data mining* che consiste nella manipolazione/trasformazione di dati grezzi in un nuovo formato maggiormente comprensibile. I dati grezzi che descrivono parametri fisico-chimici del mondo reale sono molto spesso incompleti o incoerenti e possono contenere errori di misura e/o effetti rumorosi di varia natura. Questi fattori, se non opportunamente trattati, potrebbero causare problemi di interpretazione dei risultati, nascondendo determinati comportamenti (o *trend*) ed influenzando la qualità stessa dei risultati (*Daszykowski et al., 2007; Engel et al., 2013*).

Nel caso di dati discreti, la pre-elaborazione dati viene utilizzata, in genere, per omogenizzare differenti tipologie di dati e/o esaltarne le differenze inter- ed intra- classe.

La pre-elaborazione dei dati spettrali o segnali in genere, invece, non solo è necessaria per correggere rumori strumentali, effetti di *scattering* e/o altri effetti rumorosi, ma anche per esaltare le differenze tra i diversi campioni e per risolvere al meglio le informazioni ricercate nel dominio spettrale o del segnale (*Rinnan et al., 2009*). L'obiettivo principale di questi metodi di pre-elaborazione è quello di fornire un "peso" adeguato a ciascuna variabile analizzata e a trattare quelle caratteristiche, dipendenti dalla tecnica analitica, che possono influenzare negativamente i risultati dell'analisi dei dati.

La scelta del metodo più appropriato di pre-elaborazione dei dati è strettamente correlata con le caratteristiche dei dati da analizzare e rappresenta, inoltre, una fase cruciale nell'analisi dati; in quanto, durante questa fase si corre il rischio di rimuovere variazioni significative e/o di introdurre artefatti nel set di dati.

La pre-elaborazione può essere effettuata sulle colonne del set di dati (*column-wise*), variabile per variabile, oppure sulle righe (*row-wise*), segnale per segnale ([Oliveri et al., 2010](#)).

2.3.1.1 Metodi di pre-elaborazione *column-wise*

Gli algoritmi di pre-elaborazione *column-wise* mirano a correggere le differenze sistematiche che possono essere presenti tra le variabili. I metodi di pre-elaborazione *column-wise* maggiormente utilizzati sono principalmente due: *Mean Center* e *Autoscale*.

Mean Center (MC)

In generale, gli algoritmi di *Centering* sono basati sull'equazione:

$$\mathbf{X}_{(C)} = \mathbf{X} - \mathbf{1} \bar{\mathbf{x}}' \quad (2.3.1)$$

Dove $\mathbf{X}_{(C)}$ rappresenta la matrice di dati centrati, $\mathbf{1}$ è un vettore colonna di valori unitari e $\bar{\mathbf{x}}'$ è un vettore che rappresenta il punto di riferimento di ciascuna variabile ([Eigenvector Research Inc. 2012](#)).

Il *Mean Center* (MC) è un metodo che, di per sé, non modifica la varianza dei dati e viene utilizzato per esaltare il carattere informativo della differenza presente in ciascuna variabile, annullando il contributo fornito dal valore medio. In particolare, l'algoritmo MC calcola la media di ciascuna colonna e poi la sottrae da ciascuna colonna, ovvero, per ciascuna variabile del vettore \mathbf{x}_k , sottrae il valore medio di ciascun elemento del vettore \mathbf{x}_{ik} :

$$x_{ik(MC)} = x_{ik} - \bar{x}_k \quad (2.3.2)$$

Graficamente, questa procedura corrisponde ad una traslazione dell'origine degli assi in corrispondenza del centroide dei dati.

Autoscale

L'utilizzo dell'algoritmo *Autoscale* è necessario quando vengono utilizzate variabili con differenti unità di misura, all'interno di uno stesso set di dati. Le differenze di magnitudine delle variabili analizzate potrebbero, infatti, mascherare importanti informazioni ([Eigenvector Research Inc. 2017a](#)).

Per fornire lo stesso "peso" alle variabili analizzate, ogni colonna k della matrice di dati \mathbf{X} viene scalata in modo da avere varianza unitaria. Tale metodo si basa sull'utilizzo di MC, seguito dalla divisione di ciascuna colonna (variabile) per la deviazione standard della stessa (s_k), come descritto dalla seguente equazione:

$$x_{ik(Autoscale)} = \frac{x_{ik} - \bar{x}_k}{S_k} \quad (2.3.3)$$

2.3.1.2 Metodi di pre-elaborazione row-wise

Gli algoritmi di pre-elaborazione che lavorano sulle righe della matrice sono utilizzati per correggere le differenze sistematiche che possono essere presenti nei segnali (i.e. rumori strumentali, *offset*, effetti di *scattering*, ecc.). In letteratura, sono descritti numerosi algoritmi di pre-elaborazione *column-wise* (Rinnan et al., 2009; Brown et al., 2009; Oliveri et al., 2010; Eigenvector Research Inc. 2014). Ivi sono riportati alcuni di questi metodi che sono stati utilizzati nei casi di studio considerati (Vedere capitolo IV. *Casi di studio*).

Normalize

Nelle applicazioni spettroscopiche, le differenze in scala sono dovute prevalentemente ad effetti di *scattering*, variazioni ambientali che influenzano la fonte illuminante e/o interferiscono con il *detector*. I metodi di normalizzazione del campione tentano di correggere questo tipo di effetti identificando alcuni aspetti di ciascun campione che dovrebbero essere essenzialmente costanti da un campione all'altro e correggendo il ridimensionamento di tutte le variabili in base a questa caratteristica. Il *pre-processing Normalize* utilizza differenti metodi di calcolo, come ad esempio:

- *1-Norm*: si normalizza (si divide ciascuna variabile) per la somma del valore assoluto di tutte le variabili del campione;
- *2-Norm*: si normalizza per la radice quadrata della somma dei valori al quadrato di tutte le variabili di un dato campione.

Standard Normal Variate (SNV)

L'algoritmo *Standard Normal Variate* (SNV) viene generalmente utilizzato per ridurre gli effetti di *scattering* e/o l'influenza delle differenti dimensioni delle particelle analizzate negli spettri (Barnes et al., 1989; Eigenvector Research Inc. 2017b). Lo spettro corretto dall'algoritmo SNV è dato dalla formula:

$$X_{i(SNV)} = \sqrt{\frac{\sum_{j=1}^n (X_{i,j} - \bar{x}_i)^2}{(n-1)}} \quad (2.3.4)$$

Dove n è il numero di spettri, $X_{i,j}$ è il valore della j -esima variabile dell' i -esimo campione, mentre \bar{x}_i è la media degli spettri definita dall'equazione:

$$\bar{x}_i = \frac{\sum_{j=1}^n X_{i,j}}{n} \quad (2.3.5)$$

Multiplicative Scatter Correction (MSC)

L'algorithmo *Multiplicative Scatter Correction* (MSC) cerca di correggere gli effetti di *offset* e quelli di scala (*Eigenvector Research Inc. 2017b*). Tale correzione si ottiene regredendo uno spettro misurato rispetto ad uno spettro di riferimento e quindi correggendo lo spettro misurato usando la pendenza (e possibilmente l'intercetta) di questo adattamento. In particolare, tale metodo segue la procedura, descritta di seguito.

Sia \mathbf{x} un vettore colonna corrispondente a uno spettro da standardizzare ed \mathbf{r} un vettore corrispondente allo spettro di riferimento (spesso è lo spettro medio del set di dati di calibrazione). Questi due vettori vengono corretti con l'algorithmo MC, secondo le equazioni:

$$\mathbf{x}_c = \mathbf{x} - \bar{x} \mathbf{1} \quad (2.3.6)$$

$$\mathbf{r}_c = \mathbf{r} - \bar{r} \mathbf{1} \quad (2.3.7)$$

Dove: \mathbf{x}_c e \mathbf{r}_c sono dei vettori mediati e centrati, \bar{x} e \bar{r} sono le rispettive medie ed $\mathbf{1}$ è un vettore unitario.

Successivamente, Il fattore moltiplicativo sconosciuto b viene determinato utilizzando le seguenti formule:

$$\mathbf{r}_c b = \mathbf{x}_c \quad (2.3.8)$$

$$b = (\mathbf{r}_c^T \mathbf{r}_c)^{-1} \mathbf{r}_c^T \mathbf{x}_c \quad (2.3.9)$$

Lo spettro corretto dall'algorithmo MSC, $\mathbf{X}_{i(MSC)}$, sarà quindi dato dalla formula:

$$\mathbf{X}_{i(MSC)} = \frac{\mathbf{x}_c}{b} + \bar{r} \mathbf{1} \quad (2.3.10)$$

Probabilistic Quotient Normalization (PQN)

Il *Probabilistic Quotient Normalization* (PQN) è un algorithmo di *pre-processing* robusto simile all'algorithmo MSC basato sulla mediana. Questo algorithmo applica una prima normalizzazione delle righe di \mathbf{X} e successivamente MSC basato sulla mediana per il fitting delle righe (*Eigenvector Research Inc. 2015*).

Extended Multiplicative Scatter Correction (EMSC)

L'algorithmo *Extended Multiplicative Scatter Correction* (EMSC) è utilizzato per rimuovere gli effetti additivi e moltiplicativi correlati agli effetti di *scattering* (*Martens et al., 1991*). Questo algorithmo si comporta come un filtro che fa passare solo una certa parte del segnale (*Eigenvector Research Inc. 2016a*). Con l'EMSC, il segnale \mathbf{x} , con dimensioni $\{N \times 1\}$, è modellato come:

$$\mathbf{x} = [\mathbf{x}_{ref} \mathbf{V} \mathbf{S} \mathbf{P}] \mathbf{c} \quad (2.3.11)$$

Dove \mathbf{x}_{ref} è un vettore di riferimento con dimensioni $\{N \times 1\}$ e \mathbf{V} è una matrice con dimensioni $\{N \times k_v\}$ di polinomi riferiti alla scala dell'asse. Ad esempio, se la scala dell'asse è riportata in numeri d'onda, si avrà che:

$$\mathbf{V} = [1 \ v \ v^2 \ \dots] \quad (2.3.12)$$

Invece, la matrice \mathbf{S} , con dimensioni $\{N \times k_s\}$ rappresenta la parte di segnale che è stato fatto passare dal filtro; mentre, la matrice \mathbf{P} con dimensioni $\{N \times k_p\}$ rappresenta la parte di segnale filtrata.

Il vettore \mathbf{c} contiene, invece, i coefficienti che verranno stimati utilizzando il metodo dei minimi quadrati:

$$\mathbf{c}^T = \mathbf{c}_{ref}^T + \mathbf{c}_v^T + \mathbf{c}_s^T + \mathbf{c}_p^T \quad (2.3.13)$$

I coefficienti stimati saranno poi utilizzati per “correggere” il segnale. Il segnale corretto con l’algoritmo EMSC sarà dunque dato da:

$$\mathbf{X}_{i(EMSC)} = \frac{(\mathbf{x} - \mathbf{V}\mathbf{c}_v - \mathbf{P}\mathbf{c}_p)}{\mathbf{c}_{ref}} \quad (2.3.14)$$

Detrend

Il metodo di pre-elaborazione dati denominato *Detrend* viene spesso utilizzato quando sono presenti problemi di *offset* (Barnes et al., 1989). Questo metodo cerca di adattare un polinomio ad un determinato ordine all'intero campione e successivamente lo va a sottrarre allo stesso (Eigenvector Research Inc. 2013a).

Metodi di derivazione

I metodi di derivazione sono utilizzati generalmente per: migliorare la risoluzione del segnale, separare i contributi dei picchi sovrapposti in dati spettrali, accentuare piccole differenze nella struttura di segnali quasi identici e per rimuovere *offset* di *baseline* (Oliveri et al., 2010). In particolare, la prima derivata di un segnale $y = f(x)$, ovvero $y' = dx/dy$, è rappresentativa della pendenza della tangente al segnale in ogni suo punto e consente la correzione di effetti additivi. Mentre, la derivata seconda ($y'' = d^2x/d^2y$) rappresenta la curvatura del segnale originale (appunto la derivata della pendenza) e consente la correzione sia di effetti additivi che di effetti moltiplicativi.

Metodo di derivazione Savitzky-Golay (S-G)

Il metodo di derivazione *Savitzky-Golay (S-G)* è utilizzato per rimuovere la *baseline* non significativa del campione ed è molto utile quando le variabili sono fortemente correlate tra loro (Eigenvector Research Inc. 2013a). Questo algoritmo si basa sul calcolo delle derivate delle risposte misurate rispetto alla variabile numerica (indice) o ad un'altra scala

dell'asse rilevante (i.e. lunghezza d'onda, numero d'onda ecc.). Per il calcolo delle derivate ci si basa sul metodo delle differenze finite: la derivata prima viene stimata come la differenza tra due punti successivi dello spettro, mentre la derivata seconda viene stimata calcolando la differenza tra due punti successivi derivati al primo ordine, secondo le equazioni:

$$x'_{i(S-G)} = x_i - x_{i-1} \quad (2.3.15)$$

$$x''_{i(S-G)} = x'_i - x'_{i-1} = x_{i-1} - 2x_i + x_{i+1} \quad (2.3.16)$$

Metodo di derivazione Norris-Williams (N-W)

Il metodo di derivazione *Norris-Williams* (N-W), o *gap-segment derivatives* è stato sviluppato per limitare l'inflazione dell'errore alle differenze finite (Norris, 1983; Hopkins, 2001). L'applicazione dell'algorithmo di derivazione N-W si basa su due fasi principali. La prima fase consiste nello *smoothing* degli spettri, effettuando una media lungo un determinato numero di punti (i.e. in una determinata finestra):

$$X_{smooth,i} = \sum_{k=-m}^m \frac{X_{org,i+j}}{2m+1} \quad (2.3.17)$$

Dove: m sono il numero di punti nella finestra in cui avviene lo *smoothing*, centrata nell'intorno del punto di misura i .

Nella seconda fase, per la derivata del primo ordine si prende la differenza tra due valori (su cui è stato applicato lo *smoothing*) con un determinato valore di "gap" tra di loro (maggiore di zero); mentre, per la derivata seconda si prende la differenza tra i valori, con una distanza definita pari al "gap" (maggiore di zero), su ciascuno degli estremi dell'intorno di i ed il doppio del valore nel punto i . Per cui, si avrà che:

$$X'_{i(N-W)} = X_{smooth,i+gap} - X_{smooth,i-gap} \quad (2.3.18)$$

$$X''_{i(N-W)} = X_{smooth,i-gap} - 2 \cdot X_{smooth,i} + X_{smooth,i+gap} \quad (2.3.19)$$

Smoothing

Uno dei modi più semplici per attenuare i dati fluttuanti è la media mobile (Savitzky et al., 1964). L'algorithmo di *Smoothing Savitzky-Golay* (S-G) mira ad aumentare il rapporto segnale-rumore dei segnali utilizzando un filtro passa-basso per rimuovere i componenti del rumore ad alta frequenza (Eigenvector Research Inc. 2013a). Questo filtro si basa sull'assunto che, in un segnale, le variabili adiacenti contengono informazioni simili e che quindi possono essere mediate per ridurre il rumore punto-punto.

2.3.1.3 Altri metodi di pre-elaborazione dati

Di seguito verranno descritti alcune tipologie di pre-elaborazione che possono essere ascrivibili in molteplici categorie, come ad esempio metodi di pre-elaborazione strumento-specifici (i.e. la *Splice Correction*) e i filtri multivariati (i.e. *Generalized Least Squares – Weighting*).

Splice correction (SC)

L'algoritmo *Splice Correction* consente l'eliminazione dei "salti" (*gap*) presenti nei segnali in riflettanza acquisiti con lo spettrofotometrietro FieldSpec 4® Standard-Res (per ulteriori informazioni su questo strumento si rimanda al capitolo [III. Strumenti utilizzati](#)), localizzati tra i vari domini di acquisizione dei tre detector dello strumento: VIS, SWIR 1 e SWIR 2 ([Danner et al., 2015](#)). Le transizioni critiche tra tali domini sono situate alle lunghezze d'onda $\lambda = 1000$ nm e $\lambda = 1800$ nm. La *Splice Correction* ha come obiettivo quello di compensare le differenze tra i punti di riflettanza $R_{1000\text{ nm}}$ and $R_{1001\text{ nm}}$, adattando tutti i valori a partire da 1001 nm fino ad arrivare al livello della riflettanza a 1000 nm. Matematicamente, la *Splice Correction* è definita dall'equazione:

$$R_{i,splice} = \begin{cases} R_i, & i \in [1; 1000] \\ R_i - f_{1000}, & i \in [1001; 1800] \\ R_i - f_{1800}, & i \in [1801; 2500] \end{cases} \quad (2.3.20)$$

$R_{i,splice}$ rappresentano i nuovi valori corretti, in corrispondenza delle lunghezze d'onda delle transizioni critiche. Mentre f_{1000} e f_{1800} rappresentano i "bias" che sono aggiunti ai valori originali. f_{1000} ed f_{1800} sono espressi matematicamente dalle formule (2.3.21) e (2.3.22), rispettivamente.

$$f_{1000} = R_{1001\text{ nm}} - (2 \cdot R_{1000\text{ nm}} - R_{999\text{ nm}}) \quad (2.3.21)$$

$$f_{1800} = R_{1801\text{ nm}} - (2 \cdot R_{1800\text{ nm}} - R_{1799\text{ nm}}) \quad (2.3.22)$$

Filtro Baseline

Il metodo di pre-elaborazione *Baseline* consente di adattare un polinomio ad uno specifico numero di punti, noti come punti di base ([Eigenvector Research Inc., 2013](#)). Questo metodo viene utilizzato in applicazioni spettroscopiche dove alcune variabili del segnale sono tipicamente caratterizzate dal *baseline* (ovvero il *background*).

Generalized Least Squares - Weighting (GLS-W)

L'algoritmo di pre-elaborazione *Generalized Least Squares - Weighting* (GLS-W) è un filtro che viene calcolato sulla base delle differenze tra campioni simili ([Eigenvector Research Inc.](#)

2013b). Il filtro GLS-W cerca di ridurre il peso di queste differenze tra i campioni. Le differenze sono dette anche interferenze o *clutter*.

Nel caso di un problema di classificazione, i campioni simili faranno parte della stessa classe. Una qualsiasi variazione all'interno di una classe (varianza intra-classe), può essere vista come interferenza, il che renderebbe la classificazione più difficile. L'obiettivo principale del GLS-W è quello di rimuovere maggior parte della varianza intra-classe, senza rendere le classi troppo simili tra loro (influenzando la varianza inter-classe). L'algoritmo GLS-W si basa sul calcolo della matrice di covarianza dalle differenze presenti tra campioni simili. Nel caso di un problema di trasferimento di calibrazione tra differenti strumenti, la differenza viene calcolata tra due gruppi di campioni, sottoposti a pre-elaborazione MC. Date due matrici di campioni, \mathbf{X}_1 e \mathbf{X}_2 , queste verranno pre-elaborate con MC e successivamente ne verrà calcolata la differenza:

$$\mathbf{X}_{1,mc} = \mathbf{X}_1 - \mathbf{1}\bar{x}_1 \quad (2.3.23)$$

$$\mathbf{X}_{2,mc} = \mathbf{X}_2 - \mathbf{1}\bar{x}_2 \quad (2.3.24)$$

$$\mathbf{X}_d = \mathbf{X}_2 - \mathbf{X}_1 \quad (2.3.25)$$

Dove: $\mathbf{1}$ è un vettore contenente solo valori unitari, uguale in lunghezza al numero di righe di \mathbf{X}_1 , \bar{x}_1 è la media di tutte le righe di \mathbf{X}_1 e \bar{x}_2 è la media di tutte le righe di \mathbf{X}_2 .

Successivamente, viene calcolata la matrice di covarianza \mathbf{C} (riportata nella formula 2.3.26), a cui è applicata una decomposizione ai valori singoli (come riportato in equazione 2.3.27), detta anche SVD (dall'inglese, *Singular Value Decomposition*).

$$\mathbf{C} = \mathbf{X}_d^T \mathbf{X}_d \quad (2.3.26)$$

$$\mathbf{C} = \mathbf{V} \mathbf{S}^2 \mathbf{V}^T \quad (2.3.27)$$

Dove: \mathbf{V} , vettore singolare sinistro, rappresenta gli autovettori di $\mathbf{V} \mathbf{V}^T$; \mathbf{V}^T , vettore singolare destro, rappresenta gli autovettori di $\mathbf{V}^T \mathbf{V}$; mentre, \mathbf{S} è la matrice diagonale dei valori singolari. A partire da \mathbf{S} , verrà successivamente la matrice \mathbf{D} , come:

$$\mathbf{D} = \sqrt{\frac{\mathbf{S}^2}{\alpha} + \mathbf{1}_D} \quad (2.3.28)$$

Dove $\mathbf{1}_D$ è una matrice diagonale e α è un parametro che consente di modulare l'intensità con la quale attenuare le interferenze o *clutter*. In genere, per valori elevati di α (i.e. per valori di $\alpha > 0.001$), l'effetto del filtro si attenua; mentre, per valori bassi di α (i.e. per valori di $\alpha < 0.001$), l'effetto del filtro si amplifica.

La matrice filtro $\mathbf{X}_{(GLS-W)}$ sarà quindi data da:

$$\mathbf{X}_{(GLS-W)} = \mathbf{V} \mathbf{D}^{-1} \mathbf{V}^T \quad (2.3.29)$$

2.3.2 Tecniche di analisi esplorativa e semplificazione dei dati

Negli ultimi decenni, il continuo miglioramento tecnologico ha portato allo sviluppo di strumentazione analitica sempre più potente ed automatizzata. Di pari passo con lo sviluppo di nuove tecnologie, si è assistito ad un considerevole aumento della quantità di dati che possono essere raccolti in tempi brevi, al fine di caratterizzare in modo efficiente i campioni oggetto di analisi. Quando si affronta l'analisi di una serie di dati sperimentali, ottenere una panoramica della struttura dei dati è un compito non affatto banale. In questo contesto, la *Principal Component Analysis* (PCA), o analisi delle componenti principali, può essere considerata come la base sulla quale costruire una qualsiasi procedura di analisi multivariata.

2.3.2.1 *Principal Component Analysis* (PCA)

Nell'ambito della statistica multivariata, la PCA è una tecnica utilizzata per la semplificazione dei dati da analizzare (*Beebe et al., 1998; Wold et al., 1987*). Infatti, la PCA rappresenta, in genere, la prima fase da affrontare quando si ha a che fare con dati multivariati complessi. La PCA è un potente mezzo per effettuare una analisi esplorativa dei dati, la quale consente di valutare una intera matrice di dati al fine di descrivere la struttura dei campioni e delle variabili che entrano in gioco, nonché di valutare possibili correlazioni esistenti tra i campioni e le variabili analizzate.

Lo scopo di questa tecnica di analisi multivariata è quello di ridurre il numero di variabili che descrivono una matrice di dati ad un numero minore di variabili latenti o "componenti principali" (*Principal component*, o abbreviato PC), che sono una combinazione lineare dei descrittori originali (variabili sperimentali), limitando il più possibile la perdita di informazioni. In altre parole, la PCA consente di rappresentare i dati in spazi dimensionali più piccoli, dove gli assi non sono più rappresentati dalle m variabili originali ma dalle componenti principali, definite in modo che: la prima (PC1) descrive la massima varianza dei dati, la seconda (PC2) è ortogonale alla prima e descrive la rimanente massima varianza dei dati e così via per le altre componenti principali (*Figura 2.3.1*). Ad ogni modo, la PCA non solo modella le variazioni di particolare rilevanza in un set di dati, ma anche gli errori casuali di misura. Le componenti principali descrivono la varianza spiegata in percentuali progressive sempre inferiori (i.e. la PC1 spiega il 70% di varianza dei, la PC2 spiega il 10 % di varianza, la PC3 il 5 % e così via). Le percentuali più piccole di varianza spiegata, in genere, rappresentano le differenze presenti nel set di dati che sono associate prevalentemente ad effetti rumorosi e/o ad errori casuali di varia natura.

Questa tecnica chemiometrica è in grado di estrarre i pattern dominanti di una matrice di dati X con dimensioni $\{n \times m\}$, ottenuta misurando m variabili su n campioni, in termini di prodotto di due matrici più piccole: la matrice degli *scores* e quella dei *loadings*.

Matematicamente, questo metodo consiste nella scomposizione della matrice di dati \mathbf{X} in una parte di “struttura” e una parte di “rumore”, secondo la formula:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} = \sum_{a=1}^A \mathbf{t}_a \mathbf{p}_a + \mathbf{E} \quad (2.3.30)$$

Dove \mathbf{TP}^T è la parte di “struttura” ed \mathbf{E} è la parte di “rumore”, con:

- i) \mathbf{T} che è la matrice degli *scores*, la quale tiene conto della varianza associata ai campioni;
- ii) \mathbf{P} che è la matrice dei *loadings*, la quale include informazioni della varianza associata alle variabili;
- iii) ed \mathbf{E} che è la matrice dei *residui* (*residuals*), la quale descrive variazioni non-sistematiche.

La fattorizzazione o decomposizione della matrice di dati è schematizzata in [Figura 2.3.2](#).

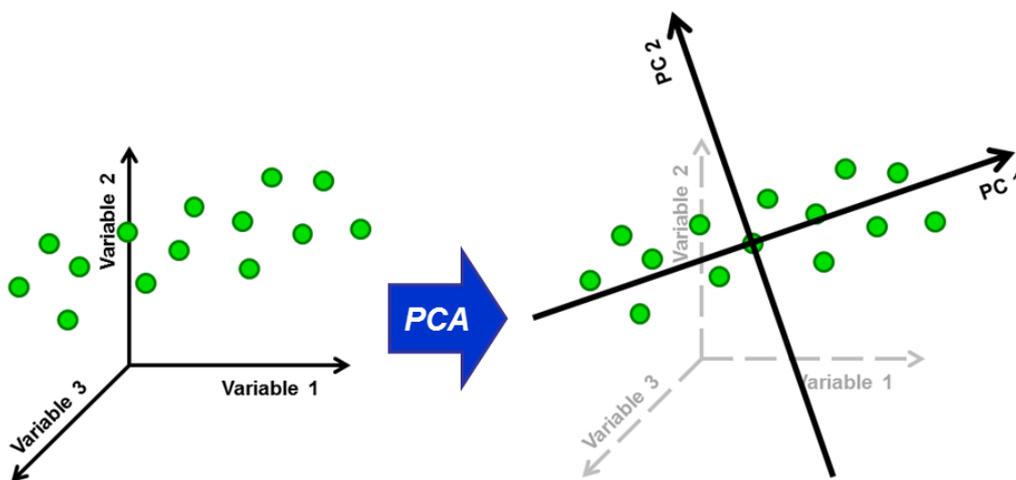


Figura 2.3.1. Definizione del nuovo spazio delle componenti principali, a seguito di analisi in componenti principali, con riferimento al dominio dei dati originale.

Entrando più nel dettaglio, i valori degli *scores* rappresentano i valori delle coordinate del campione, proiettate nel nuovo spazio delle componenti principali. Ad esempio, nel caso presentato [Figura 2.3.1](#) dove per descrivere il set di dati sono necessarie solamente 2 componenti principali, il campione i -esimo (con $1 \leq i \leq n$), proiettato nel “nuovo” spazio bidimensionale, è definito da due valori degli score: $t_{i,1}$ e $t_{i,2}$ che corrispondono alle coordinate del campione i -esimo, rispettivamente sulle componenti principali PC1 e PC2.

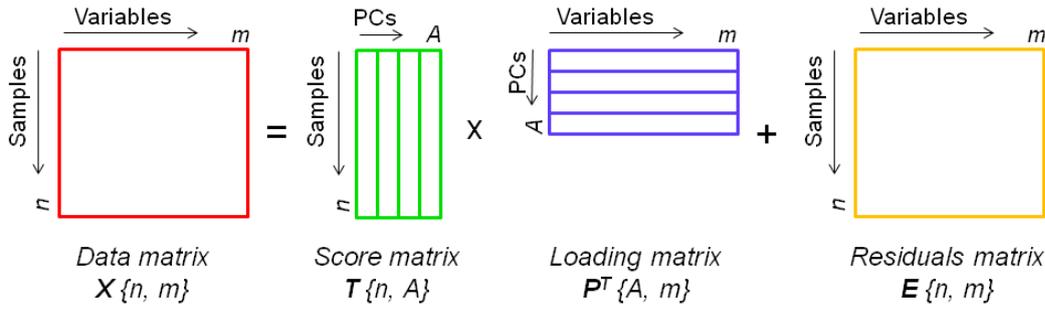


Figura 2.3.2. Fattorizzazione di una matrice di dati, secondo la tecnica della PCA.

Rappresentando graficamente la posizione di tutti i campioni nello spazio delle componenti principali (il cosiddetto *score plot*), è quindi possibile valutare la struttura dei campioni e identificare eventuali pattern presenti, come gruppi di dati caratterizzati da elementi simili (detti *cluster*) e / o possibili valori anomali (i cosiddetti *outliers*, identificabili dalle distanze relative dagli altri *scores*).

Durante un'analisi esplorativa, l'individuazione di possibili valori anomali è di fondamentale importanza, in quanto la loro presenza può influenzare in modo significativo i risultati della successiva analisi dei dati. Al fine di valutare se i campioni identificati come anomali debbano essere eliminati o esclusi dalle analisi, si deve tener conto di particolari aspetti del sistema oggetto di studio. In particolare, si devono considerare due parametri statistici: la distanza di ciascun campione della proiezione dell'iper-piano nel modello, definita come *Q residual* e la distanza della proiezione di ciascun campione sul modello PCA a partire dal centro del modello stesso, definita come *Hotelling's T²*.

I *Q residui* sono calcolati come la somma dei quadrati di ciascuna riga della matrice dei *residui E*, secondo la seguente equazione riferita al campione *i*-esimo della matrice *X*, ovvero \mathbf{x}_i :

$$Q_i = \mathbf{e}_i \mathbf{e}_i^T = \mathbf{x}_i (\mathbf{I} - \mathbf{P} \mathbf{P}^T) \mathbf{x}_i^T \quad (2.3.31)$$

Dove: \mathbf{e}_i è la *i*-esima riga della matrice *E*, *P* è la matrice dei *loading* ed *I* è la matrice identità (con opportune dimensioni, a seconda del caso considerato).

I *Q residui* con valori elevati possono voler indicare che alcuni campioni non seguano lo stesso comportamento degli altri campioni, come mostrato in [Figura 2.3.3a](#).

Mentre, il parametro *Hotelling's T²* quantifica la variabilità di un campione nel modello e fornisce informazioni sul contributo di ciascun campione nella definizione del modello stesso. In particolare, *T²* è dato dalla somma dei quadrati normalizzati dei valori degli *score*:

$$T_i^2 = \sum_{a=1}^A \frac{t_{i,a}^2}{s_a^2} = \frac{t_{i1}^2}{s_1^2} + \frac{t_{i2}^2}{s_2^2} + \dots + \frac{t_{in}^2}{s_n^2} \quad (2.3.32)$$

Dove: $t_{i,a}$ è il valore dello score riferito all' i -esimo campione per la a -esima PC, mentre s_a^2 è la varianza dell' a -esimo vettore score t_a , riferito al modello PCA.

I campioni con T^2 elevati, come quello mostrato nell'esempio in [Figura 2.3.3b](#), sono quindi quelli che hanno maggiore influenza sul modello. Quindi, gli *outliers* possono essere identificati ed esclusi dalle successive analisi, esplorando il grafico dei parametri *Hotelling's T²* e i *Q residui*.

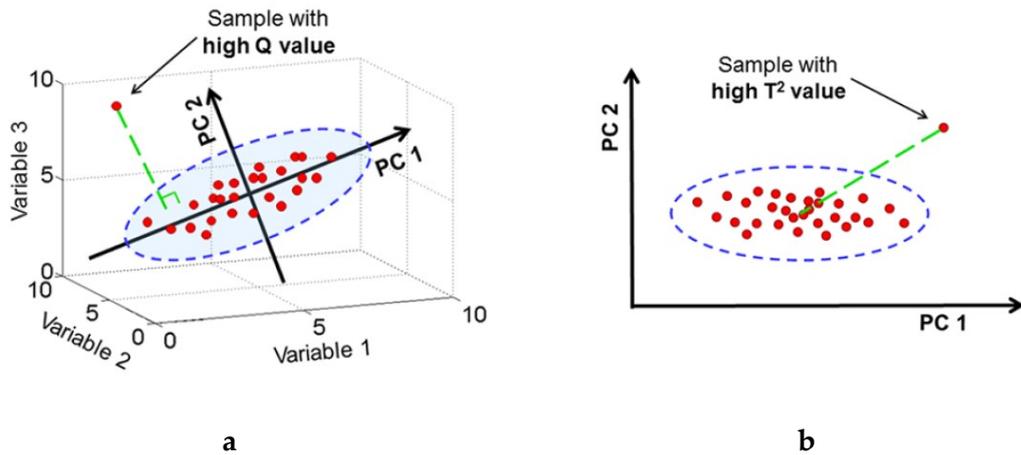


Figura 2.3.3. Individuazione di un campione con valore di *Q* residuo elevato (a) e con valore *Hotelling's T²* elevato (b).

Come accennato in precedenza, la PCA consente di analizzare anche la struttura delle variabili. La matrice dei *loadings* P , infatti, contiene l'informazione relativa alla nuova direzione rispetto a quella di ogni variabile originale per ogni PC o, in altre parole, l'informazione relativa al contributo di ogni variabile originale nel definire la direzione della componente principale considerata. Analogamente a quanto precedentemente descritto per i valori degli *scores*, è possibile rappresentare graficamente i valori dei *loadings* di due PC su uno *scatter plot*, al fine di valutare la posizione di ciascuna variabile nel nuovo spazio delle componenti principali.

I grafici dei *loading* (o *loading plot*) forniscono informazioni utili sulle correlazioni tra le variabili originali (ad esempio, le variabili che si trovano nella stessa direzione rispetto all'origine dello spazio delle componenti principali, sono correlate con il segno positivo, mentre le variabili posizionate in direzione opposta rispetto all'origine degli assi sono correlate con il segno negativo) e sulle variabili originali che hanno un certo peso nella descrizione dello spazio dei dati: le variabili con i valori dei *loadings* assoluti più elevati posseggono un maggiore peso sulle direzioni delle componenti principali. Quando le variabili originali sono molto numerose per cui non possono essere rappresentate in un grafico a dispersione (i.e. quando il set di dati è composto da segnali), i vettori *loading* possono essere riportati graficamente come una funzione del dominio originale (ad esempio, nel caso di spettri NIR, i *loadings* possono essere riportati in funzione delle

lunghezze d'onda). Ciò semplifica ulteriormente l'identificazione delle regioni di segnale più rilevanti, in quanto i profili dei *loadings* ottenuti sono direttamente confrontabili con gli spettri originali.

Inoltre, il confronto tra *scores plot* e *loading plot* consente di evidenziare le relazioni esistenti tra campioni e variabili originali, fornendo una migliore comprensione del sistema oggetto di studio. Ad esempio, quando un *cluster* di campioni si viene a generare nel primo quadrante del grafico degli *scores* su PC1-PC2, osservando il grafico dei *loading* corrispondente è possibile determinare quali variabili di questi campioni sono la causa di tale comportamento.

2.3.3 Tecniche di regressione per analisi quantitative

I metodi di regressione multivariata sono impiegati per effettuare predizioni quantitative relativamente ad una o più proprietà di un sistema oggetto di studio. Lo scopo di queste tecniche di regressione è quello di trovare la migliore relazione tra un insieme di variabili che descrive gli oggetti studiati ed un insieme di risposte misurate degli stessi oggetti ([Wise et al., 2006](#)).

La messa a punto di un qualsiasi modello di regressione multivariata richiede principalmente tre fasi: l'addestramento (*calibration* o *training*), la convalida incrociata (o *cross-validation*) su un set di campioni di calibrazione, e la validazione (*test / validation*) su un set esterno di campioni.

2.3.3.1 Multiple Linear Regression (MLR)

L'obiettivo della *Multiple Linear Regression* (MLR), conosciuta anche come *Ordinary Least Squares* (OLS), è quello di trovare una relazione tra una o più variabili indipendenti (dette predittori), ed una o più variabili dipendenti, le cosiddette risposte ([Yale University, 1997](#); [Wise et al., 2006](#)).

Nella sua forma più semplice, un modello lineare specifica la relazione (lineare) tra una variabile dipendente (risposta) y , ed un insieme di predittori, x , cosicché:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_px_p \quad (2.3.33)$$

Dove b_0 è il coefficiente di regressione per l'intercetta ed i valori b_i sono i coefficienti di regressione (per le variabili da 1 a p) calcolati dai dati.

Riscrivendo l'equazione (2.3.33) in forma matriciale, si avrà che:

$$\mathbf{X}\mathbf{b} = \mathbf{y} \quad (2.3.34)$$

Il vettore di regressione \mathbf{b} sarà determinato a partire da un insieme di misure \mathbf{X} , ed i valori noti della proprietà di interesse, \mathbf{y} .

Quindi, \mathbf{b} sarà definito come:

$$\mathbf{b} = \mathbf{X}^+ \mathbf{y} \quad (2.3.35)$$

Dove \mathbf{X}^+ è la pseudo-inversa di \mathbf{X} , definita, nel caso specifico di un modello MLR, come:

$$\mathbf{X}^+ = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \quad (2.3.36)$$

2.3.3.2 Regressione Partial Least Squares (PLS)

Il metodo di regressione *Partial Least Squares* (PLS) punta a ricercare le relazioni lineari esistenti tra una matrice descrittore \mathbf{X} e una matrice di risposta \mathbf{Y} (Geladi et al., 1986; Wold et al., 2001; Abdi, 2003). Analogamente alla PCA, la PLS si basa sul presupposto che il sistema studiato possa essere descritto da un set di variabili, chiamate variabili latenti, in inglese *Latent Variables* (LVs), che sono combinazioni lineari delle variabili originali. La regressione PLS viene utilizzata, in genere, quando a partire da un gran numero di variabili indipendenti deve essere predetto un certo numero di variabili dipendenti. L'obiettivo principale della regressione PLS è di predire la matrice delle risposte \mathbf{Y} , di dimensione $\{n \times p\}$, che memorizza le variabili dipendenti a partire da \mathbf{X} , la matrice dei predittori, con dimensioni $\{n \times m\}$ e contenente i predittori, per descrivere la struttura comune sottostante dei dati. La caratteristica principale di questo metodo di regressione è che la soluzione viene ottenuta tenendo conto delle informazioni di entrambe le matrici contemporaneamente. In particolare, entrambe le matrici \mathbf{X} e \mathbf{Y} sono scomposte individualmente in modo sostanzialmente simile a quanto avviene nella PCA. Matematicamente, la struttura principale di un modello PLS è descritta dalle formule:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (2.3.37)$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F} \quad (2.3.38)$$

Dove \mathbf{T} e \mathbf{U} rappresentano, rispettivamente, le matrici di proiezione (o degli *scores*) di \mathbf{X} e \mathbf{Y} , con dimensioni $\{n \times l\}$; mentre \mathbf{P} e \mathbf{Q} sono, rispettivamente, le matrici dei *loadings*, con dimensioni $\{m \times l\}$ e $\{p \times l\}$; le matrici \mathbf{E} e \mathbf{F} rappresentano i termini di errore.

Applicando questo processo di decomposizione, le direzioni delle variabili latenti vengono definite considerando che \mathbf{X} deve essere un buon predittore di \mathbf{Y} e che, pertanto, la variazione dell'oggetto del blocco descrittore, espressa dalla matrice di proiezione (o degli *scores*) di \mathbf{X} , \mathbf{T} può essere usata per descrivere \mathbf{Y} , invece della matrice di proiezione (o degli *scores*) di \mathbf{Y} , \mathbf{U} . Ciò può essere espresso dalla seguente formula:

$$\mathbf{Y} = \mathbf{TQ}^T + \mathbf{F} \quad (2.3.39)$$

Per migliorare il modello, cioè per spiegare \mathbf{Y} nel modo migliore possibile minimizzando la matrice \mathbf{F} dei residui, la fattorizzazione di \mathbf{X} deve essere ottimizzata in modo tale che la matrice \mathbf{T} contenga la variazione di \mathbf{X} , la quale consenta di descrivere \mathbf{Y} al meglio, o in altre parole, che la covarianza tra \mathbf{T} e \mathbf{U} sia massimizzata. Ciò si ottiene

calcolando, per ogni LV, un vettore di “peso” (*weight*), w , che tiene conto del contributo delle variabili originali alla spiegazione di Y .

Per calcolare un modello PLS, esistono diversi algoritmi, tra questi, i più diffusi sono il NIPALS (*Nonlinear Iterative Partial Least Squares*), che è un metodo iterativo (Wright, 2017), e il SIMPLS, un metodo più veloce del NIPALS ma maggiormente complesso (de Jong, 1993). Entrambi questi metodi forniscono esattamente gli stessi risultati quando Y contiene una singola variabile (i.e. matrice monovariata), mentre per matrici multivariate (ovvero contenenti più variabili) di dati Y si osservano delle leggere differenze, per quanto trascurabili.

L'equazione 2.3.39 può essere riarrangiata come l'equazione descrittiva di un modello di regressione multipla, dove la Y stimata, ovvero \hat{Y} , è data da:

$$\hat{Y} = \mathbf{XWQ}^T + \mathbf{F} = \mathbf{XB} + \mathbf{F} \quad (2.3.40)$$

In cui W è la matrice dei pesi e B rappresenta l'insieme dei coefficienti di regressione della PLS.

Per valutare l'importanza del contributo delle variabili descrittive al modello finale si devono prendere in considerazione entrambe le variabili significative alla descrizione di Y (i coefficienti di regressione con valori assoluti più significativi, b_{mk}) e le variabili rilevanti per la descrizione di X (i *loadings* con valori assoluti maggiori, p_{ak}). Le informazioni sul contributo delle variabili d'interesse possono essere sintetizzate dagli *scores* della *Variable Importance for Projection* (VIP). Il calcolo di questo parametro, per ciascuna variabile y , può essere effettuato grazie alla formula:

$$VIP_k = \sqrt{p \frac{\sum_{a=1}^A (SS(b_a t_a) (w_{ka} / \|w_a\|)^2)}{\sum_{a=1}^A (SS(b_k t_k))}} \quad (2.3.41)$$

Dove:

$$SS(b_a t_a) = b_a^2 t_a^T t_a \quad (2.3.42)$$

Generalmente, il criterio adottato per determinare se una certa variabile è effettivamente significativa è basato sul valore che assume il VIP. Secondo questo criterio, basato sul fatto che il valore medio dei quadrati dei VIP *scores* è pari ad uno, le variabili con valori $VIP > 1$ sono quelle che forniscono un contributo significativo alla previsione della variabile y . Tale criterio viene anche adottato come tecnica di selezione delle variabili.

2.3.3.3 Parametri di performance di una regressione multivariata

I due parametri principali utilizzati per valutare la bontà dei modelli di regressione sono: *Root Mean Square Error* (RMSE) e R^2 (Martens et al., 1989; Naes, 2002). Il parametro

RMSE fornisce la possibilità di valutare le differenze tra la risposta stimata \hat{Y} e la risposta osservata Y , ed è definito come:

$$RMSE_K = \sqrt{\frac{1}{N}(\hat{Y}_K - Y)^2} \quad (2.3.43)$$

Dove N è il numero dei campioni ed il pedice k indica il numero di variabili latenti. Il parametro $RMSE_K$ può riferirsi alla fase di calibrazione, di convalida incrociata o di predizione.

Nel caso dell'errore in calibrazione del *training set*, o *Root Mean Square Error of Calibration* (RMSEC), \hat{Y} rappresenta i valori calcolati dal modello per i campioni di *training* ed N è il numero di campioni presenti nel *training set*; invece, nel caso dell'errore in convalida incrociata, o *Root Mean Square Error of Cross-Validation* (RMSECV), \hat{Y} contiene i valori predetti grazie alla procedura di convalida incrociata ed N , come nel caso precedente, è il numero di campioni presenti nel *training set*; mentre, nel caso dell'errore in predizione del *test set*, o *Root Mean Square Error of Prediction* (RMSEP), N rappresenta il numero di campioni nel *test set*.

Mentre, R^2 è un parametro usato per valutare il fitting del modello:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (2.3.44)$$

Dove SS_{res} è la somma dei quadrati residui $e_i = y_i - \hat{y}_i$, ed è definito come $SS_{res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$; mentre, SS_{tot} è la somma dei quadrati totale definita come $SS_{tot} = \sum_{i=1}^n (y_i - \bar{y})^2$, con \bar{y} valore medio della risposta.

Anche in questo caso, il parametro R^2 può riferirsi alla fase di calibrazione (R_{Cal}^2), di convalida incrociata (R_{CV}^2) o di predizione (R_P^2).

Un altro interessante parametro statistico, utilizzato per valutare matematicamente la differenza sistematica che può essere presente tra i due set di risultati, ovvero quello dei risultati ottenuti con il metodo di riferimento e i risultati ottenuti in previsione (i.e. dati di previsione ottenuti da un set di dati spettroscopici) è il cosiddetto *Bias*. Il *Bias* viene calcolato come:

$$Bias_K = \sum_{i=1}^N \frac{(y - \hat{y}_K)}{N} \quad (2.3.45)$$

Dove: y è il valore di riferimento dell' i -esimo campione, \hat{y} è il valore predetto dell' i -esimo campione, N rappresenta il numero dei campioni, mentre il pedice k indica il numero di variabili latenti del modello. Anche, il parametro *Bias* può riferirsi alla fase di calibrazione, di convalida incrociata o di predizione.

2.3.4 Tecniche di classificazione supervisionate per analisi qualitative

I metodi di classificazione hanno l'obiettivo di costruire, sulla base di un certo numero di descrittori, un modello capace di individuare la classe di appartenenza degli oggetti. O in altre parole, un algoritmo di classificazione cerca di prevedere una proprietà y del campione, a partire da una matrice di descrittori X . La differenza chiave tra i metodi di calibrazione/regressione e quelli di classificazione è che, in quest'ultimo caso, la matrice di risposta Y riporta informazioni qualitative e non quantitative. L'appartenenza ad una classe è, quindi, una risposta qualitativa e classificare significa cercare un modello che metta in relazione tale risposta con le variabili indipendenti.

Per poter classificare è necessario disporre di un training set di dati nel quale ciascun oggetto è assegnato ad una classe. Spesso i set di dati sono costituiti da campioni appartenenti a diversi gruppi o classi. Dunque, le possibili classi devono essere preventivamente definite. Quindi, nella classificazione il modello attribuisce l'elemento analizzato ad una delle classi predeterminate.

Nel corso degli anni, sono stati proposti numerosi algoritmi di classificazione, tra cui, ad esempio: la *Partial Least Squares – Discriminant Analysis* - PLS-DA (Barker et al., 2003), la *k-Neighbor Neighbours* - k-NN (Gemperline, 2006), la *Linear Discriminant Analysis* o LDA, la *Quadratic Discriminant Analysis* - QDA, la *Regularized Discriminant Analysis* - RDA (Wu et al., 1996), la *Soft Independent Modeling of Class Analogy* - SIMCA (Wold, 1976), la *Gradient Boosted Tree Discriminant Analysis* - XGBoostDA (Tianqi et al., 2016), metodi basati sulla *Support Vector Machine* – SVM (Cortes et al., 1995), sulle reti neurali artificiali - ANN (Marini et al., 2008) e sulle trasformate *wavelet* (Cocchi et al., 2001).

Chiaramente, il calcolo della probabilità che un campione appartenga o meno ad una classe viene eseguito in modo diverso per ciascun metodo di classificazione.

In qualsiasi caso, la classificazione richiede principalmente tre fasi: l'addestramento (*calibration* o *training*), la convalida incrociata (o *cross-validation*) e la validazione (*test / validation*).

2.3.4.1 Matrice di confusione e parametri di performance di un classificatore

Per valutare la performance di un classificatore supervisionato, si valuta la matrice di confusione, o *confusion matrix* (Fawcett, 2006). In questa matrice, ogni riga rappresenta le istanze in una classe predetta, mentre ciascuna colonna rappresenta le istanze di una classe reale. Ogni colonna della matrice rappresenta i valori predetti, mentre ogni riga rappresenta i valori reali. L'elemento sulla riga i e sulla colonna j è il numero di casi in cui il classificatore ha classificato la classe "vera" i come classe j .

Se si considerasse un classificatore e un'istanza, avremmo quattro risultati, come mostrato in Tabella 2.3.1. In questo caso, se un'istanza positiva viene classificata come positiva si ha un *True Positive (TP)*; mentre, si ha un *False Negative (FN)* quando un'istanza positiva viene classificata come negativa; quando, invece, un'istanza negativa è

classificata come negativa si ha un *True Negative*, o *TN*; nel caso in cui, un'istanza negativa è classificata come positiva si ha un *False Positive*, o *FP*. A partire dalla matrice di confusione, possono essere calcolati differenti parametri, che forniscono informazioni sulle prestazioni del classificatore.

Tra i parametri più importanti, abbiamo: la *Sensitivity*, definita dall'equazione 2.3.46, che misura l'abilità del sistema di individuare istanze della classe positiva, ovvero la percentuale di istanze classificate positive tra tutte le istanze realmente positive, e la *Specificity*, complementare alla *Sensitivity*, ma calcolata sulla base della classe negativa con l'equazione 2.3.47.

Tabella 2.3.1. Generica matrice di confusione binaria (Fawcett, 2006).

| | | True class | |
|--------------------|---|-------------------------|-------------------------|
| | | p | n |
| Hypothesized class | Y | True Positives (TP) | False Positives (FP) |
| | N | False Negatives (FN) | True Negatives (TN) |

L'utilizzo di entrambi questi due parametri è ampiamente diffuso in campo medico/clinico. Per misurare la correttezza globale del sistema di classificazione, viene di solito usato il parametro *Accuracy*, o *Accuratezza*, calcolato secondo l'equazione 2.3.48. Un altro parametro di particolare importanza è la *Precision*, calcolato con l'equazione 2.3.49, che misura la precisione con cui è stata stimata la classe positiva, ovvero la percentuale di istanze classificate come positive che effettivamente lo sono. Altri parametri interessanti sono il *Misclassification Error*, complementare all'*Accuracy* - come riportato dalla formula 2.3.50 - ed il *Class Error*, definito dall'equazione 2.3.51.

$$Sensitivity = \frac{TP}{TP + FN} \quad (2.3.46)$$

$$Specificity = \frac{TN}{FP + TN} = 1 - Sensitivity \quad (2.3.47)$$

$$Accuracy = 1 - \frac{TP + FN}{TP + TN + FP + FN} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.3.48)$$

$$Precision = \frac{TP}{TP + FP} \quad (2.3.49)$$

$$Misclassification Error (ERR) = 1 - Accuracy \quad (2.3.50)$$

$$Class Error = 1 - \left(\frac{Sensitivity + Specificity}{2} \right) \quad (2.3.51)$$

Sia la matrice di confusione che i parametri calcolabili a partire dai valori ricavati da questa, si possono riferire alla fase di calibrazione, convalida incrociata e predizione del *test set*.

Nel caso dell'*imaging* iperspettrale la classificazione può essere effettuata secondo due logiche: la logica *pixel-based* e quella *object-based*.

Secondo la logica *pixel-based*, la classificazione viene eseguita a livello di pixel, utilizzando solo le informazioni spettrali disponibili per ogni singolo pixel. In questo caso, l'algoritmo di classificazione addestrato produrrebbe una previsione di classe per ogni singolo pixel di un'immagine.

Invece, secondo la logica *object-based*, la classificazione viene eseguita su un gruppo localizzato di pixel, tenendo conto delle proprietà spaziali di ciascun pixel in relazione tra loro. In questo caso, l'algoritmo di classificazione addestrato assegnerebbe in predizione una classe ad un gruppo di pixel.

Seguendo una logica *object-based*, una soglia del 50-60% di pixel correttamente classificati all'interno di un oggetto potrebbe essere utilizzata come discriminante per l'assegnazione di una classe. In questo caso, e più in generale per i classificatori a più *step*, le performance di classificazione possono essere descritte dagli indicatori *Recognition* (o *Accuracy*), *Error* e *Reliability*, secondo le seguenti equazioni (Silva et al., 2017):

$$Recognition = \frac{N_{Cor}}{N} \quad (2.3.52)$$

$$Error = \frac{N_{Mis}}{N} \quad (2.3.53)$$

$$Reliability = \frac{N_{Mis}}{N_{Mis} + N_{Cor}} \quad (2.3.54)$$

Dove N è il numero totale di campioni (o oggetti), mentre N_{Cor} ed N_{Mis} sono rispettivamente il totale dei campioni correttamente riconosciuti ed il totale dei campioni non riconosciuti correttamente.

2.3.4.2 Partial Least Squares – Discriminant Analysis (PLS-DA)

Tra gli algoritmi di classificazione, la PLS-DA ha avuto un ampio successo in diversi ambiti di applicazione. Questo metodo, in realtà, non è altro che una estensione della regressione PLS ai fini della classificazione, simile ad un altro metodo di classificazione chiamato *Linear Discriminant Analysis* - LDA (Barker et al., 2003; Wise et al., 2006). Nella PLS-DA viene utilizzata la regressione PLS per sviluppare un modello in grado di prevedere il numero di classe per ogni campione in studio.

In altre parole, la PLS-DA si basa essenzialmente sul calcolo di un modello PLS, che utilizza come matrice di risposta Y un insieme di variabili "fittizie" (o "dummy" variables), le quali codificano l'appartenenza ad una determinata classe (Ballabio & Consonni, 2013). Più nel dettaglio, per ogni classe modellata viene creato un vettore binario y , che assegna il valore 1 ai campioni appartenenti alla classe considerata e 0 ai campioni non appartenenti a quella classe. Una volta che il modello è stato calcolato, per ogni classe modellata viene calcolata la probabilità di ciascun campione di essere membro di quella classe. In questo contesto, quindi, un passaggio chiave è rappresentato dalla definizione del valore di soglia utilizzato per distinguere tra campioni appartenenti o meno ad una certa classe. Questo valore può essere stabilito applicando un metodo bayesiano o sulla base dei valori di errore riferiti alla convalida incrociata.

2.3.5 Calibrazione, convalida incrociata e validazione di un modello

Come già definito in precedenza, la messa a punto di modelli di regressione multivariata o di classificazione, passa per tre fasi principali: tre fasi: l'addestramento (*calibration* o *training*), la convalida incrociata (o *cross-validation*) su un set di campioni di addestramento e la validazione (*test / validation*) su un set esterno di campioni.

In genere, il set di dati contenente le misure sui campioni analizzati viene suddiviso casualmente in due parti: il 70% del dataset viene utilizzato per creare il set di calibrazione (*training set*) e il 30% viene invece utilizzato come *test set*.

Per suddividere il dataset possono essere utilizzate diverse strategie. Nel caso di spettri puntuali, ad esempio, può essere utilizzato l'algoritmo Kennard/Stone - K/S (Kennard & Stone, 1969; Eigenvector Research Inc. 2013c). La procedura di selezione del set di addestramento basata sull'algoritmo K/S segue le fasi riportate nel seguente elenco:

- i) Selezione dei campioni più lontani in termini di distanza euclidea. La distanza media normalizzata per tutti i campioni viene calcolata sulla base della formula:

$$\bar{d}_i (\text{Normalized}) = \frac{\bar{d}_i - \min(\bar{d}_i)}{\max(\bar{d}_i) - \min(\bar{d}_i)} \quad (2.3.55)$$

Dove \bar{d}_i è la distanza media calcolata con l'equazione:

$$\bar{d}_i = \frac{\sum_{j=1}^n d_{ij}}{n-1} \quad (2.3.56)$$

Con d_{ij} è la distanza tra due campioni.

- ii) Ricerca del campione con massima dissimilarità tra ciascuno dei campioni selezionati nello *step i* ed inclusione dello stesso nel *training set*.
- iii) Lo *step ii* viene ripetuto iterativamente fino al raggiungimento del numero di campioni desiderato da includere nel *training set*.

Una volta selezionato il numero di campioni necessari da inserire all'interno del *training set*, il modello viene calibrato e si passa alla fase di *cross-validation* per la scelta del giusto grado di complessità dello stesso.

Una fase importante del calcolo di un modello (i.e. PLS o PLS-DA) è la definizione della dimensionalità, ovvero la scelta del numero di variabili latenti al fine di prendere in considerazione tutta la parte informativa del sistema, trascurando il rumore. In genere, questo compito è assolto valutando l'abilità predittiva del modello per i valori dei campioni (y), non utilizzati per la definizione del modello. A questo scopo, può essere applicata una procedura di validazione "interna", la cosiddetta convalida incrociata (in inglese *cross-validation*). La convalida incrociata può essere definita in due fasi: *i*) suddivisione del set di dati, o *dataset* (n campioni), in un numero s di gruppi (definiti dall'utente) e *ii*) calcolo iterativo del modello escludendo un gruppo al fine di prevedere i valori y degli oggetti esclusi. Lo *step ii* viene ripetuto fino a che tutti gli n gruppi siano stati eliminati almeno una volta.

Esistono differenti algoritmi di *cross-validation* (*Eigenvector Research Inc. 2016b*); tra questi, gli algoritmi più diffusi sono:

- ❖ *Leave-one-out*, con il quale ciascun oggetto nel *dataset* viene usato come *test set*.
- ❖ *Venetian Blinds*, con il quale ciascun *test set* viene determinato selezionando ogni s -esimo oggetto nel *dataset*, a partire dall'oggetto 1 fino all'oggetto s .
- ❖ *Contiguous Blocks*, mediante il quale ciascun *test set* viene determinato selezionando dei blocchi continui di n/s oggetti, a partire dall'oggetto 1.
- ❖ *Random Subsets*, nel quale gli s *test set* sono determinati selezionando in maniera casuale n/s oggetti all'interno del *dataset*, in maniera tale che nessun singolo campione sia presente in più di un *test set*.

Gli schemi di selezione dei campioni utilizzati nei *test set* di questi algoritmi di *cross-validation* sono riportati in [Figura 2.3.4](#).

In genere, il numero delle variabili latenti viene scelto sulla base del parametro RMSE (dall'equazione 2.3.43). In particolare, si sceglie il numero di variabili latenti per le quali il modello abbia un valore di RMSECV il più basso possibile e che non si discosti troppo dal corrispondente valore di RMSEC.

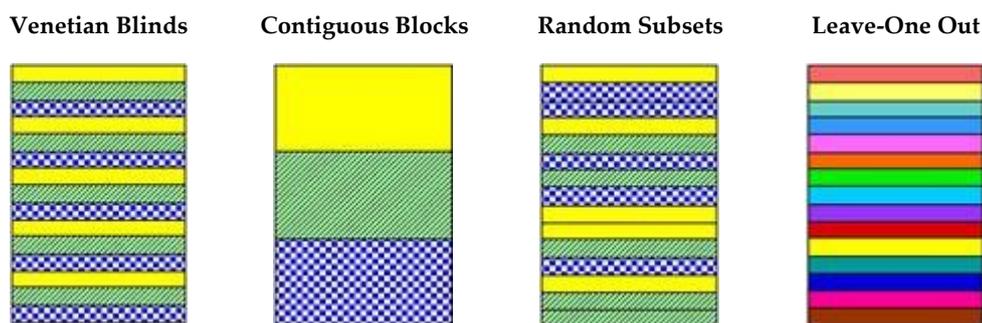


Figura 2.3.4. Schemi di selezione dei campioni utilizzati nei test set per alcuni algoritmi di cross-validation (Eigenvector Research Inc. 2016b).

Anche se la procedura di *cross-validation* fornisce una buona stima sull'efficacia del modello, tutti quanti i campioni noti sono utilizzati ai fini del calcolo del modello stesso. Una volta che è stata definita la dimensionalità del modello, la reale abilità predittiva del modello dovrà essere verificata utilizzando nuovi campioni con valori noti di y . La cosiddetta fase di validazione del modello consisterà, quindi, nell'utilizzo di un nuovo set di campioni (*test set*), differente dal set di campioni utilizzato per costruire il modello (*training set*).

2.3.6 Tecniche di selezione delle variabili significative

Come già descritto, nei metodi di calibrazione e classificazione multivariata entrano in gioco molteplici variabili sperimentali, utilizzate al fine di prevedere caratteristiche qualitative o quantitative del campione analizzato.

Ad ogni modo, la riduzione del numero di variabili necessarie ai fini di una modellazione è, nella maggior parte dei casi, una procedura utile ai fini della riduzione dei tempi nonché dei costi della sperimentazione. Rimuovendo o escludendo variabili non significative o rumorose è probabile che si ottenga un miglioramento della robustezza del modello e della capacità di previsione.

Negli ultimi anni, sono stati proposti svariati algoritmi di selezione delle variabili: i.e. *jack-knifing*, metodi basati su algoritmi genetici, metodi basati su *Selectivity Ratio* (SR), VIP (Leardi et al., 1998; Leardi et al., 2005; Andersen et al., 2010) e bootstrap-VIP (Gosselin et al., 2010).

Nel capitolo IV. *Casi di studio* sono riportati alcuni esempi applicativi degli algoritmi *interval-PLS* o *iPLS* (Nørgaard et al., 2000) e la sua implementazione per compiti di classificazione, ovvero la *interval-PLSDA*, o *iPLS-DA* (Eigenvector Research Inc. 2018), rispettivamente per la selezione delle variabili significative per i modelli di regressione e classificazione.

Questi algoritmi suddividono l'intero spettro in intervalli di larghezza uguale e calcolano modelli di regressione / classificazione aggiungendo o rimuovendo in modo iterativo gli intervalli, a seconda della strategia di ricerca applicata, in avanti (*forward*) o all'indietro (*reverse*). Nella modalità di selezione *forward*, i modelli vengono inizialmente calcolati utilizzando una regione spettrale alla volta e l'intervallo selezionato sarà quello con RMSECV minimo. L'algoritmo, quindi, aggiungerà in maniera iterativa gli intervalli di variabili a quelli scelti in precedenza, sempre sulla base del valore di RMSECV. Questa procedura viene ripetuta iterativamente fino a quando non possono essere più ottenuti miglioramenti significativi di RMSECV.

In alternativa, quando viene applicata la modalità di selezione inversa (*reverse*), un numero di modelli pari al numero di intervalli viene inizialmente calcolato considerando ogni volta tutti gli intervalli spettrali, tranne uno. Viene selezionato l'intervallo con il modello che presenta un RMSECV minimo, e ancora una volta i modelli vengono calcolati rimuovendo ogni volta uno degli intervalli rimanenti; questa procedura viene ripetuta iterativamente fino a quando non si osserva una diminuzione significativa di RMSECV.

3 STRUMENTI UTILIZZATI

In questo paragrafo verranno descritte le funzionalità, le specifiche tecniche degli strumenti utilizzati, le procedure di calibrazione da effettuare prima dell'acquisizione dei dati e la successiva gestione dei dati acquisiti.

3.1 SISTEMA IPERSPETTRALE SWIR

Il sistema iperspettrale SisuCHEMA™ XL Chemical Imaging workstation (*Specim, Spectral Imaging Ltd. 2019*) permette acquisizioni nel range dello SWIR (*Short Wave InfraRed*) in riflettanza diffusa. Tale sistema è costituito da una telecamera CCD ed uno spettrografo Inspector Specim® modello N25E (le caratteristiche ottiche dello spettrografo sono riportate in [Tabella 3.1.1](#)), operante nella regione dello spettro elettromagnetico dell'infrarosso ad onde corte (1000-2500 nm), con una risoluzione spettrale di 10 nm e campionamento spettrale/pixel di 6.3 nm, una piattaforma mobile su cui possono essere posizionati i campioni, un impianto di illuminazione a fonte alogena ed un sistema integrato di calibrazione del bianco che permette l'ottimizzazione della messa a fuoco dell'obiettivo (*Middleton Spectral Vision, 2004*).

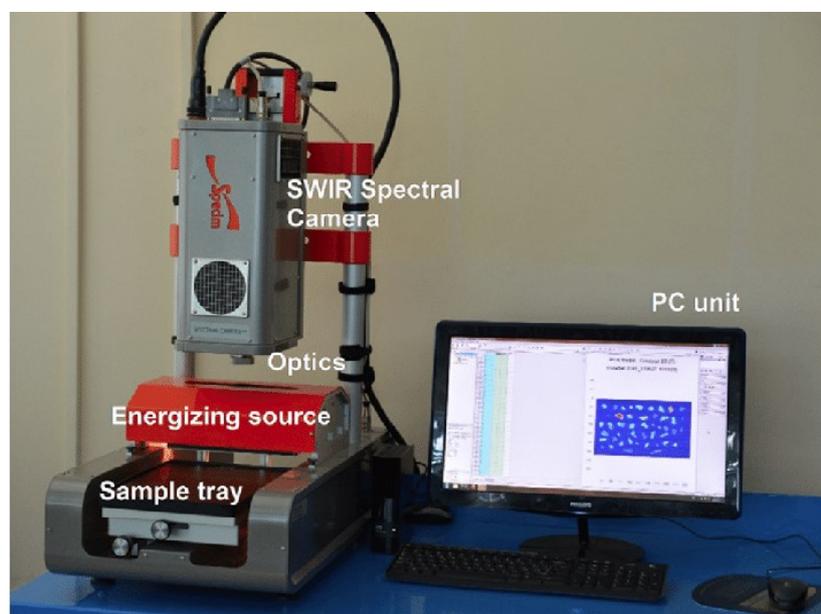


Figura 3.1.1. Il sistema iperspettrale SisuCHEMA™ XL Chemical Imaging workstation (Serranti et al., 2018).

Strumenti utilizzati

Tabella 3.1.1. Caratteristiche ottiche dello spettrografo ImSpector N25E.

| | | | | | | |
|-----------------------------------|-----------|---|--------------|------------------|--------------|-------------------|
| Spectrograph | | <i>ImSpector N25E</i> | | | | |
| Spectral Range | | <i>1000-2500 nm ±</i> | | | | |
| Spectral resolution | | <i>10 nm (30 μm slit)</i> | | | | |
| Spectral sampling/pixel | | <i>6.3 nm</i> | | | | |
| Spatial resolution | | <i>Rms spot radius <15 μm (320)</i> | | | | |
| Aberrations | | <i>Insignificant astigmatism, smile or keystone <5 μm</i> | | | | |
| Numerical aperture | | <i>F/2.0</i> | | | | |
| Slit width options | | <i>30 μm (50 or 80 μm optional)</i> | | | | |
| Effective slit length | | <i>9.6 mm</i> | | | | |
| Total efficiency (typical) | | <i>>50%, independent of polarization</i> | | | | |
| Stray light | | <i><0.5% (halogen lamp, 1400 nm notch filter)</i> | | | | |
| Field of view (mm) | | <i>15mm lens</i> | | <i>31mm lens</i> | | <i>Macro lens</i> |
| | | <i>200</i> | | <i>50</i> | <i>100</i> | <i>10</i> |
| Pixel dimension (mm) | <i>xr</i> | <i>0.625</i> | <i>0.156</i> | <i>0.312</i> | <i>0.031</i> | |
| | <i>yr</i> | <i>(x dimension in mm * 0.03) / 9.6</i> | | | | |
| Scanning speed (mm/s) | | <i>72.50</i> | <i>17.35</i> | <i>32.25</i> | <i>2.55</i> | |
| Scanning rate | | <i>100 hyperspectral line images/s (max), corresponding to</i> <i>- 3 mm/s with 30 micron pixel</i> <i>- 30 mm/s with 300 micron pixel</i> <i>- 60 mm/s with 600 micron pixel</i> | | | | |
| Typical scanning time | | <i>< 7s for single 320*320 pixel image capture with 256 spectral bands</i> | | | | |
| Illumination | | <i>SPECIM's diffuse line illumination unit</i> | | | | |
| Data format | | <i>BIL file format, Evince and ENVI compatible</i> | | | | |
| Instrument calibration | | <i>Instrument is delivered with spectral calibration. Image data is automatically calibrated to reflectance by measuring an internal standard reference target before each sample scan.</i> | | | | |

Strumenti utilizzati

La modalità di acquisizione delle immagini iperspettrali con questo sistema segue la logica *push-broom*: il sistema acquisisce l'immagine iperspettrale linea per linea, acquisendo informazioni spettrali per ogni pixel nella linea. Le dimensioni massime di acquisizione di tale sistema sono di: 297x210 mm.

Il sistema di acquisizione è controllato da un PC che tramite un software di acquisizione specializzato (ChemaDAQ™) consente di gestire le differenti unità del sistema di misura, di eseguire l'acquisizione degli spettri e di provvedere alla loro raccolta. Tale sistema iperspettrale si trova presso il Raw-Ma Lab (*Raw Materials Laboratory*) del Dipartimento di Ingegneria Chimica Materiali Ambiente (DICMA) della Sapienza - Università di Roma.

Procedura di calibrazione ed acquisizione/gestione dati

La calibrazione dello strumento viene eseguita dapprima acquisendo una immagine nera riferimento D_i , registrando un'immagine con la lente della camera completamente chiusa e, successivamente, mediante la misurazione della "immagine di riferimento bianca", W_i . Il riferimento nero viene utilizzato per rimuovere l'effetto della *dark current* del CCD. Dopo questa fase di calibrazione, viene acquisita l'immagine iperspettrale grezza (R_{0_i}) e viene quindi calcolata l'immagine in riflettanza (R_i) dell'immagine acquisita, grazie all'equazione:

$$R_i = \frac{R_{0_i} - D_i}{W_i - D_i} \quad (3.1.1)$$

La procedura di calibrazione e di acquisizione dati viene effettuata mediante il software ChemaDAQ™. Successivamente, il dato grezzo può essere elaborato preliminarmente (i.e. ritaglio dei bordi dell'immagine iperspettrale grezza) con il software Evince® e salvato in formato `.mat` per essere esportato in ambiente MATLAB (*The MathWorks, Inc. 2019*). In alternativa, il dato grezzo ottenuto con il software ChemaDAQ™ può essere importato in ambiente MATLAB con l'utilizzo di uno script scritto *ad hoc* per lo scopo: `Sisuchema_SWIR_import.m` (si rimanda alla sezione [Allegato I - MATLAB script - Sisuchema_SWIR_import.m](#)). Questo script consente di aprire i tre file `.hdr` (uno contenente l'informazione del bianco di riferimento, uno quella del nero ed uno contenente l'immagine iperspettrale grezza) ottenuti dal software ChemaDAQ™, applicare la correzione dell'immagine (equazione 3.1.1) e ricavare il dataset corretto dell'immagine iperspettrale in riflettanza.

3.2 SISTEMA IPERSPETTRALE VIS-NIR

Il sistema iperspettrale Vis-NIR utilizzato in vari *case studies*, costruito dall'azienda DV s.r.l. nell'ambito del progetto europeo W2PLASTICS, è costituito da due spettrografi (operanti nel Vis e NIR) montati su una struttura in metallo (DV s.r.l. 2013). Tale sistema è composto, inoltre, da un nastro trasportatore, la cui velocità di movimentazione può essere variata, e da un sistema illuminante, costituito da lampade alogene installate all'interno di un apposito riflettore in alluminio. L'intero sistema di acquisizione è controllato tramite un personal computer (PC). Gli spettrografi montati sul sistema sono: l'ImSpector Specim® modello V10E (Figura 3.2.2a), operante nel VIS-NIR (800-1000 nm), e l'ImSpector Specim® modello N17E (Figura 3.2.2b), operante nel NIR (1000-1700 nm).

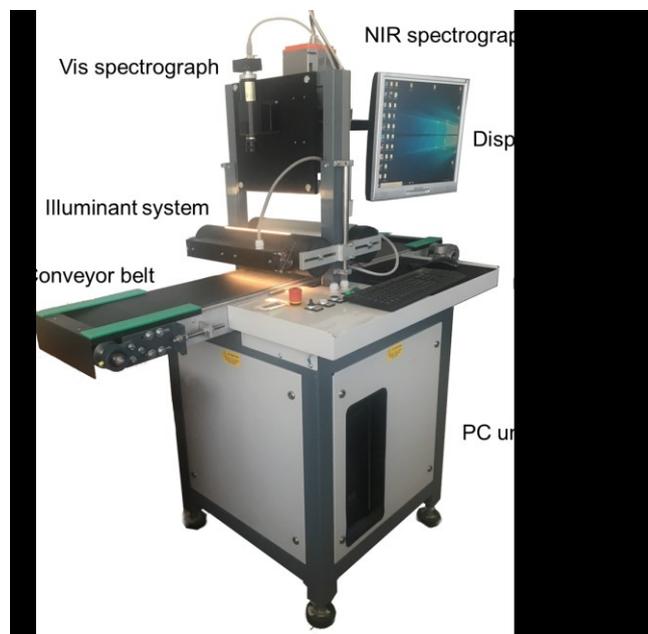


Figura 3.2.1. Sistema iperspettrale Vis-NIR (DV s.r.l. 2013).

Le caratteristiche ottiche della linea di spettrografi ImSpector V10E sono riportate in Tabella 3.2.1. Mentre, in Tabella 3.2.2 sono riportate le caratteristiche ottiche dello spettrografo ImSpector N17E. Lo schema costruttivo di uno spettrografo ImSpector™ è riportato in Figura 3.2.3.

L'intero sistema lavora in modalità *push-broom*, acquisendo informazioni spettrali contigue per ogni pixel nella linea. Il PC connesso al sistema permette acquisire, raccogliere e gestire i dati iperspettrali grazie al software Spectral Scanner™ (Ver. 2.0).

Strumenti utilizzati

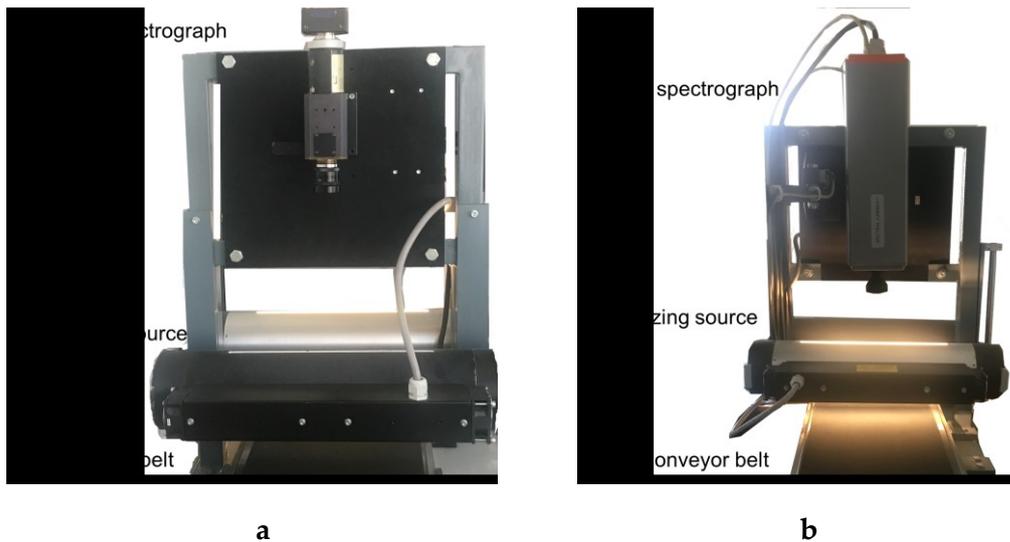


Figura 3.2.2. Configurazione degli spettrografi: a) ImInspector Specim® modello V10E, operante nel VIS-NIR (800-1000 nm), e b) ImInspector Specim® modello N17E, operante nel NIR (1000-1700 nm).

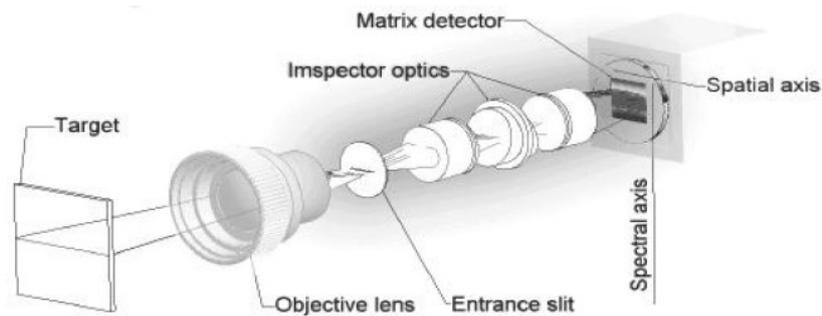


Figura 3.2.3. Schema costruttivo di uno spettrografo della serie ImInspector™.

Procedura di calibrazione ed acquisizione/gestione dati

La calibrazione degli spettrografi ImInspector Specim® modello V10E e modello N17E segue la stessa procedura del sistema iperspettrale SisuCHEMA™ XL: viene dapprima acquisita una immagine nera riferimento (D_i), registrando un'immagine con la lente della camera completamente chiusa e acquisendo, successivamente, il bianco di calibrazione (W_i) su uno standard di riferimento in ceramica (i.e. con riflettanza nominale al 99%). Il riferimento nero viene utilizzato, anche in questo caso, per rimuovere l'effetto della *dark current* del CCD. Dopo questa fase di calibrazione, viene acquisita l'immagine iperspettrale grezza (R_{0_i}) e viene quindi calcolata l'immagine in riflettanza (R_i) dell'immagine acquisita, grazie all'equazione 3.1.1. La procedura di calibrazione degli spettrografi e l'acquisizione di immagini iperspettrali può essere effettuata grazie all'ausilio del software Spectral Scanner™ (Ver. 2.0). Le immagini iperspettrali, salvate in formato *ENVI reflectance* con estensione .hdr, potranno poi essere elaborate in ambiente MATLAB.

Strumenti utilizzati

Tabella 3.2.1. Caratteristiche ottiche della linea di spettrografi ImSpector V10E (Specim, Spectral Imaging Ltd. 2016).

| | | |
|--|--|-----------------------------------|
| Spectral Range | 400-1000 nm | |
| Spectral resolution | 2.8 nm (with 30 μm slit) | |
| Dispersion | 97.5 nm/mm | |
| Image size | max 6.15 mm (spectral) * 14.2 mm (spatial) | |
| Numerical aperture | F/2.4 | |
| Spatial resolution | Rms spot radius < 9 μm (580 pixels) | |
| Aberrations | No astigmatism | |
| Bending of spectral lines across spatial axis | Smile < 1.5 μm | |
| Bending of spatial lines across spectral axis | Keystone < 1 μm | |
| Slit width, default | 30 μm (13, 18, 50, 80 and 150 μm) | |
| Slit length | 14.2 | |
| Optical input | Telecentric | |
| Efficiency | > 50%, independent of polarization | |
| Stray light | < 0.5% (halogen lamp, 633 nm notch filter) | |
| Field of view (cm) | 12.5 mm lens | 25 mm lens |
| | 15.5 | 7.8 |
| Pixel dimension (mm) | x | 0.267 |
| | yr | (x dimension in mm * 0.03) / 14,2 |

Strumenti utilizzati

Tabella 3.2.2. Caratteristiche ottiche dello spettrografo ImSpector N17E (Specim, Spectral Imaging Ltd., 2016).

| | | |
|------------------------------------|-----------|--|
| Spectrograph | | <i>ImSpector N17E</i> |
| Spectral Range | | <i>1000-1700 nm ± 10 nm</i> |
| Optical spectral resolution | | <i>5 nm (30 μm slit)</i> |
| Spectral sampling / pixel | | <i>3.3 nm</i> |
| Spatial resolution | | <i>Rms spot radius < 15 μm (320 pixels)</i> |
| Aberrations | | <i>Insignificant astigmatism, sub-pixel smile and keystone</i> |
| Numerical aperture | | <i>F/2.0</i> |
| Slit width options | | <i>30 μm (18, 50, 80 μm optional)</i> |
| Effective slit length | | <i>9.6 mm</i> |
| Total efficiency (typical) | | <i>> 50%, independent of polarization</i> |
| Stray light | | <i>< 0.5 % (halogen lamp, 1400 nm notch filter)</i> |
| Objective | | <i>15 mm focal distance</i> |
| Field of view (cm) | | <i>19.5</i> |
| Pixel dimension (mm) | <i>xr</i> | <i>0.609</i> |
| | <i>yr</i> | <i>(x dimension in mm * 0.03) / 9.6</i> |

3.3 SPETTROFOTORADIOMETRO PORTATILE VIS-SWIR FIELDSPEC® 4

Lo spettrofotoradiometro portatile ASD FieldSpec® 4 Standard-Res (Figura 3.3.1) lavora nell'intervallo dello spettro elettromagnetico compreso tra 350 e 2500 nm (VIS-SWIR), con risoluzione spettrale di 3 nm a 700 nm e di 10 nm a 1400/2100 nm (ASD Inc. 2015a; ASD Inc. 2015b). Questo spettroradiometro è costituito da un case contenente il sistema di rivelatori collegato ad un cavo in fibra ottica, alla cui estremità è montata una sonda a contatto per effettuare misurazioni in riflettanza.



Figura 3.3.1. Spettrofotoradiometro ASD FieldSpec® 4 Standard-Res e sonda a contatto per misure in riflettanza (Malvern Panalytical Ltd, 2019).

Il sistema rivelatore è costituito da tre detector separati da diversi reticoli di diffrazione. Il sistema di rivelatori è costituito da un rivelatore VNIR (512 element silicon array, funzionante nell'intervallo 350-1000 nm), il rivelatore SWIR 1 (Graded Index InGaAs-Photodiode-Two Stage TE Cooled, che opera nell'intervallo: 1001-1800 nm) e il rivelatore SWIR 2 (Graded Index InGaAs. Photodiode, Two Stage TE Cooled; funzionante nel range spettrale: 1801-2500 nm).

La sonda a contatto, progettata per effettuare misurazioni in riflettanza a contatto con materiali solidi è costituita da un sistema illuminante a lampada alogena con temperatura di colore 2901 (+/- 10 %) °K (la temperatura di colore quantifica la tonalità della luce). La dimensione dello spot della sonda è di 10 mm. La lunghezza della sonda è di circa 25.4 cm ed un peso di 0.7 kg. La sorgente illuminante è posta a circa 12° rispetto alla normale dello spot della sonda, mentre la fibra ottica è posta a circa 35° rispetto all'asse principale della sonda. In Figura 3.3.2 è riportato il diagramma costruttivo di una sonda ASD per spettrofotoradiometri FieldSpec®.

Strumenti utilizzati

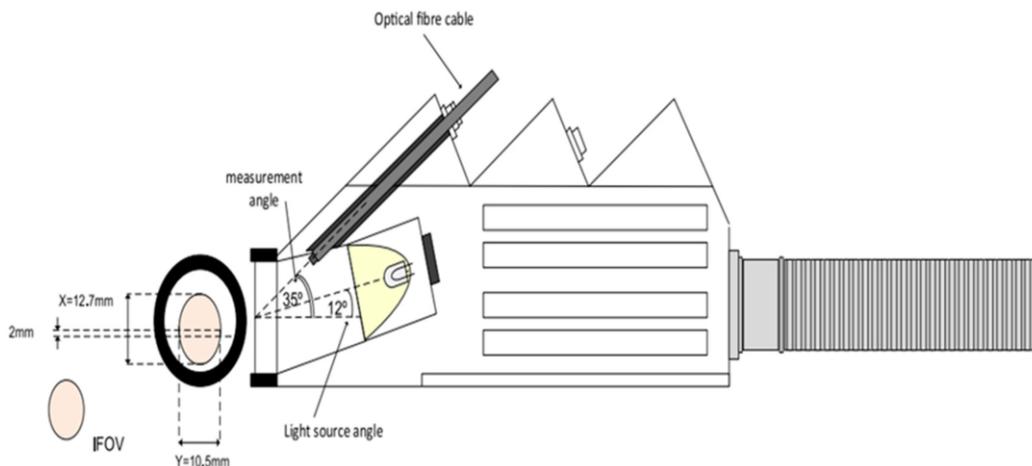


Figura 3.3.2. Diagramma costruttivo di una sonda ASD high-intensity. X e Y rappresentano altezza e larghezza della Field of View, FoV (Lamine et al., 2019).

Le acquisizioni spettrali possono essere effettuate grazie ad un PC connesso al dispositivo mediante cavo Ethernet. La procedura di calibrazione e acquisizione degli spettri può essere effettuata grazie all'ausilio del software ASD RS³, Ver. 6 (ASD Inc. 2008). Mentre, il software che consente la conversione dei file .asd in file .txt con codifica ASCII è ViewSpec Pro (Ver. 6.2.0.).

Procedura di calibrazione ed acquisizione/gestione dati

La calibrazione dello spettrofotometri ASD FieldSpec[®] 4 Standard-Res segue una procedura simile a quella degli spettrografi ImSpector Specim[®] e del sistema iperspettrale SisuCHEMA[™] XL.

In questo caso, il bianco di calibrazione (W_i) viene acquisito su uno standard di riferimento in Spectralon[®] (i.e. con riflettanza nominale al 99%). Mentre, il nero riferimento (D_i) viene richiamato dal *Dark Current calibration file* presente nelle librerie del software RS³ (ASD Inc. 2008). Lo spettro grezzo (R_{0_i}) viene quindi calcolato come spettro in riflettanza (R_i), grazie all'equazione 3.1.1. La procedura di calibrazione dello strumento e l'acquisizione degli spettri può essere effettuata grazie all'ausilio del software RS³. Gli spettri acquisiti in formato .asd potranno poi essere convertiti in file .txt, con codifica ASCII, mediante il software ViewSpec Pro (Ver. 6.2.0.). I file di testo, contenenti le informazioni sugli spettri acquisiti, saranno poi facilmente esportabili in ambiente MATLAB[®]. Una soluzione per importare tali file in ambiente MATLAB è quello dell'utilizzo della routine `fieldspec_import2.m` (si rimanda alla sezione allegati: [Allegato II - MATLAB script - fieldspec_import2.m](#)).

Strumenti utilizzati

Tabella 3.3.1. Caratteristiche ottiche e specifiche costruttive dello spettrofotometri ASD FieldSpec® 4 Standard-Res (Malvern Panalytical Ltd, 2019).

| | |
|---|--|
| Spectral Range | 350-2500 nm |
| Spectral Resolution | 3 nm @ 700 nm |
| | 10 nm @ 1400/2100 nm |
| Spectral sampling (bandwidth) | 1.4 nm @ 350-1000 nm |
| | 1.1 nm @ 1001-2500 nm |
| Scanning Time | 100 milliseconds |
| Stray light specification | VNIR 0.02%, SWIR 1 & 2 0.01% |
| Wavelength reproducibility | 0.1 nm |
| Wavelength accuracy | 0.5 nm |
| Maximum radiance | VNIR 2X Solar, SWIR 10X Solar |
| Channels | 2151 |
| Detectors | VNIR detector (350-1000 nm): 512 element silicon array |
| | SWIR 1 detector (1001-1800 nm): Graded Index InGaAs Photodiode, Two Stage TE Cooled |
| | SWIR 2 detector (1801-2500 nm): Graded Index InGaAs Photodiode, Two Stage TE Cooled |
| Input | 1.5 m fiber optic (25° field of view). Optional narrower field of view fiber optics available. |
| Noise Equivalent Radiance (NE_L) | VNIR: 1.0×10^{-9} W/cm ² /nm/sr @ 700 nm |
| | SWIR 1: 1.2×10^{-9} W/cm ² /nm/sr @ 1400 nm |
| | SWIR 2: 1.9×10^{-9} W/cm ² /nm/sr @ 2100 nm |

3.4 SPETTROMETRO PORTATILE NIR MICRONIR™

Lo spettrometro portatile JDSU MicroNIR™ 1700, mostrato in [Figura 3.4.1](#), è in grado di acquisire spettri in riflettanza nel range dello spettro elettromagnetico che va da 950 a 1650 nm (NIR). Il MicroNIR™ è basato sull'utilizzo di un *linear variable filter* (LVF) come elemento dispersivo, accoppiato con un *detector array*: 128 pixel *InGaAs photodiode array* (JDS Uniphase Corporation, 2013; JDS Uniphase Corporation, 2014). Questo strumento portatile è interfacciabile con un PC tramite cavo USB 2.0 ed integra nella sua architettura 2 lampade a incandescenza (a filamento di tungsteno). In [Figura 3.4.2](#) è mostrata l'architettura del dispositivo, mentre le specifiche tecniche sono riportate in [Tabella 3.1.1](#).

La procedura di calibrazione e l'acquisizione dati può essere effettuata grazie all'ausilio del software MicroNIR Pro.



Figura 3.4.1. Spettrometro portatile MicroNIR™ 1700.

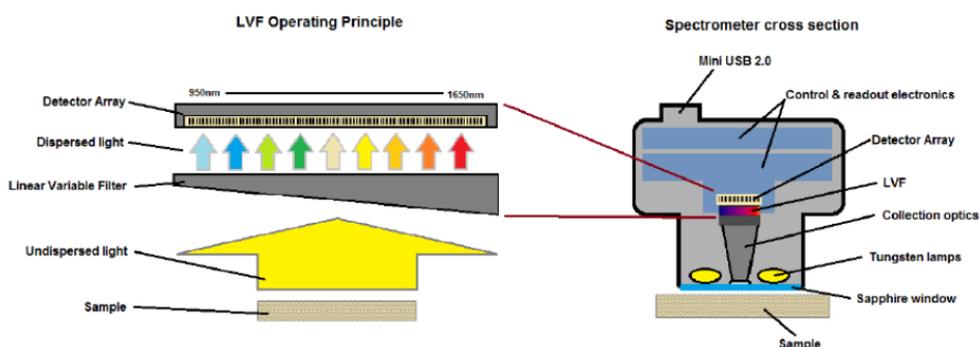


Figura 3.4.2. Principio di funzionamento e diagramma costruttivo del MicroNIR™ (JDS Uniphase Corporation, 2014).

Procedura di calibrazione ed acquisizione/gestione dati

Strumenti utilizzati

La calibrazione dello spettrometro portatile JDSU MicroNIR™ segue una procedura a quella dello spettrofotometri ASD FieldSpec® 4 Standard-Res.

In questo caso, il bianco di calibrazione (W_i) viene acquisito su uno standard di riferimento in Spectralon® (i.e. con riflettanza nominale al 99%). Mentre, la *dark current* (D_i) viene acquisita con lo spot del sensore chiuso da un tappo in plastica nera (i.e. PE). Lo spettro grezzo (R_{0i}) viene quindi calcolato come spettro in riflettanza (R_i), grazie all'equazione 3.1.1. La procedura di calibrazione dello strumento e l'acquisizione degli spettri può essere effettuata grazie all'ausilio del software MicroNIR Pro. Gli spettri acquisiti in formato .csv, saranno poi facilmente esportabili in ambiente MATLAB®. Una soluzione per importare tali file in ambiente MATLAB è quello dell'utilizzo della routine `micronir_import.m`.

Tabella 3.4.1. Caratteristiche tecniche di funzionamento del MicroNIR™ 1700.

| | |
|---------------------------------------|--|
| Input aperture dimensions | 0.5 x 2.0 mm |
| Sample working plane | 2 x 4 mm at window face |
| Dispersing element | Linear variable filter |
| Detector | 128 pixel InGaAs photodiode array |
| Pixel Size / Pitch | 30 μm x 250 μm / 50 μm |
| Wavelength range | 950 – 1650 nm |
| Pixel to pixel interval | 6.2 nm |
| Spectral bandwidth (FWHM) | <1.25% of center wavelength (1% typical). For example, at 1000 nm, resolution is <12.5 nm |
| Measurement time (typical) | 0.25 sec |
| Integration time | Minimum 100 μs ; maximum limited by dark signal only |
| Operating environment | -20 to 40°C |
| Dimensions (diameter x height) | 45 x 42 mm |
| Weight | <60 g |

3.5 SPETTROMETRO MICRO-XRF M4 TORNADO™

Il sistema M4 TORNADO™ della Bruker (mostrato in [Figura 3.5.1](#)), presente presso il *Raw-Material Lab* del Dipartimento di Ingegneria Chimica, Materiali ed Ambiente della Sapienza- Università di Roma, consente di effettuare analisi in micro-fluorescenza ai raggi X. Grazie a questo strumento è possibile realizzare mappe composizionali ad alta risoluzione dei singoli elementi presenti in un campione, o di tutti gli elementi inclusi. Infatti, è possibile effettuare l'analisi elementare dei materiali, sia in termini di distribuzione degli elementi chimici (con numero atomico $z > 11$) lungo una linea (i.e. *line scan*) che su delle aree (i.e. *mapping*). Grazie alla presenza di ottiche capillari, si ha la possibilità di generare alte intensità di fluorescenza. Le ottiche raccolgono la radiazione del tubo focalizzandola in uno spot di dimensioni inferiori ai 25 μm ([Bruker Corporation, 2015](#)). Le caratteristiche tecniche dello spettrometro Micro-XRF della Bruker sono riportati in [Tabella 3.5.1](#).

Procedura di acquisizione/gestione dati

Il software per effettuare le analisi elementali quantitative e l'estrapolazione delle mappe di distribuzione degli elementi chimici è ESPRIT (Ver. 1.5). L'ipercubo dati salvato in formato `.raw` è esportabile in ambiente MATLAB® mediante la funzione `importxrfmap.m` (vedere [Allegato IV - MATLAB script - importxrfmap.m](#)); mentre la funzione `create_datasetXRF.m` re-interpola la matrice di dati, riducendone le dimensioni, al fine di semplificare la gestione del set di dati e la procedura di analisi successiva, nonché di diminuire il tempo necessario di elaborazione dati (si veda [Allegato V - MATLAB script - create_datasetXRF.m](#)).



Figura 3.5.1. Lo spettrometro Micro-XRF M4 TORNADO (Bruker Corporation, 2015).

Strumenti utilizzati

Tabella 3.5.1. Specifiche tecniche dello spettrometro Micro-XRF M4 TORNADO della Bruker (Bruker Corporation, 2015).

| | |
|--------------------------------|--|
| Sample types | <i>Solids, particles, liquids</i> |
| Sample chamber size | <i>WxDxH: 600 mm x 350 mm x 260 mm</i> |
| Stage | <i>WxD: 330 mm x 170 mm Max. weight load: 5 kg</i> |
| Measurement media | <i>Air or oil free vacuum, ready for measurement within 100 s</i> |
| Pixel Size / Pitch | <i>30 μm x 250 μm / 50 μm</i> |
| Sample travel | |
| <i>Max. travel</i> | <i>WxDxH: 200 mm x 160 mm x 120 mm</i> |
| <i>Mapping travel</i> | <i>WxD: 190 mm x 160 mm</i> |
| <i>Travel speed</i> | <i>Up to 100 mm/s with TurboSpeed stage</i> |
| Pixel to pixel interval | <i>6.2 nm</i> |
| Sample view | <i>2 simultaneous live images from above with different magnifications for sample overview and precise positioning. Lateral fisheye camera for the sample chamber overview</i> |
| Excitation | <i>High brilliance X-ray tube with polycapillary X-ray optics</i> |
| Excitation parameters | |
| <i>Target material</i> | <i>Rh, optional: Mo, Ag, Cu, W</i> |
| <i>Tube parameters</i> | <i>50 kV, 30 W (40 W for collimator) Less than 25 μm for Mo-K with polycapillary lens</i> |
| <i>Spot size</i> | <i>Up to 5 filters</i> |
| <i>Filters</i> | |
| Detection | <i>XFlash[®] silicon drift detector</i> |
| Detector parameters | |
| <i>Sensitive area</i> | <i>30 mm² < 145 eV at 300,000 cps (count rate Cu > 400,000 cps)</i> |
| <i>Energy resolution</i> | |
| Dimensions | <i>WxDxH: 815 mm x 680 mm x 580 mm, 130 kg</i> |

4 CASI DI STUDIO

In questo capitolo sono riportati i casi di studio che hanno interessato l'attività di ricerca durante il percorso di dottorato. L'attività di ricerca svolta ha riguardato diversi casi di studio, in cui sono state applicate differenti metodiche di statistica multivariata, *machine learning* e *computer vision* a complemento dell'utilizzo di tecniche di spettroscopia classica ed iperspettrale, finalizzate alla caratterizzazione di materie prime e prodotti.

In particolare, nel campo delle **materie prime** secondarie, la ricerca si è focalizzata sul riconoscimento di diverse tipologie di materiali all'interno di un flusso di rifiuti eterogeneo (i.e. RAEE in pezzatura) e parzialmente omogeneo (i.e. packaging proveniente da raccolta differenziata della plastica), con le seguenti finalità:

- controllo qualità step by step in un processo di riciclo (classificazione di più materiali negli output analizzati);
- recupero di metalli (i.e. identificazione di Cu in schede a circuito stampato e identificazione di circuiti stampati in un flusso di RAEE);
- selezione/riciclo (i.e. impostazione di modelli a cascata per la separazione/selezione di diversi materiali) e/o contaminanti (i.e. carta e cartone da imballi in plastica).

Mentre, nel campo **agro-industriale** e dell'**industria alimentare**, la ricerca si è focalizzata sull'analisi di diversi prodotti, in particolare:

- valutazione della maturazione/invasatura di alcune tipologie di frutti (i.e. kiwi ed olive), mediante l'utilizzo di spettrometri portatili finalizzata al controllo qualità di drupe in *real time*.
- riconoscimento di prodotti finiti (i.e. frutta secca) da contaminanti, mediante l'utilizzo di tecnologie iperspettrali applicabili nell'ambito del controllo qualità;
- valutazione di parametri qualitativi e/o percettivo-gustativi di diversi prodotti agro-alimentari (i.e. kiwi, farine, grano, pasta, olive e peperoncino), finalizzata al controllo di qualità e alla valutazione della *shelf life* dei prodotti;
- valutazione della germinabilità di alcune tipologie di sementi, mediante l'utilizzo di tecnologie iperspettrali applicabili nell'ambito del controllo qualità;
- valutazione e riconoscimento di prodotti alterati (i.e. olive attaccate da mosca olearia e/o danneggiate da muffe), finalizzata al controllo di qualità in *real time*;
- valutazione della granulometria dei prodotti (i.e. peperoncino e farine), finalizzata al controllo di qualità.

4.1 MATERIE PRIME SECONDARIE

4.1.1 Caratterizzazione di un flusso di RAEE risultante da comminazione e separazione magnetica per densità

Problematica e stato dell'arte

Televisori LCD, Smartphone, macchine fotografiche digitali e Personal Computer, sono solamente alcuni degli oggetti elettronici che ci circondano durante la quotidianità. L'elettronica è ormai diventata parte integrante nella nostra vita grazie al continuo progresso tecnologico. In un contesto in continua evoluzione di pari passo con il progresso tecnologico crescono i volumi di rifiuti provenienti da apparecchiature elettriche ed elettroniche. Nondimeno, però, con il passare del tempo è migliorata anche la nostra capacità nella produzione di nuovi materiali, oltre alla capacità di caratterizzare, prevedere e sfruttare le proprietà fisiche e chimiche dei minerali. Così, è diventato possibile sviluppare nuove applicazioni che migliorano la prestazione tecnica, la durata e l'affidabilità dei prodotti, nonché la riduzione degli oneri ambientali. Nell'era dell'informazione, gli sviluppi nell'ingegneria e nella scienza dei materiali continuano a far crescere e a sostenere lo sviluppo di nuove tecnologie. Le proprietà uniche di minerali non combustibili, prodotti minerali, metalli e leghe contribuiscono alla fornitura di cibo, riparo, infrastrutture, trasporti, comunicazioni, sanità e applicazioni di difesa. Il telefono cellulare è un esempio familiare, che illustra la dipendenza da una tecnologia d'importanza globale basata proprio sui minerali, per via delle loro proprietà chimiche e fisiche e dei materiali che possono essere creati con loro. Proprio per questo, data la crescente domanda di materie prime (che per alcune può sfociare addirittura in una indisponibilità delle stesse), molti sforzi vengono fatti nel campo del riciclo delle apparecchiature a fine vita e del recupero di materie prime secondarie.

Con la sigla RAEE si indicano i Rifiuti da Apparecchiature Elettriche ed Elettroniche o RAEE (in inglese WEEE, ovvero *Waste of Electric and Electronic Equipment*), ossia quelle apparecchiature elettriche ed elettroniche (altrimenti detti AEE) giunte al fine del proprio ciclo di vita. Queste apparecchiature dette anche AEE, incluse tutte le componenti e i materiali di consumo che ne costituiscono parte integrante, diventano rifiuti quando corrispondono alla definizione degli stessi riportata nel D.lgs. 152/06 e ss.mm.ii. fondata sul concetto del "disfarsi". Si definiscono, quindi, RAEE (Rifiuti da Apparecchiature Elettriche ed Elettroniche) quelle apparecchiature elettriche ed elettroniche di cui il possessore intenda disfarsi in quanto guaste, inutilizzate, o obsolete e che ai sensi della normativa ambientale si considerano rifiuti (art. 183, comma 1, lett. a) del D. lgs. 152/2006), inclusi tutti i componenti e i materiali di consumo che sono parte integrante del prodotto, quando si assume la decisione di disfarsene. I RAEE si dividono in domestici e professionali a seconda che siano originati da nuclei domestici o da attività commerciali, industriali, istituzionali e di altro tipo.

Alcune stime indicano che annualmente sono generate circa 40 milioni di tonnellate di rifiuti elettronici, che sono all'incirca il 5% dei rifiuti solidi presenti nel mondo. A titolo di esempio, solamente nel 2016 sono state prodotte circa 44.7 milioni di tonnellate (circa 6.1 kg/abitante) di e-waste, l'equivalente di circa 4500 torri Eiffel! Questo quantitativo è destinato a crescere in futuro. Secondo ulteriori studi, nel 2021 verranno prodotte globalmente circa 52.2 milioni di tonnellate di RAEE, pari a 6.8 kg prodotti per abitante (*Robinson, 2009*).

Un dato preoccupante è quello relativo alla quantità di RAEE trattati nel mondo. Solamente il 20% dei 44.7 milioni di tonnellate di e-waste prodotte nel 2016, è stato raccolto e riciclato in maniera adeguata. Mentre, il destino del rimanente 80% di tale volume non è stato documentato. Si pensa che tale quota sia stata trattata in maniera impropria, inviata in discarica senza alcun pre-trattamento, commerciata o riciclata in altri paesi, i.e nei paesi in via di sviluppo come Cina, India, Bangladesh e Africa (*Abdelbasir et al., 2018*).

La produzione di RAEE a livello globale è destinata a crescere inesorabilmente nei prossimi anni. Questa drastica crescita è dovuta alla continua ed inesorabile commercializzazione di nuovi prodotti elettronici (si pensi ai televisori, cellulari e computer di nuova generazione) con migliori caratteristiche nelle performance che comporta un accorciamento del ciclo di vita dei prodotti stessi.

I RAEE rappresentano una categoria di rifiuti estremamente eterogenea, in termini di composizione, volume e forma. La composizione chimica di un RAEE varia a seconda di ciascun prodotto. Ad esempio, le TV LED presentano un maggior contenuto di polimeri rispetto ai microonde, che hanno un maggior contenuto di metalli. Infatti, la composizione chimica di un RAEE dipende da diversi fattori, quali, ad esempio, la tipologia di RAEE, l'anno di produzione, il marchio manifatturiero ed il paese d'origine. In generale, si possono trovare diversi metalli all'interno dei RAEE, come rame, ferro, alluminio, ottone ed altri metalli preziosi, come oro, argento e palladio, insieme a vari polimeri, come polietilene, polipropilene, poliuretano e molti altri ancora. Inoltre, un rifiuto elettronico può contenere anche materiali ceramici, come vetro ed altri materiali organici e inorganici, talvolta anche radioattivi. All'interno di un flusso di RAEE si possono ritrovare anche numerose sostanze pericolose utilizzate per la fabbricazione di varie apparecchiature, come il piombo nelle schede dei circuiti stampati ed il cadmio nei chip semiconduttori ed ancora, nichel, antimonio, arsenico e cromo esavalente, insieme ad altri composti organici come bifenili policlorurati, i clorofluorocarburi, idrocarburi policiclici aromatici e polialogenati, potrebbero causare gravi problemi ambientali, se non adeguatamente riciclati o smaltiti. A causa di tali sostanze, I RAEE sono considerati tossici per la salute umana e nocivi per l'ambiente e quindi classificati come rifiuti pericolosi. Tuttavia, come descritto precedentemente, i RAEE contengono quantità significative di materiali pregiati (come metalli - circa il 66%, metalli preziosi, plastica di alta qualità - 19% ed altre componenti) che possono essere proficuamente essere recuperate come materie prime seconde. A causa dei potenziali rischi che potrebbe causare il non corretto trattamento dei RAEE e le considerevoli "risorse" contenute in questi, risulta di fondamentale importanza il miglioramento e l'ottimizzazione dei processi di riciclaggio

dei RAEE, sia da un punto di vista economico che ambientale (*Zhang & Forsberg, 1997; Veit & Moura, 2015*).

Una sfida tecnologica, che consiste nell'implementare i processi di riciclaggio ecocompatibili ed economicamente sostenibili, partendo da una caratterizzazione approfondita di questo flusso di materiale specifico.

Obiettivi dello studio

Lo scopo del presente studio è quello di studiare le possibilità offerte dall'utilizzo combinato dell'analisi HSI nell'infrarosso e l'utilizzo dell'analisi XRF per eseguire una classificazione dei circuiti stampati e degli altri materiali, nonché identificare il rame nei circuiti stampati. In particolare, il rame presente nei circuiti stampati viene identificato con tecniche basate sull'analisi XRF e mediante modellazione chemiometrica. Mentre, le particelle di circuiti stampati sono identificate dagli altri materiali mediante tecniche HSI basate nel range dello SWIR (1000-2500 nm) e con l'utilizzo di metodi chemiometrici. Più nel dettaglio, sono stati messi a punto due modelli PLS-DA: uno per riconoscere il rame sull'ipercubo di dati XRF e l'altro per riconoscere i circuiti stampati a partire dall'ipercubo di dati HSI nello SWIR. Il diagramma di flusso delle analisi effettuate è riportato in [Figura 4.1.1](#).

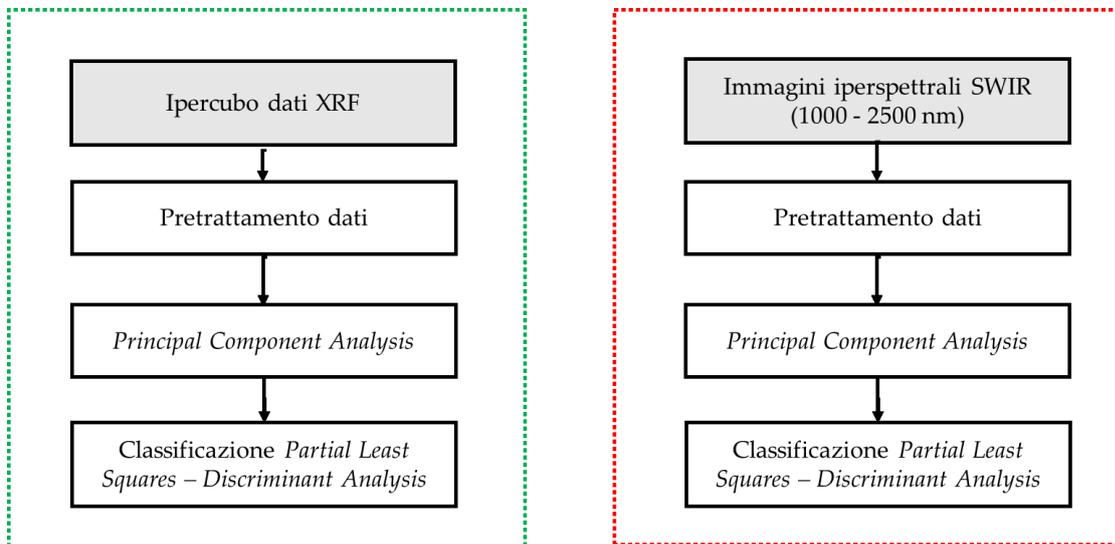


Figura 4.1.1. Diagramma di flusso delle analisi effettuate. Le analisi effettuate per riconoscere il rame sull'ipercubo di dati XRF sono incluse nel rettangolo tratteggiato in verde; mentre le analisi effettuate per riconoscere i circuiti stampati a partire dall'ipercubo di dati HSI nello SWIR sono incluse nel rettangolo tratteggiato in rosso.

Procedura

Preparazione dei campioni

Il materiale analizzato in questo studio consiste in un sottoprodotto di separazione magnetica per densità (MDS, *Magnetic Density Separation*) e comminuzione effettuata su un flusso RAEE di piccole e medie dimensioni.

Il separatore MDS può essere classificato nella categoria di separatori che utilizzano come mezzo di separazione liquidi magnetici. Cambiando lo stato di magnetizzazione del fluido magnetico, l'MDS consente la separazione simultanea di materiali con densità differenti. Tale separatore rappresenta un'ottima soluzione per concentrare minerali e metalli, nonché per effettuare separazione di polimeri differenti (Hu et al., 2011).

I sottoprodotti analizzati provenienti da separazione MDS (Figura 4.1.2) di classe granulometrica + 10 mm e trancia densimetrica 1300-2200 kg/m³ sono riportati in Figura 4.1.3.

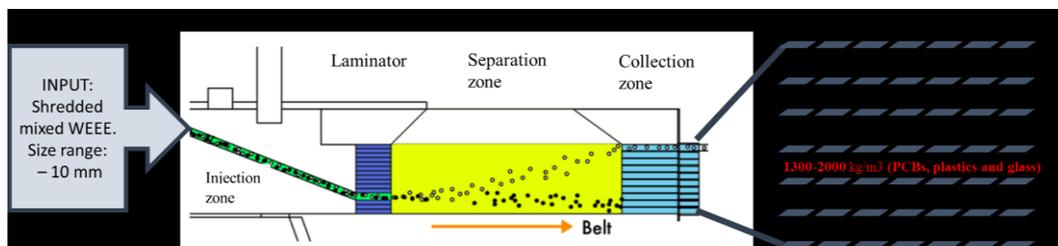


Figura 4.1.2. Schema della separazione MDS (Adattato da Hu et al., 2011).

Questi sono composti prevalentemente da circuiti stampati, plastiche, vetro/ceramiche e materiali cellulósici (Figura 4.1.4). Anche se, idealmente, in questa trancia densimetrica non dovrebbero essere presenti metalli, è ancora presente un cospicuo quantitativo di metalli (circa il 5% in peso). La frazione metallica presente nel campione analizzato è prevalentemente costituita dal rame inglobato all'interno delle particelle di circuito stampato e da quello presente nei cavi elettrici comminuiti.

Analisi XRF

L'analisi XRF è stata effettuata per valutare la composizione elementare e la concentrazione degli elementi presenti nel campione. Lo spettrometro μ -XRF utilizzato per le analisi è il Bruker M4 TORNADO (Bruker Corporation, 2013; Bruker Corporation, 2015). Sono state effettuate due scansioni: una su uno standard di rame e una sul materiale in capsula Petri. Le condizioni di acquisizione sono state le seguenti: 50 kV, 600 μ A. Sul materiale analizzato in capsula Petri è stata selezionata una ROI mediante il software HyperMap ed è stato ottenuto il report riferito ai pesi normalizzati degli elementi rilevati (si rimanda all'Allegato VI - Case study - Analisi micro-XRF di sottoprodotti MDS provenienti da RAEE).

Casi di studio

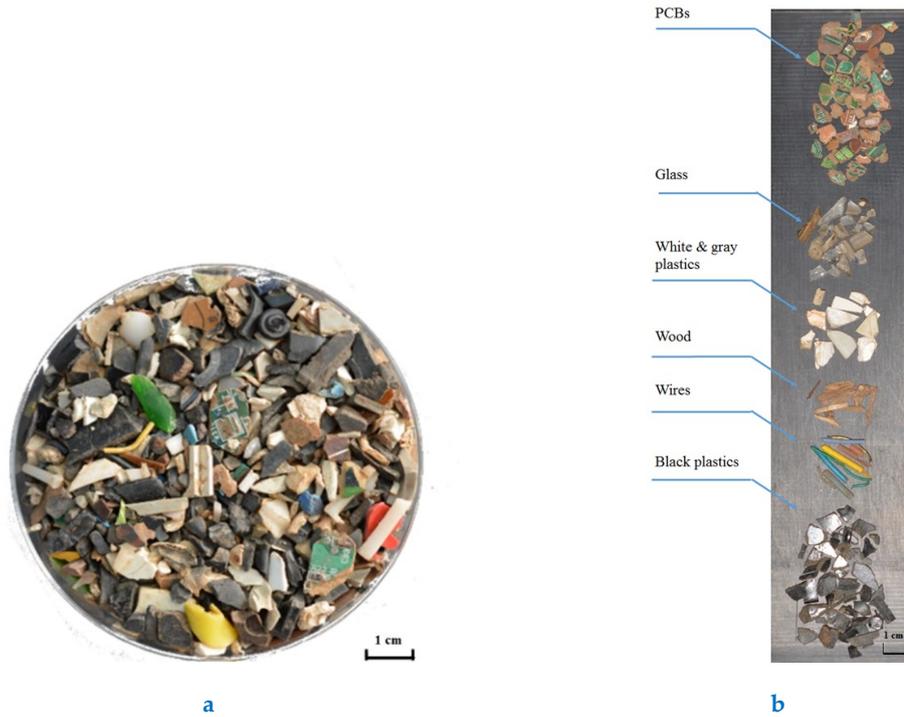


Figura 4.1.3. Sottoprodotto di separazione MDS analizzato (a) e set di addestramento per il modello PLS-DA finalizzato al riconoscimento dei circuiti stampati (b).

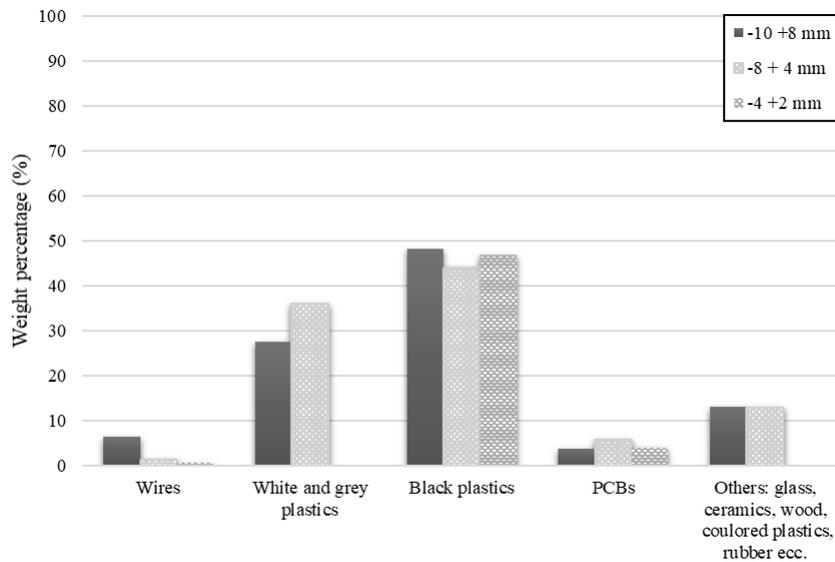


Figura 4.1.4. Composizione media in percentuale del sottoprodotto MDS analizzato, secondo classi granulometriche.

Gli ipercubi di dati ottenuti in formato `.raw` sono stati successivamente importati in ambiente MATLAB® mediante la funzione `importxrfmap.m` (vedere [Allegato IV - MATLAB script - importxrfmap.m](#)); mentre la funzione `create_datasetXRF.m` è stata utilizzata per re-interpolare la matrice di dati, riducendone le dimensioni, al fine di semplificare la gestione del set di dati e la procedura di analisi successiva, nonché per

diminuire il tempo necessario di elaborazione dati (si veda [Allegato V - MATLAB script - create_datasetXRF.m](#)). Le successive analisi effettuate sugli ipercubi sono state effettuate mediante l'ausilio dei toolbox MATLAB® PLS_toolbox e MIA_toolbox.

L'ipercubo di calibrazione è stato successivamente pre-elaborato con la seguente combinazione di algoritmi: *Normalize (1-Norm)*, *Baseline (Specified points)* e *Mean Center* ([Rinnan et al., 2009](#); [Eigenvector Research Inc., 2017b](#)). Come metodo di classificazione è stata scelta la PLS-DA ([Barker et al., 2003](#); [Wise et al., 2006](#); [Ballabio & Consonni, 2013](#)).

Il modello PLS-DA è stato calibrato utilizzando l'ipercubo dello standard di rame come set di addestramento ed è stato applicato al set di test (materiale in capsula Petri).

Analisi HSI

L'obiettivo principale dell'analisi HSI è quello di misurare le firme spettrali associate ai materiali presenti nel campione, esplorando tutti gli elementi o pixel, presenti nell'immagine acquisita.

Metodi basati sull'HSI, nel vicino infrarosso, sono stati utilizzati nell'ambito del riciclo dei RAEE ai fini dell'implementazione di selezionatori *sensor-based*; secondo differenti autori, questi metodi consentono di rilevare differenti plastiche e resine ai fini della separazione degli stessi da un flusso di RAEE ([Picon et al., 2010](#); [Masoumi et al., 2012](#)).

In questo studio, l'analisi di immagini iperspettrali è stata scelta ai fini della modellazione di un classificatore in grado di prevedere con buona accuratezza la posizione spaziale delle particelle contenenti metalli/leghe metalliche. Come set di calibrazione è stata acquisita una immagine iperspettrale raffigurante il set di addestramento (circuiti stampati, vetro, plastiche bianche e grigie, plastiche nere materiali cellulosici e cavi elettrici); mentre come set di test è stata acquisita una immagine iperspettrale sul materiale in capsula Petri. Lo strumento utilizzato per acquisire le immagini iperspettrali nell'infrarosso ad onde corte (1000-2500 nm) è il SisuCHEMA XL™ Chemical Imaging Workstation. Gli ipercubi ottenuti sono stati successivamente importati in ambiente MATLAB®, utilizzando PLS_toolbox e MIA_toolbox. Lo sfondo di ciascuna immagine è stato escluso per ridurre i tempi di calcolo e la memoria richiesta dal calcolatore.

Analisi dei risultati

Analisi quantitativa e identificazione del rame mediante analisi μ -XRF

Dal report ottenuto dall'analisi μ -XRF (si veda l'[Allegato VI - Case study - Analisi micro-XRF di sottoprodotti MDS provenienti da RAEE](#)), la concentrazione di rame presente nel materiale in capsula Petri è di circa 1.9 % in peso. La percentuale in peso dei circuiti stampati nel campione è di circa 4.7 %. Questo significa che i circuiti stampati sono ancora ricchi di rame. La mappa elementare in relativo di rame e piombo, ottenuta dal software

nativo HyperMap della BRUKER è riportato in [Figura 4.1.5a](#). Mentre, la mappa di predizione risultante dalla PLS-DA è riportata in [Figura 4.1.5b](#).

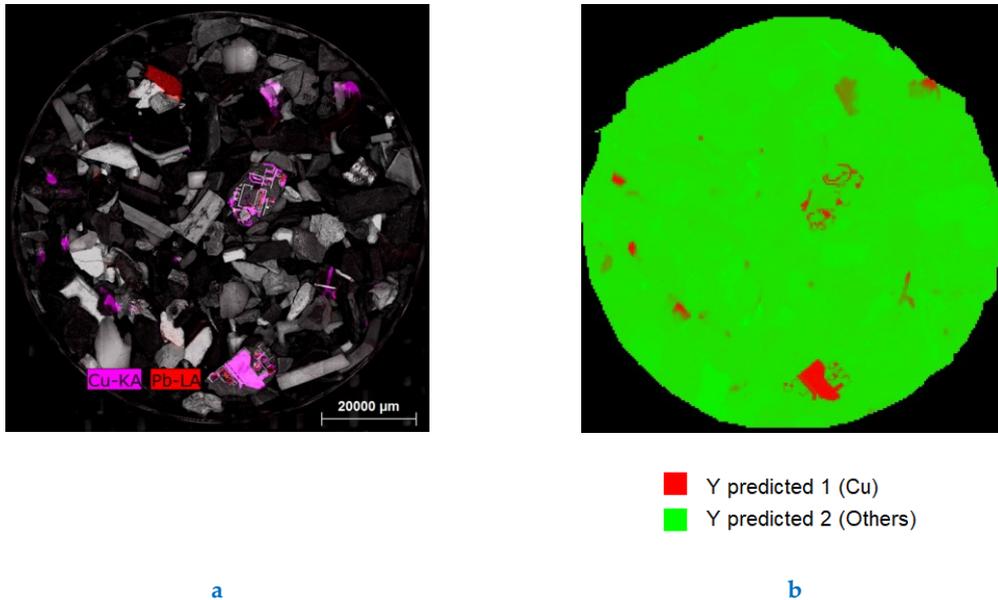


Figura 4.1.5. Mappa elementale relativa del rame e del piombo (a) e mappa di predizione PLS-DA per l'identificazione del rame applicata sull'ipercubo XRF (b).

Identificazione dei circuiti stampati mediante HSI nell'infrarosso ad onde corte

La mappa di predizione della PLS-DA per i circuiti stampati è riportata in [Figura 4.1.6](#). Come si può notare, il classificatore riesce a riconoscere tutti circuiti stampati presenti nel campione.

Discussione dei risultati

L'intervallo iperspettrale dello SWIR non consente una identificazione apprezzabile degli spettri di materiali metallici. Ad ogni modo, le mappature XRF hanno consentito l'identificazione del rame, contenuto prevalentemente nelle particelle di circuiti stampati. Il modello di classificazione PLS-DA applicato sull'immagine iperspettrale nello SWIR consente l'identificazione del materiale di base dei circuiti stampati: i.e. la resina epossidica presente sullo strato sottostante il rame e la pellicola protettiva per le saldature sovrastante tale strato.

Casi di studio

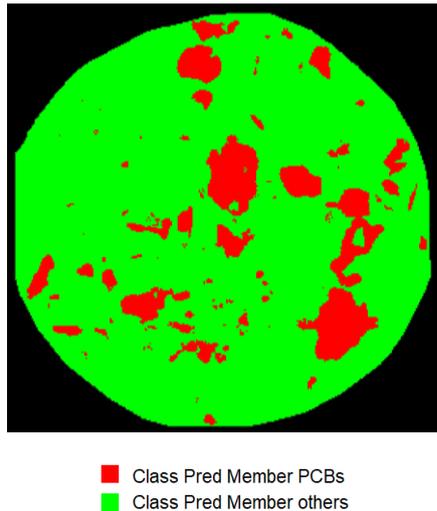


Figura 4.1.6. Mappa di predizione delle schede a circuito stampato ottenuta dalla PLS-DA applicata sull'immagine iperspettrale nello SWIR (1000-2500 nm).

Si può affermare che una tale applicazione di metodologie basate sull'HSI e l'analisi XRF apre le porte a possibili future innovazioni nel campo del riciclo dei RAEE, come ad esempio:

- I. La possibilità di mettere a punto sistemi in grado di riconoscere le particelle di circuito stampato, non solo per finalità di controllo qualità del prodotto ottenuto da un processo di comminazione e/o separazione, ma anche per finalità di separazione delle stesse particelle.
- II. La possibilità di assicurare una produzione di prodotti riciclati di qualità, anche in termini quantitativi; come in questo caso, assicurando l'ottenimento di circuiti stampati contenenti rame, grazie all'utilizzo combinato della metodologia HSI e la tecnologia XRF.

4.1.2 Identificazione a cascata di materiali provenienti da monitor a schermo piatto giunti a fine vita mediante *HyperSpectral Imaging*

Problematica e stato dell'arte

L'innovazione tecnologica e la messa in commercio di prodotti elettronici sempre più performanti, genera quantitativi sempre più cospicui di RAEE. In questo scenario, I monitor a schermo piatto giunti a fine vita (EoL, *End of Life*) rappresentano una categoria di rifiuti in continua crescita, come conseguenza della rapida evoluzione nelle tecnologie. Negli ultimi decenni, i monitor a schermo piatto giunti a fine vita provenienti da televisioni, notebook e tablet stanno aumentando in volume e continueranno ad aumentare (*Zeng et al., 2018; Salhofer et al., 2011*). Questo fatto è probabilmente dovuto al rapido cambiamento nelle tecnologie: i dispositivi elettronici diventano obsoleti e vengono rimpiazzati con prodotti caratterizzati da migliori performance (*Chancerel & Rotter, 2009; Oliveira et al., 2012; Palmieri et al., 2014*). Ad ogni modo, all'interno di questo specifico flusso di RAEE sono presenti numerosi materiali proficuamente recuperabili: i.e. metalli e leghe metalliche, plastiche di alta qualità ed altri composti.

I metalli non-ferrosi proficuamente recuperabili dai RAEE sono: il rame, il piombo, l'alluminio e lo zinco. Inoltre, i RAEE contengono anche metalli preziosi, come ad esempio l'oro, il platino, il palladio e l'argento (*UNEP 2009*). Ad esempio, i circuiti stampati sono ricchi di rame e contengono un cospicuo quantitativo di metalli preziosi (i.e. oro, argento e palladio). L'implementazione della catena di valorizzazione dei metalli (i.e. azioni di recupero), alla fine di un processo industriale è una sfida tecnologica, la quale deve essere necessariamente affrontata partendo da una caratterizzazione approfondita dello specifico flusso di materiali in esame (*Bonifazi et al., 2018c; Robinson 2009*). I RAEE sono anche ricchi di plastiche e altri materiali polimerici. Ad esempio, il policarbonato (PC) e/o il polimetilmetacrilato (PMMA) si possono spesso ritrovare come materiale costitutivo dell'unità di retroilluminazione (*Backlight Unit, BU*); più nel dettaglio possono essere ritrovati come materiale costituente il *Light Guide Panel* (LGP) degli schermi a monitor piatto (*Tarantili et al., 2010; Suresh et al., 2017; Suresh et al., 2018*) e degli schermi a cristalli liquidi (LCD, *Liquid Cristal Display*).

A causa del potenziale danno che potrebbe causare il non corretto trattamento dei RAEE e del cospicuo quantitativo di risorse ancora presenti negli stessi, è necessario implementare corrette azioni di riciclo per questo specifico flusso di rifiuti, sia da un punto di vista ambientale che da un punto di vista economico (*Veit & Bernardes, 2015*). L'impostazione di una infrastruttura di riciclo adeguata è quindi una priorità. Per progettare azioni di controllo e identificazione di materiali è necessario uno strumento che sia affidabile, efficiente e a basso costo; il quale sia in grado di valutare: la composizione del flusso di rifiuti, gli attributi fisico-chimici dei materiali di risulta. La conoscenza di tali informazioni è un requisito necessario al fine della progettazione e la messa in atto delle corrette azioni di riciclo.

Obiettivi dello studio

In questo studio, viene proposto un approccio di identificazione a cascata basato sull'utilizzo di tecniche di *HyperSpectral Imaging* nel vicino infrarosso (NIR-HSI). Più nel dettaglio, è stata messa a punto ed applicata una classificazione a 4 stadi per riconoscere diversi materiali presenti in uno specifico flusso di rifiuti RAEE: monitor a schermo piatto giunti a fine vita (EoL). Più nel dettaglio, il flusso di rifiuti studiato è costituito da un prodotto di comminazione di schermi a monitor piatto, provenienti da un impianto di trattamento RAEE.

L'approccio proposto si basa sull'HSI nel vicino infrarosso al fine di caratterizzare le differenti frazioni di materiali presenti nel campione, con l'obiettivo di definire una strategia di controllo e/o selezione. Più nel dettaglio il processo di identificazione a cascata è stato messo a punto per riconoscere le seguenti frazioni di materiali presenti nel campione analizzato: particelle metalliche, frammenti di circuiti stampati, componenti elettriche e plastiche varie. La procedura messa a punto in questo studio è schematizzata in [Figura 4.1.7](#).

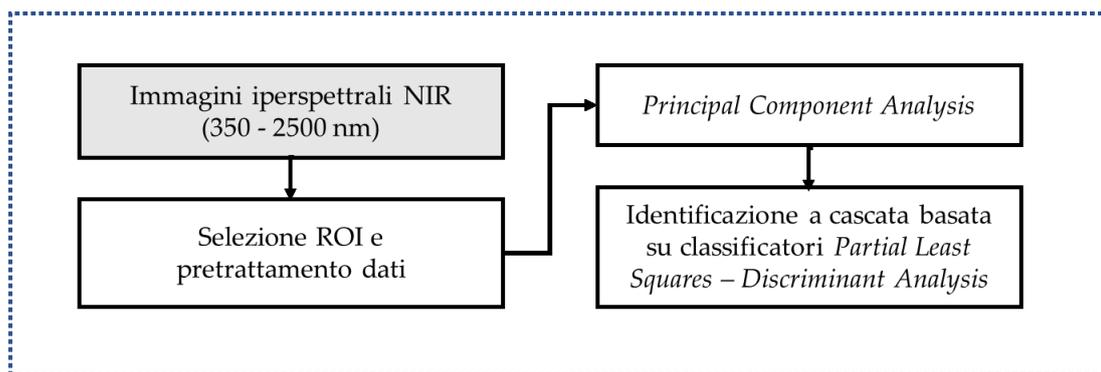


Figura 4.1.7. Diagramma di flusso delle analisi effettuate.

Procedura

Preparazione dei campioni

I campioni analizzati in questo studio provengono da una linea di comminazione di un impianto di trattamento RAEE. Più nel dettaglio, si tratta di monitor a schermo piatto EoL ([Figura 4.1.8](#)).

Casi di studio



Figura 4.1.8. Foto digitale del campione analizzato composto da monitor a schermo piatto EoL, proveniente da una linea di comminuzione di un impianto di trattamento RAEE.

La dimensione delle particelle analizzate è $-4 + 3$ cm. Il campionamento è stato eseguito secondo il metodo della quartatura. Il campione analizzato (circa 215 g) è composto prevalentemente da frammenti di LGP, Particelle di plastica nera, particelle metalliche, frammenti di circuiti stampati e componenti elettroniche ed altro (i.e. particelle polimeriche di varia natura e particelle cellulosiche). A seguito della quartatura è stata effettuata una selezione manuale in accordo con le classi di materiale identificate mediante ispezione visiva. Le categorie di materiale definite e le rispettive percentuali in peso sul totale del campione sono riportate in Figura 4.1.9.

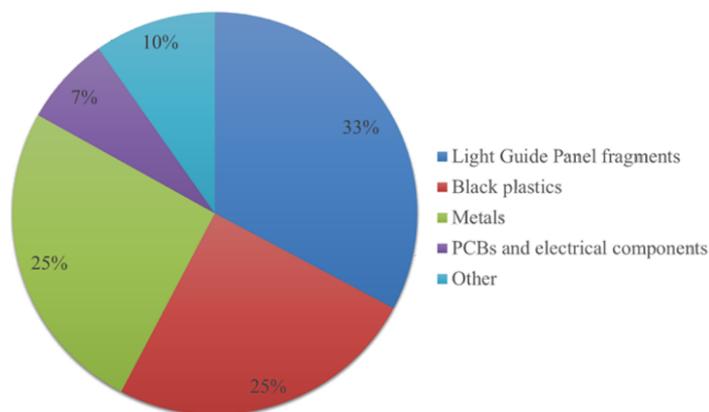


Figura 4.1.9. Composizione in percentuale delle categorie di materiale individuate nel campione analizzato.

Per la definizione del modello di classificazione a cascata è stato necessario suddividere il campione in due set: il set di addestramento per la calibrazione del modello e il set di validazione per verificare la validità del modello in classificazione. Il set di addestramento (Figura 4.1.10) è composto da 228 particelle (circa il 77 % in peso del campione), mentre il set di validazione (Figura 4.1.11) è costituito da 68 particelle (circa il 23% in peso del campione).

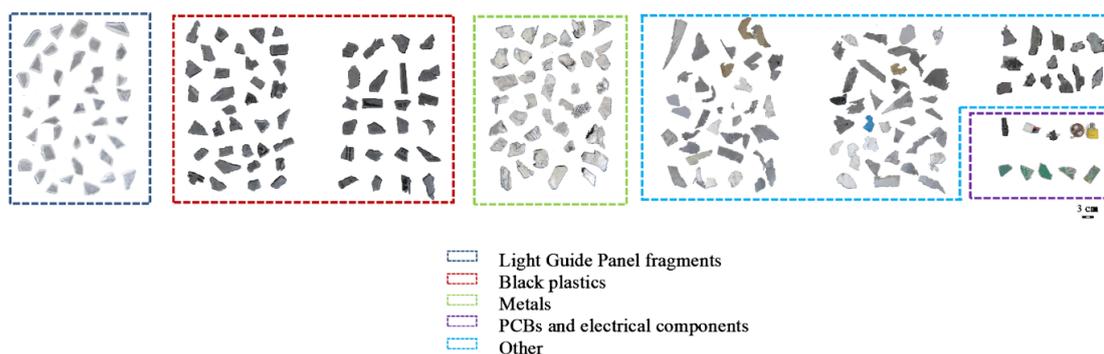


Figura 4.1.10. Le categorie di materiali utilizzate nel set di addestramento.

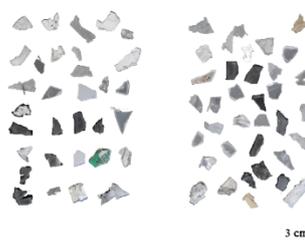


Figura 4.1.11. Immagine digitale delle particelle utilizzate nel set di validazione.

Acquisizione di immagini iperspettrali

Sono state acquisite 9 immagini iperspettrali; più nel dettaglio, 7 immagini sono state acquisite per il set di calibrazione e 3 immagini per il set di validazione. Le acquisizioni sono state effettuate mediante la NIR Spectral Camera™, equipaggiata con lo spettrografo ImSpector™ N17E (SPECIM Ltd, Finlandia) operante nel NIR (1000 – 1700 nm). Il software Spectral Scanner (Ver. 1.2) è stato utilizzato per acquisire e gestire le immagini iperspettrali. I dati acquisiti sono stati successivamente importati in ambiente MATLAB® (Ver. R2019a, The Mathworks, Inc.) ed analizzate mediante i toolbox della Eigenvector Research: PLS_Toolbox (Ver. 8.7) e MIA_toolbox.

Classificazione a cascata

Per effettuare il riconoscimento delle diverse frazioni di materiali presenti nel campione, sono stati messi a punto 4 modelli di classificazione in cascata (Figura 4.1.12). Nel primo step di classificazione vengono riconosciuti i frammenti di LGP (classe "LGP") dalle altre particelle ("Other (1)"); nel secondo step di classificazione vengono, invece, riconosciute le plastiche nere (classe "Black plastic") dalle altre particelle (classe "Other (2)"), a seguito della rimozione dei frammenti di LGP. Nel terzo step di classificazione sono classificate o come particelle metalliche, frammenti di circuiti stampati e componenti elettroniche (classe "Metals, PCBs and electrical components") oppure come altro materiale

(classe “Other (3)”). Nell’ultimo step di classificazione le particelle riconosciute come “Metals, PCBs and electrical components” sono classificate o come particelle metalliche (classe “Metals”), oppure come frammenti di circuito stampato e componenti elettronico (classe “PCBs and electrical components”).

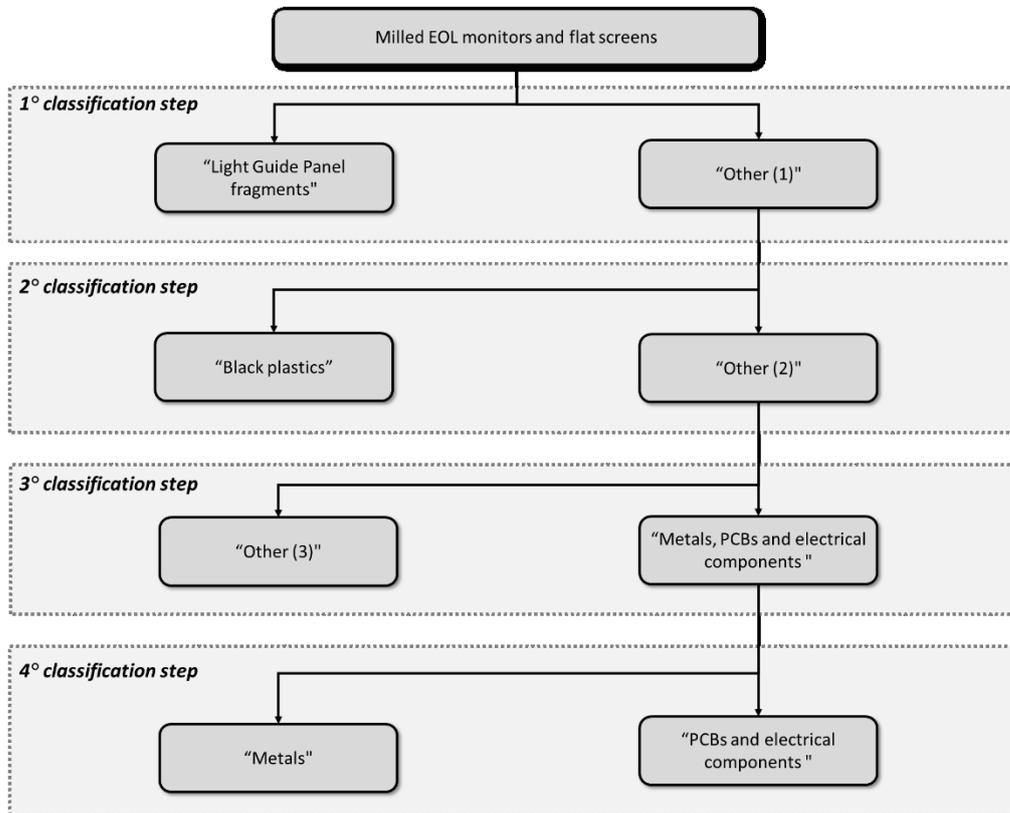


Figura 4.1.12. Diagramma di flusso della procedura di identificazione a cascata.

Per riconoscere le differenti categorie di materiali secondo la procedura di riconoscimento a cascata, ciascuno step di classificazione è stato effettuato mediante *Partial Least Squares - Discriminant Analysis* (Barker & Rayens, 2003; Ballabio & Consonni, 2013). Prima di ciascuno step di classificazione è stata applicata una combinazione di algoritmi di *pre-processing* ad hoc (Rinnan et al., 2009).

Analisi dei risultati

Identificazione dei frammenti di *Light Guide Panel*

A seguito della selezione delle ROI, sono stati estratti gli spettri medi delle classi “LGP” e “Other (1)” (Figura 4.1.13a). La combinazione scelta di algoritmi di pre-elaborazione (i.e. *Standard Normal Variate + Mean Center*), enfatizza le differenze tra le due classi (Figura 4.1.13b). I risultati della classificazione PLS-DA, in termini di mappa di predizione sono riportati in Figura 4.1.4; mentre, gli indicatori di performance del modello sono riportati

in [Tabella 4.1.1](#). Come mostrato nella mappa di predizione, tutte le particelle realmente appartenenti alla classe “LGP” sono assegnate correttamente alla propria classe di appartenenza.

Il Light Guide Panel (LGP) dell’unità di retroilluminazione (BLU) dei monitor a schermo e degli schermi a cristallo liquido (LCD) sono composti principalmente da due tipologie di polimeri ([Hwang & Ko, 2019](#); [Chen and Yu, 2007](#)): il policarbonato (PC) o dal polimetilmetacrilato (PMMA). Per riconoscere la tipologia di polimero principale presente nei frammenti di LGP è stato quindi messo a punto un modello di classificazione PLS-DA, addestrato con spettri in riflettanza di pellet in PC e PMMA vergini (si rimanda all’[Allegato VII- Case study - Riconoscimento della tipologia di polimero dei frammenti di Light Guide Panel provenienti da monitor a schermo piatto giunti a fine vita](#)). In questo caso, i frammenti di LGP sono risultati essere costituiti da PMMA.

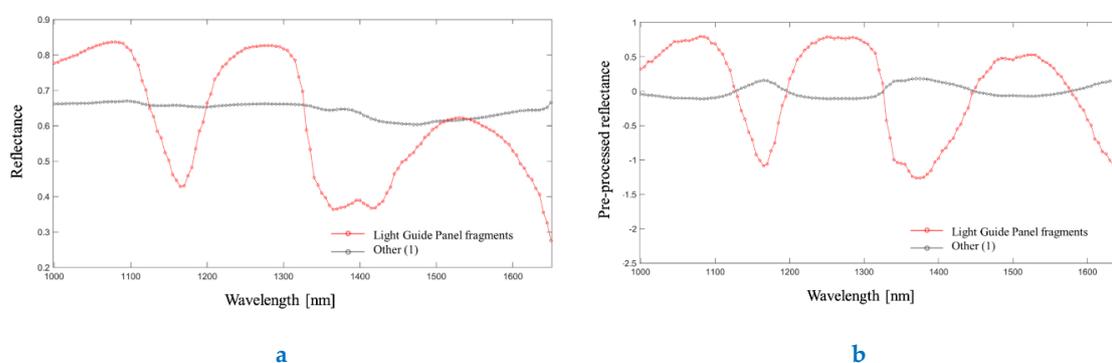


Figura 4.1.13. Spettri medi grezzi (a) e pre-elaborati (b) delle classi “LGP” e “Other (1)”

Tabella 4.1.1. Indicatori di performance relativi al primo step di classificazione.

| | <i>Class</i> | <i>Sensitivity</i> | <i>Specificity</i> |
|-------------------------|------------------|--------------------|--------------------|
| <i>Calibration</i> | <i>LGP</i> | 0.998 | 1.000 |
| | <i>Other (1)</i> | 1.000 | 0.998 |
| <i>Cross-validation</i> | <i>LGP</i> | 0.998 | 1.000 |
| | <i>Other (1)</i> | 1.000 | 0.998 |
| <i>Validation</i> | <i>LGP</i> | 0.859 | 0.999 |
| | <i>Other (1)</i> | 0.999 | 0.859 |

Casi di studio

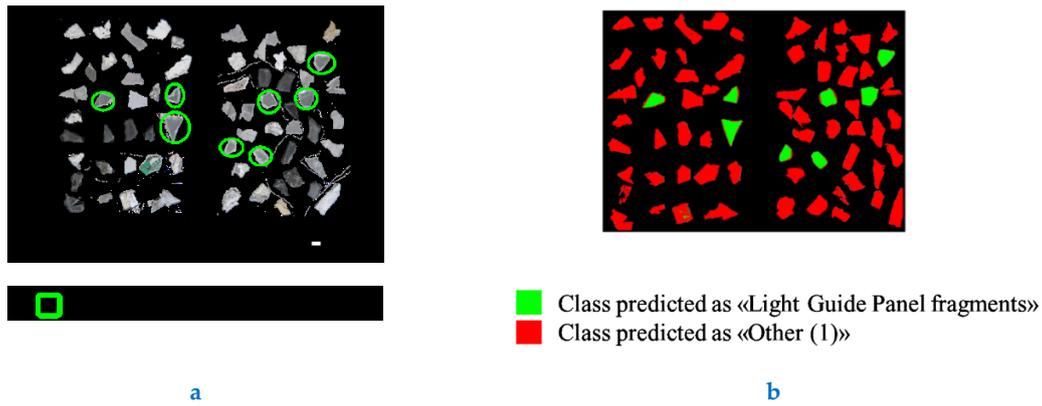


Figura 4.1.14. Immagine digitale (a) e mappa di predizione PLS-DA (b) delle classi di materiali “LGP” ed “Other (1)”.

Identificazione delle plastiche nere dagli altri materiali

In questo set-up sperimentale, è stato utilizzato l’algoritmo *Mean Center* per pre-elaborare i dati spettrali, a seguito della selezione di ROI. Gli spettri medi grezzi e pre-elaborati delle classi “*Black plastic*” ed “*Other (2)*” sono riportati in Figura 4.1.15. Al fine di classificare le due classi di materiale, è stata messa a punto una nuova classificazione PLS-DA. La mappa di predizione PLS-DA e la rispettiva immagine digitale del campione sono mostrate in Figura 4.1.16; mentre, gli indicatori di performance del modello sono riportati in Tabella 4.1.2. In questo caso tutte le particelle di plastica nera sono correttamente riconosciute dal modello.

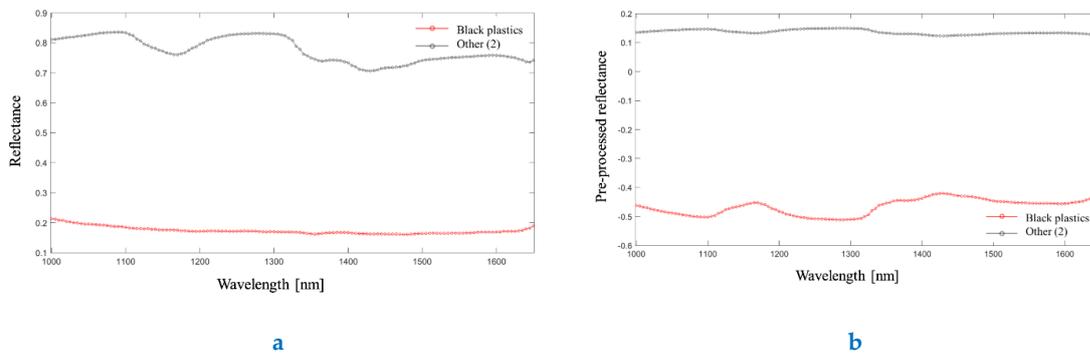


Figura 4.1.15. Spettri medi grezzi (a) e pre-elaborati (b) delle classi “*Black plastic*” e “*Other (2)*”.

Tabella 4.1.2. Indicatori di performance del secondo step di classificazione.

| | <i>Class</i> | <i>Sensitivity</i> | <i>Specificity</i> |
|-------------------------|----------------------|--------------------|--------------------|
| <i>Calibration</i> | <i>Black plastic</i> | 0.976 | 0.987 |
| | <i>Other (2)</i> | 0.987 | 0.976 |
| <i>Cross-validation</i> | <i>Black plastic</i> | 0.976 | 0.987 |
| | <i>Other (2)</i> | 0.987 | 0.976 |
| <i>Validation</i> | <i>Black plastic</i> | 0.869 | 0.992 |
| | <i>Other (2)</i> | 0.992 | 0.869 |

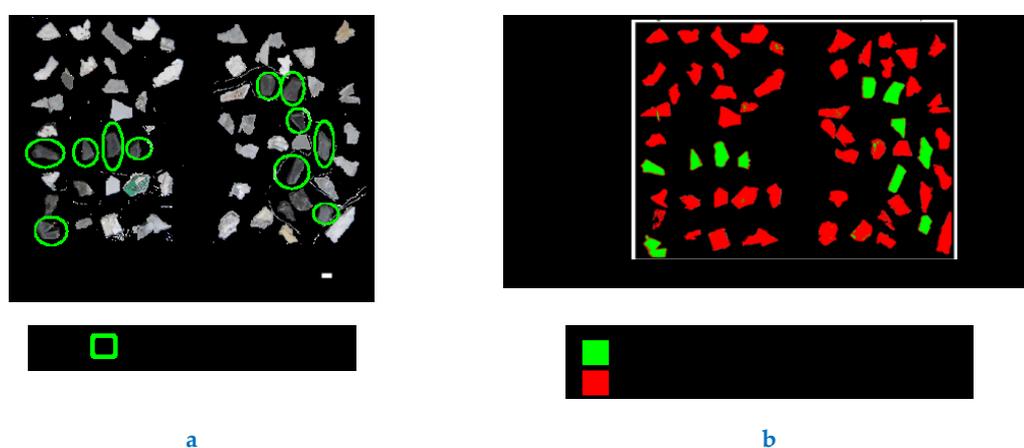


Figura 4.1.16. Immagine digitale (a) e mappa di predizione (b) delle classi di materiale “Black plastic” e “Other (2)”.

Identificazione delle particelle metalliche e dei frammenti di circuiti stampati e componenti elettroniche dagli altri materiali

In questo step di classificazione, la combinazione di algoritmi di pre-elaborazione utilizzati è stata: *Standard Normal Variate + S-G Smoothing + Mean Center*. Gli spettri medi grezzi e pre-elaborati sono riportati in Figura 4.1.17. In Figura 4.1.18, sono riportati i risultati della PLSDA in termini di mappa di predizione e la rispettiva immagine digitale del campione; mentre, in Tabella 4.1.3 sono riportati gli indicatori di performance riferiti alle classi modellate. In questo set-up sperimentale, alcuni pixel dell'immagine non vengono correttamente classificati. Ad ogni modo, se si considera esclusivamente il singolo oggetto nell'immagine (i.e. particella o frammento), più del 50 % dei pixel vengono correttamente assegnati alla reale classe di appartenenza. Si può dire che, se si implementasse una funzione in grado di assegnare un'unica classe a ciascun oggetto (i.e. viene assegnata un'unica classe ad un oggetto sulla base della percentuale di pixel correttamente classificati rispetto al totale dei pixel dell'oggetto), il modello sarebbe in grado di assegnare correttamente tutte le classi a ciascun oggetto.

Casi di studio

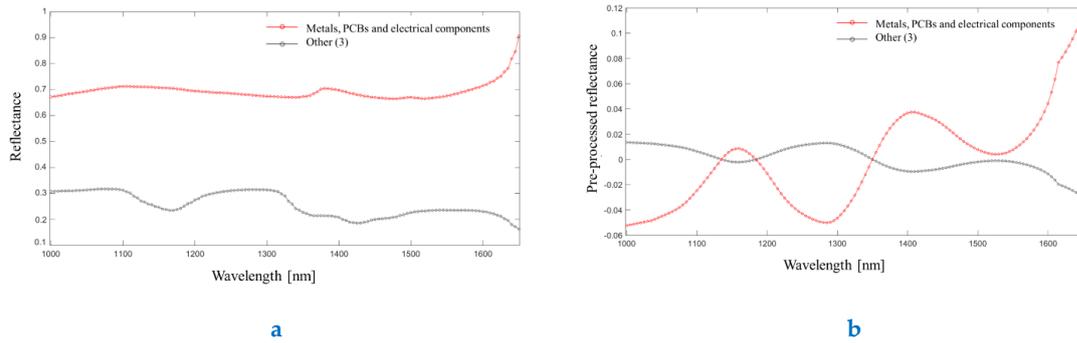


Figura 4.1.17. Spettri medi grezzi (a) e spettri medi pre-elaborati (b) delle classi di materiale “Metals, PCBs and electrical components” e “Other (3)”.

Tabella 4.1.3. Indicatori di performance del terzo step di classificazione.

| | Class | Sensitivity | Specificity |
|------------------|--|-------------|-------------|
| Calibration | Metals, PCBs and electrical components | 0.983 | 1.000 |
| | Other (3) | 1.000 | 0.983 |
| Cross-validation | Metal, PCB and electrical component | 0.983 | 1.000 |
| | Other (3) | 1.000 | 0.983 |
| Validation | Metal, PCB and electrical component | 0.755 | 0.982 |
| | Other (3) | 0.982 | 0.755 |

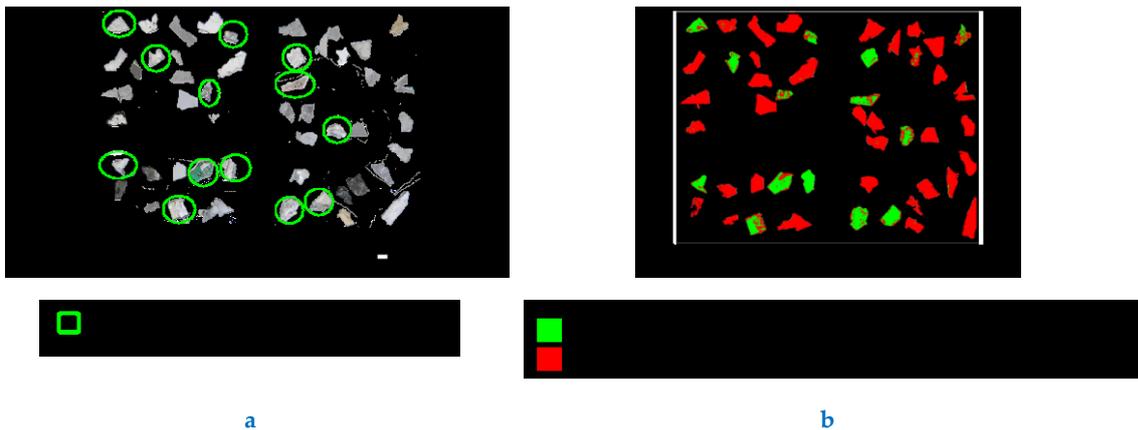


Figura 4.1.18. Immagine digitale (a) e mappa di predizione PLS-DA (b) delle classi di materiale Metals, PCBs and electrical components” e “Other (3)”.

Identificazione delle particelle metalliche dai frammenti di circuiti stampati e di componenti elettroniche

Casi di studio

L'ultimo step di classificazione è stato messo a punto per riconoscere le particelle metalliche (liberate) dai frammenti di circuito stampato e componenti elettroniche. Gli spettri medi e pre-elaborati (con la combinazione di algoritmi: *Standard Normal Variate*, *S-G Smoothing*, *Detrend* e *Mean Center*) delle classi "Metals" e "PCBs and electrical components", sono mostrati in [Figura 4.1.19](#). Mentre, in [Figura 4.1.20](#) sono riportati i risultati della classificazione PLS-DA in termini di mappa di predizione. Gli indicatori di performance del modello di classificazione sono riportati in [Tabella 4.1.4](#). In questo caso, tutte le particelle sono correttamente riconosciute e classificate. Alcuni pixel non vengono correttamente assegnati, ciò è probabilmente dovuto ai valori di riflettanza particolarmente elevati delle particelle considerate e/o a fenomeni di *scattering*.

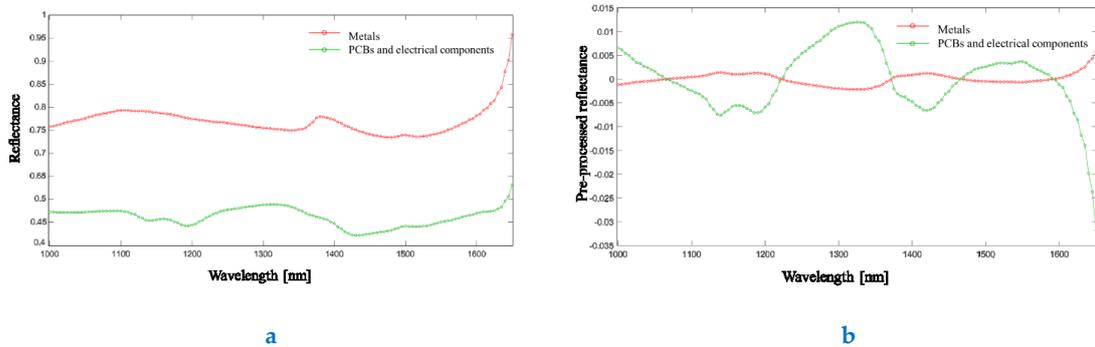


Figura 4.1.19. Spettri medi grezzi (a) e pre-elaborati (b) delle classi di materiale "Metals" e "PCBs and electrical components".

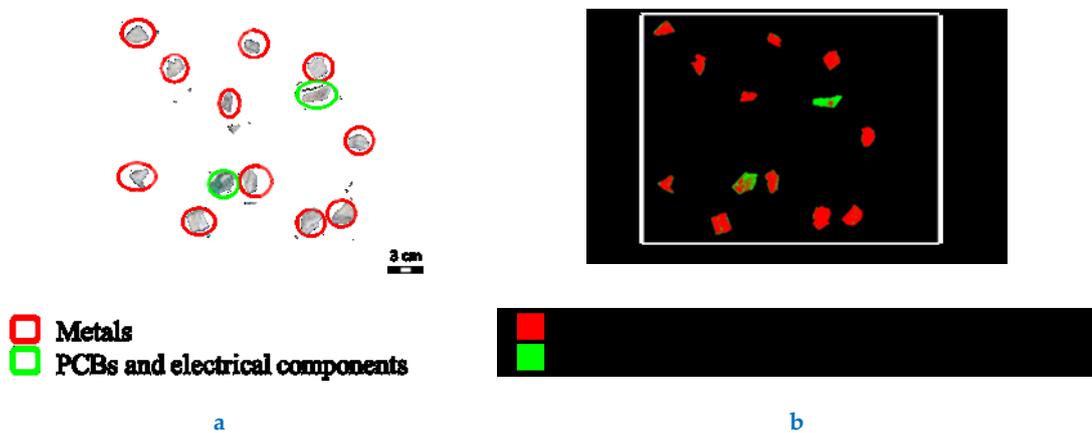


Figura 4.1.20. Immagine digitale (a) ed immagine predetta (b) delle classi di materiale "Metals" e "PCBs and electrical components".

Tabella 4.1.4. Indicatori di performance dell'ultimo step di classificazione.

| | <i>Class</i> | <i>Sensitivity</i> | <i>Specificity</i> |
|-------------------------|-------------------------------------|--------------------|--------------------|
| <i>Calibration</i> | <i>Metal</i> | 0.905 | 0.684 |
| | <i>PCB and electrical component</i> | 0.684 | 0.905 |
| <i>Cross-validation</i> | <i>Metal</i> | 0.905 | 0.684 |
| | <i>PCB and electrical component</i> | 0.684 | 0.905 |
| <i>Validation</i> | <i>Metal</i> | 0.918 | 0.693 |
| | <i>PCB and electrical component</i> | 0.693 | 0.918 |

Discussione dei risultati

In questo studio, si è riusciti a mettere a punto una procedura di identificazione a cascata per riconoscere differenti categorie di materiali presenti all'interno di un flusso di RAEE specifico: monitor a schermo piatto giunti a fine vita. Più nel dettaglio, grazie all'utilizzo di un sistema HSI operante nel NIR, sono stati messi a punto 4 classificatori in grado di identificare le classi di materiale: "LGP", "Black Plastics", "PCBs and electrical components" e "Metals". Le performance di ciascun classificatore nel riconoscere le specifiche categorie in ciascuno step sono risultate essere molto promettenti. La presenza di pixel non correttamente classificati è probabilmente dovuta a fenomeni di *scattering*, alla non-omogeneità della superficie delle particelle e alla possibile presenza di impurezze. Una logica basata su algoritmica *machine vision* è necessaria per migliorare la classificazione, come ad esempio una funzione in grado di assegnare una sola classe a seconda della percentuale di pixel correttamente assegnati in un determinato oggetto (i.e. se i pixel correttamente assegnati > 50 % dell'areale dell'oggetto, allora l'intero oggetto verrà assegnato ad una determinata classe).

Adottando tale approccio, i differenti materiali presenti in questo specifico flusso di RAEE sono correttamente riconosciuti e classificati. I risultati ottenuti possono essere utili non solo per la messa a punto di un sistema di controllo di qualità dei materiali risultanti dal processo di riciclo, ma anche per migliorare le azioni specifiche di selezione automatica specifiche per questo settore. L'approccio proposto presenta numerosi vantaggi: è rapido, non distruttivo e a basso costo. La riduzione dei costi di lavorazione è un obiettivo importante da perseguire nel settore delle materie prime secondarie, dove l'utilizzo di dispositivi costosi non può essere proposto per un processo di riciclaggio efficiente ed economicamente valido. La procedura implementata può essere proficuamente utilizzata per impostare strategie online finalizzate a migliorare l'efficienza dei processi di riciclaggio, riducendo i costi e migliorando la "qualità finale" dei prodotti recuperati. Infine, se pienamente implementato, l'approccio NIR-HSI proposto offre la possibilità di sviluppare un sistema in grado di riconoscere materiali

Casi di studio

diversi in un flusso RAEE, che può essere utilizzato non solo come nucleo analitico per eseguire il controllo di qualità, ma anche come motore di selezione.

4.1.3 Identificazione a cascata di materiali provenienti da rifiuti di imballaggi mediante *HyperSpectral Imaging*

Problematica e stato dell'arte

I rifiuti di imballaggi rappresentano una delle principali fonti per il recupero di plastiche come materie prime seconde (Bonifazi et al., 2015a). Il riciclo dei rifiuti di imballaggi prevede una complessa catena di attività. Questa può essere suddivisa in tre fasi principali: i) raccolta porta a porta o recupero da rifiuti indifferenziati, ii) selezione ed infine, iii) riciclo meccanico (Jansen et al., 2015). La complessità di tali attività è principalmente influenzata dalla complessità di questa tipologia di rifiuti proveniente da raccolta differenziata. Infatti, la presenza di potenziali contaminanti come frammenti di materiali cellululosici e poliaccoppiati di varia natura rendono questo flusso di rifiuti particolarmente eterogeneo. Questi contaminanti, inoltre, sono difficoltosi da separare mediante l'utilizzo di metodi classici (i.e. separazione gravimetrica), in quanto presentano caratteristiche fisiche simili.

Ad ogni modo, il riconoscimento di alcune tipologie di materiali presenti in un flusso di rifiuti può essere effettuato grazie alle risposte spettrali della superficie degli stessi. Tramite la risposta spettrale della superficie dei materiali analizzati, possono essere ottenute informazioni utili riguardanti le caratteristiche fisico-chimiche degli stessi. Queste informazioni possono essere sfruttate per il set-up e la messa a punto di logiche in grado di riconoscere i differenti materiali presenti in un flusso di rifiuti risultanti da specifiche azioni di selezione e/o trattamento (Bonifazi et al., 2009).

Negli ultimi anni, la maggior parte delle operazioni industriali finalizzate alla selezione automatizzata di plastiche provenienti da imballaggi sono supportate da tecnologie basate su sensori ottici. In questo panorama, le tecnologie basate sull'HSI nel vicino infrarosso stanno godendo di un ottimo successo (Hollestein et al., 2015). Infatti, l'HSI rappresenta una buona soluzione per differenti *task*, quali: caratterizzazione, classificazione e controllo di qualità di materiali e bio-materiali. Per questo tale tecnica viene utilizzata in differenti settori: i.e. beni culturali (Agresti et al., 2013; Capobianco et al., 2015), industria agro-alimentare (Kumaravelu & Gopal, 2015; Tsuchikawa & Kobori, 2015; Serranti et al., 2018a; Serranti et al., 2018b), industria chimico-farmaceutica (Larrechi & Callao, 2003; Roggo et al., 2007) e chimica analitica (Pasquini, 2003). Proprio negli ultimi anni, l'HSI sta emergendo nell'industria delle materie prime secondarie e del riciclo (Serranti et al., 2011; Serranti et al., 2013; Hu et al. 2013; Ulrici et al., 2013; Palmieri et al., 2014; Bonifazi et al., 2015a; Bonifazi et al., 2015b; Serranti et al., 2015).

L'identificazione sistematica di contaminanti e/o impurità presenti in un flusso di rifiuti provenienti da imballaggi e dei differenti materiali presenti in tale flusso di rifiuti è la chiave per certificare e classificare i materiali in alimentazione all'impianto di riciclo e per monitorare i prodotti e/o sottoprodotti in uscita all'impianto, i quali devono presentare determinati standard per essere immessi nuovamente nel mercato. L'utilizzo

di tecniche basate sull'imaging iperspettrale e machine learning consentono di raggiungere tali obiettivi.

Obiettivi dello studio

In questo studio, è stata studiata una logica di monitoraggio per la separazione automatizzata di carta, cartone, plastiche e poliaccoppiati. Più nel dettaglio, è stata implementata una logica di identificazione a cascata basata sull'HSI nel vicino infrarosso, al fine di identificare differenti materiali presenti all'interno di un flusso di rifiuti di imballaggi provenienti dalla raccolta differenziata della plastica. In particolare, il processo di identificazione a cascata (*Silva et al., 2017*) è stato impostato per riconoscere: la frazione plastica, la frazione cellulosa e i poliaccoppiati di carta e plastica.

Il modello di classificazione a cascata si basa su tre classificatori PLS-DA, impostati in due livelli: nel primo si identificano i frammenti cellulostici da quelli di plastica, nel secondo livello si identificano i poliaccoppiati a base prevalentemente cellulosa (i.e. Tetra Pak) dai frammenti di carta, e i poliaccoppiati a base plastica dai frammenti di plastiche. Il diagramma di flusso delle analisi effettuate è riportato in [Figura 4.1.21](#).

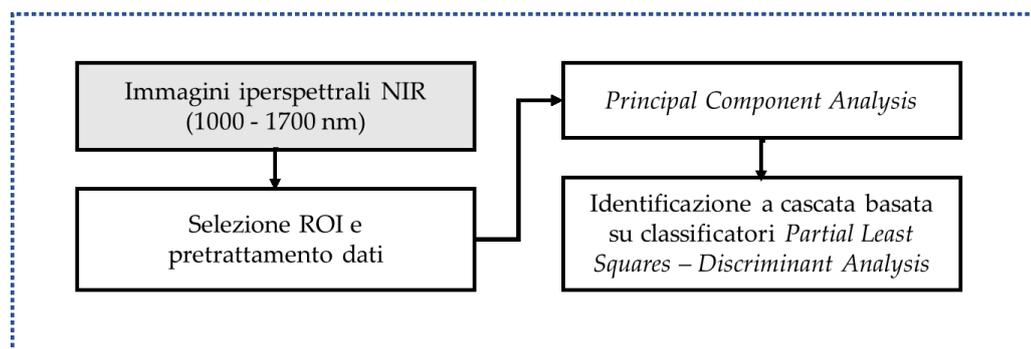


Figura 4.1.21. Diagramma di flusso delle analisi effettuate.

Procedura

Preparazione dei campioni

Il campione analizzato proviene da un impianto intermedio per la pressatura (CC, Centro Comprensionale), situato nella regione Lazio (Italia). Il campionamento è stato effettuato su una balla compattata di frammenti di imballi in plastica, risultanti da separazione manuale e semi-automatica (i.e. *oversize* di un vaglio rotativo a tamburo). Idealmente, l'output di questa linea di separazione non dovrebbe presentare frammenti di carta, cartone e/o altri materiali a base cellulosa. In realtà, invece, ciò non accade; da controlli qualità effettuati sul materiale in uscita dall'impianto si evince che giornalmente si ha in media: il 84.9 % in peso di plastica, il 3.45 % di carta (e simili) ed il 6.2 % di imballaggi con evidenti residui del contenuto. Il metodo di campionatura utilizzato è

stato quello della quartatura. Da 5 kg di questo materiale sono stati successivamente selezionati 70 frammenti.

Carta, cartone e plastica. Per calibrare il modello in grado di riconoscere i frammenti di materiali cellulosici dai frammenti di plastica, sono stati usati 40 frammenti misti di plastica, carta e cartone (Figura 4.1.22): 24 particelle di plastica (nel set sono stati inclusi anche frammenti di poliaccoppiati a base prevalente di plastica) e 16 particelle di materiali cellulosici (inclusi poliaccoppiati a base prevalente di carta. I.e. Tetra Pak®). Mentre, per la validazione sono state utilizzate 30 particelle, suddivise in due set: “Validation set 1” e “Validation set 2” (Figura 4.1.23). Nel “Validation set 1” sono state incluse 16 particelle: 11 frammenti di plastica (inclusi frammenti di poliaccoppiati a base prevalente di plastica) e 5 particelle di materiale cellulosici (inclusi poliaccoppiati a base prevalente di carta). Mentre nel “Validation set 2” sono state utilizzate 14 particelle: 9 frammenti di plastica (inclusi frammenti di poliaccoppiati a base prevalente di plastica) e 5 particelle di materiale cellulosico (inclusi poliaccoppiati a base prevalente di carta).

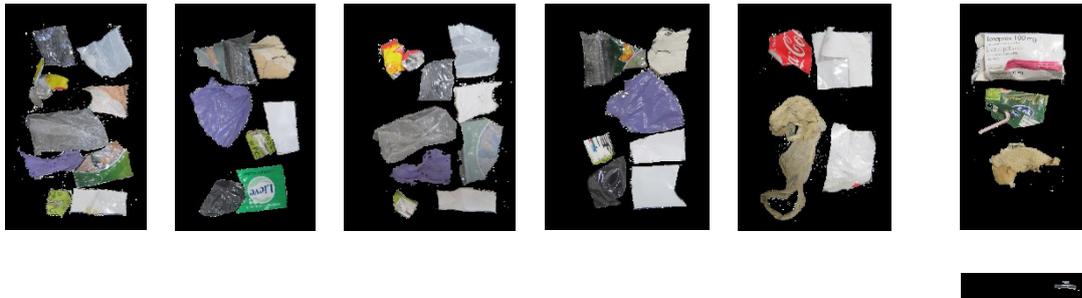


Figura 4.1.22. Immagini digitali delle particelle usate nel set di addestramento.



Figura 4.1.23. Immagini digitali delle particelle usate nel set di validazione: “Validation set 1” (a) and “Validation set 2” (b).

Poliaccoppiati a base prevalente di carta. Per riconoscere i poliaccoppiati a base prevalente di carta dagli altri frammenti di carta e cartone sono state incluse altre 5 particelle di poliaccoppiati nel set di calibrazione (Figura 4.1.24).

Casi di studio

Poliaccoppiati a base prevalente di plastica. Per riconoscere le particelle di poliaccoppiati a base prevalente di plastica dagli altri frammenti di plastica, sono stati inclusi 6 frammenti di poliaccoppiati nel set di calibrazione (Figura 4.1.25).

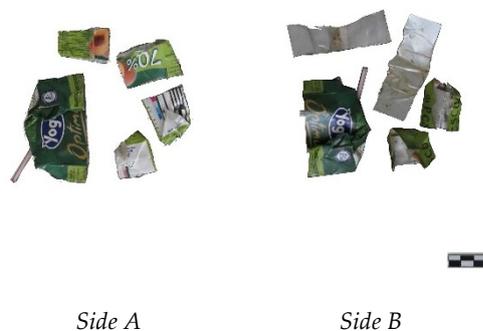


Figura 4.1.24. Immagini digitali delle particelle di poliaccoppiati a base prevalente di carta aggiunti al set di calibrazione. Side A mostra i frammenti con il lato plastica/carta/metallo/plastica esposto. Mentre, Side B mostra i frammenti con il lato plastica/metallo/carta/plastica esposto.



Figura 4.1.25. Immagini digitali delle particelle di poliaccoppiati a base prevalente di plastica. Side A mostra i frammenti con il lato plastica/metallo esposto. Mentre, Side B mostra i frammenti con il lato metallo/plastica esposto.

Acquisizione di immagini iperspettrali

Il sistema iperspettrale utilizzato in questo studio è la NIR spectral camera (Specim) che monta uno spettrografo imaging ImSpector N17E, operante nell'intervallo spettrale del NIR (1000-1700nm). Per acquisire le immagini iperspettrali, la velocità del nastro trasportatore è stata impostata su 10 mm / s (per 1000 frame).

Le 40 particelle utilizzate come set di addestramento per sviluppare un modello in grado di discriminare i frammenti di carta da particelle di polimero (di seguito denominate "Set di addestramento"), sono state acquisite in 6 immagini iperspettrali (Figura 4.1.26a).

Per quanto riguarda l'impostazione del modello per discriminare i poliaccoppiati a base prevalente di carta dagli altri materiali in cellulosa, sono state acquisite 5 frammenti di poliaccoppiati in 2 immagini iperspettrali (Figura 4.1.27a): un'immagine che mostra le particelle dal lato plastica / carta / metallo / plastica (*Side A*), mentre l'altra immagine che mostra il lato plastica / metallo / carta / plastica (*Side B*).

Invece, per impostare il modello per identificare frammenti di poliaccoppiati a base prevalente di plastica dalle altre particelle di plastica, sono stati acquisiti 6 frammenti di poliaccoppiati in 4 immagini iperspettrali su 2 lati (Figura 4.1.28a): 2 immagini con il lato plastica / metallo delle particelle esposto (*Side A*) e 2 immagini con il lato metallo / plastica delle particelle rivolto verso il *detector* (*Side B*).

Mentre, per preparare il set di dati su cui convalidare i modelli, 16 particelle incluse in "*Validation set 1*" sono state scansionate in 2 immagini iperspettrali (Figura 4.1.29a) e 14 particelle incluse in "*Validation set 2*" sono state acquisite in altre 2 immagini iperspettrali (Figura 4.1.29b).

Pre-elaborazione dati ed analisi in componenti principali

Ciascuna immagine iperspettrale acquisita è stata inizialmente ridotta da 950-1700 nm a 1000-1650 nm, escludendo le parti più rumorose degli spettri. Successivamente, è stato rimosso lo sfondo a tutte le immagini iperspettrali acquisite. Come mostrato nella Figura 4.1.26, sono state selezionate 39 ROI tra le 6 immagini iperspettrali del "set di addestramento": 23 ROI di particelle di plastica ("*Mixed polymers*") e 16 ROI di particelle di carta ("*Paper & cardboard*"). Gli spettri dei 39 ROI sono stati quindi estratti e concatenati per creare un unico set di dati (per le classi "*Mixed polymers*" e "*Paper & cardboard*").

Mentre, per la calibrazione del modello di classificazione dei poliaccoppiati a base prevalente di carta e delle altre particelle di cellulosa, sono state estratte 15 ROI dalle due immagini iperspettrali concatenate, come mostrato in Figura 4.1.27: 8 ROI da *Side A* (lato plastica/carta/metallo/plastica) e 7 ROI da *Side B* (lato plastica/metallo/carta/plastica). Gli spettri da ROI estratti sono stati utilizzati per creare un nuovo set di dati per la classe "*Paper & cardboard (laminated card)*".

La Figura 4.1.28 mostra le 4 immagini iperspettrali concatenate raffiguranti particelle di poliaccoppiati in plastica inclusi nel set di calibrazione e le 30 ROI estratte (14 ROI sul lato A e 16 ROI sul lato B: il lato A è il lato in plastica / metallo, mentre il lato B è il metallo / plastica). Gli spettri da ROI estratti sono stati utilizzati per creare un nuovo set di dati per la classe "*Polymer (Laminated plastic)*". Mentre le immagini iperspettrali del "*Validation set 1*" sono state concatenate solo insieme, come mostrato nella Figura 4.1.29. La stessa procedura è stata applicata al "*Validation set 2*".

Casi di studio

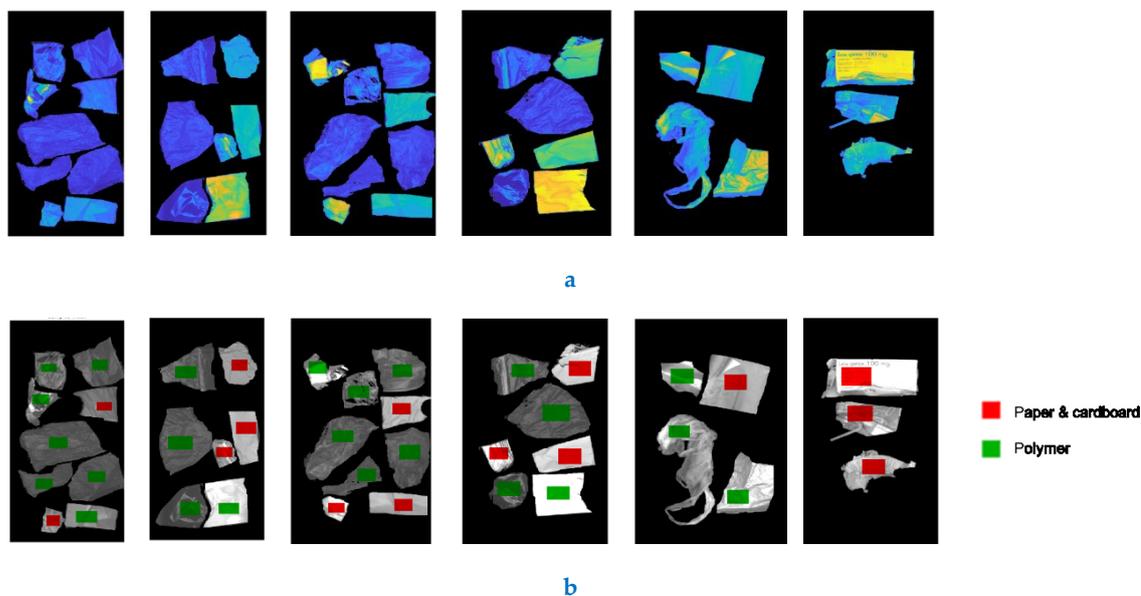


Figura 4.1.26. “Calibration set”: Immagini iperspettrali (a) e ROI delle classi “Paper & cardboard” e “Polymer” (b).

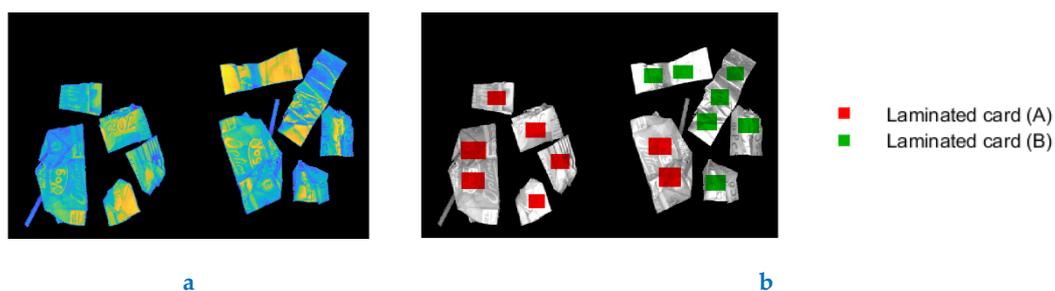


Figura 4.1.27. Poliaccoppiati a base prevalente di carta: a) ipercubo delle particelle e b) ROI selezionate (side A è il lato plastica/carta/metallo/plastica, mentre side B è il lato plastica/metallo/carta/plastica).

Ciascun set di calibrazione è stato pre-elaborato con una combinazione di *pre-processing*. La pre-elaborazione dei dati non è solo necessaria per correggere gli effetti di artefatti rumorosi, ma anche per migliorare le differenze che si verificano tra i cluster di dati. Sono state esplorate diverse strategie di pre-elaborazione: è stata quindi scelta quella che ha dato una risposta migliore in termini di decomposizione dei dati. La combinazione di algoritmi di pre-elaborazione utilizzata è *Standard Normal Variate (SNV) + Smoothing + Mean Center*.

Per identificare e rimuovere eventuali *outlier* dai set di calibrazione è stata utilizzata l'analisi in componenti principali (Wold *et al.*, 1987).

Casi di studio

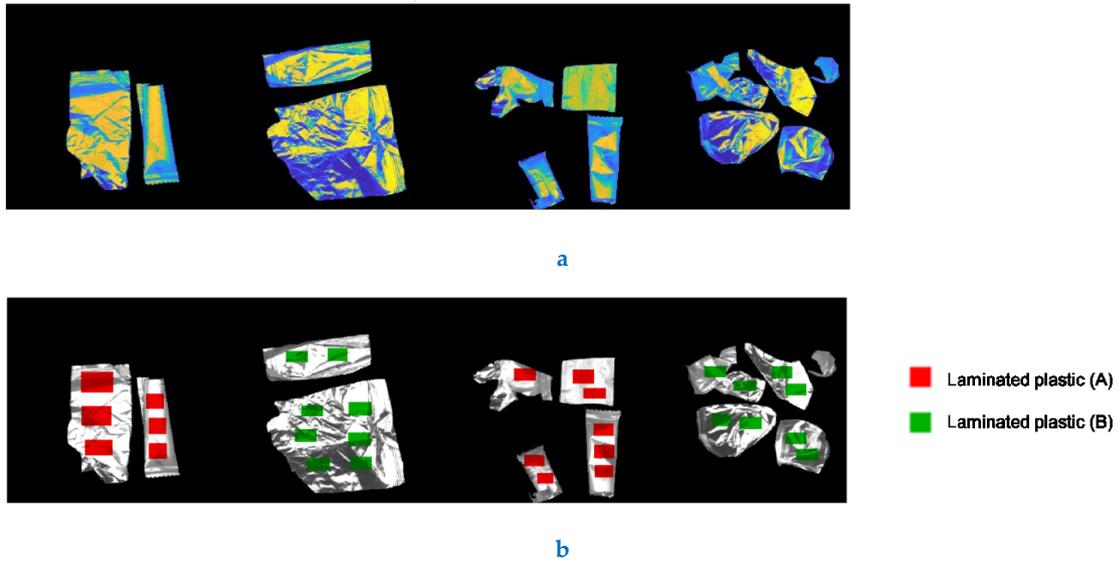


Figura 4.1.28. Frammenti di poliaccoppiati a base prevalente di plastica (Side A è il lato plastica/metallo, mentre Side B è il lato metallo/plastica): a) immagini iperspettrali concatenate e b) ROI selezionate.

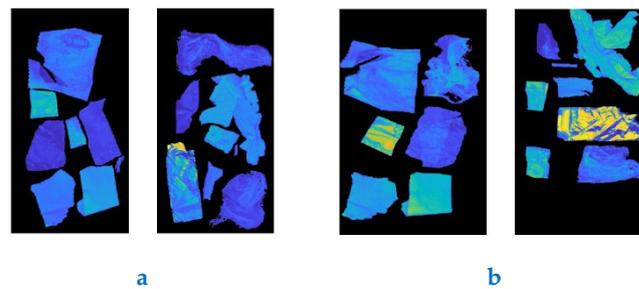


Figura 4.1.29. Immagini iperspettrali: Validation set 1" (a) e "Validation set 2" (b).

Classificazione a cascata

Per riconoscere i frammenti di plastica da quelli di carta e cartone, e successivamente per identificare i frammenti di poliaccoppiati (poliaccoppiati a base prevalente di plastica e a base prevalente di carta) è stata messa a punto una procedura di classificazione ad albero basata su classificatori PLS-DA, come mostrato in Figura 4.1.30 (Barker et al., 2003; Wise et al., 2006; Ballabio & Consonni, 2013).

Nella prima fase di classificazione vengono riconosciuti i frammenti di materiale celluloso ("Paper & Cardboard") e i frammenti di materiale plastico/polimerico ("Polymer"). Nella seconda fase di classificazione, vengono riconosciuti i frammenti di poliaccoppiato a base prevalente di carta ("Paper & cardboard (Laminated card)") dal restante materiale celluloso ("Paper & cardboard 2") a partire dalla classe "Paper & cardboard"; mentre, a partire dalla classe identificata "Polymer" vengono riconosciuti i frammenti di poliaccoppiati a base prevalente di plastica ("Polymer (Laminated plastic)") dalle altre particelle di plastica ("Polymer 2").

Casi di studio

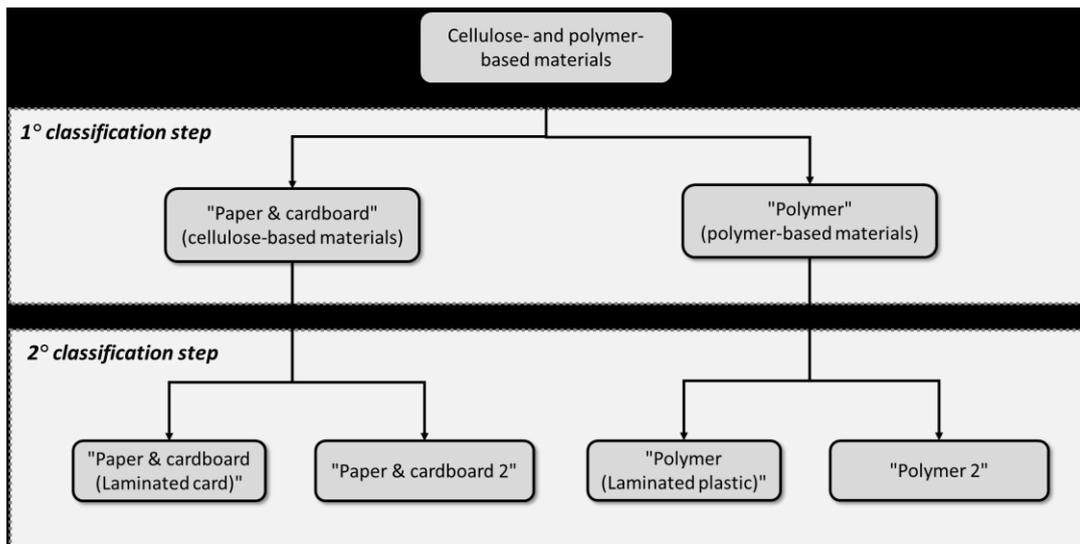


Figura 4.1.30. Logica di identificazione a cascata per la separazione di carta, cartone, poliaccoppiati a base prevalente di carta e poliaccoppiati a base prevalente di plastica.

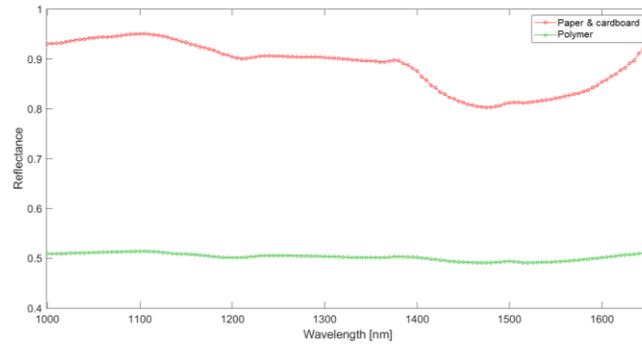
Il primo modello è stato addestrato per riconoscere le classi generali *"Paper & cardboard"* e *"Polymer"* utilizzando gli spettri estratti da ROI sul set di calibrazione (Figura 4.1.26b). Per scegliere la giusta complessità del modello è stato utilizzato l'algoritmo di *cross-validation Venetian-blinds*. In questo caso, sono state scelte 3 variabili latenti. Il modello è stato successivamente validato su *"Validation set 1"* e su *"Validation set 2"* per valutarne l'abilità predittiva. Nella seconda fase della procedura, il modello PLS-DA per discriminare le particelle di poliaccoppiati a base prevalente di carta (*"Paper & cardboard (Laminated card)"*) dalle altre particelle cellulosiche (*"Paper & cardboard 2"*) è stato addestrato utilizzando le ROI selezionate per la classe *"Paper & cardboard (Laminated card)"* (Figura 4.1.27b). In questo caso sono state utilizzate 7 LV. Il modello è stato validato sugli ipercubi *"Validation set 1"* e su *"Validation set 2"*, a seguito della rimozione delle particelle di plastica. Mentre, il modello PLS-DA per identificare i poliaccoppiati a base prevalente di plastica (*"Polymer (Laminated plastic)"*) dalle altre particelle di plastica (*"Polymer 2"*) è stato addestrato a partire dagli spettri estratti dalle ROI mostrate in Figura 4.1.28b. In questo caso, sono state scelte 5 LV. Questa volta, il modello è stato validato sugli ipercubi *"Validation set 1"* e su *"Validation set 2"*, a seguito della rimozione delle particelle a base di materiale cellulosico.

Analisi dei risultati

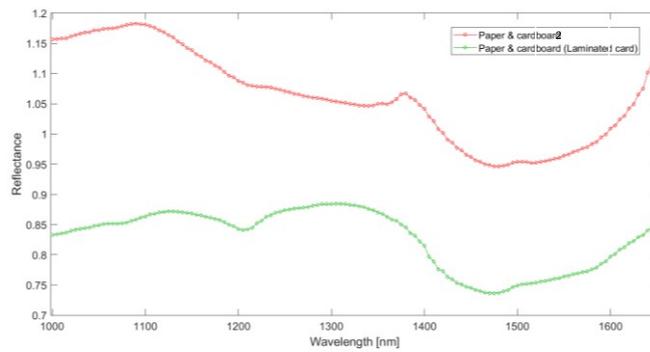
Analisi esplorativa

In Figura 4.1.31 sono mostrati gli spettri medi estratti da ROI delle classi *"Paper & cardboard"* - *"Polymer"*, *"Paper & cardboard 2"* - *"Paper & cardboard (Laminated card)"* e delle classi *"Polymer 2"* - *"Polymer (Laminated plastic)"*.

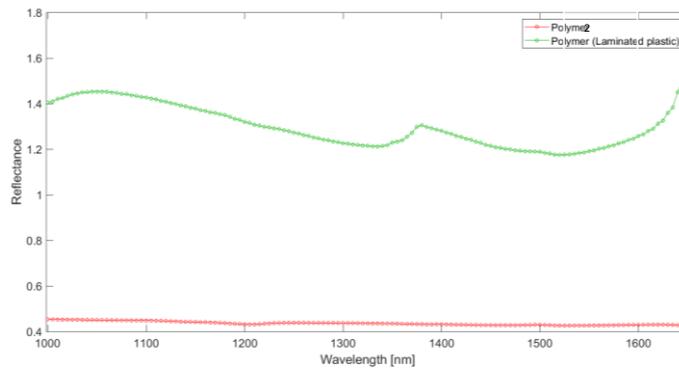
Casi di studio



a



b



c

Figura 4.1.31. Spettri medi grezzi per: a) Set di calibrazione "Paper & cardboard" - "Polymer", b) set di calibrazione "Paper & cardboard 2" - "Paper & cardboard (Laminated card)" e c) set di calibrazione "Polymer 2" - "Polymer (Laminated plastic)". La riflettanza è riportata in scala arbitraria non normalizzata.

L'analisi in componenti principali ha mostrato una certa capacità di auto-raggruppamento delle classi analizzate (si rimanda all'*Allegato VIII- Case study – Packaging*). Ad ogni modo sono presenti problemi di sovrapposizione degli scores probabilmente dovuti a fenomeni di *scattering* e riflessione sulla superficie bianca

altamente riflettente di alcune particelle cellulosiche e alla rugosità superficiale dei frammenti di poliaccoppiati in plastica.

Identificazione dei frammenti carta, cartone e di plastica

I dettagli delle performance di classificazione del modello “*Paper & cardboard*” - “*Polymer*” in calibrazione, *cross-validation* e predizione sono riportati in Tabella 4.1.5. Il modello validato è in grado di riconoscere tutte le particelle presenti nel “*Validation set 1*” (come si può notare in Figura 4.1.32) e 13/14 particelle nel “*Validation set 2*” (93% del totale, come mostrato in Figura 4.1.33): una particella di poliaccoppiato a base prevalente di plastica (cerchiata in blu in Figura 4.1.33b), è stata classificata per oltre il 50% dei pixel come “*Paper & cardboard*”.

Identificazione dei poliaccoppiati a base prevalente di carta dagli altri frammenti di materiale cellulosico

Le performance ottenute in classificazione per le classi “*Paper & cardboard*” - “*Paper & cardboard (Laminated card)*” sono riportate in Tabella 4.1.6. Il modello validato è stato in grado di riconoscere tutti gli oggetti nel “*Validation set 1*” (come si può vedere in Figura 4.1.34) e nel “*Validation set 2*” (Figura 4.1.35).

Tabella 4.1.5. Dettagli del modello di classificazione “*Paper & cardboard*” - “*Polymer*” nelle fasi di calibrazione, *cross-validation* e validazione. (*) Err: Misclassification Error.

| | <i>Class</i> | <i>Sensitivity</i> | <i>Specificity</i> | <i>Err*</i> | <i>Precision</i> | <i>Accuracy</i> | <i>Class Error</i> |
|-------------------------|------------------------------|--------------------|--------------------|-------------|------------------|-----------------|--------------------|
| Calibration | <i>Paper & cardboard</i> | 1.000 | 0.970 | 0.021 | 0.938 | 0.980 | 0.015 |
| | <i>Polymer</i> | 0.970 | 1.000 | 0.021 | 1.000 | 0.980 | 0.015 |
| Cross-validation | <i>Paper & cardboard</i> | 1.000 | 0.970 | 0.021 | 0.938 | 0.980 | 0.015 |
| | <i>Polymer</i> | 0.970 | 1.000 | 0.021 | 1.000 | 0.980 | 0.015 |
| Validation set 1 | <i>Paper & cardboard</i> | 0.952 | 0.993 | 0.020 | 0.985 | 0.980 | 0.028 |
| | <i>Polymer</i> | 0.993 | 0.952 | 0.020 | 0.978 | 0.980 | 0.028 |
| Validation set 2 | <i>Paper & cardboard</i> | 0.958 | 0.896 | 0.083 | 0.829 | 0.917 | 0.073 |
| | <i>Polymer</i> | 0.896 | 0.958 | 0.083 | 0.976 | 0.917 | 0.073 |

Casi di studio

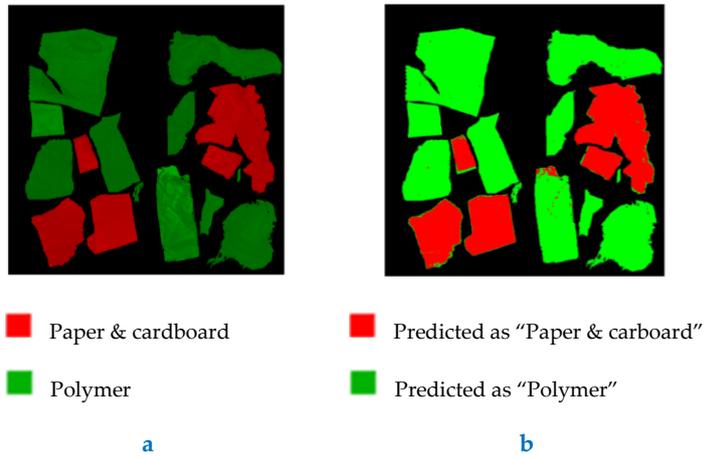


Figura 4.1.32. "Validation set 1", Classificazione "Paper & cardboard" - "Polymer": classi reali (a) e classi predette dal modello PLS-DA (b).

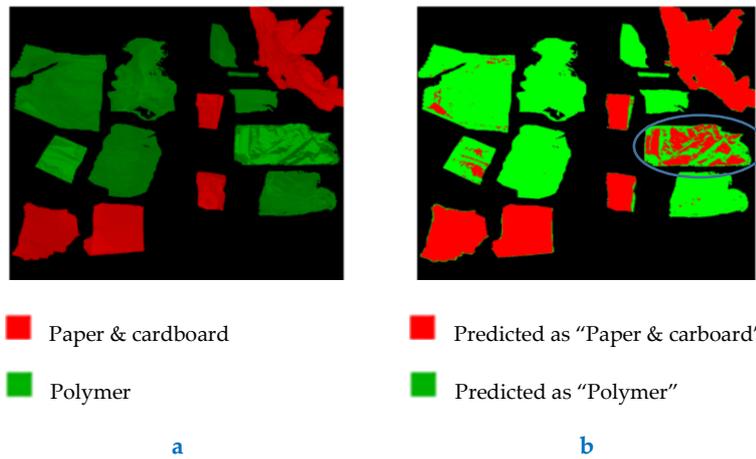
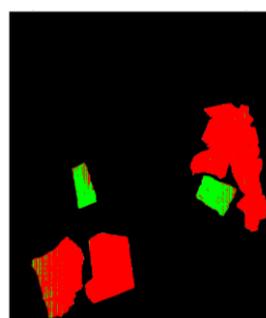


Figura 4.1.33. "Validation set 2", Classificazione "Paper & cardboard" - "Polymer": classi reali (a) e classi predette dal modello PLS-DA (b). La particella cerchiata in blu è stata classificata erroneamente per più del 50% dei pixel.

Casi di studio

Tabella 4.1.6. Dettagli del modello di classificazione “Paper & cardboard 2” - “Paper & cardboard (Laminated card)” nelle fasi di calibrazione, cross-validation e validazione. (*) Err: Misclassification Error.

| | Class | Sensitivity | Specificity | Err* | Precision | Accuracy | Class Error |
|------------------|---|-------------|-------------|-------|-----------|----------|-------------|
| Calibration | <i>Paper & cardboard 2</i> | 0.882 | 0.937 | 0.099 | 0.965 | 0.901 | 0.090 |
| | <i>Paper & cardboard (Laminated card)</i> | 0.937 | 0.882 | 0.099 | 0.804 | 0.901 | 0.090 |
| Cross-validation | <i>Paper & cardboard 2</i> | 0.881 | 0.937 | 0.100 | 0.965 | 0.901 | 0.091 |
| | <i>Paper & cardboard (Laminated card)</i> | 0.937 | 0.881 | 0.100 | 0.804 | 0.901 | 0.091 |
| Validation set 1 | <i>Paper & cardboard 2</i> | 0.983 | 0.887 | 0.031 | 0.982 | 0.969 | 0.065 |
| | <i>Paper & cardboard (Laminated card)</i> | 0.887 | 0.983 | 0.031 | 0.894 | 0.969 | 0.065 |
| Validation set 2 | <i>Paper & cardboard 2</i> | 0.997 | 0.954 | 0.009 | 0.992 | 0.991 | 0.024 |
| | <i>Paper & cardboard (Laminated card)</i> | 0.954 | 0.997 | 0.009 | 0.983 | 0.991 | 0.024 |



■ Paper & cardboard 2

■ Predicted as “Paper & cardboard 2”

■ Paper & cardboard (Laminated card)

■ Predicted as “Paper & cardboard (Laminated card)”

a

b

Figura 4.1.34. “Validation set 1”, classificazione “Paper & cardboard 2” - “Paper & cardboard (Laminated card)”: classi reali (a) e classi predette dal modello PLS-DA (b).

Casi di studio

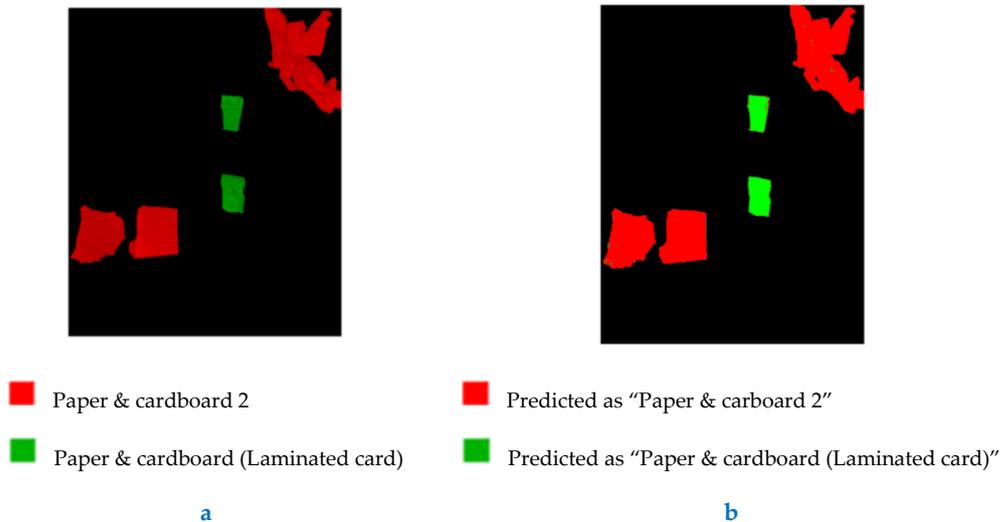


Figura 4.1.35. “Validation set 2”, classificazione “Paper & cardboard 2” - “Paper & cardboard (Laminated card)”: classi reali (a) e classi predette dal modello PLS-DA (b).

Identificazione dei poliaccoppiati a base prevalente di plastica dagli altri frammenti di plastica

In Tabella 4.1.7 sono riportati i risultati del modello di classificazione per le classi “Polymer 2” - “Polymer (Laminated plastic)” in termini di indicatori di performance. In validazione, il modello è stato in grado di riconoscere 10/11 particelle nel “Validation set 1” (91% degli oggetti presenti nel set, come si può notare in Figura 4.1.36): un frammento di poliaccoppiato a base prevalente di plastica (cerchiato in blu in Figura 4.1.36b) è stato classificato per più del 50% di pixel come “Polymer 2”. Ad ogni modo, il modello riesce a classificare correttamente tutti gli oggetti presenti nel “Validation set 2” (Figura 4.1.37).

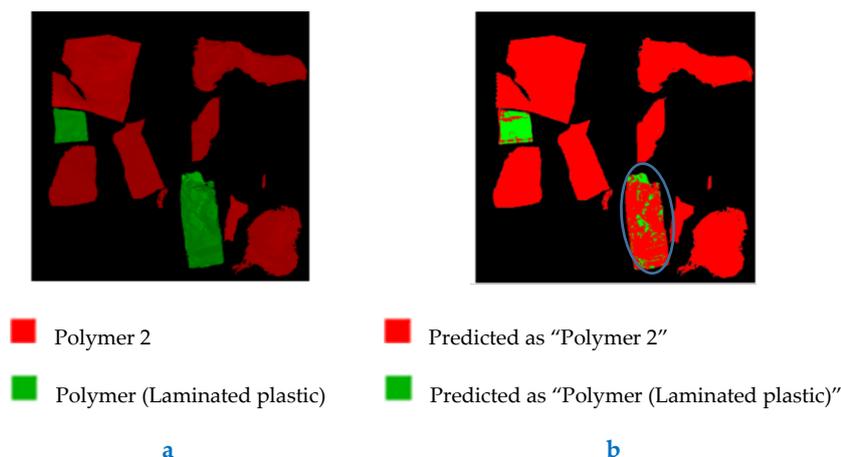


Figura 4.1.36. “Validation set 1”, classificazione “Polymer 2” - “Polymer (Laminated plastic)”: classi reali (a) e classi predette dal modello PLS-DA (b). La particella cerchiata in blu è stata classificata erroneamente per più del 50% dei pixel.

oggetti presenti nei due set di validazione). Nella seconda fase, il modello di classificazione “*Paper & cardboard 2*” - “*Paper & cardboard (Laminated card)*” è stato in grado di identificare tutti gli oggetti (10/10), mentre il modello “*Polymer*” - “*Polymer (Laminated plastic)*” è stato in grado di classificare 19/20 oggetti presenti nei set di validazione. Le particelle non correttamente classificate sono frammenti di poliaccoppiati a base prevalente di plastica. Questo fatto è probabilmente legato ad effetti di *scattering* causati dalle superfici altamente riflettenti dei frammenti di imballaggi piegati. Approcciando il problema con una logica pixel-based, infatti, questo fatto appare maggiormente evidente, in quanto le zone maggiormente riflettenti e quelle in ombra dei frammenti vengono classificate a volte in maniera erronea. In predizione, l’intera procedura di classificazione raggiunge una *Recognition* ed una *Reliability* di 0.933, ed un errore pari a 0.066.

Da quanto visto, un approccio HSI può rappresentare una ottima soluzione, affidabile e a basso costo al fine di identificare sistematicamente contaminanti e materiali composti all’interno di flussi di rifiuti di plastica. Tale approccio può essere anche visto come una logica di controllo qualità online o come nucleo analitico per la selezione dei materiali di questo specifico flusso di rifiuti. Si può dire che l’utilizzo di tecniche basate sull’HSI nel NIR(1000-1700 nm) applicato ad un flusso di rifiuti di imballaggi può aprire le porte a future innovazioni nel settore delle materie prime e del riciclo, come: i) la possibilità di sviluppare un sistema in grado di riconoscere la plastica e / o le sostanze inquinanti, al fine di essere utilizzato non solo come motore di selezione, ma anche come nucleo analitico per eseguire un controllo di qualità su prodotti e sottoprodotti di diverse fasi di produzione e ii) la possibilità di garantire una produzione affidabile del prodotto riciclato in termini sia di qualità che di quantità.

4.2 PRODOTTI AGRO-ALIMENTARI

4.2.1 Valutazione qualitativa della piccantezza di campioni di peperoncino essiccato mediante spettrofotometria puntuale

Problematica e stato dell'arte

Il peperoncino (*Capsicum annuum*), facente parte della famiglia delle *Solanaceae*, è una tipologia di spezia ampiamente utilizzata in tutto il mondo per la sua versatilità (*Guzmán & Bosland, 2017*). Tale spezia è in grado di esaltare i sapori, grazie alle proprie caratteristiche tipiche, quali piccantezza, colore ed aroma. Dal punto di vista commerciale, la piccantezza è uno tra gli attributi più importanti del peperoncino. Tale caratteristica del peperoncino è dovuta al gruppo chimico dei capsacinoidi (*Park et al., 2008*).

La piccantezza di un peperoncino si indica utilizzando la scala di Scoville, che è riferibile al contenuto di capsaicina (16 unità di Scoville sono equivalenti a una parte di capsaicina per milione). Il grado Scoville o *Scoville heat unit* (SHU) è quindi l'unità di misura della piccantezza, che è direttamente correlata con il contenuto di capsacinoidi presenti all'interno di un peperoncino. La piccantezza può essere determinata tramite un test organolettico (*Othman et al., 2011*). Sulla base di questo esame organolettico, i peperoncini possono essere classificati secondo la piccantezza: non-piccante (0-700 SHU), leggermente piccante (700-3'000 SHU), moderatamente piccante (3'000-25'000 SHU), molto piccante (25'000-70'000 SHU) e altamente piccante (> 80'000 SHU). Ad ogni modo, il test organolettico non è accurato ed è soggettivo, in quanto dipende perlopiù dal gusto dal soggetto che esegue il test.

In tempi più recenti, la tecnica maggiormente usata per misurare il contenuto di capsacinoidi è quella che prevede l'utilizzo della cromatografia liquida ad alta prestazione, in inglese *High Performance Liquid Chromatography* (HPLC). L'analisi quantitativa dei capsacinoidi viene effettuata in laboratorio e prevede prevalentemente due fasi: estrazione e separazione tramite l'utilizzo di un solvente e successivamente l'analisi HPLC (*Panda & Suresh, 2018*). Tale test non misura direttamente le unità di Scoville, ma le unità di piccantezza ASTA (*American Spice Trade Association*) che sono poi convertite in unità Scoville. Ad ogni modo, tale tecnica richiede tempo ed è costosa. Una procedura di analisi, scalabile a livello di produzione industriale nel campo alimentare, dovrebbe essere veloce, accurata e riproducibile (*Lim et al., 2015*). In questo scenario, l'utilizzo della spettroscopia Vis-NIR (*Visible e Near InfraRed*) potrebbe consentire il raggiungimento di tali obiettivi. La spettroscopia ottica coadiuvata da tecniche chemiometriche deve il suo successo alla facilità di utilizzo e al fatto di poter essere poco dispendiosa in termini di tempo, tanto che può essere utilizzata anche in applicazioni per il monitoraggio on-line.

Nel campo della scienza e tecnologie agroalimentari, alcuni studi hanno dimostrato come sia possibile utilizzare la spettroscopia UV, visibile e ad infrarosso assistite da

tecniche chemiometriche per valutare il contenuto di capsacinoidi, umidità ed il colore dei peperoncini (*Bae et al., 1998; Lim et al., 2012*).

Obiettivi dello studio

In questo studio, al fine di mettere a punto modelli in grado di prevedere il grado di piccantezza ed altri parametri qualitativi del *Capsicum annuum* essiccato in polvere, è stato utilizzato uno spettrofotoradiometro operante nel range Vis-SWIR (350-2500 nm) per acquisire spettri in riflettanza su peperoncini essiccati frantumati e macinati.

Gli spettri in riflettanza acquisiti sono stati successivamente portati in regressione con la tecnica PLS (*Partial Least Squares*), per la previsione dei valori di piccantezza dei campioni analizzati. Sono stati altresì calibrati modelli di regressione PLS con gli spettri Vis-SWIR per il tenore di umidità ed il contenuto in ceneri.

Infine, è stato calibrato e validato un modello di classificazione *Partial Least Squares - Discriminant Analysis* (PLS-DA) per riconoscere le diverse classi di piccantezza, al fine di comparare i risultati ottenuti portando in regressione gli spettri Vis-SWIR con i valori di piccantezza. Il diagramma di flusso esemplificativo delle analisi effettuate è riportato in [Figura 4.2.1](#).

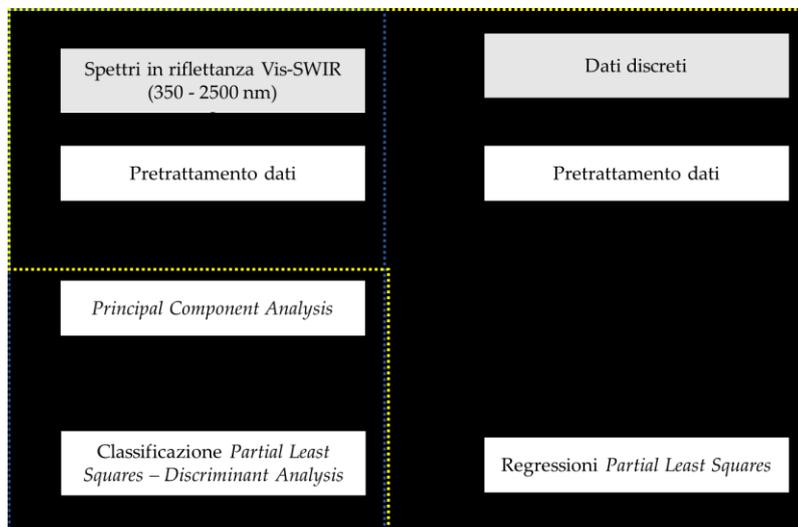


Figura 4.2.1. Diagramma di flusso delle analisi effettuate. La procedura effettuata per mettere a punto la classificazione è inscritta all'interno del quadrato tratteggiato in blu, mentre gli step effettuati per mettere a punto le regressioni *Partial Least Squares* sono all'interno dell'area tratteggiata in giallo.

Procedura

Preparazione dei campioni

In questo studio, come già definito in precedenza, sono stati analizzati peperoncini rossi essiccati con differenti valori di piccantezza, riportati in grado Scoville o *Scoville heat*

unit (SHU), provenienti da differenti regioni italiane. I campioni forniti in polvere e macinati sono stati disposti in capsule Petri. Si tratta di 5 campioni di peperoncino rosso essiccato in polvere e 2 campioni di peperoncino macinato. Per i due campioni F ed S_2, i valori di umidità e contenuto di ceneri non sono disponibili, per questa ragione non sono stati considerati all'interno dei modelli di regressione calibrati per questi due parametri.

In [Tabella 4.2.1](#), sono riportati i valori di piccantezza, contenuto di umidità e contenuto in ceneri. Inoltre, sono state preparate 3 miscele di peperoncino in polvere, utilizzando i campioni A ed F. Il campione Mix 1 è stato preparato utilizzando il 25 % in peso di F ed il 75 % in peso di A; mentre, il campione Mix 2 è stato preparato utilizzando 50 % in peso di F ed il 50 % in peso di A; infine, il campione Mix 3 è stato preparato utilizzando il 75 % in peso di F ed il 25 % in peso di A. In una condizione ideale, le miscele dei campioni A ed F dovrebbero avere valori intermedi di piccantezza. I valori di piccantezza in gradi Scoville delle tre miscele in polvere determinate per interpolazione sono riportate in [Tabella 4.2.2](#). I campioni in capsula Petri sono riportati in [Figura 4.2.2](#).

Tabella 4.2.1. Valori di piccantezza, contenuto di umidità e contenuto percentuale in ceneri di ciascun campione di peperoncino fornito.

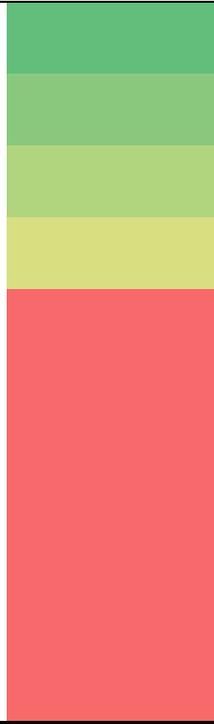
| <i>Identificativo campione</i> | <i>Tipologia di campione</i> | <i>Piccantezza (SHU)</i> | <i>Contenuto di umidità (%)</i> | <i>Contenuto in ceneri (%)</i> |
|--------------------------------|-----------------------------------|--------------------------|---------------------------------|--------------------------------|
| A | <i>Ground Dried Chili Pepper</i> | 40100 | 8.1 | 10.4 |
| B | <i>Ground Dried Chili Pepper</i> | 48000 | 7.97 | 12.3 |
| C | <i>Ground Dried Chili Pepper</i> | 48000 | 10.42 | 9 |
| E | <i>Ground Dried Chili Pepper</i> | 48000 | 9.52 | 8.1 |
| F | <i>Ground Dried Chili Pepper</i> | 48000 | - | - |
| S_1 | <i>Crushed Dried Chili Pepper</i> | 48000 | 10.24 | 12.2 |
| S_2 | <i>Crushed Dried Chili Pepper</i> | 48000 | - | - |

Acquisizione di spettri in riflettanza

Per ciascun campione sono stati acquisiti 20 spettri in riflettanza (per un totale di 160 spettri), con spettrofotoradiometro portatile ASD FieldSpec 4 Standard – Res, utilizzando l'apposita sonda a contatto collegata allo strumento. L'acquisizione dei dati è stata effettuata tramite il software RS³ ([ASD Inc. 2015b](#)).

Casi di studio

Tabella 4.2.2 Valori di piccantezza di ciascun campione di peperoncino rosso (inclusi i preparati in miscela). La scala di colori da verde a rosso è da intendersi come scala relativa dei valori di piccantezza: dal minimo (verde) fino al massimo (rosso).

| Identificativo campione | Tipologia di campione | Piccantezza (SHU) | Piccantezza (scala di colori) |
|----------------------------|-----------------------------------|-------------------|--|
| A | Ground Dried Chili Pepper | 40100 |  |
| Mix 1: 25 wt% F + 75 wt% A | Ground Dried Chili Pepper Mixture | 42075 | |
| Mix 2: 50 wt% F + 50 wt% A | Ground Dried Chili Pepper Mixture | 44050 | |
| Mix 3: 75 wt% F + 25 wt% A | Ground Dried Chili Pepper Mixture | 46025 | |
| B | Ground Dried Chili Pepper | 48000 | |
| C | Ground Dried Chili Pepper | 48000 | |
| E | Ground Dried Chili Pepper | 48000 | |
| F | Ground Dried Chili Pepper | 48000 | |
| S_1 | Crushed Dried Chili Pepper | 48000 | |
| S_2 | Crushed Dried Chili Pepper | 48000 | |

Esportazione dati

Gli spettri acquisiti con software RS³ in formato .asd sono stati esportati in file di testo ASCII, .txt, grazie al software ViewSpec Pro (Ver. 6.2.0.). Il file così ottenuto è stato successivamente importato in ambiente MATLAB® (MATLAB R2016b; ver. 9.1.0.441655) come “dataset object” utilizzando una routine (script MATLAB), scritta appunto per lo scopo: `fieldspec_import2.m` (si veda [Allegato II - MATLAB script - fieldspec_import2.m](#)). Il file importato in ambiente MATLAB è stato poi analizzato utilizzando il toolbox PLS_toolbox (ver. 8.2.1). Le classi di piccantezza sono state settate tramite l'apposita interfaccia del PLS_toolbox. I dati necessari per le successive analisi (i.e. piccantezza, contenuto di umidità e contenuto in ceneri) sono stati importati anch'essi in ambiente MATLAB.

Pre-elaborazione dati

Il primo algoritmo di *pre-processing* applicato ai dati spettrali è stato il cosiddetto “splice correction” ([Danner et al., 2015](#)) per eliminare i “salti” (*gap*) presenti nei segnali in riflettanza acquisiti con lo spettrofotoradiometro FieldSpec 4® Standard-Res, localizzati tra i vari domini di acquisizione dei tre detector dello strumento: VIS, SWIR 1 e SWIR 2. Successivamente, la combinazione di *pre-processing* applicata per la correzione degli

spettri è stata: SNV per la correzione degli effetti di *scattering*, la derivata 1° basata sull'algoritmo di *Savitzky–Golay* e MC, per centrare le colonne dei dati (*Rinnan et al., 2009*).

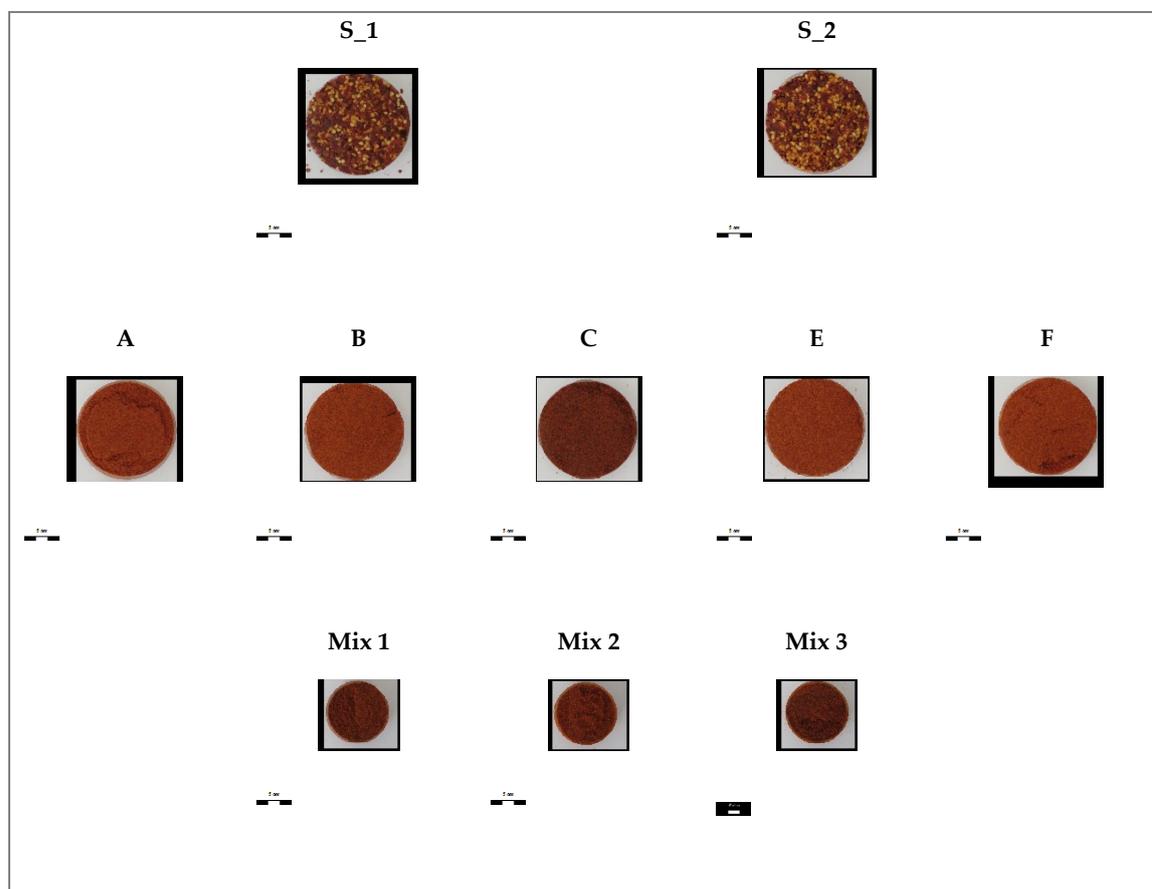


Figura 4.2.2. Campioni di peperoncino frantumato (S_1 ed S_2), macinato (A, B, C, E ed F) e le miscele di peperoncino macinato (Mix 1, Mix 2 e Mix 3).

Analisi esplorativa dei dati

L'analisi in componenti principali (o *Principal Component Analysis*, PCA) è stata utilizzata come tecnica per la semplificazione dei dati spettrali d'origine, secondo i valori di piccantezza e tipologia di campione (*Wold et al., 1987*). I dati sono stati dapprima pre-processati e successivamente è stata effettuata una analisi in componenti principali. Il numero di componenti principali, o *Principal Component* (PC), è stato scelto esplorando i grafici degli autovalori. Gli *outlier* sono stati identificati ed esclusi dalle successive analisi, esplorando il grafico dei parametri Hotelling T² e i residui Q.

Modelli di regressione

Il metodo di regressione PLS è stato utilizzato per valutare la correlazione tra gli spettri Vis-SWIR (Dataset X) ed i valori di piccantezza, contenuto in cenere e contenuto di umidità (dataset Y). Sono stati, perciò, messi a punto tre modelli di regressione (*Geladi et al., 1986; Wold et al., 2001; Abdi, 2003*).

In primo luogo, ciascun dataset è stato suddiviso in 2 parti: 70 % è stato utilizzato per calibrare il modello, mentre il restante 30 % è stato utilizzato come set di validazione. La suddivisione del dataset è stata effettuata con l'algoritmo *Kennard-Stone* (*Kennard & Stone, 1969*). In secondo luogo, i modelli sono stati calibrati e cross-validati mediante l'utilizzo dell'algoritmo *Venetian Blinds* (*Eigenvektor Research, Inc. 2016*), per valutare la complessità ottimale del modello e scegliere il numero di variabili latenti (LV). Infine, i modelli sono stati testati con l'utilizzo dei set di validazione.

Nel primo modello PLS, i valori di piccantezza di tutti i campioni sono stati utilizzati in **Y** e come dataset **X** sono stati utilizzati gli spettri in riflettanza dei campioni. Nel secondo modello, sono stati utilizzati i valori del contenuto di umidità in percentuale come **Y** (questa volta utilizzando i soli campioni A, B, C e E) ed i corrispondenti spettri come dataset **X**. Nel terzo modello, invece, sono stati utilizzati i valori del contenuto in ceneri come **Y** (per i campioni A, B, C e E) ed i corrispondenti spettri sono stati, invece, utilizzati nel dataset **X**. Nel primo caso, sono state usate 5 variabili latenti, mentre nel secondo e terzo caso sono state usate 4 variabili latenti.

Classificazione

Per classificare i dati spettrali dei campioni in base alla loro classe di piccantezza è stato messo a punto un modello PLS-DA (*Wise et al., 2006; Wise et al., 2008; Ballabio & Consonni, 2013*). A partire dagli spettri in riflettanza, il modello è stato addestrato per riconoscere le seguenti classi di piccantezza: 40100 SHU, 42075 SHU, 44050 SHU, 46025 SHU, 46025 SHU e 48000 SHU. Al fine di settare il modello PLS-DA, anche in questo caso, il set di dati è stato suddiviso casualmente in due parti, utilizzando l'algoritmo di *Kennard-Stone* (KS). Per calibrare il modello è stato utilizzato il 70 % del dataset spettrale **X**, mentre il rimanente 30 % è stato utilizzato per la validazione del modello.

Analisi dei risultati

Analisi esplorativa

Come si può evincere dalla [Figura 4.2.3b](#), gli spettri medi in riflettanza dei campioni analizzati, mediati per il relativo valore di piccantezza, presentino delle forme differenti. Questa variabilità è maggiormente apprezzabile con l'analisi in componenti principali. Dall'analisi esplorativa si è evinto come a seconda dei valori di piccantezza dei campioni analizzati si vengono a generare diverse nuvole di *scores* (si veda [Allegato IX- Case study - Peperoncino](#)). Inoltre, in PCA si vengono a generare anche addensamenti di *scores* riferibili alla granulometria dei campioni analizzati ([Figura 4.2.4](#)).

Casi di studio

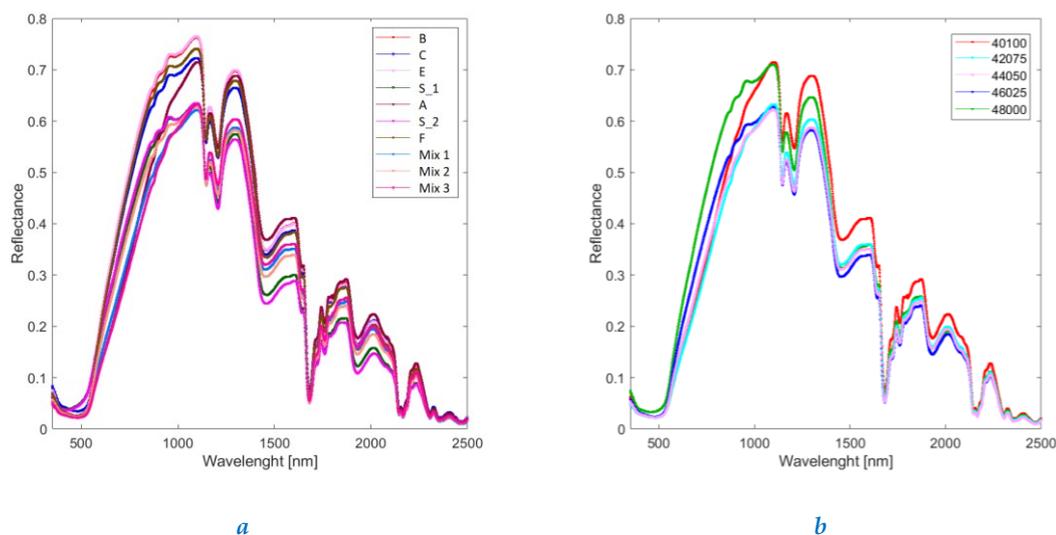


Figura 4.2.3. Spettri medi in riflettanza (450-2500 nm) di: tutti i campioni (peperoncino essiccato macinato e tritato) etichettati per denominazione del campione (a) e grado di piccantezza - SHU (b).

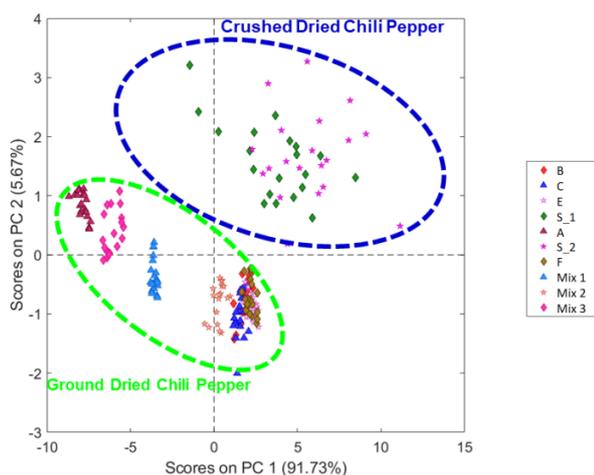


Figura 4.2.4. Scores su PC 1 vs scores su PC 2 (spettri di riflettanza Vis-SWIR: 450-2500 nm) per tutti i campioni (peperoncino essiccato macinato e tritato), etichettati per denominazione del campione.

Regressioni Partial Least Squares

Per quanto riguarda i modelli di regressione, quello relativo alla piccantezza si adatta maggiormente ai valori analizzati rispetto alle altre regressioni effettuate, con $R^2_P = 0.983$ e un RMSEP = 364 SHU (Figura 4.2.5). Portando in regressione i dati spettrali per il contenuto in ceneri si ottiene un $R^2_P = 0.674$ ed un RMSEP = 0.895 %; mentre, per quanto riguarda il contenuto di umidità si è ottenuto un $R^2_P = 0.796$ ed un RMSEP = 0.425 %. Nella Tabella 4.2.3 sono mostrati i risultati relativi alle performance di regressione.

Casi di studio

Gli *scatter plot* riferiti alle regressioni effettuate, relativamente al contenuto in ceneri e al contenuto di umidità (valori predetti in funzione di quelli misurati), sono riportati nella sezione allegati (*Allegato IX- Case study - Peperoncino*).

Tabella 4.2.3. Risultati di regressione per contenuto in ceneri (%) - Ash content (%), contenuto di umidità (%) - Moisture content (%) e piccantezza - Pungency (SHU).

| Intervallo spettrale | Applicazione | Intervallo valori in Y | LV | R ² _{Cal} | R ² _{CV} | R ² _P | RMSEC | RMSECV | RMSEP |
|----------------------|----------------------|------------------------|----|-------------------------------|------------------------------|-----------------------------|-------|--------|-------|
| 450-2500 nm | Ash content (%) | 8 - 12.5 % | 4 | 0.789 | 0.677 | 0.674 | 0.7 | 0.9 | 0.9 |
| 450-2500 nm | Moisture content (%) | 7.5 - 10.5 % | 4 | 0.867 | 0.772 | 0.796 | 0.4 | 0.5 | 0.4 |
| 450-2500 nm | Pungency (SHU) | 40100 - 4800 SHU | 5 | 0.982 | 0.978 | 0.984 | 366 | 409 | 364 |

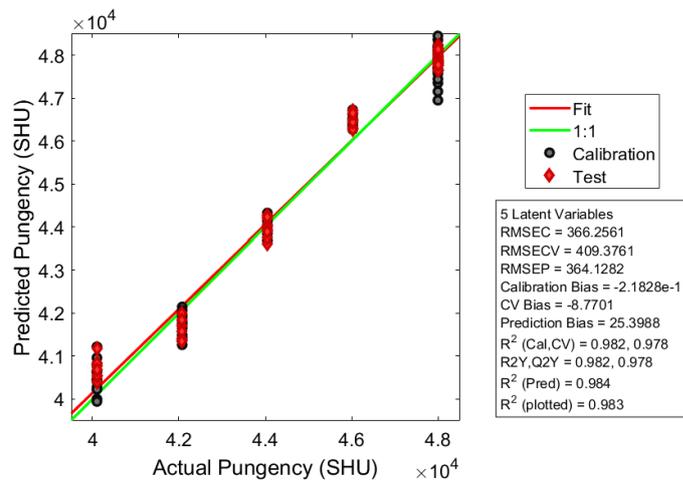


Figura 4.2.5. Diagramma di regressione PLS per la piccantezza (SHU) effettuata con gli spettri Vis-SWIR (FieldSpec 4; 450-2500 nm).

Classificazione *Partial Least Squares – Discriminant Analysis*

Come si può notare in *Tabella 4.2.4*, anche i risultati in predizione del modello di classificazione PLS-DA mostrano una spiccata capacità di predire ciascuna classe di piccantezza, per le quali il modello è stato addestrato, ottenendo per ciascuna classe una sensibilità ed una specificità (in predizione) prossime all'unità (per i grafici relativi alle posizioni del *boundary* di discriminazione si rimanda alla sezione allegati (*Allegato IX- Case study - Peperoncino*)).

Tabella 4.2.4. Parametri prestazionali ricavati dalla tabella di confusione riferita al modello PLS-DA in predizione.

| Classi di piccantezza | <i>Sensitivity</i> | <i>Specificity</i> | Numero di spettri | <i>Error Rate</i> | <i>Precision</i> | <i>Accuracy</i> | <i>Class Error</i> |
|-----------------------|--------------------|--------------------|-------------------|-------------------|------------------|-----------------|--------------------|
| 40100 | 1.000 | 1.000 | 8 | 0.000 | 1.000 | 1.000 | 0.000 |
| 42075 | 1.000 | 1.000 | 5 | 0.000 | 1.000 | 1.000 | 0.000 |
| 44050 | 1.000 | 1.000 | 10 | 0.000 | 1.000 | 1.000 | 0.000 |
| 46025 | 1.000 | 1.000 | 5 | 0.000 | 1.000 | 1.000 | 0.000 |
| 48000 | 1.000 | 1.000 | 20 | 0.000 | 1.000 | 1.000 | 0.000 |

Discussione dei risultati

L'analisi in componenti principali ha mostrato come i differenti campioni di peperoncino analizzati si raggruppano in base ai valori di piccantezza ed in base alla granulometria. Le regressioni PLS hanno inoltre evidenziato l'esistenza di una correlazione tra gli spettri Vis-SWIR ed il contenuto di umidità o e tra gli spettri Vis-SWIR ed il contenuto in ceneri. Tramite l'utilizzo di un modello PLS basato su spettri Vis-SWIR si può, quindi, prevedere la piccantezza di un peperoncino; ciò è riscontrabile sulla base di quanto evinto dai parametri di *fitting* e di errore ($R^2_P = 0.983$ and $RMSEP = 364$ SHU). La capacità nel prevedere le classi di piccantezza è stata dimostrata anche dalla validazione del modello di classificazione PLS-DA. Da quanto emerso, si evince che la spettroscopia Vis-SWIR in riflettanza ben si adatta alla valutazione qualitativa di alcuni parametri come il contenuto di umidità od il contenuto in ceneri, o parametri che fino a qualche tempo fa potevano essere considerati maggiormente oggettivi, come il grado di piccantezza. L'applicazione di tali tecniche può essere funzionale sia in ambito industriale, per applicazioni di controllo qualità del prodotto (a vari livelli, i.e. come evinto da PCA è possibile valutare anche la granulometria del campione), che per applicazioni di valutazione speditiva su campo o di monitoraggio della qualità nel momento della messa in commercio del prodotto finito. La piccantezza di un peperoncino può essere quindi valutata in modo più rapido rispetto alle tecniche tradizionali effettuate in laboratorio, grazie all'utilizzo di uno spettrofotometro Vis-SWIR portatile. Inoltre, è stato dimostrato che l'utilizzo di uno spettrometro portatile coadiuvato da tecniche chemiometriche, può essere funzionale al monitoraggio di peperoncini essiccati rossi, al fine di soddisfare le esigenze dei consumatori (i.e. grado di piccantezza) e controllarne la qualità (i.e. contenuto di umidità, contenuto in ceneri e granulometria).

4.2.2 Analisi qualitativa di pasta surgelata mediante spettrofotometria puntuale

Problematica e stato dell'arte

Il mercato mondiale della pasta è in continua crescita, per ragioni che sono riconducibile ai bisogni dei consumatori, in termini di facilità di consumo, per ragioni di salute e benessere. La pasta rappresenta uno dei principali e tradizionali prodotti alimentari italiani ed è molto apprezzata anche all'estero. L'Italia è infatti il principale produttore (3.0 milioni di tonnellate) e consumatore (circa 28 kg/anno pro capite) di pasta secca. Il Bel Paese è anche il principale esportatore di pasta secca all'estero (circa 1.5 milioni di tonnellate). In particolare, negli ultimi anni l'enorme flusso di vendite all'estero corrisponde a circa la metà della produzione nazionale di pasta ([Tamburini et al., 2016](#)).

Ad oggi, l'insieme delle analisi chimico-fisiche che devono essere effettuate al fine di garantire la qualità del prodotto ha un impatto notevole in termini economici sulle aziende produttrici, e comunque i risultati di tali analisi non sono disponibili nell'immediato. Proprio per questo, negli ultimi tempi, sempre molte più aziende che lavorano nel settore della pasta stanno iniziando ad utilizzare metodi basati sulla spettroscopia NIR combinata a metodologie tipiche della chemiometria multivariata che consente rapidità di analisi e misura simultanea di diversi parametri ([Zardetto, 2005](#); [Groß et al., 2011](#); [Bevilacqua et al., 2013](#)).

Obiettivi dello studio

In questo studio, uno spettrofotoradiometro operante nel range Vis-SWIR (350-2500 nm) è stato utilizzato per acquisire spettri in riflettanza su campioni di pasta surgelati, con due differenti gradi di salatura, destinati al mercato statunitense. Gli spettri sono stati acquisiti in condizioni congelato e scongelato (lasciati riposare circa due ore a temperatura ambiente). A partire dagli spettri acquisiti è stata effettuata una analisi in componenti principali per valutare le differenze tra i due campioni a differenti salature (Mezze Penne e Pennette 72). Successivamente, è stato calibrato e validato un modello di classificazione *Partial Least Squares - Discriminant Analysis* (PLS-DA) per riconoscere le diverse classi di residuo secco dei campioni. Il diagramma di flusso esemplificativo delle analisi effettuate è riportato in [Figura 4.2.6](#).

Procedura

Preparazione dei campioni

I campioni di pasta analizzati in questo studio sono stati forniti da una azienda sita in Cisterna di Latina (Lazio, Italia). Si tratta di due tipologie di pasta pronte, con due differenti gradi di salatura: 6 campioni di Mezze penne e 6 campioni di Pennette 72. Circa 200 gr di ciascun campione a pezzatura è stato disposto in piatti di plastica. I dati relativi al residuo secco calcolato in % sono riportati in [Tabella 4.2.5](#). I campioni sono stati

analizzati in due tempi: surgelati (appena tirati fuori dalla ghiacciaia) e scongelati (dopo due ore circa di riposo a temperatura ambiente, a circa $T=26^{\circ}\text{C}$), come mostrato in [Figura 4.2.7](#).

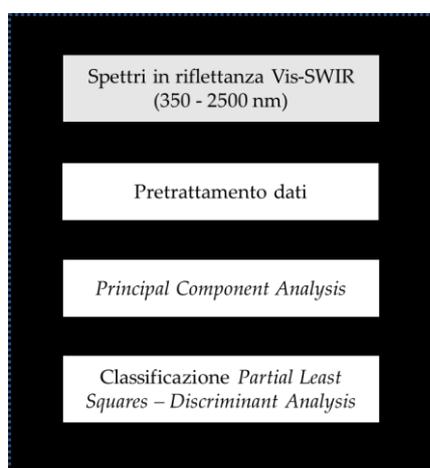


Figura 4.2.6. Diagramma di flusso delle analisi effettuate. La procedura effettuata per mettere a punto la classificazione è inscritta all'interno del quadrato tratteggiato in blu.

Acquisizione di spettri in riflettanza

Per ciascun campione sono stati acquisiti 50 spettri in riflettanza (per un totale di 600 spettri per ciascun tempo di acquisizione), in due tempi di acquisizione: surgelati (appena tirati fuori dalla ghiacciaia) e scongelati (dopo due ore circa di riposo a temperatura ambiente, a circa $T=26^{\circ}\text{C}$). Le acquisizioni sono state effettuate in due modalità: a contatto e a distanza. In entrambi i casi è stato utilizzato uno spettrofotoradiometro portatile ASD FieldSpec 4 Standard – Res equipaggiato con sonda a contatto ([ASD Inc., 2015b](#)). Nel caso delle acquisizioni a contatto, lo spot della sonda è stato posto a contatto con la superficie del campione. Nel caso delle acquisizioni a distanza, la sonda è stata posta a circa 5 cm di distanza dal piano di scansione per avere una area di campionamento (A) di circa 3.9 cm^2 , considerando una *Field of View* (FOV) del cavo di fibra ottica di circa $\alpha = 25^{\circ}$.

L'area di campionamento è dipendente dalla distanza (d) tra capo della fibra ottica e superficie del campione, secondo la formula:

$$A = \pi \left(\tan \frac{\alpha}{2} \cdot d \right)^2 \quad (4.2.1)$$

Il set-up realizzato per effettuare le acquisizioni a distanza fissa, è riportato in [Figura 4.2.8](#). L'acquisizione dei dati è stata effettuata tramite il software RS³ ([ASD Inc., 2008](#)).

Casi di studio



Figura 4.2.7. Immagini digitali di alcuni dei campioni di pasta analizzati. Mezzepenne surgelate (a) e scongelate (b) e Pennette 72 congelate (c) e scongelate (d).

Tabella 4.2.5. Dati di residuo secco in % dei campioni di pasta analizzati.

| Denominazione | Tipologia | Residuo secco [%] |
|----------------|-------------|-------------------|
| Mezze_penne_01 | Mezze penne | 42.8 |
| Mezze_penne_02 | Mezze penne | 42.8 |
| Mezze_penne_03 | Mezze penne | 42.8 |
| Mezze_penne_04 | Mezze penne | 47.5 |
| Mezze_penne_05 | Mezze penne | 47.5 |
| Mezze_penne_06 | Mezze penne | 47.5 |
| Pennette_72_01 | Pennette 72 | 46.7 |
| Pennette_72_02 | Pennette 72 | 46.7 |
| Pennette_72_03 | Pennette 72 | 46.7 |
| Pennette_72_04 | Pennette 72 | 46.7 |
| Pennette_72_05 | Pennette 72 | 46.7 |
| Pennette_72_06 | Pennette 72 | 46.7 |

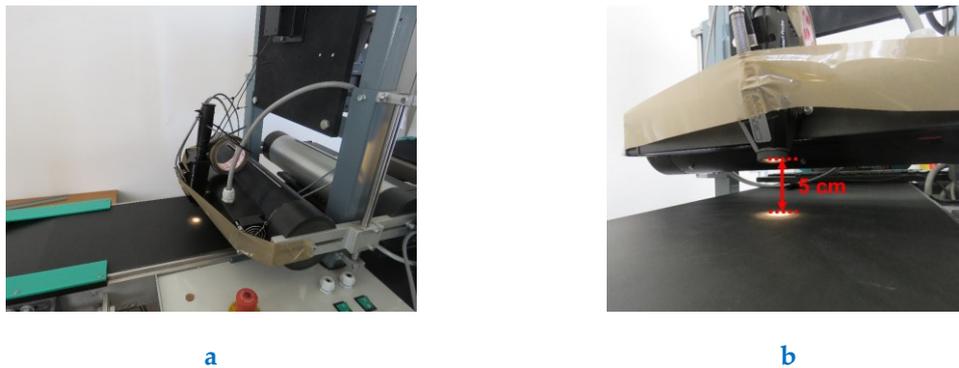


Figura 4.2.8. Set-up di acquisizione a distanza per spettrofotometri portatile ASD FieldSpec 4 Standard – Res (a) e particolare della sonda posta a distanza fissa dal piano di acquisizione (b).

Esportazione dati

Gli spettri acquisiti con software RS3 in formato “.asd” sono stati esportati in file di testo ASCII (*American Standard Code for Information Interchange*, ovvero Codice Standard Americano per lo Scambio di Informazioni) “.txt”, grazie al software ViewSpec Pro (Ver. 6.2.0.). I file così ottenuti sono stati successivamente importati in ambiente MATLAB (MATLAB R2016b ver. 9.1.0.441655) come “dataset object” utilizzando una routine (script MATLAB), scritta appunto per lo scopo: “fieldspec_import2.m”. I file importati in ambiente MATLAB sono stati poi analizzati utilizzando il toolbox PLS_toolbox (ver. 8.2.1). Le denominazioni dei campioni (tenendo conto anche delle informazioni relative alla modalità di acquisizione) e le classi di residuo secco sono state settate tramite l’apposita interfaccia.

Pre-elaborazione dati

Il primo algoritmo di *pre-processing* applicato ai dati spettrali è stato il cosiddetto “splice correction” (Danner et al., 2015) al fine di eliminare i “salti” (*gap*) localizzati tra i vari domini di acquisizione dei tre detector dello strumento FieldSpec 4 Standard-Res. Successivamente, i dati spettrali (350-2500 nm) sono stati ridotti alle lunghezze d’onda 450-2500 nm (VIS-SWIR).

Sono state esplorate diverse strategie di *pre-processing*. Infine, è stata scelta la combinazione di *pre-processing* che ha fornito una risposta migliore in termini di decomposizione dei dati. La combinazione di *pre-processing* utilizzata è stata: *Multiplicative Scatter Correction* (MSC) e *Mean Center* (MC). MSC è stato usato come filtro per la correzione di effetti di *scattering* e MC per centrare le colonne dei dati rispetto ad uno zero (Rinnan et al., 2009).

Analisi esplorativa dei dati

L’analisi in componenti principali è stata utilizzata per effettuare una analisi esplorativa dei dati spettrali, secondo i valori di residuo secco e la tipologia di campione (Beebe et al., 1998; Wold et al., 1987). Sono state considerate 5 casistiche, a seconda della condizione fisica del campione (surgelato o scongelato) ed in base alla condizione di

acquisizione: 1) campioni surgelati acquisiti con sonda a contatto; 2) campioni surgelati acquisiti con sonda a distanza; 3) campioni scongelati acquisiti con sonda a contatto; 4) campioni scongelati acquisiti con sonda a distanza; 5) campioni surgelati e scongelati acquisiti sia con sonda a contatto che con sonda a distanza.

Per ciascun dataset analizzato, il numero di componenti principali è stato scelto esplorando i grafici degli autovalori. Gli *outlier* sono stati identificati ed esclusi dalle successive analisi, esplorando il grafico dei parametri Hotelling T² e i residui Q.

Classificazione

Per classificare i dati spettrali dei campioni in base alle classi di tenori in ceneri sono stati messi a punto 4 modelli PLS-DA (*Barker et al., 2003; Wise et al., 2006; Ballabio & Consonni, 2013*), relativamente alle seguenti casistiche: campioni surgelati acquisiti con sonda a contatto, campioni surgelati acquisiti con sonda a distanza, campioni scongelati acquisiti con sonda a contatto e campioni scongelati acquisiti con sonda a distanza. A partire dagli spettri in riflettanza, I modelli sono stati addestrati per riconoscere le seguenti classi di tenori in ceneri: 42.8 %, 46.7 % e 47.5 %. Al fine di settare i modelli PLS-DA, in ciascuna delle casistiche analizzate, il set di dati relativo è stato suddiviso casualmente in due parti, utilizzando l'algoritmo di *Kennard-Stone* (*Kennard & Stone, 1969*). Per calibrare ciascun modello è stato utilizzato il 70 % del dataset spettrale X, mentre il rimanente 30 % è stato utilizzato per la validazione del modello. I modelli sono stati cross-validati mediante l'utilizzo dell'algoritmo *Venetian Blinds* (*Eigenvektor Research, Inc. 2016*), per valutare la complessità ottimale del modello e scegliere il numero di variabili latenti (LV). Infine, i modelli sono stati testati con l'utilizzo dei set di validazione.

Analisi dei risultati

Spettri medi in riflettanza

Come si può notare comparando le firme spettrali (*Figura 4.2.9*), gli spettri acquisiti dei campioni surgelati e di quelli scongelati differiscono perlopiù negli intorni relativi a 1600-1800 nm. Mentre, per quanto riguarda le differenze spettrali relative alle condizioni di acquisizione, esse sono perlopiù dovute a variazioni di intensità. Nel caso dei campioni scongelati acquisiti a distanza, queste differenze sono maggiormente evidenti.

Gli spettri medi in riflettanza dei campioni di pasta analizzati congelati e scongelati, acquisiti con sonda a contatto e a distanza, identificati per tipologia di campione e tenore in ceneri sono riportati nell'allegato *Allegato X- Case study – Pasta surgelata*.

Casi di studio

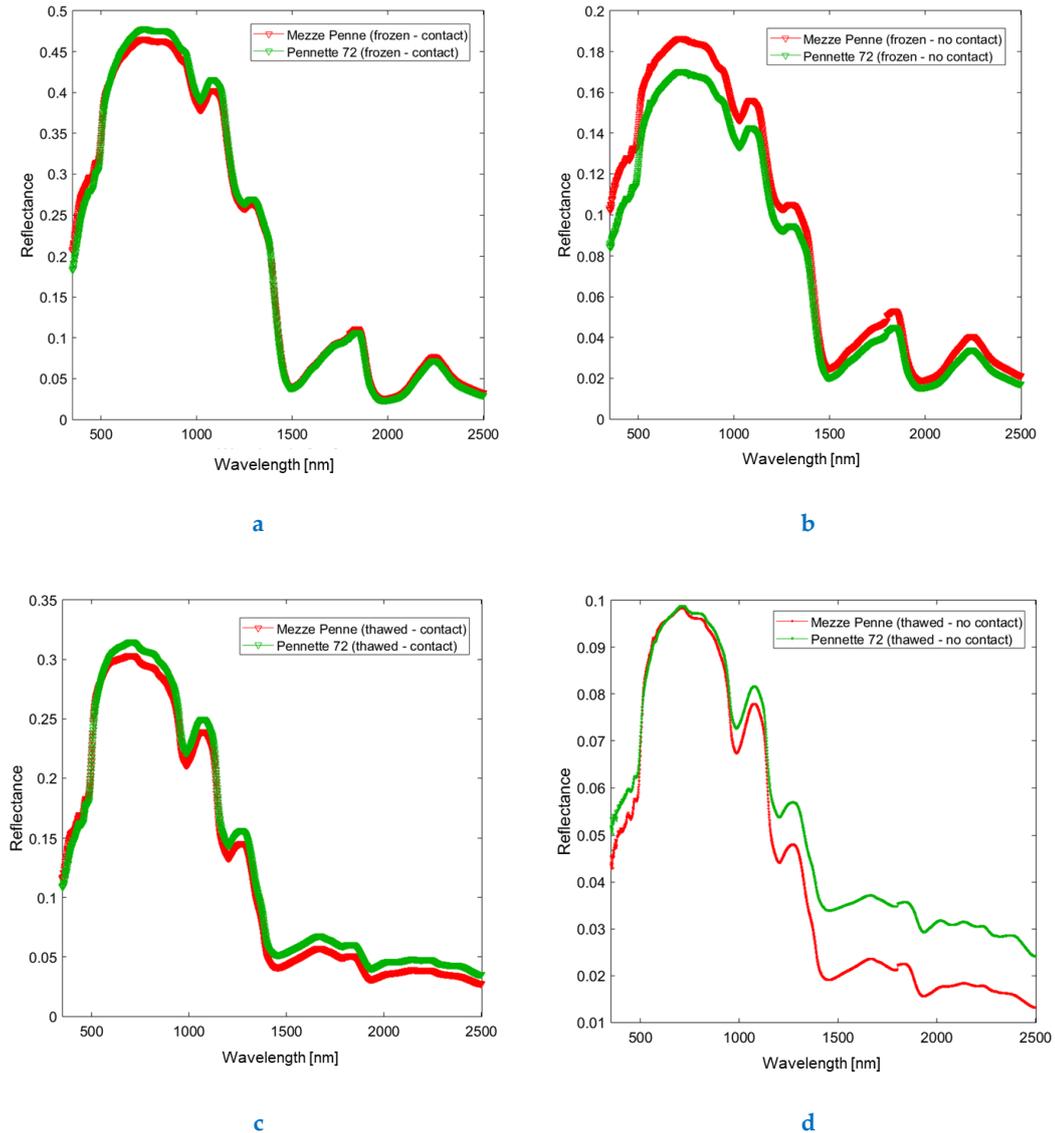


Figura 4.2.9. Spettri medi in riflettanza (350-2500 nm): campioni surgelati acquisiti con sonda a contatto (a) e a distanza (b) e campioni scongelati acquisiti con sonda a contatto (c) e a distanza (d).

Analisi esplorativa

Da quanto evinto dall'analisi in componenti principali (si rimanda all'allegato *Allegato X- Case study – Pasta surgelata*), la differenza di salatura è apprezzabile in entrambi i casi di acquisizione per le due differenti condizioni fisiche del campione. Mentre, la differenza del contenuto in ceneri tra i campioni sembrerebbe essere maggiormente apprezzabile nel caso delle acquisizioni con sonda a contatto: le nuvole degli scores dei campioni acquisiti a distanza sembrano essere maggiormente sovrapposti. In Figura 4.2.10 è mostrato lo *scores plot* relativo alla PCA applicata agli spettri acquisiti su campioni surgelati e scongelati, acquisiti con sonda a contatto e a distanza, denominati secondo la propria tipologia. In questo caso, compaiono due cluster di dati, relativamente alla condizione fisica dei campioni analizzati (i.e. essere congelati o scongelati).

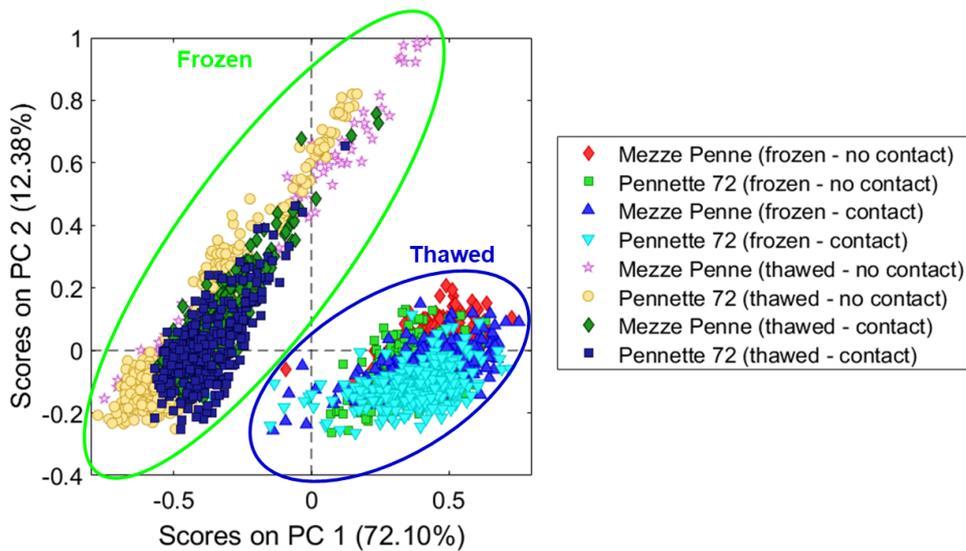


Figura 4.2.10. Scores su PC 1 vs scores su PC 2 (spettri di riflettanza Vis-SWIR: 350-2500 nm) per: campioni surgelati e scongelati acquisiti con sonda a contatto e a distanza.

Gli scores degli spettri acquisiti sui campioni scongelati appaiono nettamente separati dai campioni surgelati; mentre, non sono evidenti pattern relativi alle condizioni di acquisizione (i.e. acquisizione con sonda a distanza e a contatto).

Classificazione Partial Least Squares – Discriminant Analysis

Le metriche prestazionali relative alla matrice di confusione della classificazione PLS-DA sono mostrate nella Tabella 4.2.6. Come riportato, la sensibilità e la specificità (in predizione) di ogni classe sono promettenti. I risultati in predizione mostrano una determinata capacità di predire ciascuna classe di tenore in ceneri, per le quali il modello è stato addestrato. Le posizioni del *boundary* discriminante per ciascuna classe prevista di tenore in ceneri sono invece riportate nell'allegato *Allegato X- Case study – Pasta surgelata*.

Discussione dei risultati

L'analisi in componenti principali ha mostrato come i differenti campioni di pasta analizzati si raggruppino in base ai valori di tenore in ceneri o in base alla salatura. Inoltre, gli scores degli spettri esaminati si addensano secondo le condizioni fisiche del campione e non rispetto alle condizioni di acquisizioni, a dimostrare che le acquisizioni a distanza e quelle a contatto presentano caratteristiche similari e differiscono perlopiù di un *offset* riferibile all'intensità del segnale acquisito. La capacità nel prevedere le classi di tenore in ceneri è stata dimostrata anche con la validazione dei 4 modelli di classificazione PLS-DA.

Tabella 4.2.6. Parametri prestazionali ricavati dalla tabella di confusione riferita ai 4 modelli PLS-DA in predizione per le classi di tenore in ceneri.

Casi di studio

| Condizione del campione / metodo di acquisizione | Classe di tenore in ceneri (%) | <i>Sensitivity</i> | <i>Specificity</i> | Numero di spettri | <i>Error Rate</i> | <i>Precision</i> | <i>Accuracy</i> | <i>Class Error</i> |
|--|--------------------------------|--------------------|--------------------|-------------------|-------------------|------------------|-----------------|--------------------|
| Campioni surgelati / sonda a contatto | 42.8 | 0.614 | 0.905 | 57 | 0.186 | 0.745 | 0.814 | 0.241 |
| | 46.7 | 0.906 | 0.966 | 96 | 0.066 | 0.967 | 0.934 | 0.064 |
| | 47.5 | 0.767 | 0.850 | 30 | 0.164 | 0.500 | 0.836 | 0.192 |
| campioni scongelati / sonda a contatto | 42.8 | 0.977 | 0.949 | 43 | 0.044 | 0.875 | 0.956 | 0.037 |
| | 46.7 | 0.987 | 1.000 | 75 | 0.006 | 1.000 | 0.994 | 0.006 |
| | 47.5 | 0.881 | 0.992 | 42 | 0.038 | 0.974 | 0.962 | 0.064 |
| Campioni surgelati / sonda a distanza | 42.8 | 0.614 | 0.905 | 57 | 0.186 | 0.745 | 0.814 | 0.241 |
| | 46.7 | 0.906 | 0.966 | 96 | 0.066 | 0.967 | 0.934 | 0.064 |
| | 47.5 | 0.767 | 0.850 | 30 | 0.164 | 0.500 | 0.836 | 0.192 |
| Campioni scongelati / sonda a distanza | 42.8 | 0.976 | 0.975 | 42 | 0.025 | 0.953 | 0.975 | 0.025 |
| | 46.7 | 1.000 | 1.000 | 55 | 0.000 | 1.000 | 1.000 | 0.000 |
| | 47.5 | 0.917 | 0.990 | 24 | 0.025 | 0.957 | 0.975 | 0.047 |

Tali tecniche consentirebbero di prevedere il contenuto in ceneri, sia nel momento in cui il prodotto è surgelato, che a seguito dello scongelamento dello stesso. I modelli di classificazione PLS-DA calibrati e validati con gli spettri acquisiti a distanza, dimostrano come con tale strumento sia possibile effettuare una valutazione speditiva del tenore in ceneri anche senza diretto contatto con il campione.

Da quanto emerso, si evince che la spettroscopia Vis-SWIR in riflettanza ben si adatta alla valutazione qualitativa di alcuni parametri come il contenuto in ceneri o il grado di salatura. Si può dire che l'applicazione di tali tecniche può essere funzionale sia in ambito industriale, per applicazioni di controllo qualità del prodotto (a vari livelli, i.e. è possibile valutare anche la salatura del prodotto, come evinto da PCA), che per applicazioni di controllo qualità applicabili in maniera speditiva nel momento della messa in commercio del prodotto stesso. I risultati hanno inoltre mostrato come il tenore in ceneri di un prodotto surgelato possa essere valutato in modo più rapido rispetto alle tecniche tradizionali, grazie all'utilizzo di uno spettrofotometro Vis-SWIR portatile.

4.2.3 Classificazione mediante *HyperSpectral Imaging* di nocciole buone e relativi scarti

Problematica e stato dell'arte

La nocciola, il frutto del nocciolo, rappresenta una delle materie prime più importanti per l'industria della pasticceria e del cioccolato (*Ozdemir & Akinci, 2004*). Attualmente, la produzione mondiale di nocciole ammonta a circa 700'195 t. i maggiori produttori di nocciole sono: Turchia, Italia, Georgia e Stati Uniti. La Turchia produce annualmente circa 450'000 t di nocciole, ovvero circa il 64 % della produzione mondiale di nocciole. Gli altri paesi produttori di nocciole sono l'Italia con 75'456 t, gli Stati Uniti con 32'659 t e la Georgia con 37'400 t l'anno. La nocciola, comunemente annoverata fra la frutta secca, è una importante fonte di vitamina E ed è considerato un alimento calorico (*Yağci & Göğüş, 2008*) in quanto è molto ricco in lipidi (50-60 % di grassi), di proteine (15-19 %), carboidrati (15-17 %) e di acqua (11 %).

In genere, il valore di mercato della frutta secca è fortemente condizionato dalla loro qualità. La qualità di un lotto di frutta secca viene in genere valutato in termini di freschezza del prodotto, presenza/assenza di difetti, presenza/assenza di muffe e/o insetti ed altri parassiti e la presenza di contaminanti (i.e. presenza di gusci, pericarpi/involucro fogliacei, piccole pietre, ramoscelli, ecc.). Le nocciole possono essere soggette all'attacco di muffe, insetti, uccelli, roditori ed altri parassiti. L'alterazione della nocciola da parte di questi agenti può comportare la fissazione di residui nocivi e/o sgradevoli al gusto (i.e. sapore amaro), fino all'irrancidimento del frutto stesso. La qualità della nocciola viene valutata con riferimento al guscio legnoso e al seme del frutto stesso. In particolare, il guscio deve essere: i) intatto e ben formato, ii) pulito (non devono essere presenti corpi estranei attaccati al guscio che superino più del 5 % della superficie del guscio stesso), iii) non deve presentare macchie o discolorazioni con contrasto pronunciato che superino più del 25 % della superficie del guscio stesso. Mentre, il seme deve essere: i) privo di segni di irrancidimento, ii) sufficientemente sviluppato (in particolare, il seme deve riempire almeno il 50 % dell'interno del guscio legnoso), iii) privo di porzioni essiccate e/o indurite (la porzione deve essere meno del 25 % della superficie del seme), iv) non deve presentare macchie o discolorazioni con contrasto pronunciato che superino più del 25 % della superficie del seme stesso e v) ben formato. Il prodotto, nella sua interezza, deve essere: i) completamente essiccato, ii) pieno, iii) privo di muffe visibili ad occhio nudo, iv) esente dalla presenza di parassiti viventi, qualunque sia il loro stato di sviluppo, v) privo di danni causati da parassiti, ciò include anche l'assenza di insetti morti e delle loro feci, vi) assenza di odori e/o sapori estranei.

Gli attributi morfologici e morfometrici rappresentano una tra le principali caratteristiche che condizionano la qualità del prodotto. Infatti, questi attributi sono in qualche modo correlati con la crescita del frutto stesso, e di conseguenza ad i difetti fisici dovuti a fattori naturali (i.e. presenza di insetti e/o organismi infestanti, muffe, ecc.) e/o a fattori antropici o artificiali (come, ad esempio, la rottura del guscio di una nocciola a

causa della movimentazione di un lotto di nocciole). Ad ogni modo, un approccio basato sugli attributi morfologici e morfometrici non sempre è sufficiente ad effettuare un controllo qualità completo della nocciola. In quanto, la presenza di muffe e/o zone scolorite della nocciola non sempre alterano tali attributi. Perciò, si devono perseguire strategie di controllo e/o selezione alternative in grado di investigare le caratteristiche superficiali ed interne del frutto (*Serranti et al., 2012*).

In questo scenario, adottando tecniche di imaging iperspettrale, nell'intervallo di lunghezza d'onda 1000-2500 nm (SWIR) si può mettere a punto un sistema in grado di riconoscere le nocciole buone dai contaminanti presenti in un lotto (i.e. gusci di nocciole e nocciole marce con evidenti segni di deterioramento).

Obiettivi dello studio

Gli obiettivi del presente studio sono stati, quindi: i) il riconoscimento di gusci di nocciola, nocciole buone e nocciole marce a partire da immagini iperspettrali e ii) la selezione delle variabili importanti ai fini della classificazione. La panoramica globale delle analisi effettuate è mostrata in [Figura 4.2.11](#). Per effettuare il riconoscimento dei campioni in esame è stata dapprima acquisita una immagine iperspettrale (Dataset) in laboratorio. Lo sfondo dell'immagine iperspettrale è stato escluso dalle successive analisi e sono state settate le classi dei campioni, tramite interfaccia software. Il Dataset è stato poi diviso in due parti: dataset di calibrazione (contenente il 70 % delle particelle appartenenti a ciascuna classe) e dataset di validazione (contenente il 30 % delle particelle appartenenti a ciascuna classe). Il dataset di calibrazione è stato trattato con diverse combinazioni di algoritmi di *pre-processing* per eliminare pixel rumorosi, attenuare fenomeni di *scattering* e migliorare le differenze inter- ed intra-classe. La *Principal Component Analysis* (PCA) è stata utilizzata per effettuare un'analisi esplorativa e valutare le giuste combinazioni di *pre-processing* da applicare al dataset di calibrazione per la successiva classificazione. Il metodo di discriminazione per le classi settate è la *Partial Least Squares – Discriminant Analysis* (PLS-DA). Per ciascuna combinazione di algoritmi di *pre-processing* utilizzati, i modelli di classificazione, calibrati con dataset di calibrazione, sono stati validati sul dataset di validazione. Successivamente, è stata effettuata la selezione delle variabili importanti ai fini della classificazione tramite l'algoritmo *Interval Partial Least Squares Discriminant Analysis* (Interval-PLSDA o I-PLSDA). Tale procedura è stata applicata per tutti i modelli di classificazione. Infine, sono stati calibrati nuovi modelli di classificazione PLSDA. Tali modelli sono stati calibrati utilizzando esclusivamente le variabili selezionate dalla I-PLSDA e validati sul dataset di validazione.

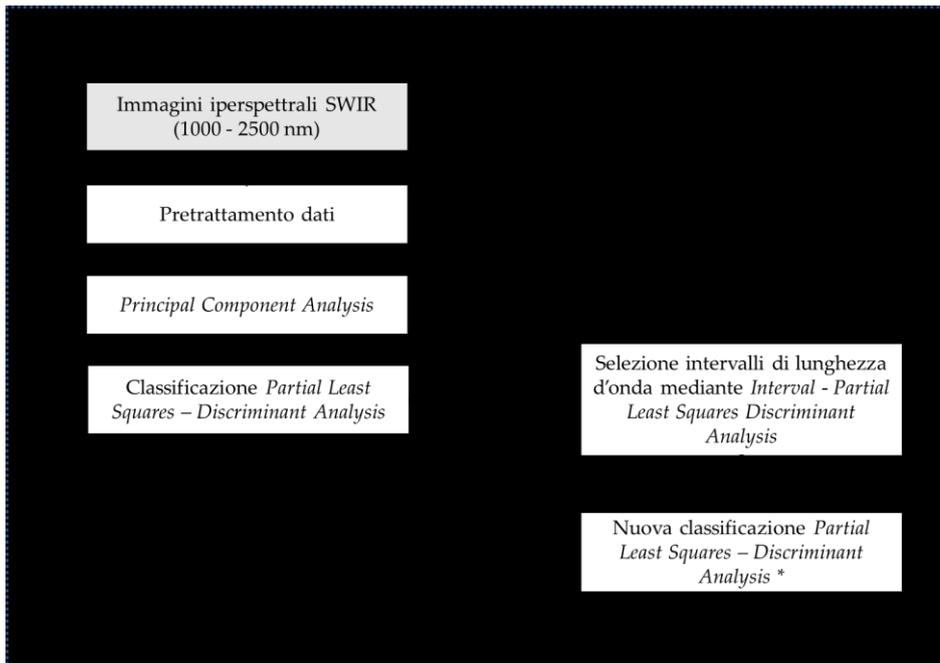


Figura 4.2.11. Diagramma di flusso delle analisi effettuate. *La nuova classificazione è effettuata con gli intervalli di lunghezza d'onda importanti per la classificazione, discriminati dall'algoritmo Interval- Partial Least Squares Discriminant Analysis (I-PLSDA).

Procedura

Preparazione dei campioni

I campioni analizzati (Figura 4.2.12) sono stati suddivisi in tre classi: gusci di nocciola, nocciole buone, nocciole marce.

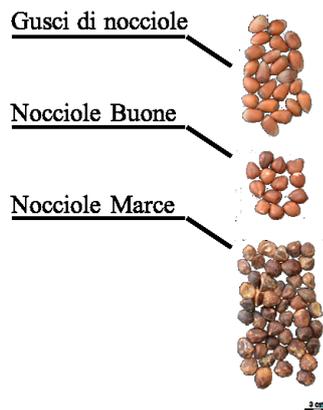


Figura 4.2.12. Campioni di nocciole buone, nocciole marce e gusci di nocciola.

Acquisizione di immagini iperspettrali

Lo strumento utilizzato per effettuare l'acquisizione delle immagini iperspettrali in riflettanza è il SisuCHEMA XL™ (Specim, Finlandia) operante nel range dello spettro elettromagnetico compreso tra 1000 nm e 2500 nm con una risoluzione spettrale di 6.2 nm.

L'ottica utilizzata per le acquisizioni è la 31 mm, la quale consente di acquisire immagini di larghezza pari a 50 mm. Utilizzando questa configurazione, la velocità di scansione del sistema è pari a 18 mm/s. Più nel dettaglio, è stata acquisita un'immagine iperspettrale: dataset Nocciole (Figura 4.2.13b).

Esportazione e preparazione dati

L'immagine iperspettrale acquisita con il sistema SisuCHEMA^{XL} è stata importata in ambiente MATLAB (ver. 9.3.0.713579; R2017b). Le successive analisi chemiometriche sono state effettuate in ambiente MATLAB con l'ausilio dei toolbox PLS_toolbox (ver. 8.6 - 21315) e MIA_toolbox (ver. 3.0.5 - 21317). Lo sfondo dell'immagine iperspettrale è stato escluso dalle successive analisi. Inoltre, il dataset è stato ridotto da 942.2-2542 nm a 1000-2500 nm per rimuovere la parte più rumorosa degli spettri di ciascun pixel, dovuta al detector dello strumento.

Il dataset Nocciole, a seguito del settaggio delle classi (Figura 4.2.13c), è stato suddiviso in due dataset: i) dataset di calibrazione (contenente circa il 70 % delle particelle per ciascuna classe) e ii) dataset di validazione (contenente circa il 30 % delle particelle per ciascuna classe). Il dataset di calibrazione è mostrato in Figura 4.2.14a, mentre quello di validazione è mostrato in Figura 4.2.14b.

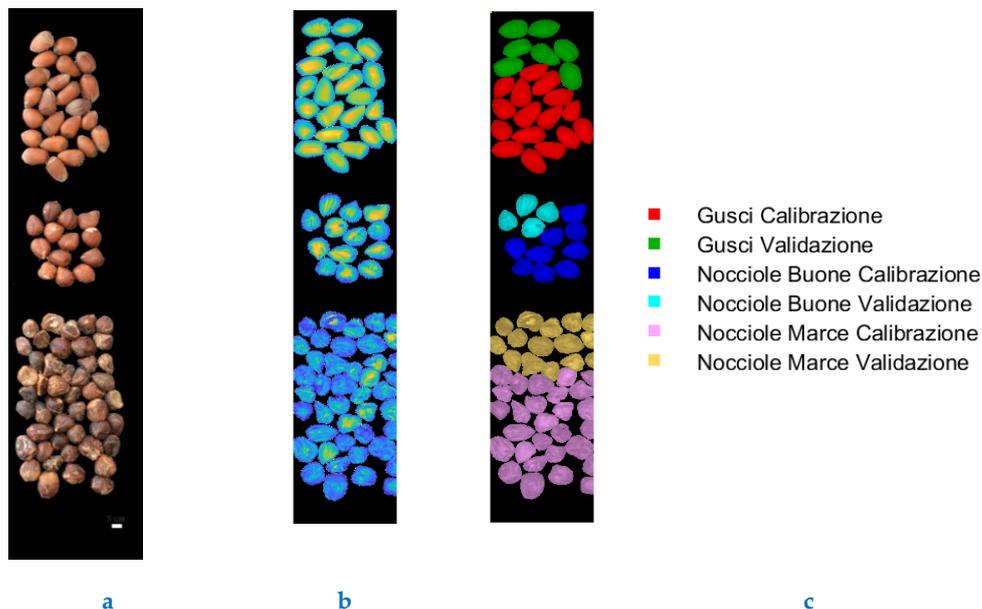


Figura 4.2.13. Immagine RGB dei campioni (a), Immagine iperspettrale in riflettanza media (b) e dataset con classi settate (c).

Pre-elaborazione dati

Prima di passare all'analisi esplorativa è stato necessario applicare una combinazione di algoritmi di *pre-processing*. La pre-elaborazione dati non solo è necessario per

correggere effetti di *scattering* e/o altri effetti rumorosi, ma anche per esaltare le differenze spettrali tra i diversi campioni.

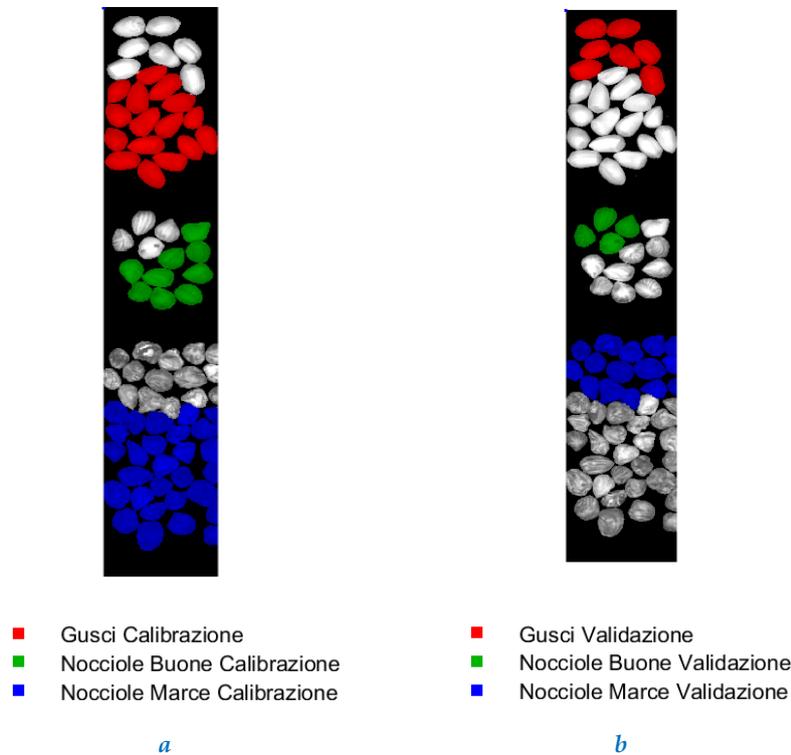


Figura 4.2.14. Dataset di calibrazione (a) e dataset di validazione (b).

Sono state valutate differenti strategie di pre-elaborazione dati, in differenti combinazioni. Le tre combinazioni di *pre-processing* utilizzate per i dati iper-spettrali sono state:

1. *Standard Normal Variate* + *Mean Center* (SNV + MC).
2. *Detrend* + *Standard Normal Variate* + *Mean Center* (DETREND + SNV + MC).
3. *Generalized Least Squares Weighting* (con α pari a 0.01) + *Mean Center* (GLS-W + MC).

Standard Normal Variate (SNV) è stato utilizzato per scalare le differenze spettrali dovute a fenomeni di *scattering* o alla variazione di sensibilità strumentale, dovute al detector e/o alla fonte luminosa (Rinnan et al., 2009). Mentre, *Detrend* è stato utilizzato al fine di rimuovere variazioni che potrebbero generare risposte non-lineari, in quanto particolarmente utile quando è presente un offset nei dati spettrali. Invece, l'algoritmo *Generalized Least Squares Weighting* (GLS-W) è stato scelto nella terza combinazione di *pre-processing*, al fine di attenuare le differenze fra gruppi di campioni (Eigenvector Research Inc. 2013b).

Classificazione

Il metodo di classificazione PLS-DA ([Wise et al., 2006](#); [Wise et al., 2008](#); [Ballabio & Consonni, 2013](#)) è stato applicato per valutare la capacità predittiva degli spettri SWIR nel classificare i campioni studiati. Nel caso specifico, le classi modellate sono state: *Nocciole buone*, *Nocciole marce* e *Gusci di nocciole*. Come algoritmo di cross – validazione, per valutare la complessità del modello e scegliere un adeguato numero di variabili latenti, è stato utilizzato *venetian blinds* ([Eigenvector Research, Inc. 2016](#)).

Per ciascuna delle tre differenti combinazioni di *pre-processing*, il modello PLS-DA è stato costruito a partire dal dataset di calibrazione ([Figura 4.2.14a](#)), *cross-* validato e validato sul set di validazione ([Figura 4.2.14b](#)).

Selezione delle variabili mediante Interval – PLSDA e nuova classificazione

La Interval Partial Least Squares Discriminant Analysis (I-PLSDA o Interval-PLSDA) è un algoritmo in grado di selezionare un sottoinsieme di variabili che fornirà una previsione simile o superiore rispetto a quella fornita utilizzando tutte le variabili di un set di dati ([Nørgaard et al., 2000](#)). Per ciascuna combinazione di *pre-processing*, è stato utilizzato l'algoritmo Interval-PLSDA in modalità in "avanti" (*forward*), settando un numero di 4 intervalli con dimensione della "finestra" pari a 9 (corrispondente a 56.2 nm). A seguito della selezione delle variabili per il dataset Nocciole, è stato settato nuovamente un modello PLS-DA includendo esclusivamente gli intervalli di lunghezza d'onda identificati dall'I-PLSDA. Tale procedura è stata ripetuta per i tre casi.

Analisi dei risultati

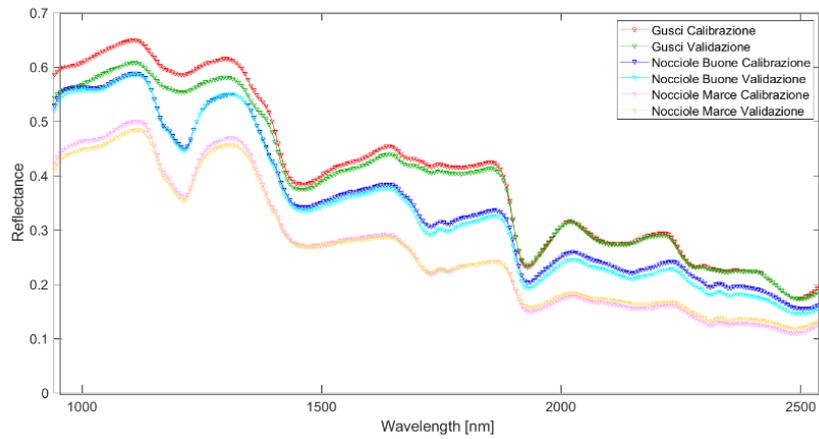
Spettri medi in riflettanza

Gli spettri medi in riflettanza per il dataset Nocciole, il dataset Nocciole di calibrazione ed il dataset Nocciole di validazione sono mostrati rispettivamente in [Figura 4.2.15a](#), [Figura 4.2.15b](#) e [Figura 4.2.15c](#). Come si può notare, gli spettri medi delle Nocciole Buone, Nocciole Marce e dei Gusci di Nocciole, presentano differenze negli intorni di 1300 nm, 1950 nm e 2300 nm. Mentre, gli spettri medi pre-processati con le tre differenti combinazioni di algoritmi di *pre-processing* sono riportati in [Figura 4.2.16](#).

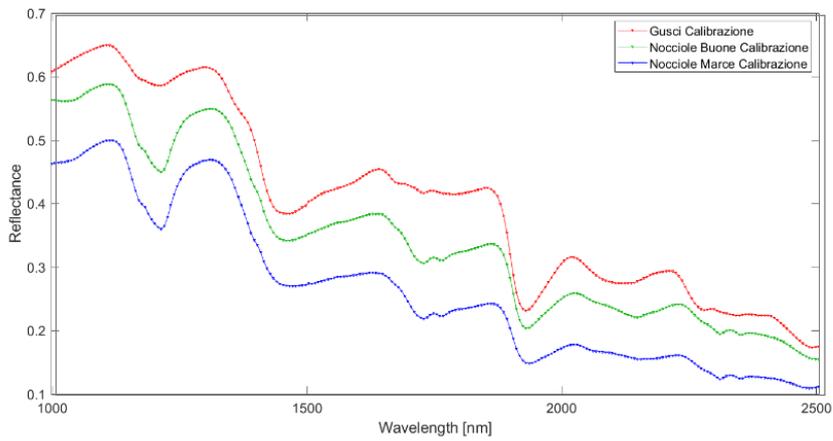
Analisi esplorativa

La variabilità spettrale intra-classe appare maggiormente evidente nelle analisi PCA. In [Figura 4.2.17](#) sono riportati gli score plot risultanti dalle PCA effettuate per le tre combinazioni di *pre-processing* sul dataset di calibrazione.

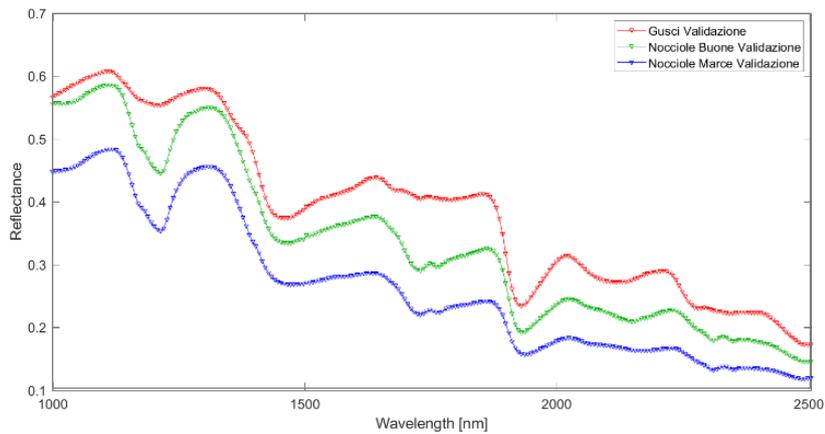
Casi di studio



a



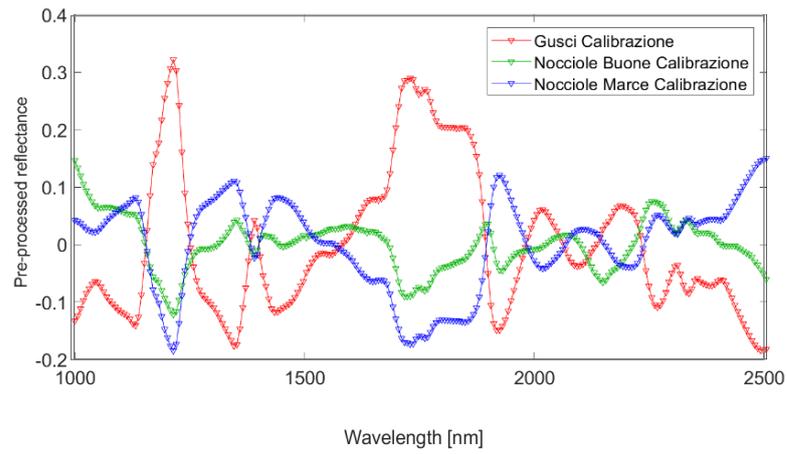
b



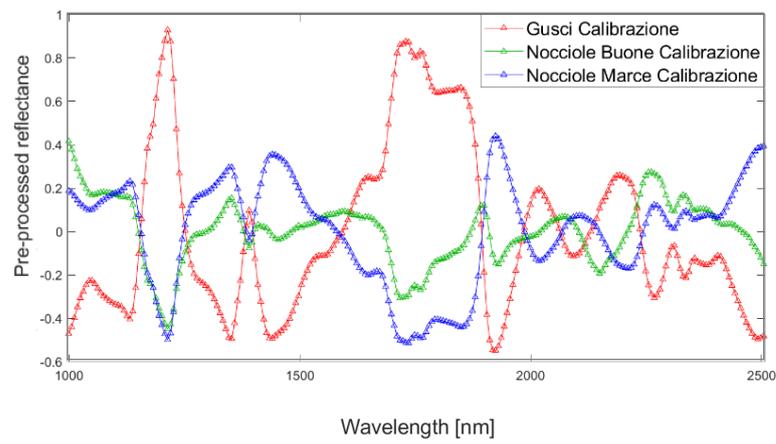
c

Figura 4.2.15. Spettri medi di riflettanza dell'intero dataset (a), del dataset di calibrazione (b) e del dataset di validazione (c), distinti per classi.

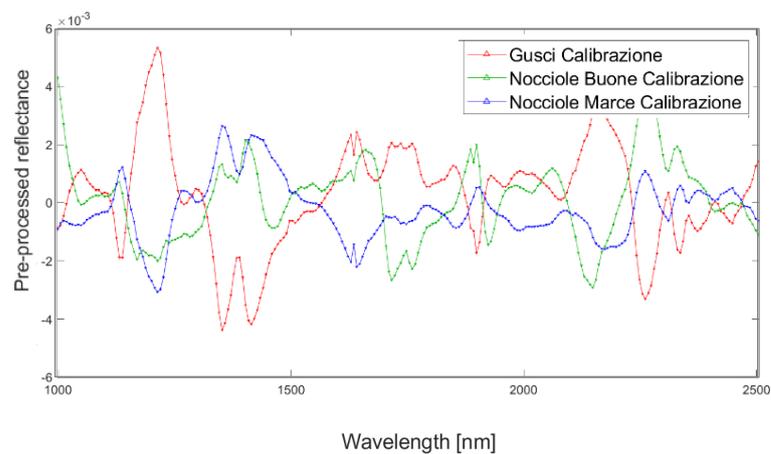
Casi di studio



a



b



c

Figura 4.2.16. Spettri medi pre-processati del dataset di calibrazione con algoritmo: SNV + MC (a), DETREND + SNV + MC (b), GLS-W + MC (c).

Casi di studio

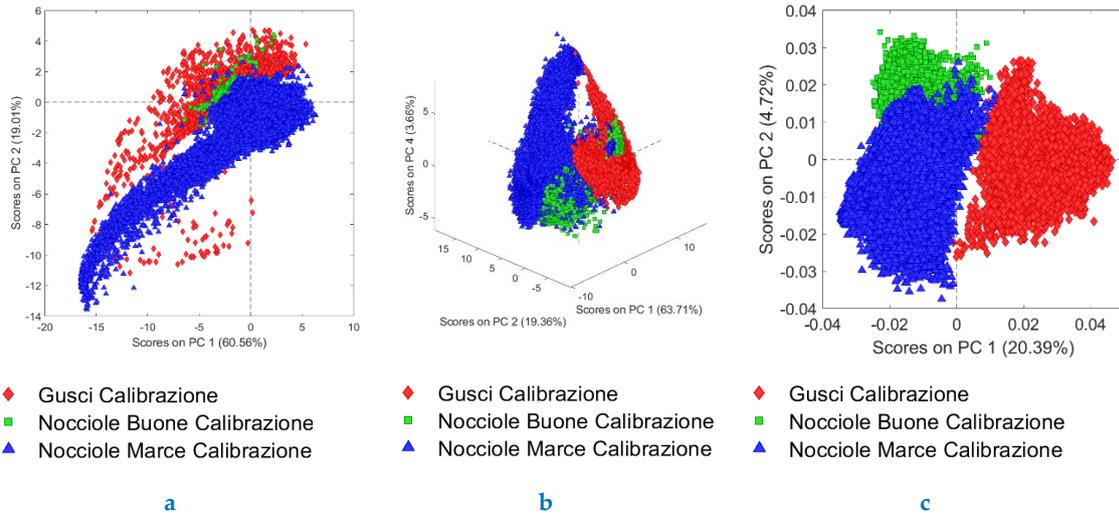


Figura 4.2.17. Scores plot: Scores on PC1 vs Scores on PC 2 per SNV + MC (a), Scores on PC 1 vs Scores on PC2 vs Scores on PC3 per DETREND +SNV+ MC (b), Scores on PC 1 vs Scores on PC 2 per GLS-W + MC (c).

Classificazione Partial Least Squares – Discriminant Analysis

I risultati delle classificazioni PLS-DA in termini di Sensitività e Specificità, per i tre differenti casi, sono riassunti in Tabella 4.2.7. Le mappe di previsione (*most probable*) per i tre modelli sono riportate in Figura 4.2.18. In questo caso, il migliore risultato in termini di sensitività e specificità è stato ottenuto utilizzando la combinazione di algoritmi di *pre-processing* GLS-W + MC. Le performance di classificazione relative ad i singoli casi analizzati sono riportati nell'allegato *Allegato XI- Case study – Nocciole*.

Tabella 4.2.7. Performance di classificazione in predizione delle classificazioni effettuate per le tre differenti combinazioni di *pre-processing*.

| Classe | Caso SNV + MC | | Caso DETREND + SNV + MC | | Caso GLSW + MC | |
|----------------|------------------|-------------|----------------------------|-------------|-------------------|-------------|
| | Sensitivity | Specificity | Sensitivity | Specificity | Sensitivity | Specificity |
| Gusci | 0.997 | 1.000 | 0.999 | 0.999 | 0.999 | 0.998 |
| Nocciole Buone | 0.658 | 0.972 | 0.695 | 0.977 | 0.949 | 0.976 |
| Nocciole Marce | 0.953 | 0.893 | 0.960 | 0.905 | 0.956 | 0.983 |

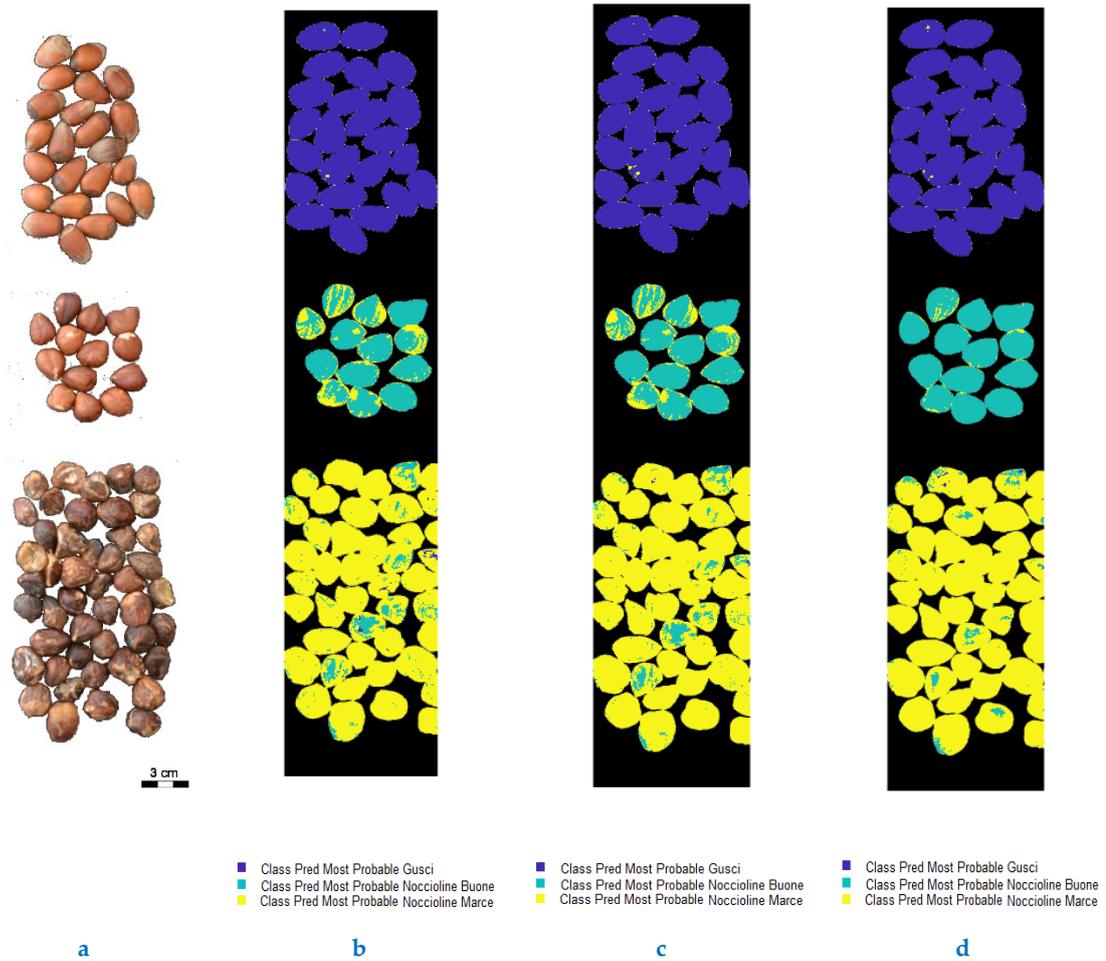


Figura 4.2.18. Immagine RGB dei campioni (a) e mappa di previsione risultanti dalla classificazione PLS-DA (class predicted most probable): caso SNV+MC (b), caso DETREND + SNV + MC (c), caso GLSW +MC (d).

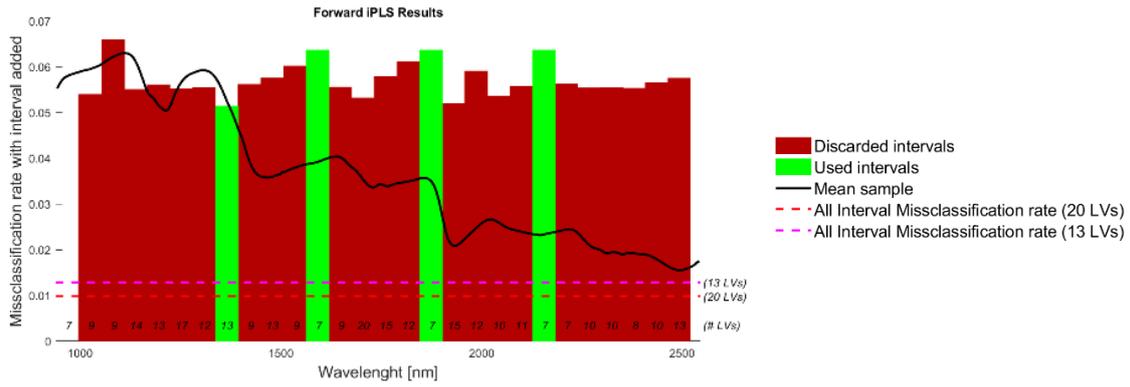
Applicazione dell'Interval PLSDA e selezione delle variabili importanti per la classificazione

Gli intervalli delle variabili selezionate dall'algoritmo I-PLSDA, per ciascuno dei tre casi, sono riportati in Figura 4.2.19 e in Tabella 4.2.8.

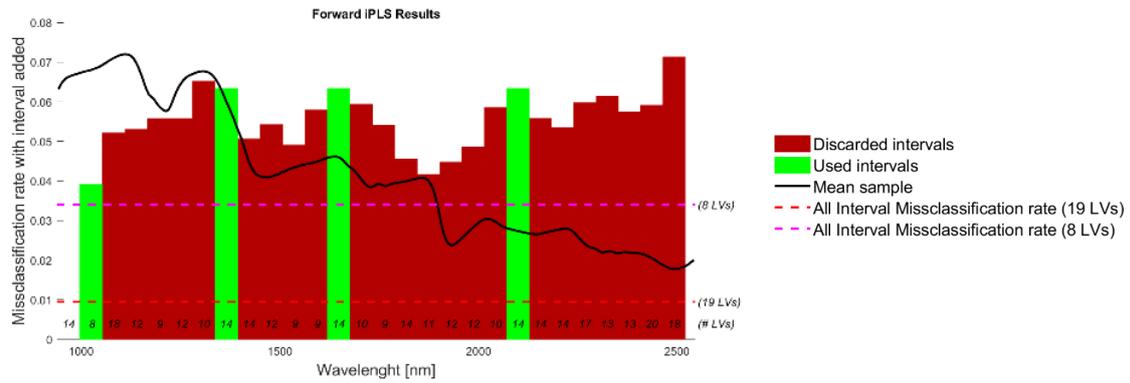
Classificazione PLSDA con variabili selezionate

I risultati delle classificazioni PLSDA (utilizzando esclusivamente gli intervalli di variabili selezionate) in termini di Sensitività e Specificità, per i tre differenti casi, sono riassunti in Tabella 4.2.9. Mentre, le mappe di previsione (*most probable*) per i tre modelli sono riportate in Figura 4.2.20. Le performance di classificazione relative ad i singoli casi analizzati sono riportati nell'allegato *Allegato XI- Case study – Nocciole*.

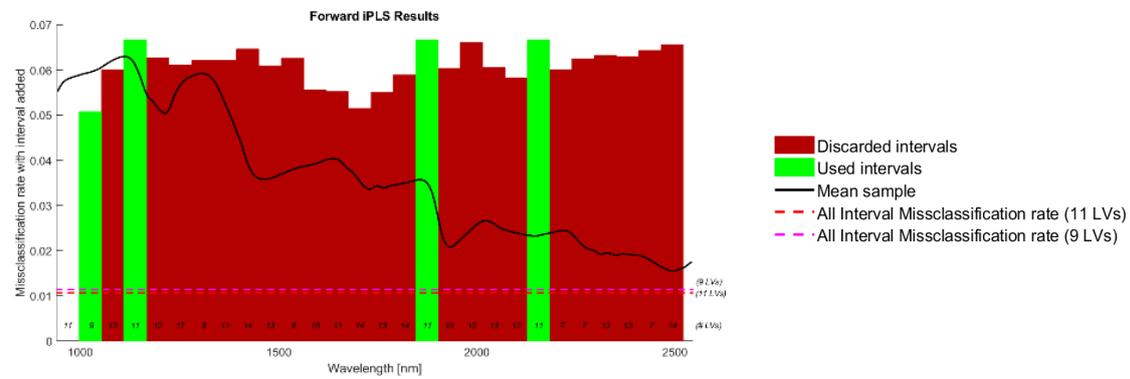
Casi di studio



a



b



c

Figura 4.2.19. Risultato della selezione delle variabili mediante I-PLSDA: caso SNV+MC (a), caso DETREND + SNV + MC (b), caso GLSW +MC (c).

Casi di studio

Tabella 4.2.8. Intervalli di variabili selezionate con Interval-PLSDA, distinti per i tre casi.

| Caso | Caso | Caso |
|----------------|--------------------|----------------|
| SNV + MC | DETREND + SNV + MC | GLS-W + MC |
| | 1000 – 1049 nm | 1000 – 1049 nm |
| | | 1112 – 1163 nm |
| 1339 - 1390 nm | 1339 – 1390 nm | |
| 1566 – 1616 nm | | |
| | 1622 - 1673 nm | |
| 1848 - 1899 nm | | 1848 - 1899 nm |
| | 2074 - 2124 nm | |
| 2130 - 2180 nm | | 2130 - 2180 nm |

In questo caso, il migliore risultato in termini di sensitività e specificità è stato ottenuto utilizzando la combinazione di algoritmi di *pre-processing* DETREND + SNV + MC.

Tabella 4.2.9. Sensitività e Specificità in predizione per i modelli PLSDA, distinte per casi.

| Classe | Caso | | Caso | | Caso | |
|----------------|----------|-------------|--------------------|-------------|-----------|-------------|
| | SNV + MC | Specificity | DETREND + SNV + MC | Specificity | GLSW + MC | Specificity |
| Gusci | 0.948 | 0.987 | 0.992 | 0.997 | 0.977 | 0.991 |
| Nocciole Buone | 0.664 | 0.969 | 0.824 | 0.980 | 0.665 | 0.982 |
| Nocciole Marce | 0.955 | 0.884 | 0.965 | 0.942 | 0.961 | 0.883 |

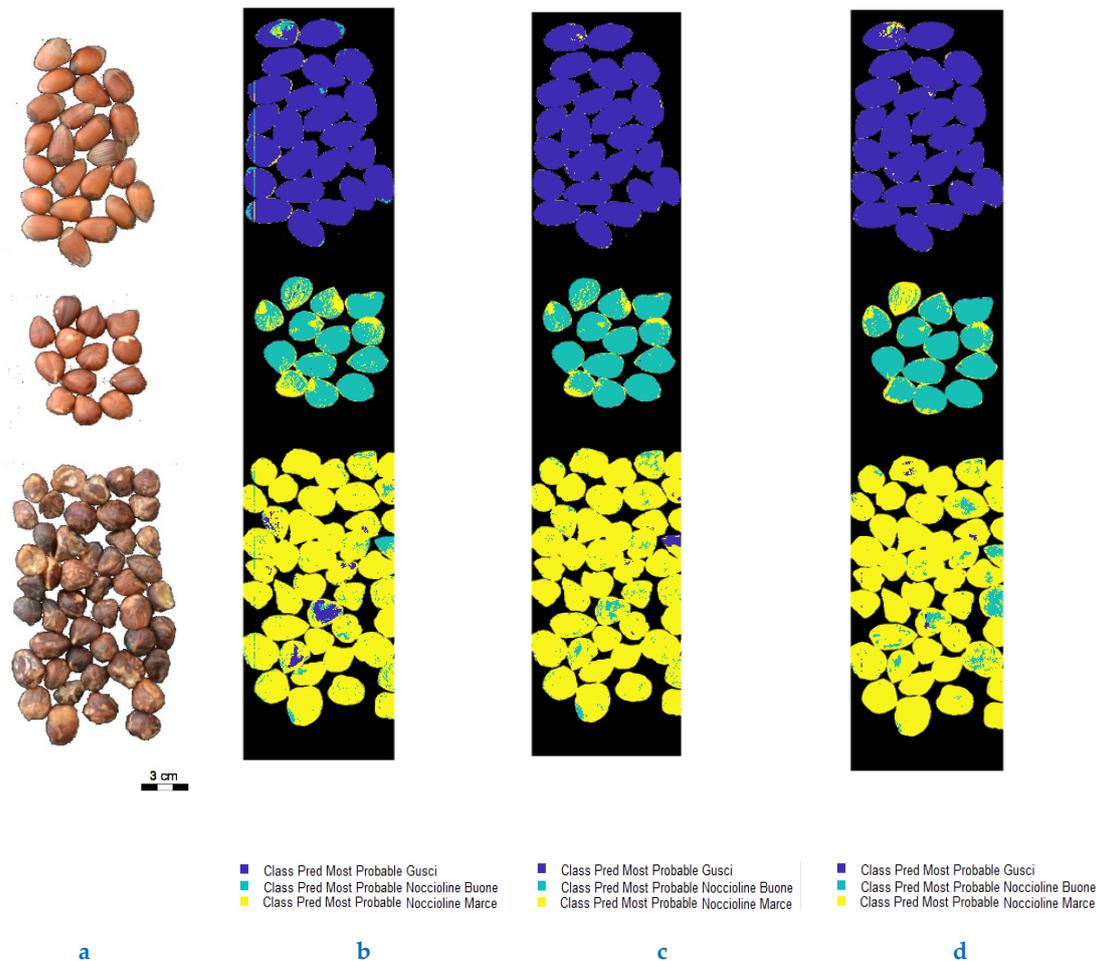


Figura 4.2.20. Immagine RGB dei campioni (a) e mappe di predizione risultanti dalla classificazione PLS-DA (class predicted most probable): caso SNV+MC (b), caso DETREND + SNV + MC (c), caso GLSW+MC (d).

Discussione dei risultati

Le classificazioni PLS-DA, calibrate nell'intervallo di lunghezza tipico dello SWIR (1000-2500 nm), mostrano una ottima capacità di riconoscimento dei campioni di nocciole buone dai gusci e dalle nocciole da scartare. Utilizzando gli spettri in riflettanza nell'intervallo dello SWIR, i migliori risultati in termini di capacità predittiva sono stati ottenuti utilizzando la combinazione di *pre-processing* GLSW + MC. A seguito della selezione delle variabili importanti per la classificazione, sono state messe appunto nuovi modelli di classificazione utilizzando i valori di riflettanza inclusi negli intervalli selezionati dall'algoritmo I-PLSDA. In questo caso i migliori risultati, in termini di capacità predittiva, sono stati ottenuti con la combinazione di algoritmi DETREND + SNV + MC, utilizzando gli intervalli: 1000 – 1049 nm, 1339 – 1390 nm, 1622 - 1673 nm, 2074 - 2124 nm. Ad ogni modo, i risultati ottenuti nei tre casi considerati, esplorando le tre diverse strategie di *pre-processing* dati, mostrano una certa capacità predittiva per le classi modellate. Questo fatto è riscontrabile dai valori di sensibilità e specificità ottenuti in predizione.

Casi di studio

Da quanto evinto, le tecniche di imaging iperspettrale ben si prestano ad effettuare un rilevamento non-distruttivo, veloce e a basso costo della qualità delle nocciole da immettere in commercio. Una tale strategia di controllo, se scalata a livello industriale/aziendale, può consentire di risolvere problemi di selezione e controllo dei prodotti da scartare, che solitamente sono difficilmente identificabili con tecniche di imaging convenzionali.

4.2.4 Classificazione mediante *HyperSpectral Imaging* di pistacchi buoni e relativi scarti

Problematica e stato dell'arte

Il pistacchio, il frutto dell'albero pistacchio (*Pistacia vera*), è una importante materia prima per l'industria della pasticceria e del cioccolato. Nel 2012, la produzione mondiale di pistacchi si aggirava attorno a 1'002'334 t. Il principale produttore di pistacchi era l'Iran con 472'097 t, seguito dagli Stati Uniti d'America con 231'000 t e la Turchia con 150'000 t prodotte. Mentre, l'Italia, settimo produttore al mondo, si attestava su una produzione di circa 2'850 t. Storicamente, in Italia sono presenti diverse coltivazioni di nicchia, tra questi i pistacchi di Bronte e Adriano, presenti sulle pendici dell'Etna e tutelate dal marchio DOP "Pistacchio verde di Bronte". Questa particolare tipologia di frutta a gusci contiene sali minerali e molte vitamine ed è formato da circa l'83 % da lipidi, 12 % da proteine e per il 5 % da carboidrati.

In genere, il valore di mercato della frutta secca è fortemente condizionato dalla loro qualità, che condiziona inoltre la possibilità di ottenere marchi di qualità. La qualità di un lotto di frutta secca viene in genere valutato in termini di freschezza del prodotto, presenza/assenza di difetti, presenza/assenza di muffe e/o insetti ed altri parassiti e la presenza di contaminanti (i.e. presenza di gusci, malli, piccole pietre, ramoscelli, ecc.).

I pistacchi possono essere soggetti all'attacco di muffe, insetti, uccelli, roditori ed altri parassiti. L'alterazione del pistacchio dovuta all'effetto di questi agenti può comportare la fissazione di residui nocivi e/o sgradevoli al gusto (i.e. sapore amaro), fino all'irrancidimento del frutto stesso.

Gli attributi morfologici e morfometrici rappresentano una tra le principali caratteristiche che condizionano la qualità del prodotto. Infatti, questi attributi sono in qualche modo correlati con la crescita del frutto stesso, e di conseguenza ad i difetti fisici dovuti a fattori naturali (i.e. presenza di insetti e/o organismi infestanti, muffe, ecc.) e/o a fattori antropici o artificiali (come, ad esempio, la rottura del guscio del pistacchio a causa di azioni di movimentazione). Ad ogni modo, un approccio basato sugli attributi morfologici e morfometrici non sempre è sufficiente ad effettuare un controllo qualità completo del pistacchio. In quanto, la presenza di muffe e/o zone scolorite non sempre alterano tali attributi. Perciò, si devono perseguire strategie di controllo e/o selezione alternative in grado di investigare le caratteristiche superficiali ed interne del frutto ([Serranti et al., 2012](#)).

In questo scenario, adottando tecniche di imaging iperspettrale, nell'intervallo di lunghezza d'onda 1000-2500 nm (SWIR) si può mettere a punto un sistema in grado di riconoscere i pistacchi buoni dai contaminanti presenti in un lotto (i.e. pietre, bastoncini, malli e gusci di pistacchio e pistacchi da scartare).

Obiettivi dello studio

Gli obiettivi del presente studio sono stati, quindi: i) il riconoscimento di pietre, pistacchi buoni, gusci di pistacchio, malli, pistacchi malati e bastoncini a partire da immagini iper-spetttrali e ii) la selezione delle variabili importanti ai fini della classificazione. La panoramica globale delle analisi effettuate è mostrata in [Figura 4.2.21](#). Per effettuare il riconoscimento dei campioni in esame è stata dapprima acquisita una immagine iperspettrale (Dataset) in laboratorio. Lo sfondo dell'immagine iperspettrale è stato escluso dalle successive analisi e sono state settate le classi dei campioni, tramite interfaccia software. Il Dataset è stato poi diviso in due parti: dataset di calibrazione (contenente il 70 % delle particelle appartenenti a ciascuna classe) e dataset di validazione (contenente il 30 % delle particelle appartenenti a ciascuna classe). Il dataset di calibrazione è stato trattato con diverse combinazioni di algoritmi di *pre-processing* per eliminare pixel rumorosi, attenuare fenomeni di *scattering* e migliorare le differenze inter- ed intra-classe. La *Principal Component Analysis* (PCA) è stata utilizzata per effettuare un'analisi esplorativa e valutare le giuste combinazioni di *pre-processing* da applicare al dataset di calibrazione per la successiva classificazione. Il metodo di discriminazione per le classi settate è la *Partial Least Squares – Discriminant Analysis* (PLS-DA). Per ciascuna combinazione di algoritmi di *pre-processing* utilizzati, i modelli di classificazione, calibrati con dataset di calibrazione, sono stati validati sul dataset di validazione. Successivamente, è stata effettuata la selezione delle variabili importanti ai fini della classificazione tramite l'algoritmo *Interval Partial Least Squares Discriminant Analysis* (Interval-PLSDA o I-PLSDA). Tale procedura è stata applicata per tutti i modelli di classificazione. Infine, sono stati calibrati nuovi modelli di classificazione PLSDA. Tali modelli sono stati calibrati utilizzando esclusivamente le variabili selezionate dalla I-PLSDA e validati sul dataset di validazione.

Procedura

Preparazione dei campioni

I campioni analizzati ([Figura 4.2.22](#)) sono stati suddivisi in tre classi: *Pietre, Pistacchi buoni, Gusci di pistacchio, Malli, Pistacchi malati e Bastoncini*.

Acquisizione di immagini iperspettrali

Lo strumento utilizzato per effettuare l'acquisizione delle immagini iperspettrali in riflettanza è il SisuCHEMA XL™ (Specim, Finlandia) operante nel range dello spettro elettromagnetico compreso tra 1000 nm e 2500 nm, avente una risoluzione spettrale di 6.2 nm. In questo caso, l'ottica utilizzata per le acquisizioni è la 31 mm, la quale consente di acquisire immagini di larghezza pari a 50 mm con velocità di scansione pari a 18 mm/s. Più nel dettaglio, è stata acquisita un'immagine iperspettrale: dataset Pistacchi ([Figura 4.2.23b](#)).

Casi di studio

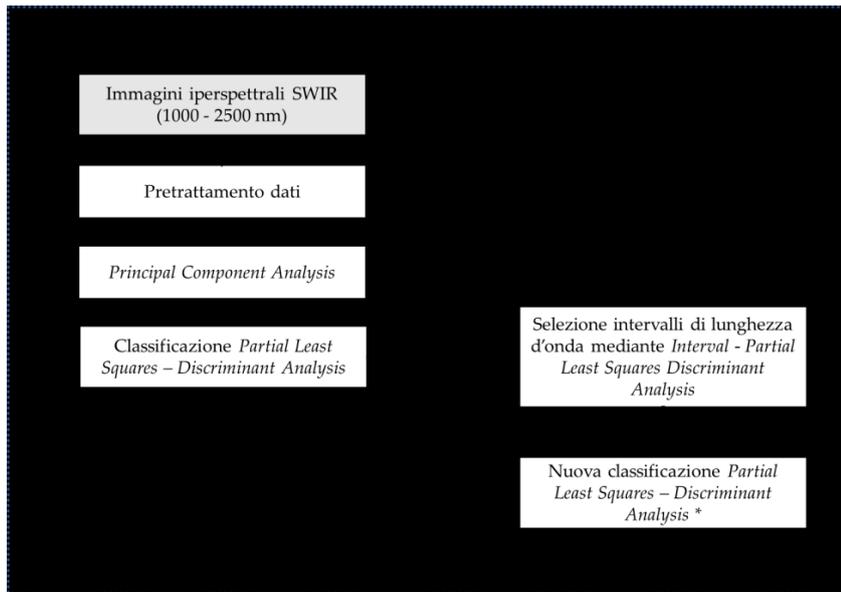


Figura 4.2.21. Diagramma di flusso delle analisi effettuate. *La nuova classificazione è effettuata con gli intervalli di lunghezza d'onda importanti per la classificazione, discriminati dall'algoritmo Interval - Partial Least Squares Discriminant Analysis (I-PLSDA).

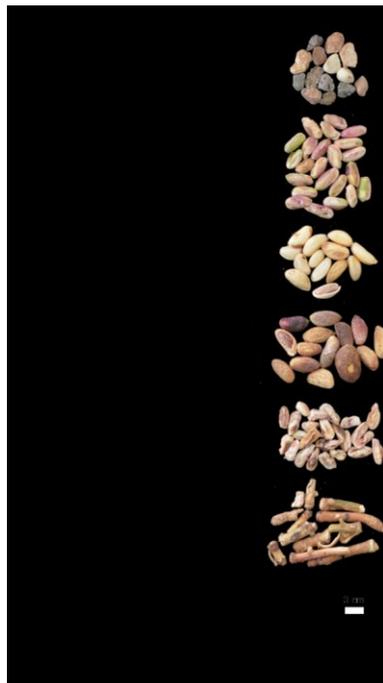


Figura 4.2.22. Campioni di Pistacchi buoni e relativi contaminanti (Pietre, Gusci di pistacchio, Mali, Pistacchi malati e Bastoncini).

Esportazione e preparazione dati

L'immagine iperspettrale acquisita con il sistema SisuCHEMA XL™ è stata importata in ambiente MATLAB (ver. 9.3.0.713579; R2017b). Le successive analisi chemiometriche sono state effettuate in ambiente MATLAB con l'ausilio dei toolbox PLS_toolbox (ver. 8.6 - 21315) e MIA_toolbox (ver. 3.0.5 - 21317). Lo sfondo dell'immagine iperspettrale è stato

escluso dalle successive analisi. Inoltre, il dataset è stato ridotto da 942.2-2542 nm a 1000-2500 nm per rimuovere la parte più rumorosa degli spettri di ciascun pixel, dovuta al detector dello strumento. L'ipercubo, a seguito del settaggio delle classi (Figura 4.2.23c), è stato suddiviso in due dataset: i) dataset di calibrazione (contenente circa il 70 % delle particelle per ciascuna classe) e ii) dataset di validazione (contenente circa il 30 % delle particelle per ciascuna classe). Il dataset di calibrazione è mostrato in Figura 4.2.24a, mentre quello di validazione è mostrato in Figura 4.2.24b.

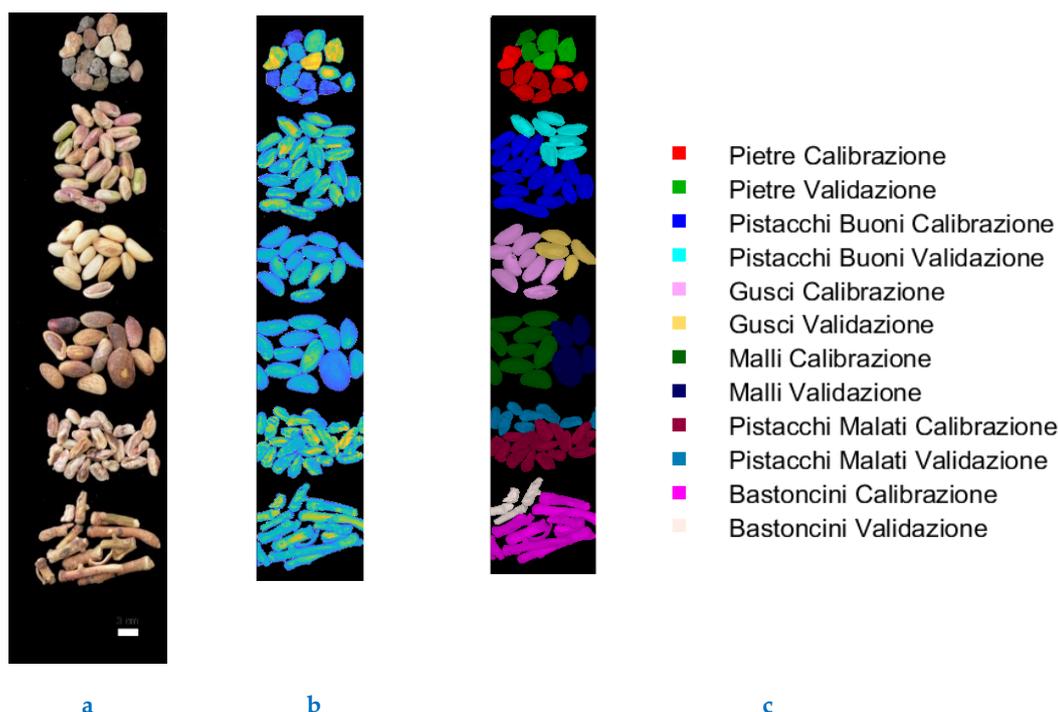


Figura 4.2.23. Immagine RGB dei campioni (a), Immagine iperspettrale in riflettanza media (b) e dataset con classi settate (c).

Pre-elaborazione dati

Prima di passare all'analisi esplorativa è stato necessario applicare una combinazione di algoritmi di *pre-processing*. Sono state quindi valutate differenti strategie di pre-elaborazione dati, in differenti combinazioni. Le tre combinazioni di algoritmi di *pre-processing* applicate ai dati iper-spettrali sono state:

1. *Standard Normal Variate + Mean Center (SNV + MC)*.
2. *Detrend + Standard Normal Variate + Mean Center (DETREND + SNV + MC)*.
3. *Generalized Least Squares Weighting (con α pari a 0.01) + Mean Center (GLS-W + MC)*.

Casi di studio

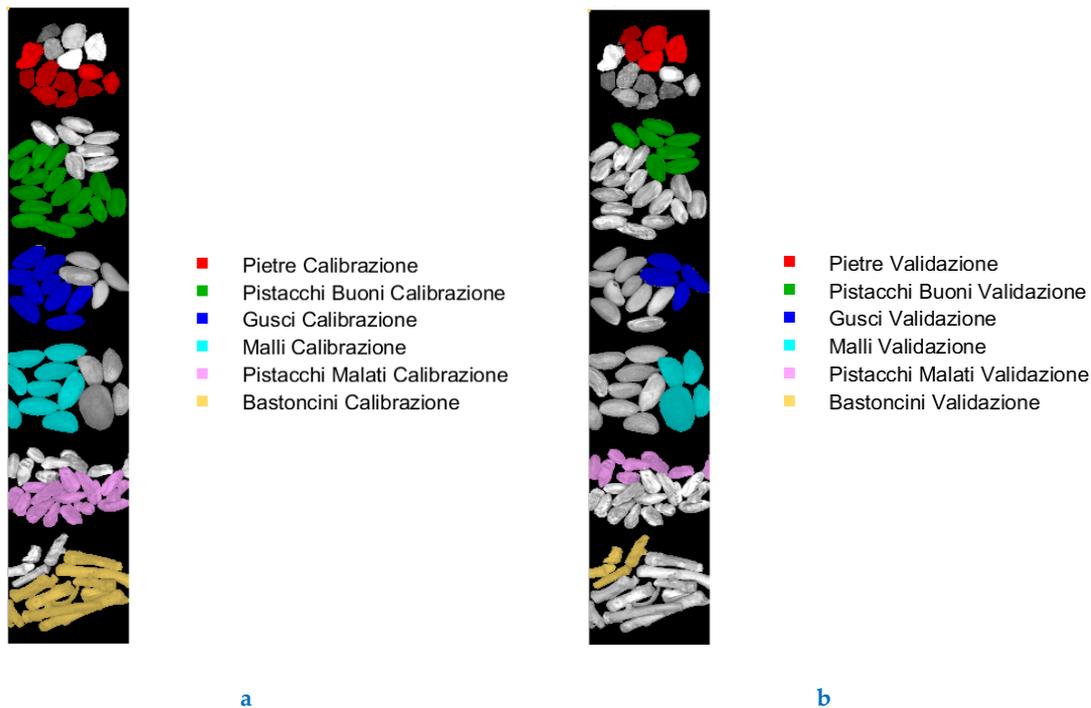


Figura 4.2.24. Dataset di calibrazione (a) e dataset di validazione (b).

Analogamente a quanto visto nel caso di studio [4.2.3 Classificazione mediante hyperspectral imaging di nocciole buone e relativi scarti](#), *Standard Normal Variate* (SNV) è stato utilizzato per scalare le differenze spettrali dovute a fenomeni di *scattering* o alla variazione di sensibilità strumentale, dovute al detector e/o alla fonte luminosa ([Rinnan et al., 2009](#)). Mentre, *Detrend* è stato utilizzato al fine di rimuovere variazioni che potrebbero generare risposte non-lineari, in quanto particolarmente utile quando è presente un *offset* nei dati spettrali. Invece, l'algoritmo *Generalized Least Squares Weighting* (GLS-W) è stato scelto nella terza combinazione di *pre-processing*, al fine di attenuare le differenze fra gruppi di campioni ([Eigenvector Research Inc. 2013b](#)).

Analisi esplorativa dei dati

La *Principal Component Analysis* è stata effettuata per esplorare la variabilità intra-classe ed interclasse ([Wold et al., 1987](#)). Le componenti principali sono state scelte valutando il grafico degli autovalori e della varianza catturata.

Per il Dataset Nocciole, sono state scelte:

- I. tre componenti principali, nel caso SNV + MC;
- II. quattro componenti principali, nel caso DETREND + SNV + MC;
- III. tre componenti principali, nel caso GLS-W + MC.

Gli *outlier* sono stati identificati ed esclusi dalle successive analisi, esplorando il grafico dei parametri Hotelling T^2 e i residui Q .

Classificazione

Il metodo di classificazione PLS-DA ([Wise et al., 2006](#); [Wise et al., 2008](#); [Ballabio & Consonni, 2013](#)) è stato applicato per valutare la capacità predittiva degli spettri SWIR nel classificare i campioni studiati. Nel caso specifico, le classi modellate sono state: *Pietre*, *Pistacchi buoni*, *Gusci di pistacchio*, *Malli*, *Pistacchi malati* e *Bastoncini*. Come algoritmo di cross – validazione, per valutare la complessità del modello e scegliere un adeguato numero di variabili latenti, è stato utilizzato *venetian blinds*. Per ciascuna delle tre differenti combinazioni di *pre-processing*, il modello PLS-DA è stato costruito a partire dal dataset di calibrazione ([Figura 4.2.24a](#)), cross- validato e validato, applicando il modello sul dataset di validazione ([Figura 4.2.24b](#)).

Analisi dei risultati

Spettri medi in riflettanza

Gli spettri medi in riflettanza per il dataset Pistacchi, il dataset Pistacchi di calibrazione ed il dataset Pistacchi di validazione sono mostrati rispettivamente in [Figura 4.2.25a](#), [Figura 4.2.25b](#) e [Figura 4.2.25c](#). Come si può notare, gli spettri medi della classe Pietre, presentano una firma spettrale visivamente differente e maggiormente accentuata rispetto agli altri spettri medi. Differenza, questa, che si accentua maggiormente con l'applicazione della procedura di *pre-processing*. Gli spettri medi pre-processati con le tre differenti combinazioni di algoritmi di *pre-processing* sono riportati in [Figura 4.2.26](#).

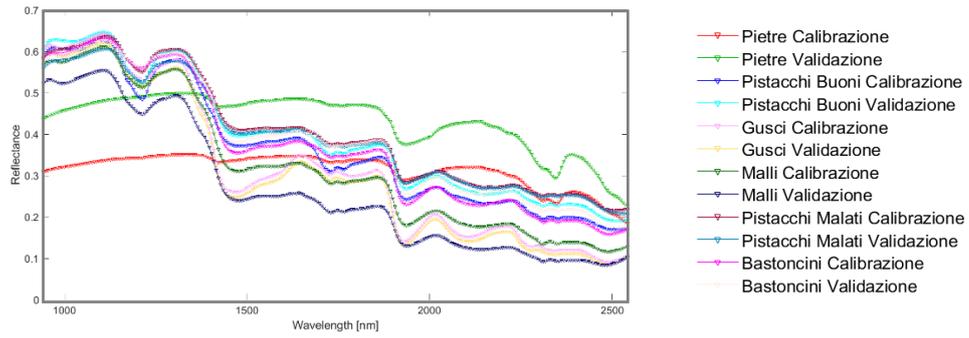
Analisi esplorativa

La variabilità spettrale intra-classe appare maggiormente evidente nelle analisi PCA. In [Figura 4.2.27](#) sono riportati gli score plot risultanti dalle PCA effettuate per le tre combinazioni di *pre-processing* sul dataset di calibrazione.

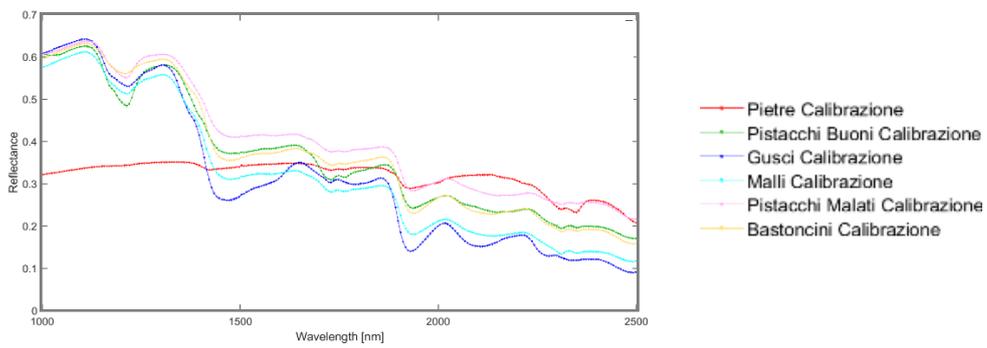
Classificazione Partial Least Squares – Discriminant Analysis

I risultati delle classificazioni PLS-DA in termini di Sensitività e Specificità, per i tre differenti casi, sono riassunti in [Tabella 4.2.10](#). Mentre, le mappe di previsione (*most probable*) per i tre modelli sono riportate in [Figura 4.2.28](#). I parametri di performance dei singoli modelli di classificazione sono riportati nella sezione “Allegati” (si veda [Allegato XII- Case study – Pistacchi](#)). In questo caso, il migliore risultato in termini di sensitività e specificità è stato ottenuto utilizzando la combinazione di algoritmi di *pre-processing* GLS-W +MC.

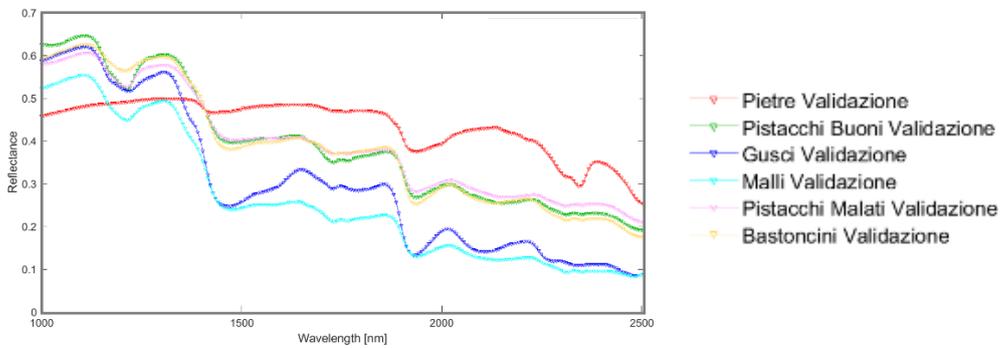
Casi di studio



a



b



c

Figura 4.2.25.. Spettri medi di riflettanza dell'intero dataset (a), del dataset di calibrazione (b) e del dataset di validazione (c), distinti per classi.

Casi di studio

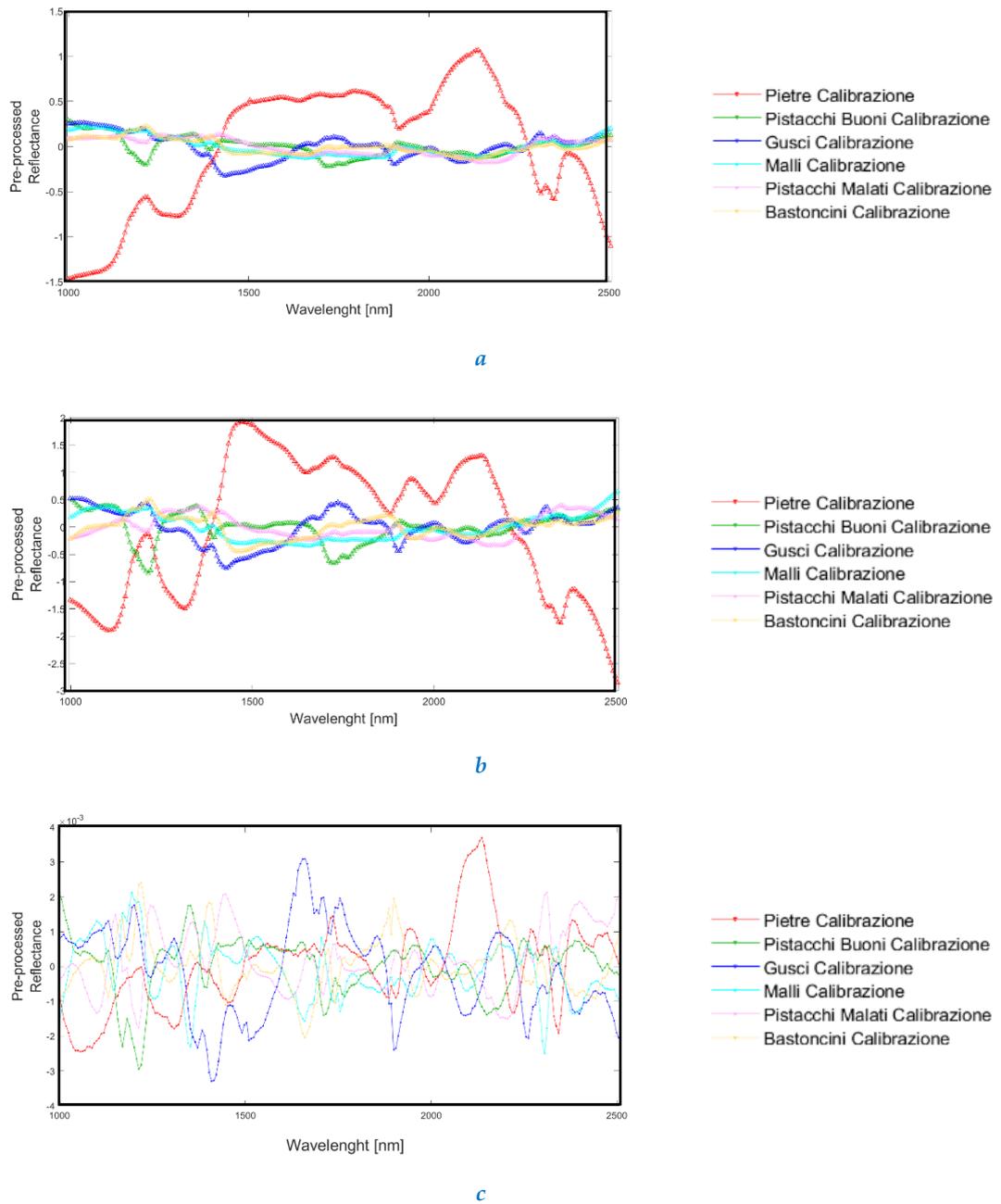


Figura 4.2.26. Spettri medi pre-processati del dataset di calibrazione con algoritmo: SNV + MC (a), DETREND + SNV + MC (b), GLS-W + MC (c).

Casi di studio

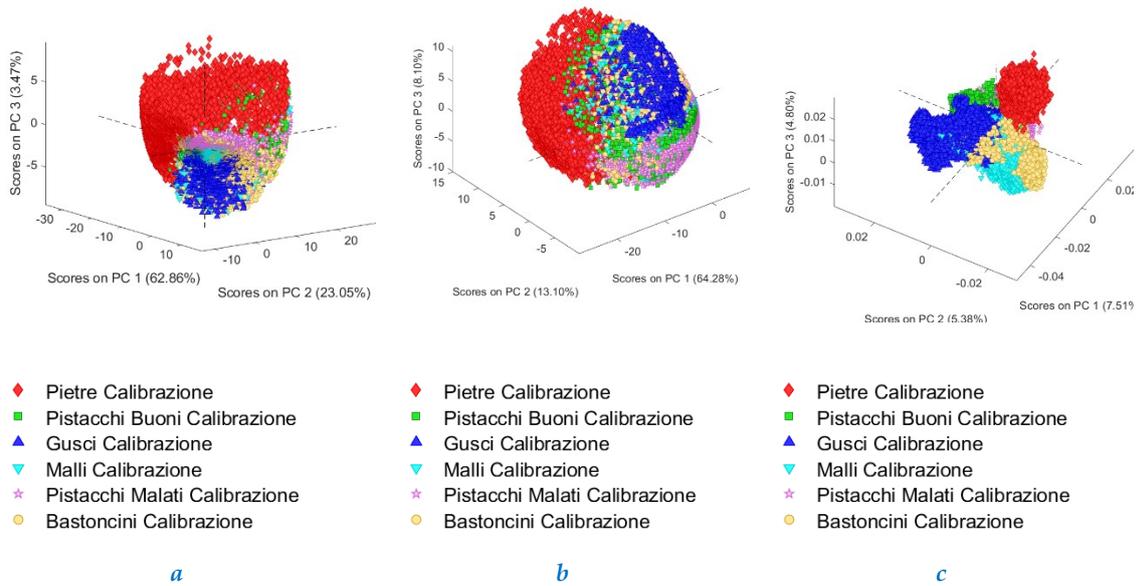


Figura 4.2.27. Scores plot: Scores on PC 1 vs Scores on PC 2 vs Scores on PC 3 per SNV + MC (a), Scores on PC 1 vs Scores on PC 2 vs Scores on PC 3 per DETREND + SNV + MC (b), Scores on PC 1 vs Scores on PC 2 per GLS-W + MC vs Scores on PC 3 (c).

Applicazione dell'Interval PLSDA e selezione delle variabili importanti per la classificazione

Gli intervalli delle variabili selezionate dall'algoritmo I-PLSDA, per ciascuno dei tre casi, sono riportati in Tabella 4.2.11. In Figura 4.2.29 è mostrato il risultato della selezione delle variabili mediante I-PLSDA.

Tabella 4.2.10. Performance di classificazione in predizione delle classificazioni effettuate per le tre differenti combinazioni di pre-processing.

| Classe | Caso SNV + MC | | Caso DETREND + SNV + MC | | Caso GLS-W + MC | |
|------------------|------------------|-------------|----------------------------|-------------|--------------------|-------------|
| | Sensitivity | Specificity | Sensitivity | Specificity | Sensitivity | Specificity |
| Pietre | 0.992 | 0.999 | 1.000 | 0.993 | 1.000 | 0.964 |
| Pistacchi Buoni | 0.707 | 0.937 | 0.779 | 0.936 | 0.871 | 0.983 |
| Gusci | 0.912 | 0.987 | 0.900 | 0.971 | 0.908 | 0.992 |
| Melli | 0.776 | 0.964 | 0.804 | 0.946 | 0.929 | 0.981 |
| Pistacchi Malati | 0.480 | 0.927 | 0.566 | 0.939 | 0.853 | 0.962 |
| Bastoncini | 0.749 | 0.907 | 0.592 | 0.955 | 0.768 | 0.993 |

Casi di studio

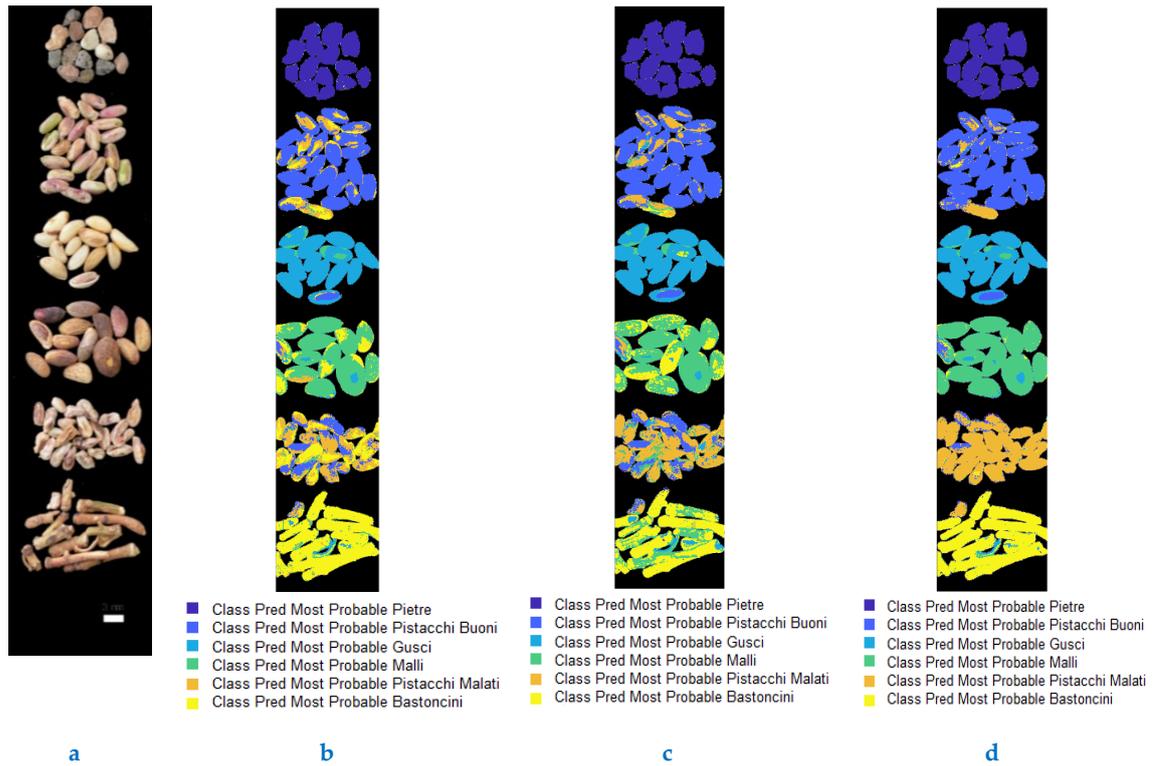


Figura 4.2.28. Immagine RGB dei campioni (a) e mappa di predizione risultanti dalla classificazione PLS-DA (class predicted most probable): caso SNV+MC (b), caso DETREND + SNV + MC (c), caso GLSW +MC (d).

Tabella 4.2.11. Intervalli di variabili selezionate con Interval-PLSDA, distinti per i tre casi.

| Caso | Caso | Caso |
|--------------|--------------------|--------------|
| SNV + MC | DETREND + SNV + MC | GLS-W + MC |
| 1000-1049 nm | 1123-1163 nm | 1123-1163 nm |
| 1169-1220 nm | | |
| 1220-1277 nm | | 1226-1276 nm |
| 1340-1390 nm | 1340-1390 nm | 1340-1390 nm |
| | 1736-1786 nm | 1736-1786 nm |
| | 2417-2461 nm | |

Casi di studio

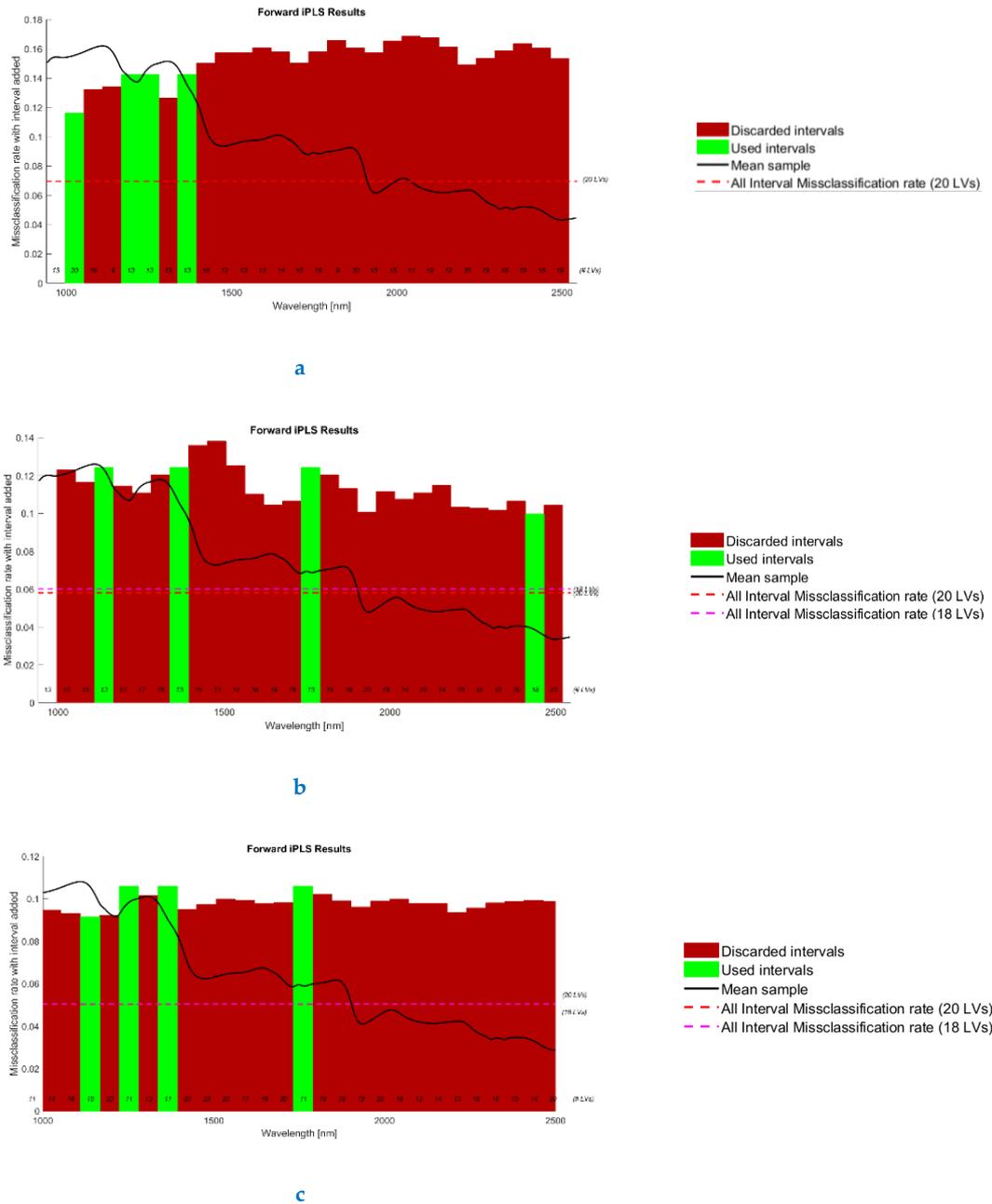


Figura 4.2.29. Risultato della selezione delle variabili mediante I-PLSDA: caso SNV+MC (a), caso DETREND + SNV + MC (b), caso GLSW +MC (c).

Classificazione PLSDA con variabili selezionate

I risultati delle classificazioni PLS-DA (utilizzando esclusivamente gli intervalli di variabili selezionate) in termini di Sensitività e Specificità, per i tre differenti casi, sono riassunti in Tabella 4.2.12. Mentre, le mappe di previsione (*most probable*) per i tre modelli sono riportate in Figura 4.2.30.

Casi di studio

Anche in questo caso, il migliore risultato in termini di sensitività e specificità è stato ottenuto utilizzando la combinazione di algoritmi di *pre-processing* GLS-W + MC. I parametri di performance dei singoli modelli di classificazione sono riportati nella sezione “Allegati” (si veda *Allegato XII- Case study – Pistacchi*).

Tabella 4.2.12. Sensitività e Specificità in predizione per i modelli PLSDA, distinte per casi.

| Classe | Caso SNV + MC | | Caso DETREND + SNV + MC | | Caso GLS-W + MC | |
|------------------|------------------|-------------|----------------------------|-------------|--------------------|-------------|
| | Sensitivity | Specificity | Sensitivity | Specificity | Sensitivity | Specificity |
| Pietre | 0.995 | 0.989 | 0.926 | 0.984 | 0.974 | 0.971 |
| Pistacchi Buoni | 0.878 | 0.915 | 0.868 | 0.922 | 0.767 | 0.964 |
| Gusci | 0.904 | 0.976 | 0.904 | 0.991 | 0.885 | 0.987 |
| Malli | 0.858 | 0.957 | 0.868 | 0.922 | 0.785 | 0.965 |
| Pistacchi Malati | 0.374 | 0.964 | 0.469 | 0.952 | 0.728 | 0.934 |
| Bastoncini | 0.687 | 0.953 | 0.516 | 0.966 | 0.658 | 0.945 |

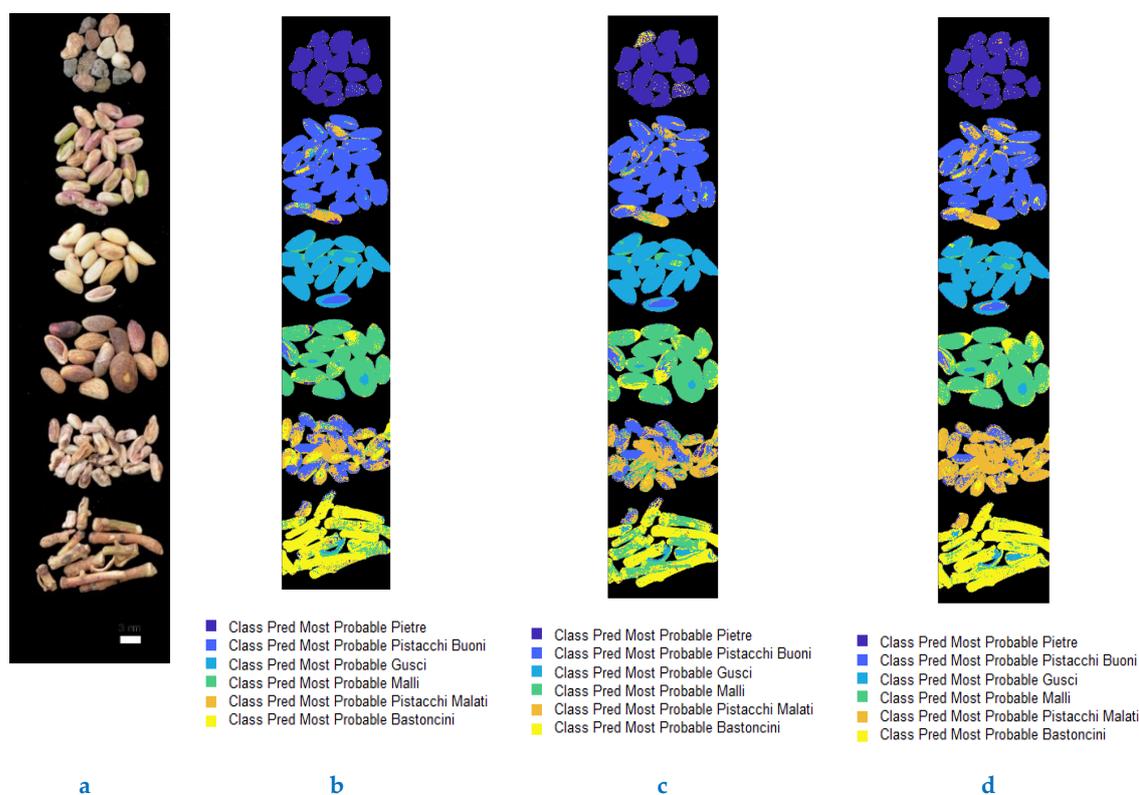


Figura 4.2.30. Immagine RGB dei campioni (a) e mappe di predizione risultanti dalla classificazione PLS-DA (class predicted most probable): caso SNV+MC (b), caso DETREND + SNV + MC (c), caso GLSW +MC (d).

Discussione dei risultati

Le classificazioni PLS-DA, calibrate nell'intervallo di lunghezza tipico dello SWIR (1000-2500 nm), mostrano una ottima capacità di riconoscimento dei campioni di pistacchi buoni dalle classi di contaminanti analizzati. Utilizzando gli spettri in riflettanza nell'intervallo dello SWIR, i migliori risultati in termini di capacità predittiva sono stati ottenuti utilizzando la combinazione di *pre-processing* GLSW + MC. A seguito della selezione delle variabili importanti per la classificazione, sono state messe appunto nuovi modelli di classificazione utilizzando i valori di riflettanza inclusi negli intervalli selezionati dall'algoritmo I-PLSDA. In questo caso, i migliori risultati in termini di capacità predittiva sono stati ottenuti con la combinazione di algoritmi SNV + MC, per quanto riguarda in particolare il riconoscimento di pistacchi buoni, utilizzando gli intervalli: 1000-1049 nm, 1169-1220 nm, 1220-1277 nm e 1340-1390 nm. Ad ogni modo, i risultati ottenuti nei tre casi considerati, esplorando le tre diverse strategie di *pre-processing* dati, mostrano una certa capacità predittiva per le classi modellate. Questo fatto è riscontrabile dai valori di sensibilità e specificità ottenuti in predizione. Da quanto evinto, le tecniche di imaging iperspettrale ben si prestano ad effettuare un rilevamento non-distruttivo, veloce e a basso costo della qualità dei pistacchi da immettere in commercio. Una tale strategia di controllo, se scalata a livello industriale/aziendale, potrebbe consentire la risoluzione di problemi di selezionamento e controllo dei prodotti da scartare, che solitamente sono difficilmente identificabili con tecniche di imaging convenzionali.

4.2.5 Studio sulla germinabilità di semi mediante *HyperSpectral Imaging* nell'infrarosso ad onde corte

Problematica e stato dell'arte

L'analisi delle sementi rappresenta una componente essenziale all'interno del processo di certificazione. Le sementi, per essere certificati, e di conseguenza per essere messe in commercio, devono possedere determinati requisiti relativi alla loro identità e purezza varietale, al loro valore biologico e fisico, al loro stato sanitario. Con l'analisi di germinabilità si determina la percentuale di semi puri capaci di produrre germogli normali, cioè plantule le cui strutture essenziali abbiano uno sviluppo equilibrato e sano e siano quindi in grado di sviluppare, in condizioni favorevoli di coltura, piante capaci di compiere l'intero loro ciclo vegetativo e riproduttivo. La facoltà germinativa riportata sui certificati di analisi indica "la percentuale di semi che hanno prodotto plantule normali nelle condizioni e nei limiti di tempo stabiliti dai metodi ufficiali di analisi." (*Spreafico, 1978*). Le prove di germinazione hanno una durata variabile a seconda della specie che va dai pochi giorni (7 per trifogli, erbe mediche, lattuga, zucche e canapa; 10-15 giorni per cereali, peperoni, barbabietole, pomodori, fagioli e piselli) fino ad arrivare ad un mese (28 giorni per il prezzemolo).

Dalla necessità di garantire la qualità e la germinabilità della produzione sementiera fino al momento dell'utilizzazione nasce il bisogno di sviluppare tecniche di analisi rapide e non invasive per il controllo di qualità delle sementi.

Negli ultimi anni, tecniche di Imaging multispettrale ed iperspettrale nel visibile, (450 - 950 nm) e nel vicino infrarosso (950-1650 nm) sono state studiate al fine di: valutare la qualità in maniera non distruttiva e non invasiva e selezionare semi sulla base della propria vitalità (*Olesen et al., 2011a; Olesen et al., 2011b*); valutare il deterioramento delle sementi (*Dell'Aquila, 2006*) ed identificare i danni abiotici (*Wang et al., 2002; Agelet et al., 2012; Jia et al., 2016*) e biotici (*Pasikatan & Dowell, 2001*) nelle sementi. Esiste in letteratura anche un certo numero di studi dell'applicazione della spettroscopia ad infrarosso per prevedere la germinazione di semi di soia (*Al-Amery, 2018*) per prevedere il tasso di germinabilità dei semi di pomodoro (*Wang & Wang, 2013*).

Obiettivi dello studio

In questo studio è stata valutata la possibilità di utilizzare tecniche basate sull'utilizzo di HSI nel range spettrale 1000-2500 nm per il riconoscimento di semi germinabili o non germinabili. Sono state analizzate tre tipologie di sementi (Melanzana, zucca e cocomero) per valutarne la germinabilità tramite approccio standard e a partire da immagini iperspettrali nello SWIR sono stati calibrati e validati modelli di classificazione PLS-DA per riconoscere i semi non germinabili (morti a distanza temporale di un mese dall'acquisizione) da semi germinabili (germinati a distanza temporale di un mese

dall'acquisizione). La panoramica globale delle analisi effettuate è mostrata in [Figura 4.2.31](#).

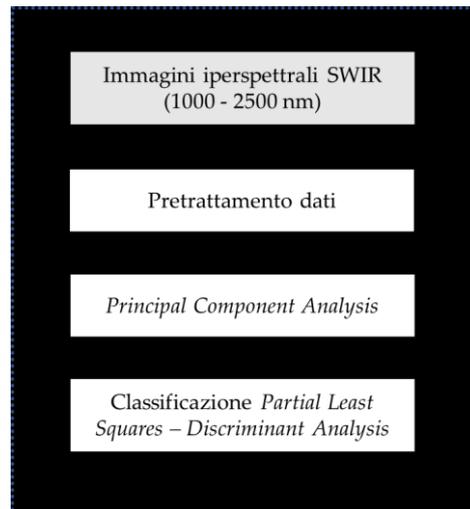


Figura 4.2.31. Diagramma di flusso delle analisi effettuate. La procedura effettuata per mettere a punto la classificazione è inscritta all'interno del quadrato tratteggiato in blu.

Procedura

Preparazione dei campioni

I campioni di sementi analizzati ([Figura 4.2.32](#)) sono:

1. 100 Semi di melanzana - *Solanum Melongena* ([Figura 4.2.32a](#)),
2. 100 Semi di zucca - *Cucurbita Pepo* ([Figura 4.2.32b](#)),
3. 100 Semi di cocomero – *Citrullus Lanatus* ([Figura 4.2.32c](#)).

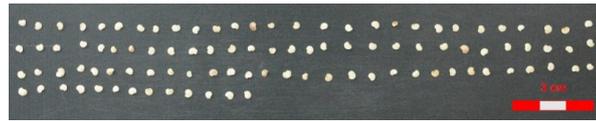
Acquisizione di immagini iperspettrali

Lo strumento utilizzato per effettuare l'acquisizione delle immagini iperspettrali in riflettanza è il SisuCHEMA XL™ (Specim, Finlandia) operante nel range 1000-2500 nm con una risoluzione spettrale di 6.2 nm. L'ottica utilizzata per le acquisizioni è la 31 mm, la quale consente di acquisire immagini di larghezza pari a 50 mm ed una velocità di scansione pari a 18 mm/s.

Più nel dettaglio, sono state acquisite tre immagini iperspettrali, raffiguranti i 100 semi di ciascuna specie analizzata:

- I. Dataset Melanzana (100 Semi di melanzana) - *Solanum Melongena* ([Figura 4.2.33 b](#));
- II. Dataset Zucca (100 Semi di zucca) - *Cucurbita Pepo* ([Figura 4.2.34 b](#));
- III. Dataset Cocomero (100 Semi di cocomero) – *Citrullus Lanatus* ([Figura 4.2.35 b](#)).

Casi di studio



a



b



c

Figura 4.2.32. I campioni analizzati: semi di *Solanum Melongena* (a), *Cucurbita Pepo* (b) e *Citrullus Lanatus* (c).

Esportazione dati e creazione dataset

Le elaborazioni dati e le successive analisi chemiometriche sono state effettuate in ambiente MATLAB (MATLAB R2017b ver. 9.3.0.713579) con il supporto dei toolbox PLS_toolbox (ver. 8.6 – 21315) e MIA_toolbox (ver. 3.0.5 – 21317). Lo sfondo di ciascuna immagine iperspettrale è stato escluso. Inoltre, i dataset sono stati ridotto da 942.2-2542 nm a 1000-2500 nm per rimuovere la parte più rumorosa degli spettri di ciascun pixel, dovuta al detector dello strumento.

Le classi di germinabilità sono state valutate tramite ispezione visiva a distanza di un mese dall'acquisizione. Le classi settate sulle immagini sono:

- "0" = non germinabile / non germinato (seme non germinato a distanza di un mese dall'acquisizione),
- "1" = germinabile / germinato (seme germinato a distanza di un mese dall'acquisizione).

Casi di studio

I dataset, a seguito del settaggio delle classi (Figura 4.2.33 c, Figura 4.2.34 c e Figura 4.2.35 c), sono stati suddivisi in:

1. set di calibrazione (contenente circa il 70 % delle particelle per ciascuna classe),
2. set di validazione (contenente circa il 30 % delle particelle per ciascuna classe).

In Tabella 4.2.13 sono riportati i numeri di semi germinati e non germinati a distanza di circa un mese dall'acquisizione e la suddivisione degli stessi.

I set di calibrazione e validazione per Dataset Melanzana - *Solanum Melongena*, Dataset Zucca - *Cucurbita Pepo* e Dataset Cocomero - *Citrullus Lanatus* sono riportati rispettivamente in Figura 4.2.36 a, Figura 4.2.36 b e Figura 4.2.36 c.

Tabella 4.2.13. Numero di semi germinati e non germinati, distinti per specie e set di appartenenza. (*) "0" = non germinabile / non germinato (seme non germinato a distanza di un mese dall'acquisizione); "1" = germinabile / germinato (seme germinato a distanza di un mese dall'acquisizione).

| | <i>Solanum Melongena</i> (Melanzana) | | <i>Cucurbita Pepo</i> (Zucca) | | <i>Citrullus Lanatus</i> (Cocomero) | |
|---------------------|---|----|----------------------------------|----|--|----|
| Condizione* | 0 | 1 | 0 | 1 | 0 | 1 |
| Totale | 29 | 71 | 23 | 77 | 5 | 95 |
| Set di calibrazione | 25 | 61 | 20 | 66 | 4 | 81 |
| Set di validazione | 4 | 10 | 3 | 11 | 1 | 14 |

Pre-elaborazione dati

Sono state testate differenti combinazioni di algoritmi di pre-elaborazione di dati spettrali. Tra queste, la combinazione scelta di algoritmi di *pre-processing* è la seguente: *Image - Smooth*, *GLS Weighting* (effettuato sulle classi, con $\alpha=0.0001$), *Smoothing* e *Mean Center* (Rinnan et al., 2009; Eigenvector Research Inc. 2013b).

Analisi esplorativa dei dati

L'analisi in componenti principali (o *Principal Component Analysis*, PCA) è stata effettuata sui dataset di calibrazione per esplorare la variabilità intra-classe (Wold et al., 1987). Le componenti principali sono state scelte valutando il grafico degli autovalori e della varianza catturata. Gli *outlier* sono stati identificati ed esclusi dalle successive analisi, esplorando il grafico dei parametri Hotelling T² e i residui Q.

Casi di studio

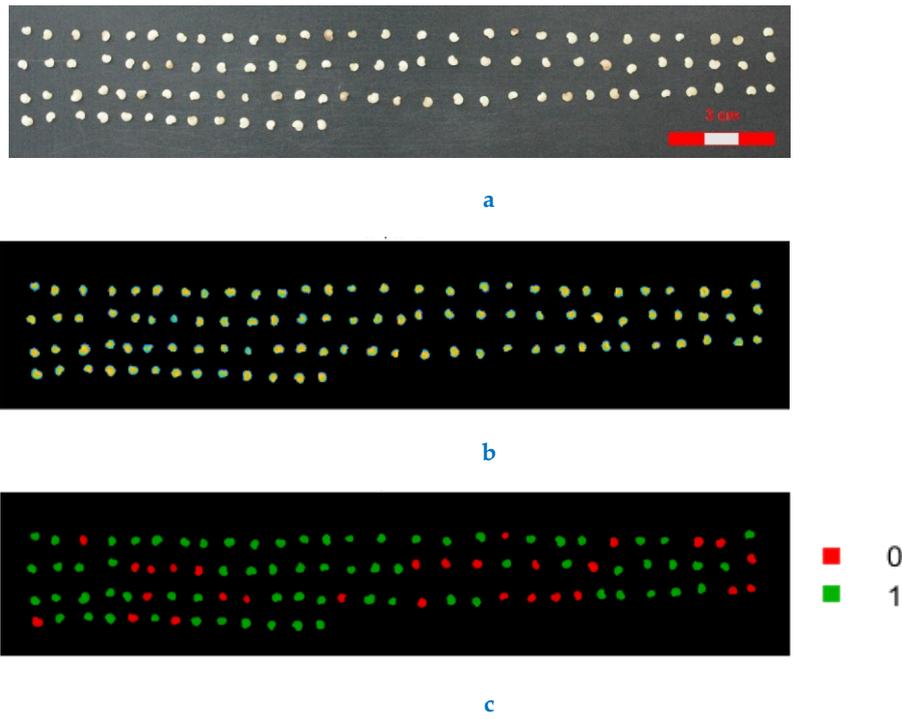


Figura 4.2.33. *Solanum Melongena* (Melanzana): Immagine RGB dei campioni (a), Immagine iper-spettrale in riflettanza media (b) e dataset con classi settate (c). Note della legenda: "0" = non germinato; "1" = germinato.

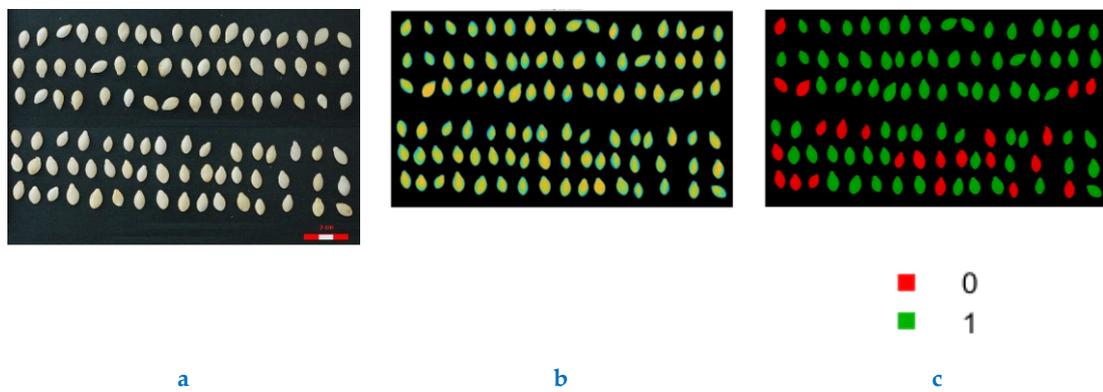


Figura 4.2.34. *Cucurbita Pepo* (Zucca): Immagine RGB dei campioni (a), Immagine iper-spettrale in riflettanza media (b) e dataset con classi settate (c). Note della legenda: "0" = non germinato, "1" = germinato.

Casi di studio

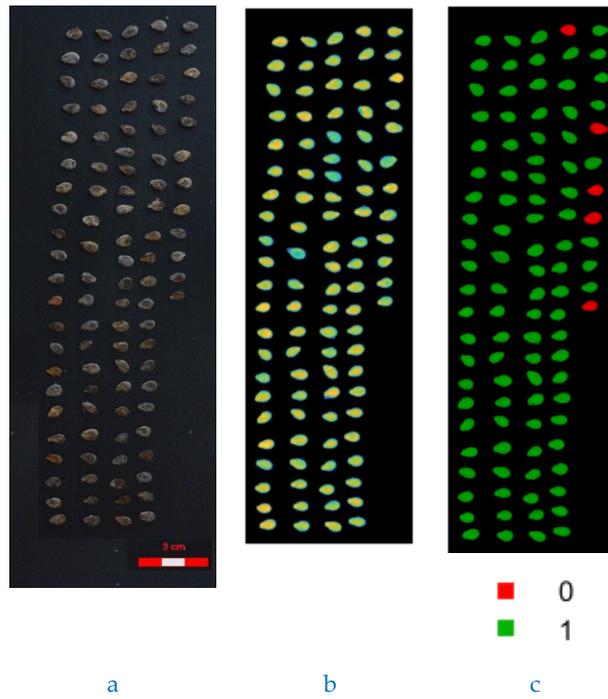


Figura 4.2.35. *Citrullus Lanatus* (Cocomero): Immagine RGB dei campioni (a), Immagine iper-spettrale in riflettanza media (b) e dataset con classi settate (c). Note della legenda: "0" = non germinato, "1" = germinato.

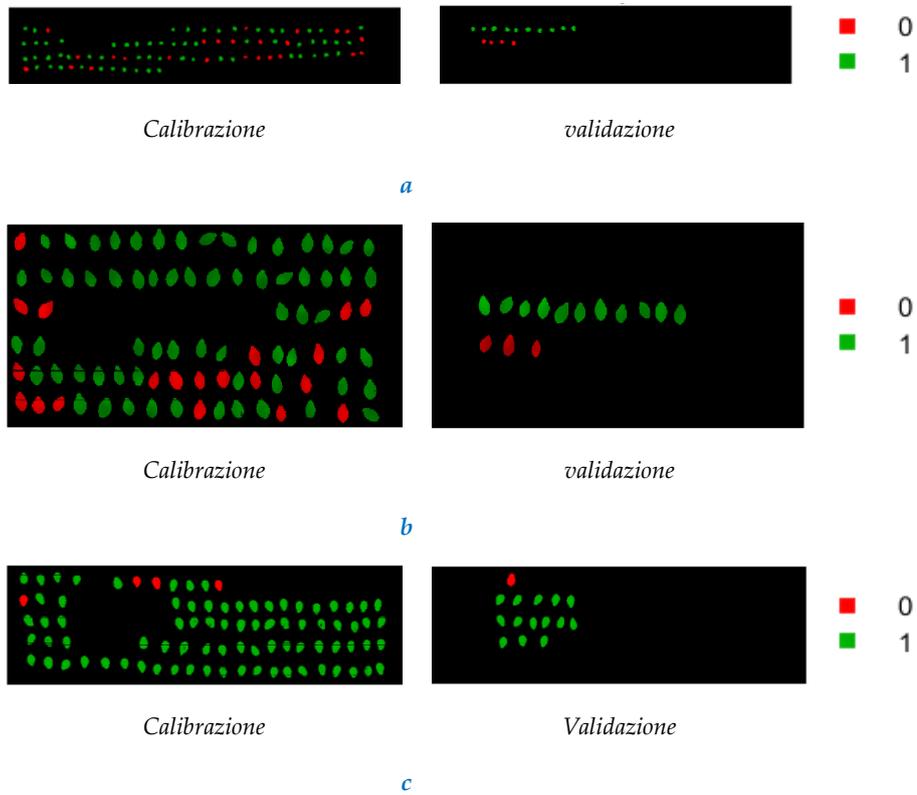


Figura 4.2.36. Set di calibrazione e set di validazione per: Melanzana - *Solanum Melongena* (a), Zucca - *Cucurbita Pepo* (b) e Cocomero - *Citrullus Lanatus* (c). Note della legenda: "0" = non germinato, "1" = germinato.

Classificazione

Il metodo di classificazione PLS-DA ([Wise et al., 2006](#); [Wise et al., 2008](#); [Ballabio & Consonni, 2013](#)) è stato applicato per valutare la capacità predittiva degli spettri SWIR nel classificare i campioni studiati. Sono stati messi a punto 3 modelli di classificazione PLS-DA, uno per ciascuna tipologia di sementi analizzata (zucca, cocomero e melanzana), per distinguere le due classi: "0" = seme non germinabile / non germinato (seme non germinato a distanza di un mese dall'acquisizione) e "1" = seme germinabile / germinato (seme germinato a distanza di un mese dall'acquisizione).

Come algoritmo di cross – validazione, per valutare la complessità del modello e scegliere un adeguato numero di variabili latenti, è stato utilizzato l'algoritmo *venetian blinds* ([Eigenvektor Research Inc. 2016b](#)).

Per ciascuno dei tre casi, il modello PLS-DA è stato costruito a partire dal dataset di calibrazione e validato applicandolo sul dataset di validazione.

Analisi dei risultati

Analisi esplorativa

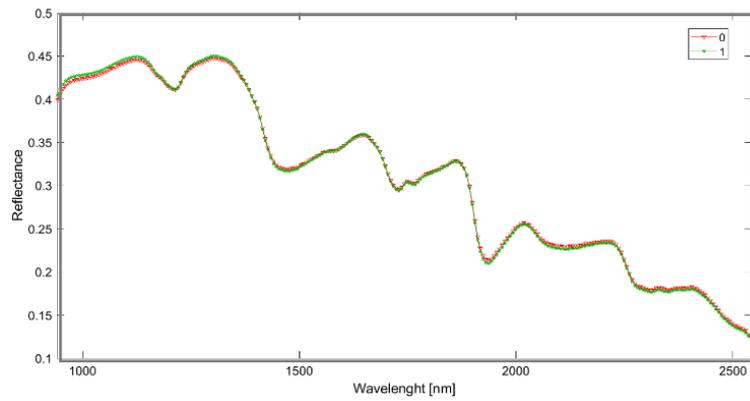
Gli spettri medi in riflettanza dei dataset di calibrazione sono riportati in [Figura 4.2.37](#). Mentre, i risultati della PCA applicata ai dataset di calibrazione sono riportati nella sezione "Allegati" ([Allegato XIII- Case study - Sementi](#)).

Classificazione Partial Least Squares – Discriminant Analysis

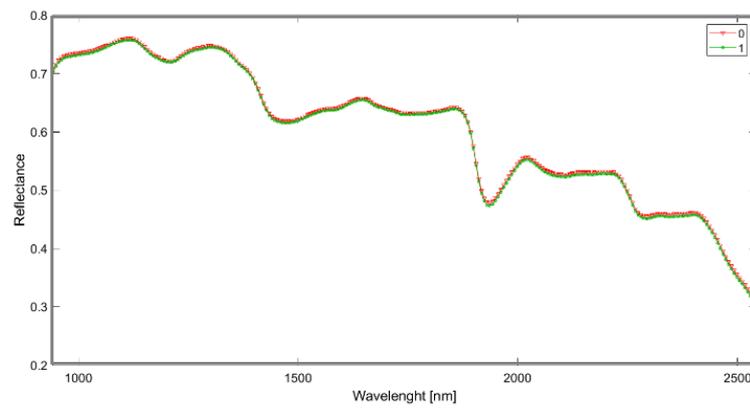
I risultati delle classificazioni PLS-DA in termini di Sensitività e Specificità, per i tre differenti casi, sono riassunti in [Tabella 4.2.14](#). Le mappe di previsione (*pred member*) per i tre modelli *Solanum Melongena* (Melanzana), *Cucurbita Pepo* (Zucca) e *Citrullus Lanatus* (Cocomero), sono riportate rispettivamente in [Figura 4.2.38](#), [Figura 4.2.39](#) e [Figura 4.2.40](#).

Da quanto si può evincere dai risultati di classificazione, tramite PLS-DA si riescono a riconoscere (in predizione): circa 2/4 semi non germinabili nel caso della Melanzana - *Solanum Melongena* ([Figura 4.2.38b](#)), 0 semi non germinabili Zucca - *Cucurbita Pepo* ([Figura 4.2.39b](#)) e la totalità dei semi non germinabili nel caso del Cocomero – *Citrullus Lanatus* ([Figura 4.2.40b](#)).

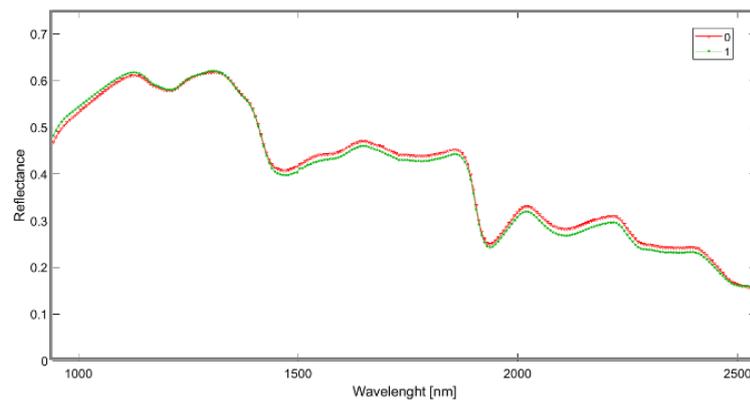
Casi di studio



a



b



c

Figura 4.2.37. Spettri medi di riflettanza delle classi di germinabilità delle sementi analizzate: Melanzana - *Solanum Melongena* (a), Zucca - *Cucurbita Pepo* (b) e Cocomero - *Citrullus Lanatus* (c). Note della legenda: 0 = non germinato, 1= germinato.

Casi di studio

Tabella 4.2.14. Sensitività e Specificità in calibrazione (Cal), cross-validazione (CV) e predizione (Pred) per i modelli PLS-DA, discriminanti le classi di germinabilità, delle tipologie di sementi: *Cucurbita Pepo* (Zucca), *Solanum Melongena* (Melanzana) e *Citrullus Lanatus* (Cocomero). (*) "0" = non germinabile / non germinato (seme non germinato a distanza di un mese dall'acquisizione); "1" = germinabile / germinato (seme germinato a distanza di un mese dall'acquisizione).

| Tipologia sementi | <i>Solanum Melongena</i> | | <i>Cucurbita Pepo</i> | | <i>Citrullus Lanatus</i> | |
|--------------------|--------------------------|-------|-----------------------|-------|--------------------------|-------|
| | (Melanzana) | | (Zucca) | | (Cocomero) | |
| Classe* | (0) | (1) | (0) | (1) | (0) | (1) |
| Sensitività (Cal) | 0.972 | 0.958 | 0.939 | 0.916 | 0.963 | 0.909 |
| Specificità (Cal) | 0.958 | 0.972 | 0.916 | 0.939 | 0.909 | 0.963 |
| Sensitività (CV) | 0.699 | 0.745 | 0.730 | 0.665 | 0.825 | 0.813 |
| Specificità (CV) | 0.745 | 0.699 | 0.665 | 0.730 | 0.813 | 0.825 |
| Sensitività (Pred) | 0.474 | 0.959 | 0.020 | 0.530 | 0.725 | 0.981 |
| Specificità (Pred) | 0.959 | 0.474 | 0.530 | 0.020 | 0.981 | 0.725 |

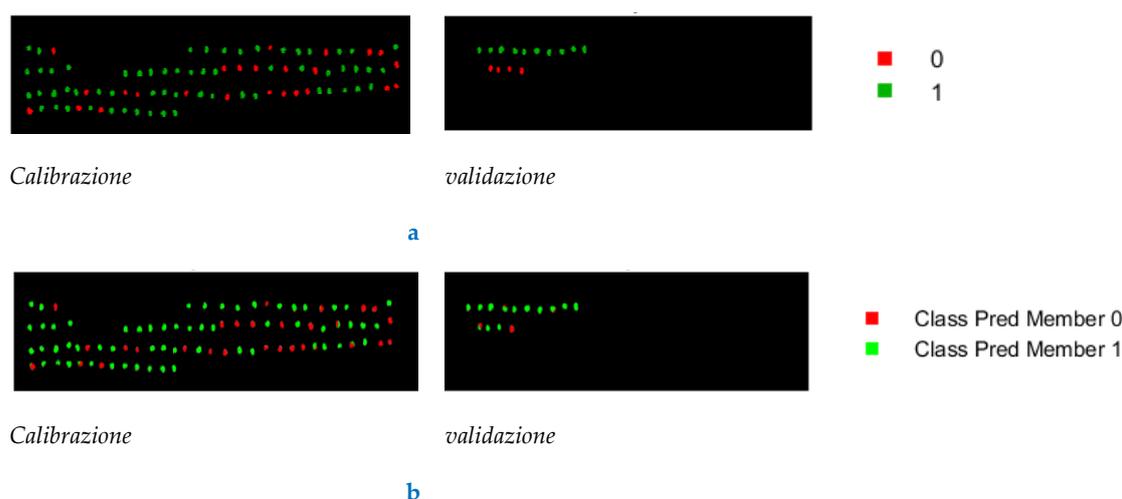


Figura 4.2.38. Set di calibrazione e set di validazione (a) e mappa di predizione (pred member) per Melanzana - *Solanum Melongena* (b), in calibrazione e validazione. Note della legenda: 0 = non germinato, 1= germinato.

Discussione dei risultati

I risultati ottenuti hanno mostrato come sia possibile riconoscere semi non germinabili di *Solanum Melongena* (Melanzana) e *Citrullus Lanatus* (Cocomero) mediante l'utilizzo di tecniche iperspettrali nel range dell'infrarosso ad onde corte (1000-2500 nm). Per quanto riguarda, invece, le sementi non germinabili di *Cucurbita Pepo* (Zucca), queste non vengono riconosciute dal modello di classificazione PLS-DA. In questo caso, emerge la necessità di effettuare ulteriori indagini per la messa a punto di un nuovo modello di classificazione: l'utilizzo di un maggior numero di campioni o di un modello di

Casi di studio

classificazione non lineare (i.e. SVMDA) potrebbero migliorare le performance di classificazione.

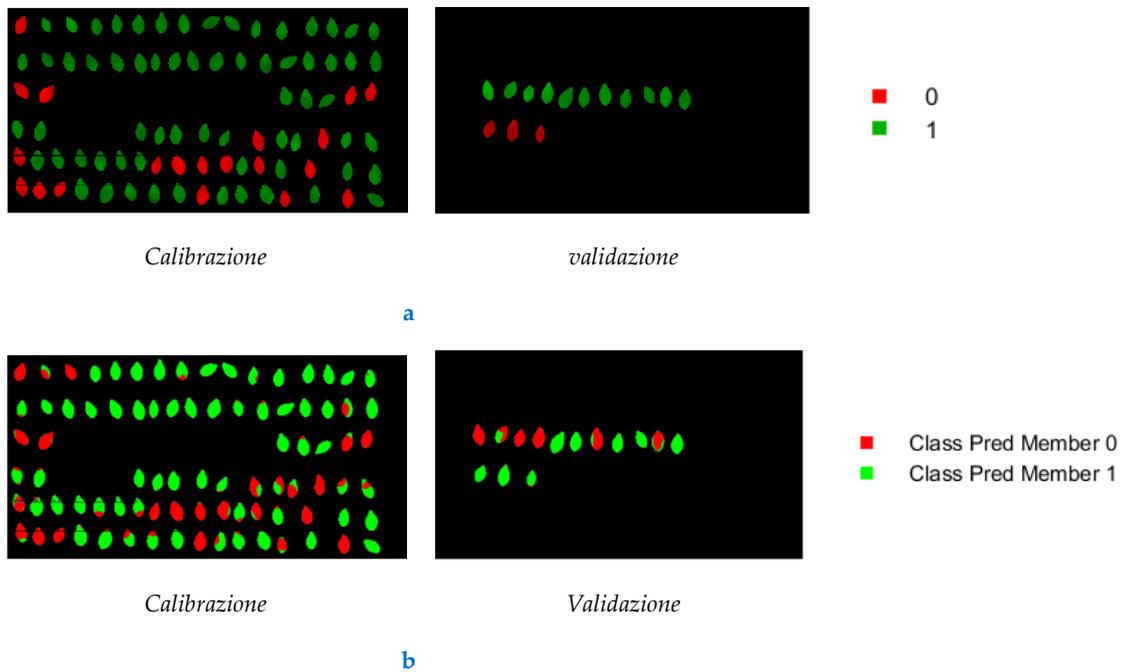


Figura 4.2.39. Set di calibrazione e set di validazione (a) e mappa di predizione (pred member) per Zucca - Cucurbita Pepo (b), in calibrazione e validazione. Note della legenda: 0 = non germinato, 1= germinato.

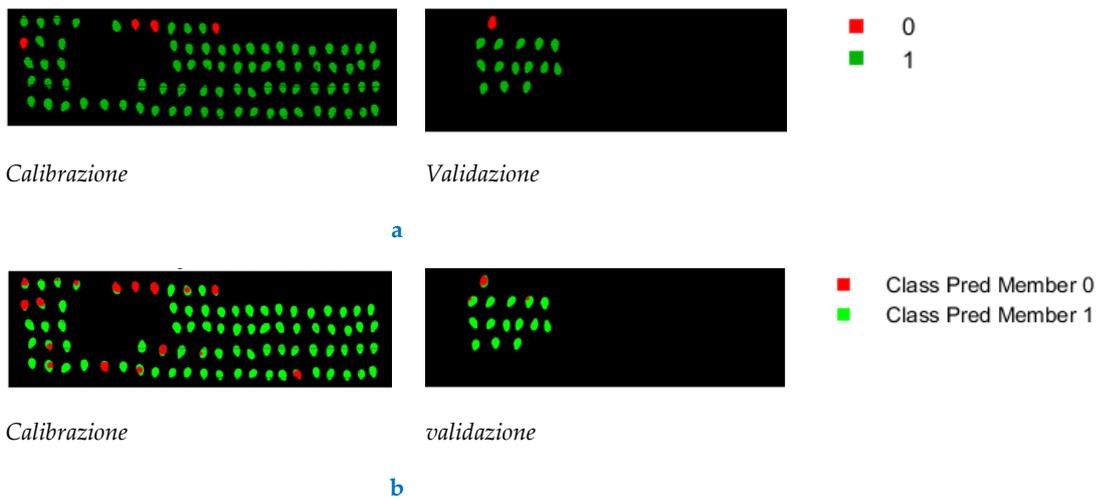


Figura 4.2.40. Set di calibrazione e set di validazione (a) e mappa di predizione (pred member) per Cocomero - Citrullus Lanatus (b), in calibrazione e validazione. Note della legenda: 0 = non germinato, 1= germinato.

4.2.6 Valutazione del processo di maturazione dell'oliva itrana mediante l'utilizzo di un sistema iperspettrale da laboratorio ed uno spettroradiometro portatile

Problematica e stato dell'arte

Gli uliveti rappresentano una delle coltivazioni più diffuse al mondo. I siti più importanti di produzione sono localizzati nell'area mediterranea. L'oliva, frutto dell'ulivo, deve la propria notorietà alle proprie caratteristiche organolettiche e salutistiche (i.e. l'olio di oliva è diventato l'alimento grasso principale della Dieta Mediterranea), tanto è che il consumo di questa drupa è diffuso in tutto il mondo, non solo nelle aree del mediterraneo (*Bonifazi & Serranti, 2011*). La produzione delle olive da tavola è di circa 2 Mt/anno. La produzione dell'Unione Europea è di circa un terzo della produzione globale di olive da tavola. Egitto e Turchia seguono l'Unione Europea per produzione, rispettivamente, con il 20 % ed il 12 % della produzione annuale globale (*IOC, 2009*). Uno degli aspetti cruciali della produzione di olive è la valutazione del processo di invaiatura per le finalità di raccolta-stoccaggio e/o trasformazione-vendita. La valutazione del processo di maturazione dell'oliva è di estrema importanza ai fini della determinazione dei tempi di raccolta; in quanto, il tempo di raccolta è fattore che influenza le caratteristiche organolettiche dell'oliva, nonché le rese in olio (se olive da frangere), le quali a loro volta influenzano il valore commerciale del prodotto finito (*Lorente et al., 2011*).

L'utilizzo della spettroscopia ad infrarosso per determinare alcuni attributi qualitativi delle drupe di olivo risale agli anni '80, ma in tempi recenti questa tecnologia sta trovando sempre più spazio in applicazioni industriali, nei paesi maggiormente sviluppati (*Nicolai et al., 2007*). Negli ultimi anni, sono stati studiati metodi non-distruttivi basati sulla spettroscopia ad infrarosso per la valutazione di alcuni parametri qualitativi fisici e chimici delle olive. Tra questi, vi sono studi incentrati sull'utilizzo di strumenti portatili per la valutazione di parametri qualitativi come il contenuto fenolico, il contenuto di olio ed umidità (*Stella et al., 2015*).

Obiettivi dello studio

Uno degli obiettivi principali di questo studio è stato, dunque, quello di valutare i diversi stadi di maturazione dell'oliva "Itrana" utilizzando tecniche iperspettrali e spettroscopiche al fine di sviluppare un sistema di ispezione/valutazione da supporto al processo decisionale che porta alla raccolta delle olive da tavola. La cultivar di olive "Itrana" è tipica della regione Lazio (Italia) ed è originario della città di Gaeta. I siti di produzione maggiori sono situati nella città di "Itri", una città collinare localizzata in prossimità della costa del Mar Mediterraneo e ad altre città situate sulle catene montuose dei Monti Lepini, Monti Ausoni e dei Monti Aurunci (la cosiddetta catena montuosa dei Monti Volsci, facenti parte dell'Antiappennino Laziale). Le caratteristiche principali di questa tipologia di drupa sono la forma ovoidale irregolare ed il colore tendente al rosso-

viola quando matura. Il periodo di maturazione tipico è tra i mesi di febbraio-marzo. Generalmente, il periodo di raccolta dell'oliva "Itrana" è novembre.

In questo studio, è stato utilizzato un sistema HSI ed uno spettroradiometro portatile per la valutazione del processo di invaiatura dell'oliva "Itrana". Grazie all'utilizzo di alcune tecniche chemiometriche sono stati messi a punto dei modelli in grado di riconoscere lo stato di maturazione (i.e. tempo di raccolta). Tali tecniche consentirebbero una valutazione speditiva sullo stato di maturazione delle olive, fornendo un nuovo metodo di monitoraggio delle coltivazioni (i.e. grazie all'utilizzo di strumentazione portatile, è possibile monitorare i frutti quando sono ancora sull'albero). Inoltre, siccome la qualità delle olive dipende fortemente dal tempo di raccolta, l'agricoltore ne potrà trarre benefici economici, assicurando l'immissione in mercato di un prodotto finale di qualità, destinato alla tavola del consumatore.

Si pensi, inoltre, che l'acquisizione di spettri in riflettanza con spettrofotometri è più veloce rispetto ad una scansione in logica *push-broom* con spettrografo da laboratorio; in aggiunta, le analisi effettuate con i dati acquisiti con spettrofotometri possono essere affidabili come quelle di uno spettrografo da banco ai fini della caratterizzazione. Il controllo di qualità di grandi quantità di olive potrebbe essere effettuato, quindi, anche senza effettuare analisi in laboratorio, ma direttamente su campo. Queste tecniche non distruttive potrebbero migliorare il processo decisionale relativamente al processo di raccolta.

L'affidabilità e l'utilità di uno spettrofotometro portatile, operante nel Vis-SWIR, sono state messe a confronto con un sistema HSI da laboratorio, al fine di identificare gli stadi di invaiatura dell'oliva itrana. Per fare ciò, è stato necessario l'utilizzo di tecniche chemiometriche. La PCA è stata effettuata a scopo esplorativo. Una regressione MLR è stata messa a punto per valutare la correlazione tra l'indice di maturazione, valutato per l'insieme di drupe (in base alle località di provenienza) a ciascun tempo di raccolta, ed il tempo di raccolta stesso. La tecnica di regressione PLS è stata invece utilizzata per cercare di prevedere lo stato di invaiatura delle drupe oggetto di studio, a partire dagli spettri in riflettanza estrapolati, utilizzando l'indice di maturazione (MI - *Maturity Index*) ed i tempi di raccolta. In [Figura 4.2.41](#) è stata schematizzata la procedura messa a punto per impostare le regressioni PLS, con i dati spettrali estrapolati da immagini iperspettrali e dagli spettri acquisiti con spettrofotometro portatile.

Procedura

Preparazione dei campioni

Le drupe di ulivo oggetto di studio appartengono alla cultivar "Itrana", detta anche "Trana" o "Esperia". La cultivar dell'oliva "Itrana" è tipica del Lazio (Italia). I frutti di ulivo sono stati raccolti in tre differenti zone dei Monti Volsci (formazione composta da tre catene montuose: Monti Lepini al nord, Monti Ausoni al centro e Monti Aurunci a sud):

- Sermoneta (Monti Lepini),

Casi di studio

- Sonnino (Monti Ausoni),
- Gaeta (Monti Aurunci).

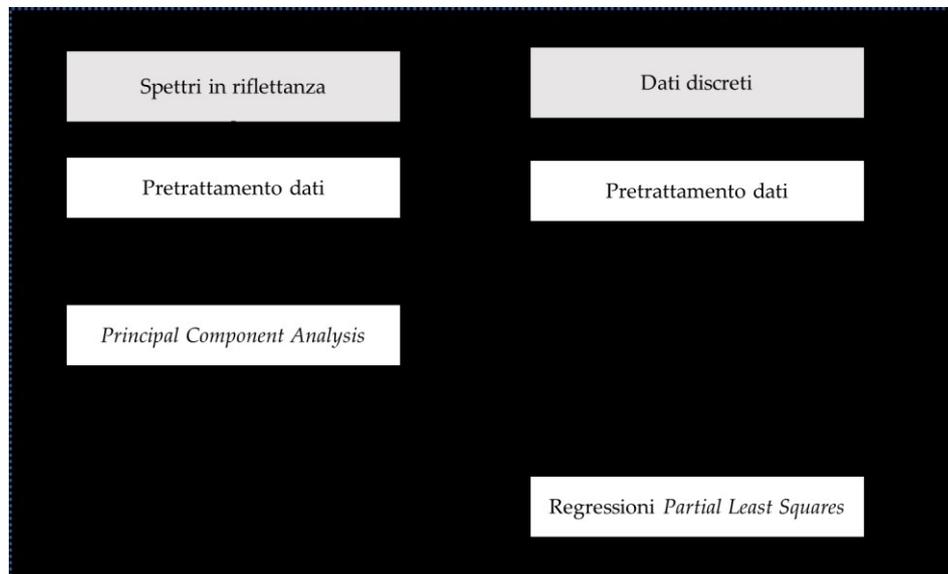


Figura 4.2.41. Diagramma di flusso di alcune delle analisi effettuate.

Il campionamento delle drupe è stato ripetuto regolarmente una volta a settimana per 8 settimane (Tempo di raccolta/*Time of harvest*: Time 0, Time 1, Time 2, Time 3, Time 4, Time 5, Time 6 e Time 7), per ciascuna zona. Per ciascun tempo e ciascuna zona, le olive sono state selezionate manualmente in base alla dimensione, valutate mediante ispezione visiva: piccole, medie e grandi. Le olive più grandi sono state disposte in capsule Petri (Figura 4.2.42) per poi essere analizzate.

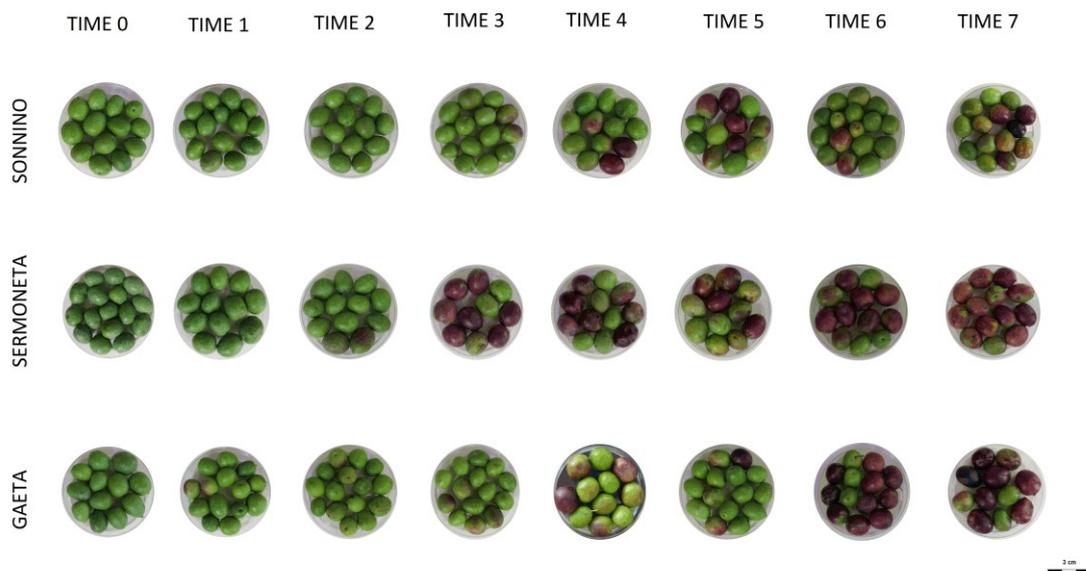


Figura 4.2.42. Campioni analizzati di olive Itrane, suddivisi per tempo di raccolta e località di provenienza.

Dapprima, è stato valutato l'indice di maturazione con uno dei metodi colorimetrici classici (metodo visivo) per mappare il processo di maturazione, successivamente sono state effettuate le analisi iperspettrali. In entrambi i casi, i dati raccolti sono stati analizzati statisticamente per validare e verificare la correttezza delle misurazioni effettuate.

Valutazione dell'indice di maturazione

L'indice di maturazione dell'oliva è un parametro che fornisce informazioni per la scelta del tempo di raccolta, dato che non tutti i frutti in una coltivazione maturano insieme (*The Olive Oil Source, 2017a; Ferguson & Sibett, 1994*). La determinazione dell'indice di maturazione è basata sull'ispezione visiva del colore delle drupe (*UC Cooperative Extension, 2006*). I frutti sono selezionati in base al colore e separati secondo il colore dell'esocarpo, per ciascun colore viene associato un punteggio (come mostrato in [Tabella 4.2.15](#)). Il numero di frutti appartenenti a ciascuna classe viene moltiplicato per il punteggio relativo alla classe di colore appartenente (da 0 a 7). L'indice di maturazione è calcolato in base alla formula:

$$MI = \frac{(1 \times N_1 + 2 \times N_2 + 3 \times N_3 + 4 \times N_4 + 5 \times N_5 + 6 \times N_6 + 7 \times N_7) \times N_{FRUITS}}{100} \quad (4.2.2)$$

Dove N_i è il numero delle olive ed N_{FRUITS} è il numero totale di drupe considerate. La seguente procedura è stata seguita per le olive campionate per ciascun tempo di raccolta.

Secondo recenti studi, l'indice di maturazione MI dei frutti può essere valutato anche utilizzando tecniche di analisi di immagine, applicate a tecniche di Machine Vision, basate su algoritmi di segmentazione sul colore (*Guzmán, 2015*).

Per valutare la correlazione esistente tra il tempo di raccolta ed indice di maturazione è stata calibrata una retta di regressione con il metodo MLR. La tecnica MLR ricerca di modellare la relazione tra le variabili esplicative e la risposta cercando di "fittare" i dati osservati Y con una equazione lineare (*Yale University, 1997; Wise et al., 2006*). Gli indici di maturazione valutati a seconda dei tempi sono stati organizzati all'interno di un dataset ed utilizzati come X . Mentre, il dataset contenente le informazioni sui tempi di raccolta ($T_0 - T_7$) è stato usato come Y . I dataset sono stati suddivisi in due parti: il 70 % è stato usato come calibrazione del modello, mentre il 30 % è stato utilizzato per la validazione. Il metodo di convalida incrociata utilizzato in questo caso è il *leave one-out* (*Eigenvector Research Inc. 2016b*). Questa procedura effettuata per valutare la correlazione tra i tempi di raccolta e l'indice di maturazione è stata messa a punto in ambiente MATLAB (MATLAB R2016b ver. 9.1.0.441655; The MathWorks, Inc., Natick, MA, USA), con l'aiuto del *toolbox* PLS_toolbox (ver. 8.2.1) della Eigenvector Research, Inc.

Tabella 4.2.15. Punteggi dell'indice di maturazione.

| <i>Colore dell'oliva</i> | Verde intenso | Giallo o giallo-verde | Giallo-verde con più della metà dell'oliva con macchie rossastre o violacee | Più della metà dell'oliva di colore rosso o viola | Viola leggero tendente al nero con macchie verdi chiare | Nero e viola fino quasi ad arrivare al nocciolo | Nero e colore scuro fino al nocciolo |
|--------------------------|---------------|-----------------------|---|---|---|---|--------------------------------------|
| <i>Punteggio</i> | 0 | 1 | 2 | 3 | 4 | 6 | 7 |
| <i>Numero di olive</i> | NO | N1 | N2 | N3 | N4 | N6 | N7 |

Acquisizione di spettri in riflettanza con spettrofotoradiometro portatile

Per ciascuna oliva analizzata sono stati acquisiti 3 spettri in riflettanza con spettrofotoradiometro portatile ASD FieldSpec 4® Standard – Res mediante l'ausilio di una sonda a contatto (ASD Inc. 2015b). L'acquisizione dei dati è stata effettuata tramite il software RS³ (ASD Inc. 2008).

Esportazione e preparazione dati

Le immagini iperspettrali acquisite con il sistema SisuCHEMA XL (Specim, Spectral Imaging Ltd. 2019) sono state convertite in file con estensione .mat, tramite interfaccia software con ChemaDAQ, per essere poi importate in ambiente MATLAB (MATLAB R2016b ver. 9.1.0.441655). Le immagini acquisite sono state successivamente concatenate tra loro per creare un unico ipercubo, grazie all'ausilio del toolbox MIA_toolbox (Ver. 3.0.3); le successive elaborazioni del dataset sono state effettuate con l'ausilio dei toolbox MIA_toolbox (Ver. 3.0.3) e PLS_toolbox (ver. 8.2.1). A partire dall'ipercubo di dati sono state selezionate delle ROI (Region of Interest) per ciascun tempo di raccolta, convertite successivamente in spettri ai fini dell'analisi esplorativa e per la calibrazione del modello di regressione. Mentre, gli spettri acquisiti con software RS³ in formato .asd sono stati esportati in file di testo ASCII (American Standard Code for Information Interchange, Codice Standard Americano per lo Scambio di Informazioni), .txt, grazie al software ViewSpec Pro (Ver. 6.2.0.). I file così ottenuti sono stati successivamente importati in ambiente MATLAB come "dataset object" utilizzando una routine (script MATLAB), scritta appunto per lo scopo: fieldspec_import2.m. I file importati in ambiente MATLAB sono stati poi analizzati utilizzando il toolbox PLS_toolbox (ver. 8.2.1).

Le classi relative alla provenienza dei campioni e al tempo di raccolta sono state settate tramite l'apposita interfaccia per ciascun dataset spettrale; i dati necessari per le successive analisi (i.e. tempi di raccolta ed i punteggi dell'MI) sono stati anch'essi importati in ambiente MATLAB.

Pre-elaborazione dati

L'intervallo di lunghezza d'onda considerato dell'ipercubo generato a partire dai dati del SisuCHEMA XL (1000 - 2540 nm) è stato ridotto a 1000 - 2500 nm. La stessa procedura è stata applicata agli spettri acquisiti con lo spettrometro portatile, ai fini della rimozione delle parti più rumorose dovute ai detector. Perciò, il dataset spettrale del FieldSpec 4 (350 - 2500 nm) è stato ridotto a 400 - 2500 nm. A partire dai due dataset, sono stati scelti 100 spettri per ciascuno: 100 spettri dalle ROI estratte dal dataset del SisuCHEMA XL™ e 100 spettri a partire dal dataset del FieldSpec 4.

Prima di passare all'analisi esplorativa è stato necessario applicare una combinazione di algoritmi di *pre-processing*. La pre-elaborazione dei dati è stata necessaria non solo per correggere effetti di *scattering* e/o altri effetti rumorosi, ma anche per esaltare le differenze spettrali tra i diversi campioni. Per entrambi i nuovi dataset, sono state valutate differenti strategie di pre-elaborazione dei dati, in differenti combinazioni. Infine, è stata scelta la combinazione di *pre-processing* che ha fornito una risposta migliore in termini di decomposizione dei dati. La combinazione di *pre-processing* utilizzata è stata: *Standard Normal Variate* (SNV) per la correzione di effetti di *scattering* (Rinnan et al., 2009), *Generalized Least Squares Weighting* (GLS-W, applicato sulle classi settate, con $\alpha = 0.001$) per esaltare le caratteristiche inter-classe (Eigenvector Research Inc. 2013b) e *Mean Center* (MC), per centrare le colonne dei dati rispetto ad uno zero (Wise et al., 2006; Wise et al., 2008).

Analisi esplorativa dei dati

L'analisi in componenti principali (o *Principal Component Analysis*, PCA) è stata effettuata sui dataset di calibrazione per esplorare la variabilità intra-classe ed inter-classe, secondo i tempi di raccolta e gli indici di maturazione dei rispettivi campioni (Wold et al., 1987).

I dati sono stati dapprima pre-elaborati e successivamente sono state effettuate analisi in componenti principali. Sono state effettuate due PCA, una per il set di dati estratti da ROI relativi al SisuCHEMA XL™ ed una con i dati del FieldSpec 4. Il numero di componenti principali, o *Principal Component* (PC), è stato scelto esplorando i grafici degli autovalori. Sono state scelte 5 componenti principali per i dati del SisuCHEMA XL. Mentre, per i dati del FieldSpec 4 sono state scelte 4 componenti principali. Gli *outlier* sono stati identificati ed esclusi dalle successive analisi, esplorando il grafico dei parametri Hotelling T² e i residui Q.

Modelli di regressione

Il metodo di regressione PLS (Geladi et al., 1986; Wold et al., 2001; Abdi, 2003) è stato utilizzato per valutare la correlazione tra gli spettri Vis-SWIR (Dataset X) ed i tempi di maturazione, nonché dei punteggi relativi all'indice di maturazione. Sono stati, perciò, messi a punto due modelli di regressione per ciascuna tipologia di dato (due regressioni per i dati del SisuCHEMA XL e due regressioni per i dati del FieldSpec 4).

In primo luogo, ciascun dataset è stato suddiviso in 2 parti: 70 % è stato utilizzato per calibrare il modello, mentre il restante 30 % è stato utilizzato come set di validazione. La

suddivisione dei dataset è stata effettuata con l'algoritmo *Kennard-Stone* (*Kennard & Stone, 1969*). In secondo luogo, i modelli sono stati calibrati e cross-validati mediante l'utilizzo dell'algoritmo "*random sub-sets*" (*Eigenvector Research Inc. 2016b*) per valutare la complessità ottimale del modello e scegliere il numero di variabili latenti (LV). Per ciascun modello sono state scelte 4 variabili latenti. Infine, i modelli sono stati testati con l'utilizzo dei set di validazione.

Nei primi due modelli sono stati utilizzati i valori del tempo di raccolta (i.e. Time 0 - Time 7: valori numerici da 0 a 7) come **Y**, mentre come **X** sono stati utilizzati rispettivamente gli spettri in riflettanza estratti da ROI del SisuCHEMA XL e gli spettri del FieldSpec 4. Negli altri due modelli sono stati utilizzati i punteggi dell'indice di maturazione (da 0 a 6) come **Y**, mentre come **X** sono stati utilizzati rispettivamente gli spettri in riflettanza estratti da ROI del SisuCHEMA XL e gli spettri del FieldSpec 4.

Analisi dei risultati

Correlazione tra indice di maturazione e tempo di raccolta

I punteggi ottenuti relativi agli indici di maturazione, ottenuti mediante analisi visiva effettuata sull'insieme di campioni per ciascun tempo di raccolta a seconda delle località di provenienza, sono riportati nella sezione "Allegati" (*Allegato XIV- Case study – Olive: inoaiatura*).

Dai risultati ottenuti dalla regressione MLR, effettuata per le variabili indice di maturazione (MI) e tempo di raccolta, si evince che l'MI sia strettamente correlato al tempo di raccolta: con $R^2= 0.844$ in predizione e $RMSEP = 1.435$ MI (si veda *Allegato XIV- Case study – Olive: inoaiatura*).

Spettri medi in riflettanza

In [Figura 4.2.43a](#) è mostrato l'ipercubo, composto dalle immagini iperspettrali concatenate acquisite con il sistema HSI SisuCHEMA XL™. Mentre, in [Figura 4.2.43b](#) ed in [Figura 4.2.43c](#) è riportato l'ipercubo con le classi settate rispettivamente riferite alla località di provenienza – tempo di raccolta e al tempo di raccolta.

Casi di studio

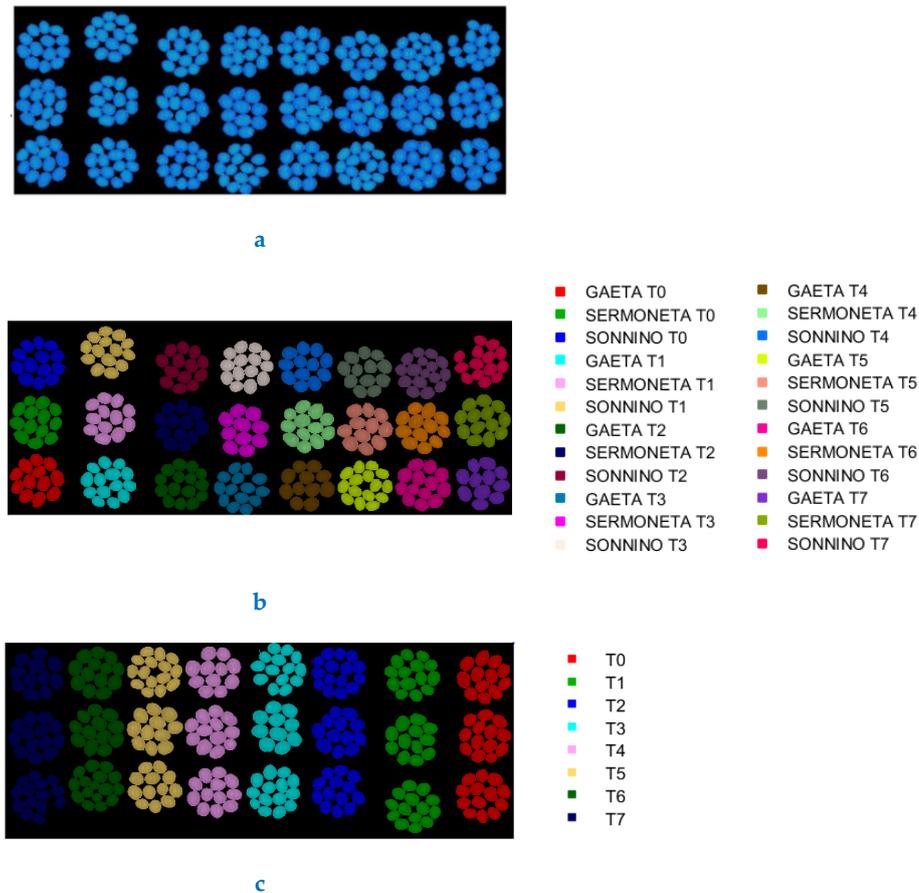


Figura 4.2.43. Immagini iperspettrali concatenate insieme relative ad i campioni di olive itrana: *a)* ipercubo grezzo; *b)* ipercubo con classi relative alla località di provenienza ed ai tempi di raccolta e *c)* ipercubo con classi relative al tempo di raccolta.

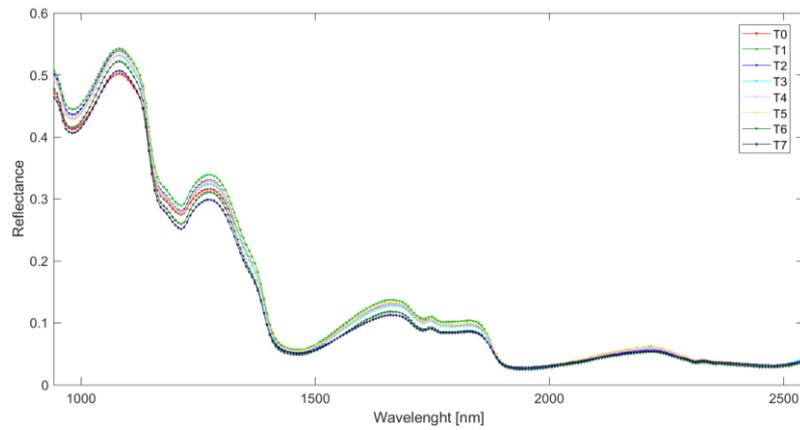
Gli spettri medi in riflettanza relativi alle ROI selezionate sull'ipercubo delle immagini acquisite con SisuCHEMA XL™ sono riportati in Figura 4.2.44. Mentre, in Figura 4.2.45 sono riportati gli spettri medi in riflettanza Vis-SWIR, acquisiti con spettrometro, con riferimento alle classi relative al tempo di raccolta e ai punteggi dell'indice di maturazione.

Da come si può notare in Figura 4.2.45a e Figura 4.2.45b, la zona di maggiore variazione morfologica degli spettri è inclusa nella zona del Visibile (350-750 nm): ciò rispecchia la variazione di colore delle drupe con il tempo di raccolta e dell'indice di maturazione.

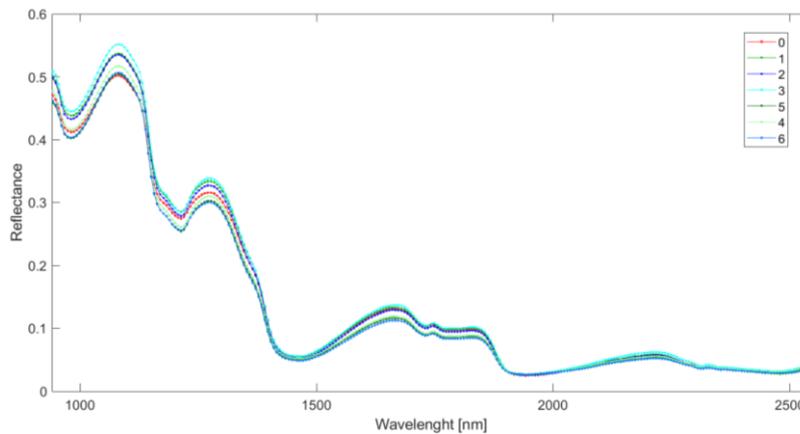
Analisi esplorativa

Ciò che si evince dalle analisi in componenti principali è il raggruppamento degli *scores* in base delle classi riferite ai tempi di raccolta ed ai punteggi MI. Questo avviene sia per le PCA effettuate sui dati del SisuCHEMA XL che per quelle effettuate con i dati del FieldSpec 4.

Casi di studio



a

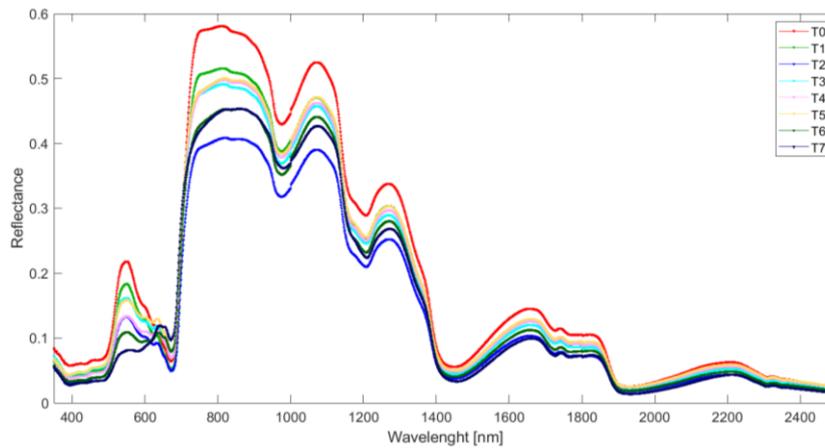


b

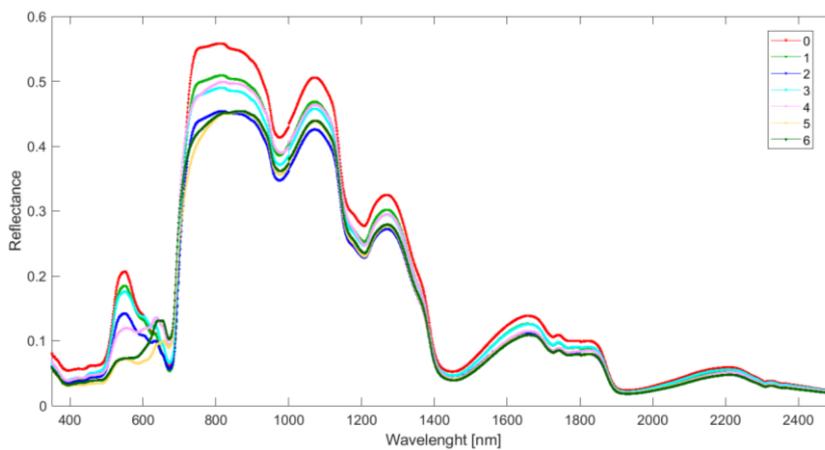
Figura 4.2.44. Spettri medi delle drupe analizzate relativi alle ROI estratte dall'ipercubo del SisuCHEMA XL™ (1000-2500 nm) a differenti tempi di raccolta (a) e a differenti indici di maturazione (b).

Ad ogni modo, diversamente da quanto avviene nel caso della PCA effettuata con i dati del SisuCHEMA XL relativamente alle classi di punteggio MI, gli scores della PCA effettuata con i dati del FieldSpec (400-2500 nm) relativi alle classi di punteggio MI, appaiono maggiormente raggruppati (si veda *Allegato XIV- Case study – Olive: invaiatura*).

Casi di studio



a



b

Figura 4.2.45. Spettri medi in riflettanza Vis-SWIR (350-2500 nm), acquisiti con spettrometro FieldSpec 4, delle classi relative al tempo di raccolta (a) ed ai punteggi dell'indice di maturazione (b).

Regressioni Partial Least Squares

I risultati in termini di fitting e RMSE delle regressioni PLS, effettuate per la previsione dei tempi di raccolta e dell'indice di maturazione, sono riportati in [Tabella 4.2.16](#); mentre, i relativi *scatter plot* sono riportati in [Figura 4.2.46](#). Le migliori performance per quanto riguarda i modelli di regressioni PLS sono state ottenute con i dati del FieldSpec 4 per prevedere i tempi di raccolta, ottenendo un $R^2_P = 0.912$ ed un $RMSEP = 0.674$ settimane (circa 5 giorni); invece, per quanto riguarda la regressione PLS effettuata con i dati del SisuCHEMA per prevedere i tempi di raccolta è stato ottenuto un R^2_P pari a 0.682 e un $RMSEP = 1.419$ (circa 10 giorni). Le regressioni effettuate per prevedere l'indice di maturazione si attestano su valori di fitting più bassi: per quanto riguarda i valori del FieldSpec 4 sono stati ottenuti un $R^2_P = 0.665$ ed un $RMSEP = 0.923$ MI; mentre, per quanto riguarda i dati del SisuCHEMA XLTM sono stati ottenuti un $R^2_P = 0.461$ ed un $RMSEP = 1.348$ MI. Il fatto che i valori di fitting siano generalmente più elevati nel caso delle regressioni effettuate con i dati del FieldSpec 4 potrebbe risiedere nell'utilizzo del più

Casi di studio

ampio intervallo spettrale considerato (350 - 1000 nm); basti pensare che, nel caso del SisuCHEMA, il range spettrale investigato spazia tra 1000 e 2500 nm ed in questo caso non è incluso il visibile. Inoltre, il punteggio assegnato durante la valutazione dell'indice di maturazione si basa su una analisi visiva soggettiva.

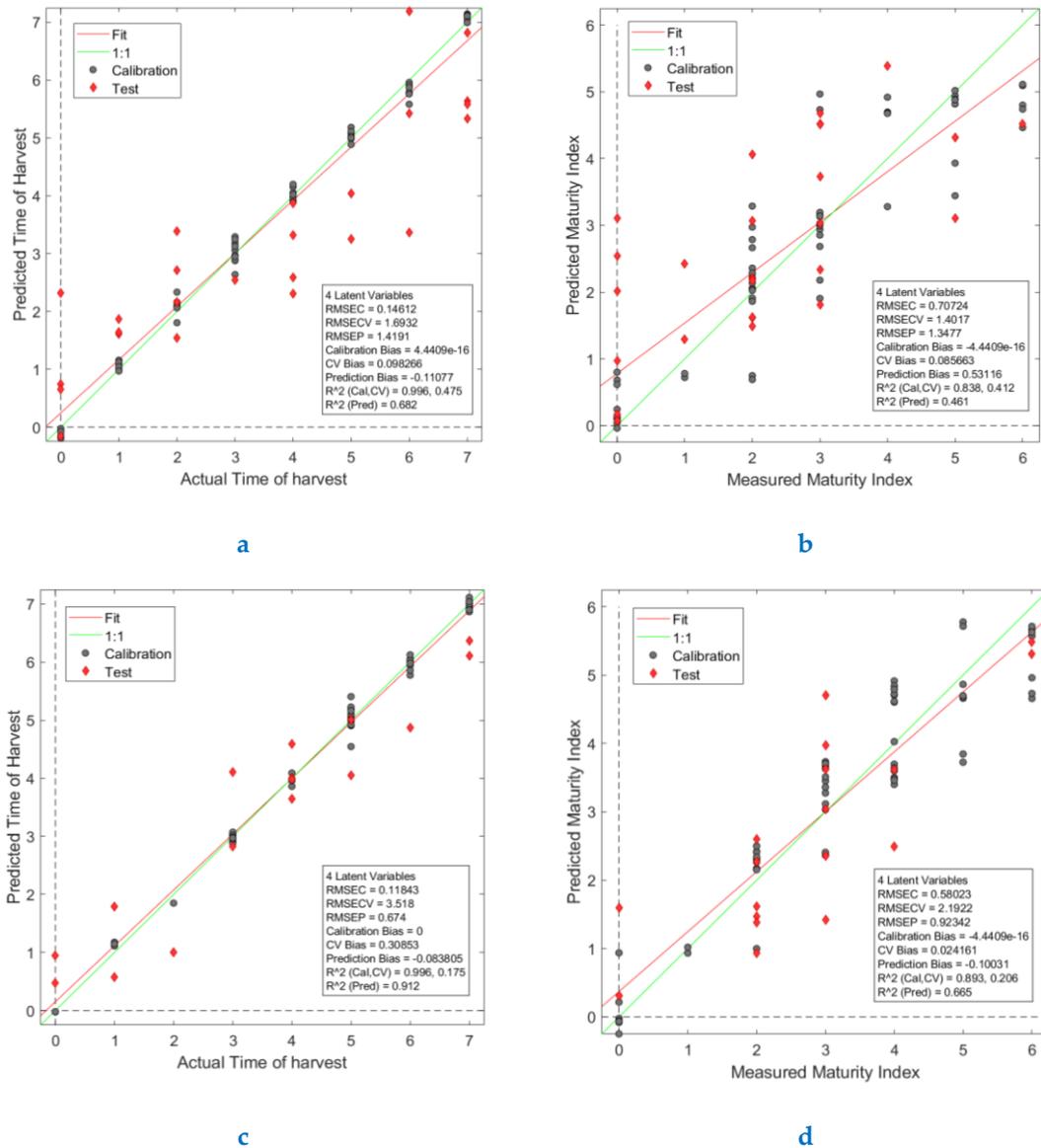


Figura 4.2.46. Scatter plot (valori stimati vs valori reali) dei modelli di regressione, calcolati sui dati del SisuCHEMA XL (1000 – 2500 nm), per prevedere i tempi di raccolta (a) e l'indice di maturazione (b), e scatter plot dei modelli di regressione, calcolati sui dati del FieldSpec 4 (400 – 2500 nm), per prevedere: i tempi di raccolta (c) e l'indice di maturazione (d).

Discussione dei risultati

In questo studio è stata valutata la variabilità degli spettri in riflettanza delle drupe di ulivo della cultivar itrana (Vis-SWIR e SWIR) per diversi tempi di raccolta. È stato dimostrato che esiste una forte correlazione tra i tempi di raccolta e l'indice di

maturazione dei campioni di olive analizzati. La regressione MLR per l'MI mostra una capacità accettabile di predire il tempo di raccolta, con un R^2 in predizione pari a 0.844 ed un errore quadratico medio in Predizione (RMSEP) pari a 1.435 MI. È stato quindi eseguito un confronto in termini di capacità predittiva dello stato di invaiatura della drupa di ulivo tra uno spettrometro portatile Vis-SWIR ed un sistema iperspettrale da laboratorio operante nello SWIR. I risultati delle regressioni PLS applicate per la previsione del tempo di raccolta mostrano che tali strumenti sono comparabili. I risultati di calibrazione e previsione PLS, calcolati sui dati spettrali, hanno mostrato che esiste una certa correlazione tra gli spettri dei campioni di oliva ed il tempo di raccolta; inoltre, esiste correlazione anche tra gli spettri dei campioni di oliva e i punteggi dell'MI (anche se più debole di quella esistente tra spettri e tempi di raccolta), relativi a ciascun campione.

Tabella 4.2.16. Risultati delle regressioni PLS effettuate per la previsione dei tempi di raccolta e dell'indice di maturazione.

| Dataset | Intervallo spettrale [nm] | Applicazione | LV | R^2_{Cal} | R^2_{cv} | R^2_P | RMSEC | RMSECV | RMSEP |
|-----------------|---------------------------------|----------------------|----|-------------|------------|---------|-------|--------|-------|
| SisuCHEMA XL | 1000 – 2500 | tempi di raccolta | 4 | 0.996 | 0.475 | 0.682 | 0.146 | 1.693 | 1.419 |
| SisuCHEMA XL | 1000 – 2500 | MI | 4 | 0.838 | 0.412 | 0.461 | 0.707 | 1.402 | 1.347 |
| FieldSpec 4 | 400 – 2500 | tempi di raccolta | 4 | 0.996 | 0.175 | 0.912 | 0.118 | 3.518 | 0.674 |
| FieldSpec 4 | 400 – 2500 | MI | 4 | 0.893 | 0.206 | 0.665 | 0.58 | 2.192 | 0.923 |

Seguendo questo approccio basato su tecnologie iperspettrali nel Vis-SWIR potrebbe essere possibile definire strategie innovative che siano da supporto al produttore per fornire informazioni circa lo stato di invaiatura della drupa dell'ulivo. Tali strategie, se scalate a livello aziendale – industriale, potrebbero inserirsi come strumenti decisionali finalizzati alla scelta del periodo di raccolta e garantire, quindi, un maggior ritorno economico al produttore. Tant'è che i risultati ottenuti mostrano come, utilizzando uno spettrofotometro Vis-SWIR portatile, l'indice di maturità (MI) potrebbe essere valutato in modo più rapido rispetto ai metodi tradizionali. L'utilizzo di uno spettrofotometro portatile, affiancato all'uso di tecniche chemiometriche, consentirebbe quindi il monitoraggio dei frutti d'ulivo durante l'invaiatura, anche qualora i frutti si trovassero ancora sull'albero.

4.2.7 Identificazione di olive attaccate da mosca olearia mediante l'utilizzo di un sistema iperspettrale da laboratorio ed uno spettroradiometro portatile

Problematica e stato dell'arte

Gli uliveti rappresentano una delle coltivazioni più diffuse al mondo. I siti più importanti di produzione sono localizzati nell'area mediterranea. L'oliva, frutto dell'ulivo, deve la propria notorietà alle proprie caratteristiche organolettiche e salutistiche (Bonifazi & Serranti, 2011), tanto è che il consumo di questa drupa è diffusa in tutto il mondo non solo nelle aree del Mediterraneo. La produzione delle olive da tavola è di circa 2 Mt/anno. La produzione dell'Unione Europea è di circa un terzo della produzione globale di olive da tavola. Egitto e Turchia seguono l'Unione Europea per produzione, rispettivamente, con il 20 % ed il 12 % della produzione annuale globale (IOC, 2009). Uno degli aspetti cruciali per la produzione di olive può essere considerato quello relativo al controllo qualità della drupa, ad esempio la valutazione dei danni da grandinata (Moschetti et al., 2016) e dell'attacco di mosche olearie e/o altre patologie (Moschetti et al., 2015).

La mosca delle olive rappresenta uno dei pericoli più gravi per le cultivar situate nelle regioni in cui tale insetto trova il suo habitat naturale. Tale insetto, tramite deposizione di uova all'interno della drupa, tende ad influenzare sia la qualità che la quantità della produzione di olive. Una seria minaccia, questa, per gli olivicoltori ed i produttori agricoli. L'infestazione da mosca dell'olivo (*Bactrocera oleae*) è considerata la calamità più dannosa che possa colpire le cultivar di ulivo nell'Europa meridionale, nel Nord Africa, nel Medio Oriente e nel Nord America. Le mosche delle olive possono sopravvivere in climi freddi, ma in genere preferiscono regioni calde e secche (i.e. Grecia, Italia, Spagna, Messico e California). La temperatura ottimale per lo sviluppo della mosca della dell'olivo è tra 20 ° e 30 ° C (68 ° - 86 ° F). Questi insetti in genere infestano gli ulivi nelle aree urbane, i quali non vengono mantenuti attivamente. Questi alberi, situati in zone urbanizzate, possono essere considerati un'importante fonte di infestazione per gli uliveti commerciali (Olive Oil Source, 2017b).

Quando un frutto di ulivo viene attaccato dalla mosca, le larve tendono a proliferare all'interno della drupa e quindi a influenzare le proprietà organolettiche della stessa (Pariante et al., 2018). Il prodotto finale deve essere un prodotto di qualità, che dovrebbe andare a finire sul tavolo dei consumatori. Da qui la necessità di istituire ed utilizzare un metodo semplice, non invasivo ed affidabile per riconoscere le larve delle mosche da oliva. Infatti, l'identificazione delle larve nella fase iniziale dell'incubazione e la valutazione del loro sviluppo all'interno delle drupe sono punti di studio particolarmente importanti nel settore olivicolo-oleario (Moschetti et al., 2015).

La sfida, quindi, è quella di utilizzare tecniche di spettroscopia ad infrarosso per mettere a punto un metodo affidabile per il controllo di qualità delle olive direttamente

sul campo e per estenderlo ad azioni di controllo di qualità "on-line" da applicare nella catena di valorizzazione dei prodotti raccolti.

Obiettivi dello studio

Lo scopo principale di questo studio è stato quello di identificare le olive attaccate da mosca olearia appartenenti alla cultivar "Itrana" provenienti da tre differenti località di provenienza, tenendo conto della variabilità della drupa a seconda dello stato di invaiatura. In questo studio, l'attenzione si è focalizzata sulle tematiche concernenti lo sviluppo di un sistema di monitoraggio delle olive, basato su tecniche di spettroscopia convenzionale e di *HyperSpectral Imaging* nell'infrarosso, finalizzato a fornire importanti informazioni nel processo decisionale di raccolta-macinazione.

Per perseguire questo scopo sono state utilizzate diverse tecniche chemometriche. Più nel dettaglio, la *Principal Component Analysis* (PCA) è stata scelta per eseguire analisi esplorativa dei dati e la *Partial Least Squares - Discriminant Analysis* (PLS-DA) è stata scelta con il fine di classificare i frutti di ulivo attaccati o non attaccati dalla mosca. Lo schema delle analisi effettuate è riportato in [Figura 4.2.47](#). L'approccio proposto sembra essere abbastanza efficiente sia per eseguire una "diagnosi precoce" sullo stato di salute della drupa in campo sia per essere utilizzato all'interno di un impianto di trasformazione per monitorare la qualità dei prodotti che devono essere trasformati.

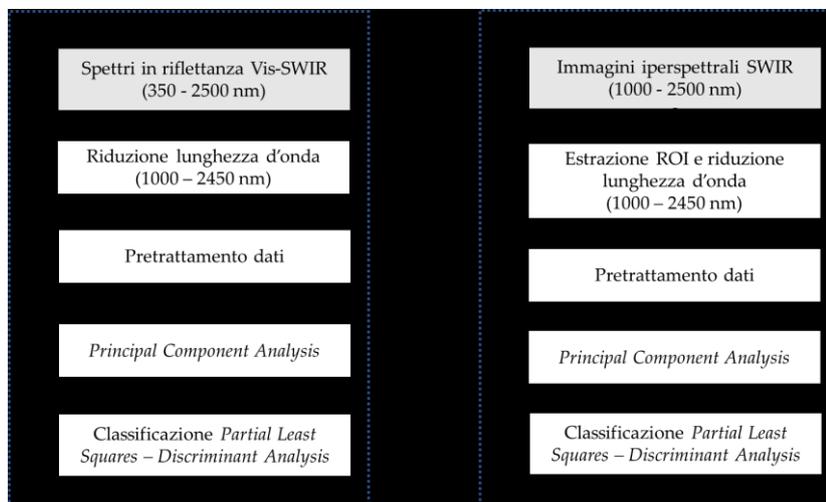


Figura 4.2.47. Diagramma di flusso delle analisi effettuate. La procedura effettuata per mettere a punto la classificazione è inscritta all'interno del quadrato tratteggiato in blu.

Procedura

Preparazione dei campioni

Le drupe di ulivo, oggetto di studio, appartengono alla cultivar "Itrana", detta anche "Trana" o "Esperia". La cultivar dell'oliva "Itrana" è tipica del Lazio (Italia). Le

caratteristiche principali di questa tipologia di drupa sono la forma ovoidale irregolare ed il colore tendente al rosso-viola al momento della maturazione completa. Il periodo di maturazione tipico è tra i mesi di febbraio-marzo. Generalmente, il periodo di raccolta dell'oliva "Itrana" è novembre.

I frutti di ulivo sono stati raccolti in tre differenti zone dei Monti Volsci (formazione composta da tre catene montuose: Monti Lepini al nord, Monti Ausoni al centro e Monti Aurunci a sud):

- I. Sermoneta (Monti Lepini),
- II. Sonnino (Monti Ausoni),
- III. Gaeta (Monti Aurunci).

Il campionamento delle drupe è stato ripetuto regolarmente una volta a settimana per 8 settimane (Tempo di raccolta/*Time of harvest*: *Time 0, Time 1, Time 2, Time 3, Time 4, Time 5, Time 6 e Time 7*), per ciascuna zona. Per ciascun tempo e ciascuna zona, le olive sono state selezionate manualmente in base alla dimensione: piccole, medie e grandi. Tra le olive più grandi sono state scelte le olive con evidenti segni della mosca olearia (i.e. drupe minate con presenza di punture ed evidenti segni di deterioramento e drupe con chiara evidenza della presenza della larva/pupa dell'insetto (*Bactrocera oleae*) ed alcune drupe intatte. Le drupe selezionate, mostrate in [Figura 4.2.48](#), sono poi state disposte in un apposito contenitore per le successive analisi spettrali.

Acquisizione di spettri in riflettanza con spettrofotoradiometro portatile

Per ciascuna oliva analizzata sono stati acquisiti 3 spettri in riflettanza con spettrofotoradiometro portatile ASD FieldSpec 4[®] Standard – Res, operante nel range 350-2500 nm, mediante l'ausilio di una sonda a contatto ([ASD Inc. 2015b](#)). L'acquisizione dei dati è stata effettuata tramite il software RS3 ([ASD Inc. 2008](#)).

Acquisizione di immagini iperspettrali

Lo strumento utilizzato per effettuare l'acquisizione delle immagini iperspettrali in riflettanza è il SisuCHEMA XL[™] (Specim, Finlandia) operante nel range dello spettro elettromagnetico compreso tra 1000 nm e 2500 nm, con una risoluzione spettrale di 6.2 nm. L'ottica utilizzata per le acquisizioni è la 31 mm, la quale consente di acquisire immagini di larghezza pari a 50 mm. Utilizzando questa configurazione, la velocità di scansione del sistema è pari a 18 mm/s.

Più nel dettaglio, sono state acquisite 22 immagini iperspettrali inquadrando il lato A delle drupe e 22 immagini iperspettrali inquadrando il lato B delle drupe. Per ciascuna acquisizione è stata scansionata anche una drupa intatta, senza segni di attacco da parte della mosca olearia, da includere poi nelle successive analisi. Il lato A (*side A*) è il lato della drupa caratterizzato dalla presenza di punture di mosca olearia e/o con evidenti segni di deterioramento dell'esocarpo. Il lato B (*side B*) è il lato opposto delle corrispondenti drupe.

Casi di studio

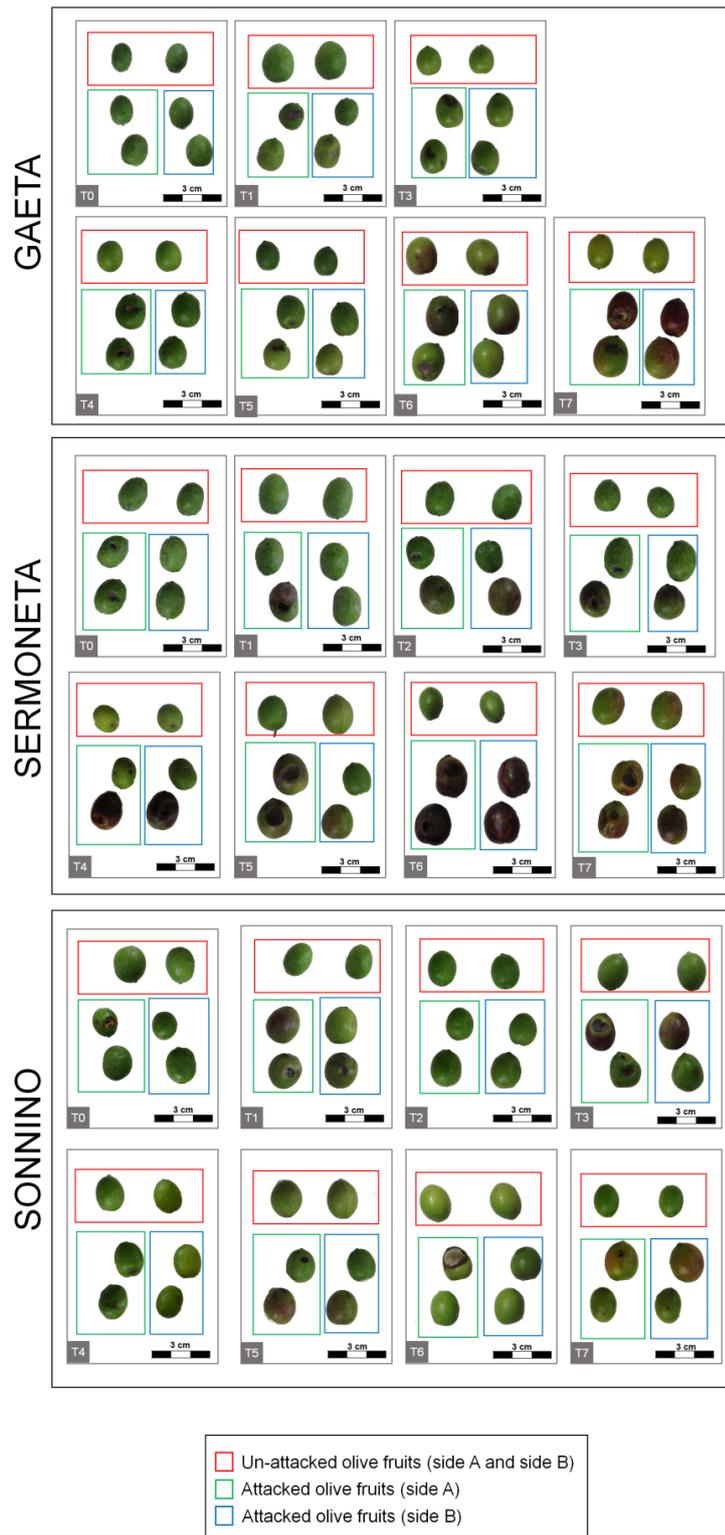


Figura 4.2.48. Immagine digitale dei campioni di olive attaccate e non attaccate da mosca olearia, per differenti tempi di raccolta per località differenti. Il lato A (side A) è lato della drupa caratterizzato dalla presenza di punture di mosca olearia e/o con evidenti segni di deterioramento dell'esocarpo. Il lato B (side B) è il lato opposto delle corrispondenti drupe. Per ciascun tempo, la drupa all'interno del rettangolo rosso è rappresentativa delle olive intatte, senza segni di deterioramento o punture di mosca olearia.

Esportazione e preparazione dati

Gli spettri acquisiti con software RS3 in formato “.asd” sono stati esportati in file di testo ASCII “.txt” mediante l’utilizzo del software ViewSpec Pro (Ver. 6.2.0.). I file così ottenuti sono stati successivamente importati in ambiente MATLAB (MATLAB R2016b ver. 9.1.0.441655) come “dataset object” utilizzando una routine (script MATLAB), scritta appunto per lo scopo: `fieldspec_import2.m`. A seguito delle scansioni con il sistema iperspettrale, le immagini iperspettrali acquisite con il sistema SisuCHEMA XL™ sono state anch’esse esportate in ambiente MATLAB. I file importati in ambiente MATLAB sono stati poi analizzati utilizzando il toolbox PLS_toolbox (ver. 8.2.1).

Lo sfondo di ciascuna immagine iperspettrale è stato escluso, per effettuare le successive analisi. Successivamente, le immagini iperspettrali acquisite, sono state concatenate insieme per generare dei nuovi dataset: *GAETA* (che include le immagini iperspettrali delle drupe provenienti da Gaeta, a differenti tempi di raccolta), *SERMONETA* (che include le drupe provenienti da Sermoneta, a differenti tempi di raccolta), *SONNINO* (che include le drupe provenienti da Sonnino, a differenti tempi di raccolta) e *ALL THE ZONES* (che include le drupe provenienti da Gaeta, Sermoneta e Sonnino, a differenti tempi di raccolta). Le classi relative alla condizione dell’oliva (Oliva attaccata da mosca, “*Attacked olive fruits*” e oliva non attaccata da mosca, “*Un-attacked olive fruits*”) e alla località di provenienza (Gaeta, Sermoneta e Sonnino) sono state settate tramite l’apposita interfaccia del PLS_toolbox ([Wise et al., 2008](#)).

Pre-elaborazione dati

I dataset contenenti le informazioni acquisite con SisuCHEMA XL™ sono stati ridotti di dimensioni: da 942.2-2542 nm a 1000-2500 nm. Questa procedura è stata effettuata per rimuovere la parte più rumorosa degli spettri di ciascun pixel, dovuta al detector dello strumento. A partire da questi ipercubi, sono state successivamente estratte delle ROI.

Mentre, l’intervallo di lunghezza d’onda considerato dei dati del FieldSpec 4 (350 – 2500 nm) è stato ridotto a 400 – 2500 nm ai fini della rimozione delle parti più rumorose dovute al sistema di rivelamento, per una prima analisi; successivamente il dataset è stato portato a 1000 – 2500 nm, per effettuare un confronto con i dati del SisuCHEMA XL™.

Prima di passare all’analisi esplorativa è stato necessario applicare una combinazione di algoritmi di *pre-processing*. La pre-elaborazione dati non solo è necessaria per correggere effetti di *scattering* e/o altri effetti rumorosi, ma anche per esaltare le differenze spettrali tra i diversi campioni ([Rinnan et al., 2009](#)). Sono state valutate differenti strategie di pre-elaborazione dati, in differenti combinazioni. Infine, è stata scelta la combinazione di *pre-processing* che ha fornito una risposta migliore in termini di decomposizione dei dati. La combinazione di *pre-processing* utilizzata è stata: *Standard Normal Variate* (SNV) per la correzione di effetti di *scattering*, *Generalized Least Squares Weighting* (GLS-W, applicato sulle classi settate, con $\alpha = 0.001$) per esaltare le caratteristiche inter-classi ([Eigenvector Research Inc. 2013b](#)) e *Mean Center* (MC) per centrare le colonne dei dati rispetto ad uno zero ([Rinnan et al., 2009](#); [Eigenvector Research Inc., 2012](#)).

Analisi esplorativa dei dati

L'analisi in componenti principali è stata effettuata per esplorare la variabilità intra-classe, secondo la condizione della drupa (Wold *et al.*, 1987). Le componenti principali sono state scelte valutando il grafico degli autovalori e della varianza catturata. Gli *outlier* sono stati identificati ed esclusi dalle successive analisi, esplorando il grafico dei parametri Hotelling T^2 e i residui Q.

Classificazione

Il metodo di classificazione PLS-DA (Wise *et al.* 2006; Wise *et al.* 2008; Ballabio & Consonni, 2013) è stato applicato per valutare la capacità predittiva degli spettri Vis-SWIR nel classificare i campioni studiati. Nel caso specifico, sono stati messi a punto modelli di classificazione PLS-DA a due classi per discriminare le drupe attaccate da mosca e le olive intatte. Le classi modellate sono state: "Attacked olive fruits", ovvero le olive attaccate da mosca (per i dataset riferiti al SisuCHEMA XL, tale classe include "Attacked olive fruits – side A" e "Attacked olive fruits – side B") e "Un-attacked olive fruits", ovvero le olive non attaccate da mosca. Per ciascuna tipologia di dato (FieldSpec 4 e SisuCHEMA XL), sono stati messi a punto 4 modelli di classificazione: uno per ciascuna località di provenienza (Gaeta, Sermoneta e Sonnino) ed uno considerando tutti i campioni delle tre località di provenienza.

Per ciascuna classificazione PLS-DA, i dataset considerati (Gaeta, Sermoneta, Sonnino e tutte tre località insieme) sono stati suddivisi casualmente in due parti, utilizzando l'algoritmo Kennard-stone (Kennard & Stone, 1969): il 70 % del dataset è stato utilizzato per la calibrazione, mentre il 30 % rimanente è stato usato in validazione. Come algoritmo di cross – validazione, per valutare la complessità del modello e scegliere un adeguato numero di variabili latenti, è stato utilizzato l'algoritmo *venetian blinds* (Eigenvector Research Inc. 2016b).

Per ciascuno dei casi analizzati, il modello PLS-DA è stato costruito a partire dal dataset di calibrazione e validato applicandolo sul dataset di validazione.

Analisi dei risultati

Le differenze tra drupe attaccate dalla mosca ed olive sane sono state dapprima valutate mediante l'analisi comparata degli spettri in riflettanza media e successivamente mediante la PCA. In Figura 4.2.49 sono riportati gli spettri medi acquisiti con spettrofotoradiometro nell'intervallo di lunghezza d'onda che va da 1000 nm a 2500 nm (nell'Allegato XV- Case study – Olive: mosca olearia, sono riportati gli spettri medi grezzi nell'intervallo 350-2500 nm). Mentre, in Figura 4.2.50 sono mostrati gli spettri relativi alle ROI estratte dall'ipercubo generato concatenando le immagini iperspettrali acquisite con il sistema SisuCHEMA XL TM.

Comparando gli spettri medi in riflettanza, come riportato in Figura 4.2.49 ed in Figura 4.2.50, si nota chiaramente una certa variabilità tra le firme spettrali delle drupe attaccate da mosca olearie e quelle sane.

Casi di studio

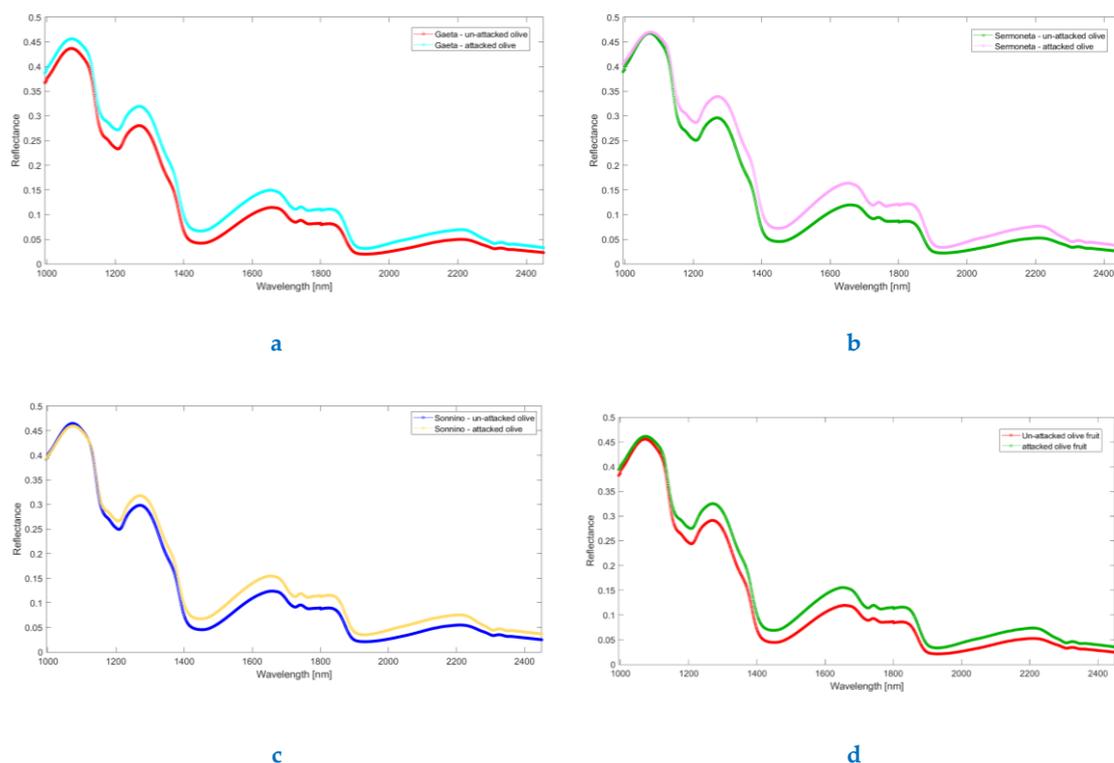


Figura 4.2.49. Spettri medi grezzi in riflettanza, acquisiti con spettrometro FieldSpec 4 (1000 – 2500 nm), delle drupe attaccate da mosca (Attacked olive fruits) e drupe sane (un-attacked olive fruits) per: a) Gaeta, b) Sermoneta, c) Sonnino e d) tutt'e tre le località insieme.

Analisi esplorativa

La variabilità notata tra drupe attaccate da mosca olearie e quelle sane comparando gli spettri medi in riflettanza (Figura 4.2.49 ed in Figura 4.2.50), appare maggiormente evidente nelle analisi PCA. Come si può notare negli *scores plot* relativamente ai dataset dello spettrofotometri portatile, mostrati in Figura 4.2.51, tale variabilità è ben spiegata con la prima componente principale (PC 1): appaiono evidenti due nubi di *scores* distinte relative alle due classi di olive. Nella sezione “Allegati”, *Allegato XV- Case study – Olive: mosca olearia*, sono riportati i risultati delle PCA effettuate sugli spettri, acquisiti con FieldSpec 4, nell’intervallo 350-2500 nm.

Per quanto riguarda, invece, le PCA effettuate sui dati del SisuCHEMA XL, i pattern sottostanti la struttura dei dati non sono completamente evidenti, soprattutto per le olive provenienti da Gaeta (Figura 4.2.52). Ad ogni modo, per quanto riguarda le strutture di dati relative alle olive provenienti da Sermoneta e Sonnino (Figura 4.2.52a e Figura 4.2.52b), si evidenzia la predominanza di due *cluster*, riferibili alle classi: “un-attacked olive fruit” e “attacked olive fruit (side A e side B)”.

Casi di studio

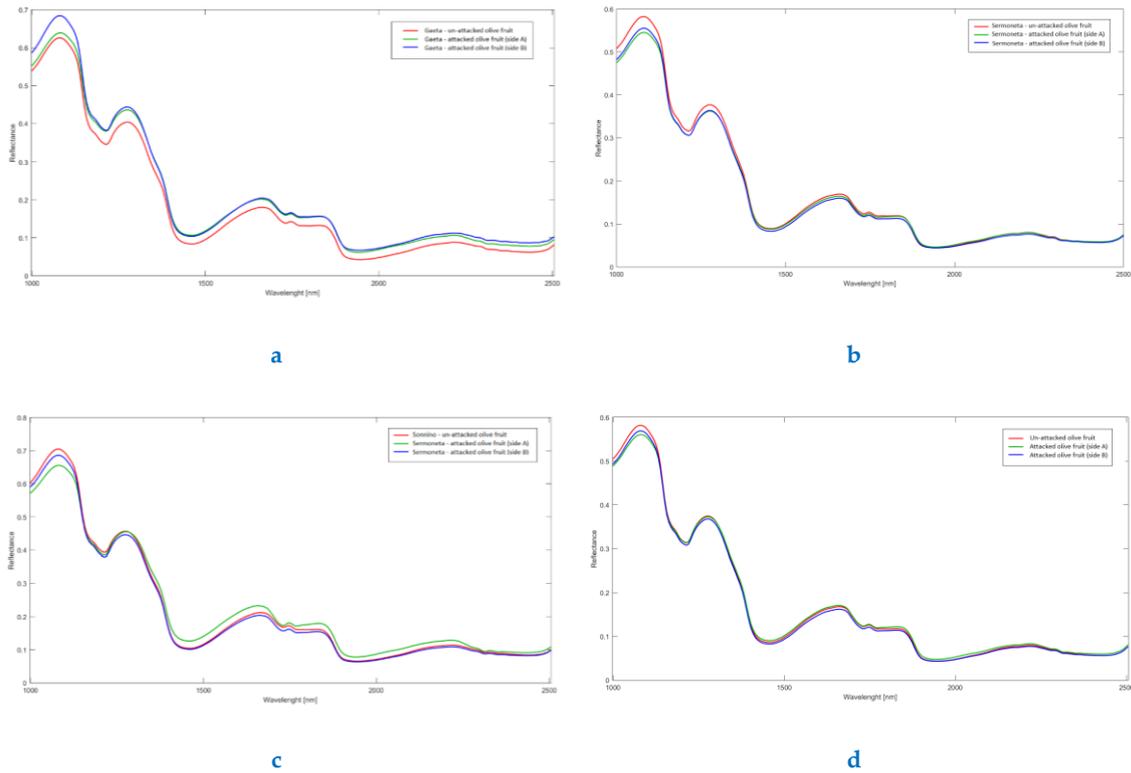
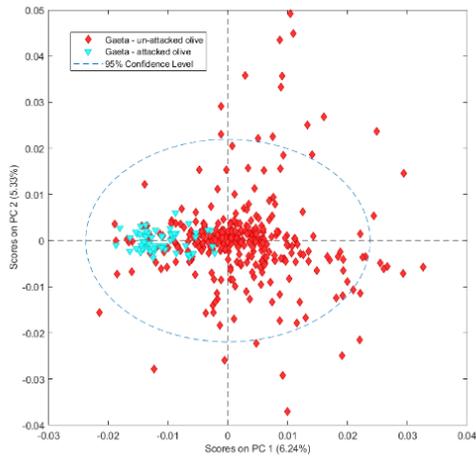


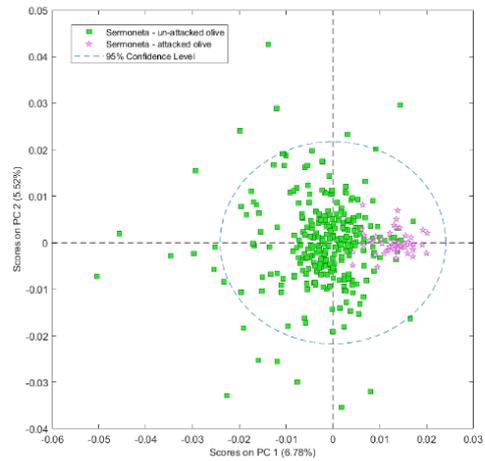
Figura 4.2.50. Spettri medi grezzi in riflettanza, estratti da ROI a partire dall'ipercubo generato concatenando le immagini iperspettrali acquisite con il sistema SisuCHEMA XL™ (1000–2500 nm), delle drupe attaccate da mosca (Attacked olive fruits – side A e Attacked olive fruits – side B) e drupe sane (un-attached olive fruits) per: a) Gaeta, b) Sermoneta, c) Sonnino e d) tutt'e tre le località insieme.

Mentre, il dataset relativo alle tre località mostra un comportamento intermedio tra le due situazioni (Figura 4.2.52c). Nel caso dei dati del SisuCHEMA XL, anche se gli *scores* delle tre classi sono perlopiù sovrapposti uno sull'altro, la presenza di due pattern dominanti possono essere considerati come una evidenza della differenza tra le firme spettrali: la drupa attaccata da mosca olearia può essere scansionata sia sul lato A che sul lato B, in analisi in componenti principali le firme spettrali nell'intervallo di lunghezza d'onda 1000-2500 nm si sovrapporranno.

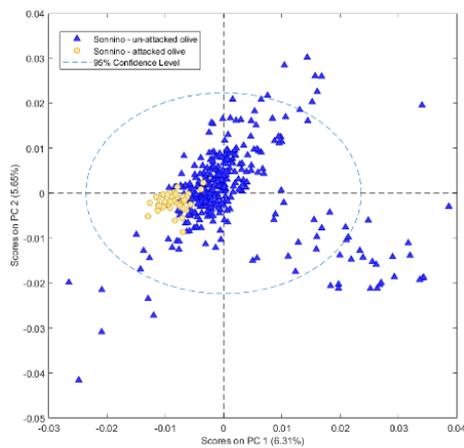
Casi di studio



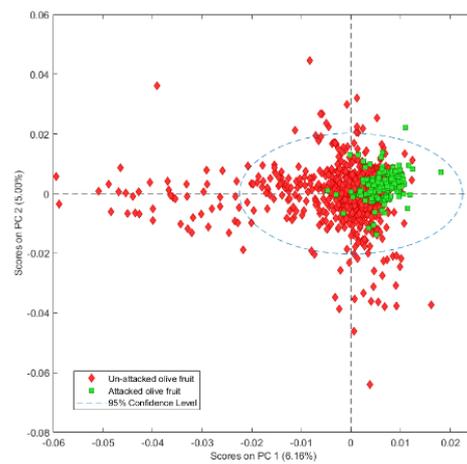
a



b



c



d

Figura 4.2.51. Grafici degli Scores sulle prime due componenti principali (PC1 vs. PC 2), relative alle PCA effettuate sugli spettri del FieldSpec 4 (1000-2500 nm), delle drupe attaccate da mosca (Attacked olive fruits) e drupe sane (un-attacked olive fruits) per: a) Gaeta, b) Sermoneta, c) Sonnino e d) tutt'e tre le località insieme.

Casi di studio

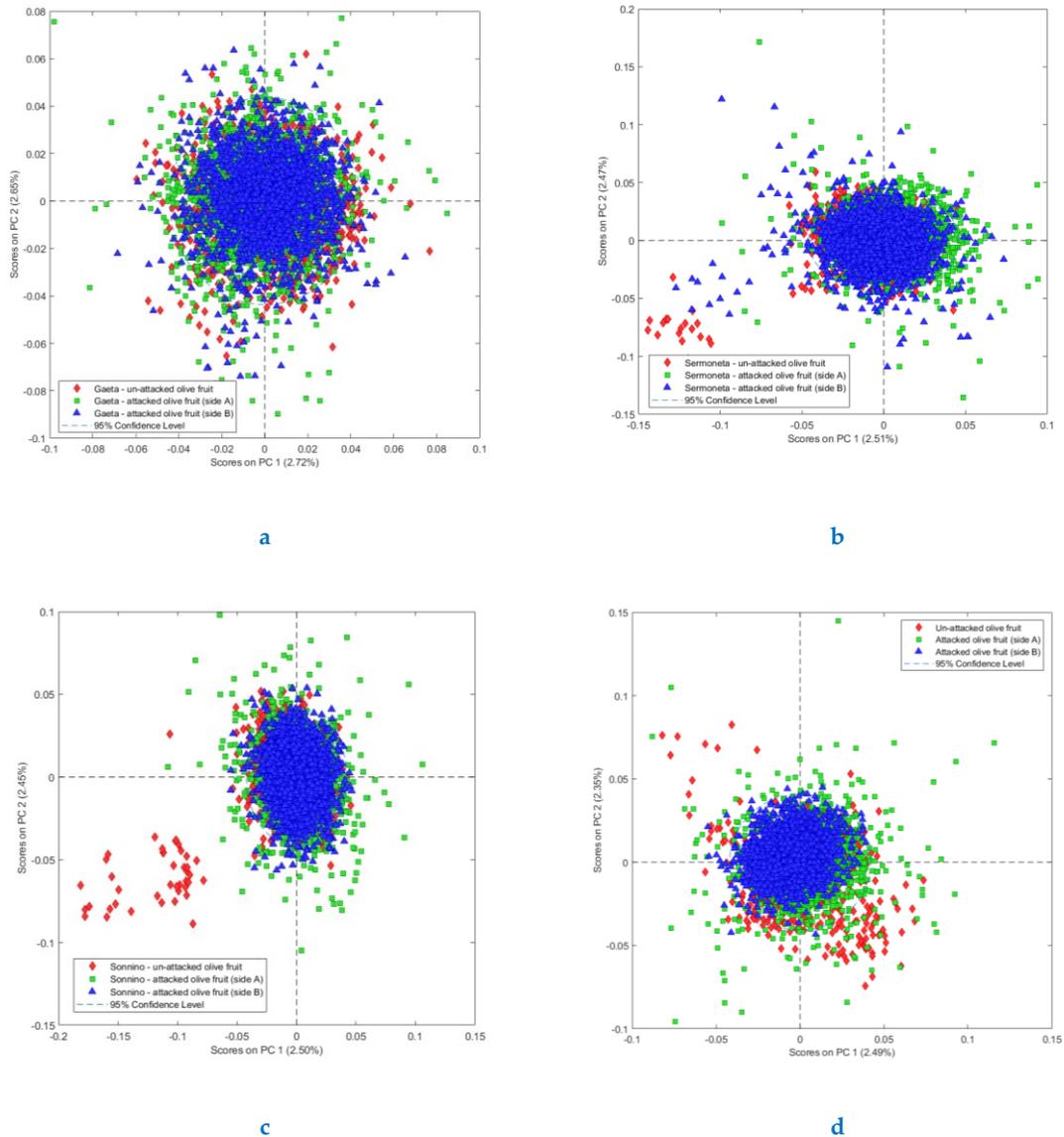


Figura 4.2.52. Grafici degli Scores sulle prime due componenti principali (PC1 vs. PC 2), relative alle PCA effettuate sugli spettri del SisuCHEMA XL (1000-2500 nm), delle drupe attaccate da mosca (Attacked olive fruits) e drupe sane (un-attacked olive fruits) per: **a**) Gaeta, **b**) Sermoneta, **c**) Sonnino e **d**) tutt'e tre le località insieme.

Classificazione Partial Least Squares – Discriminant Analysis

Nel caso dei dati del FieldSpec 4 nell'intervallo 1000-2500 nm, come si può notare in Tabella 4.2.17 dove sono riportati gli indicatori di performance in predizione, i modelli di classificazione riescono a discriminare con buona precisione le due classi di olive. Questo può essere notato sia dagli *scores plot* delle prime due variabili latenti, che dai grafici relativi alla posizione del *boundary* dei classificatori (riportati in Figura 4.2.53 and Figura 4.2.54). Le migliori performance in classificazione si ottengono con il modello riferito al dataset SONNINO (Figura 4.2.53b). Il modello che considera le drupe di tutte e tre le località presenta, invece, una minore abilità in classificazione, rispetto ai modelli che

Casi di studio

considerano la singola zona (Figura 4.2.54). I risultati delle classificazioni effettuate con i dati del FieldSpec 4 nell'intervallo 350-2500 nm sono riportati nell' *Allegato XV- Case study – Olive: mosca olearia*.

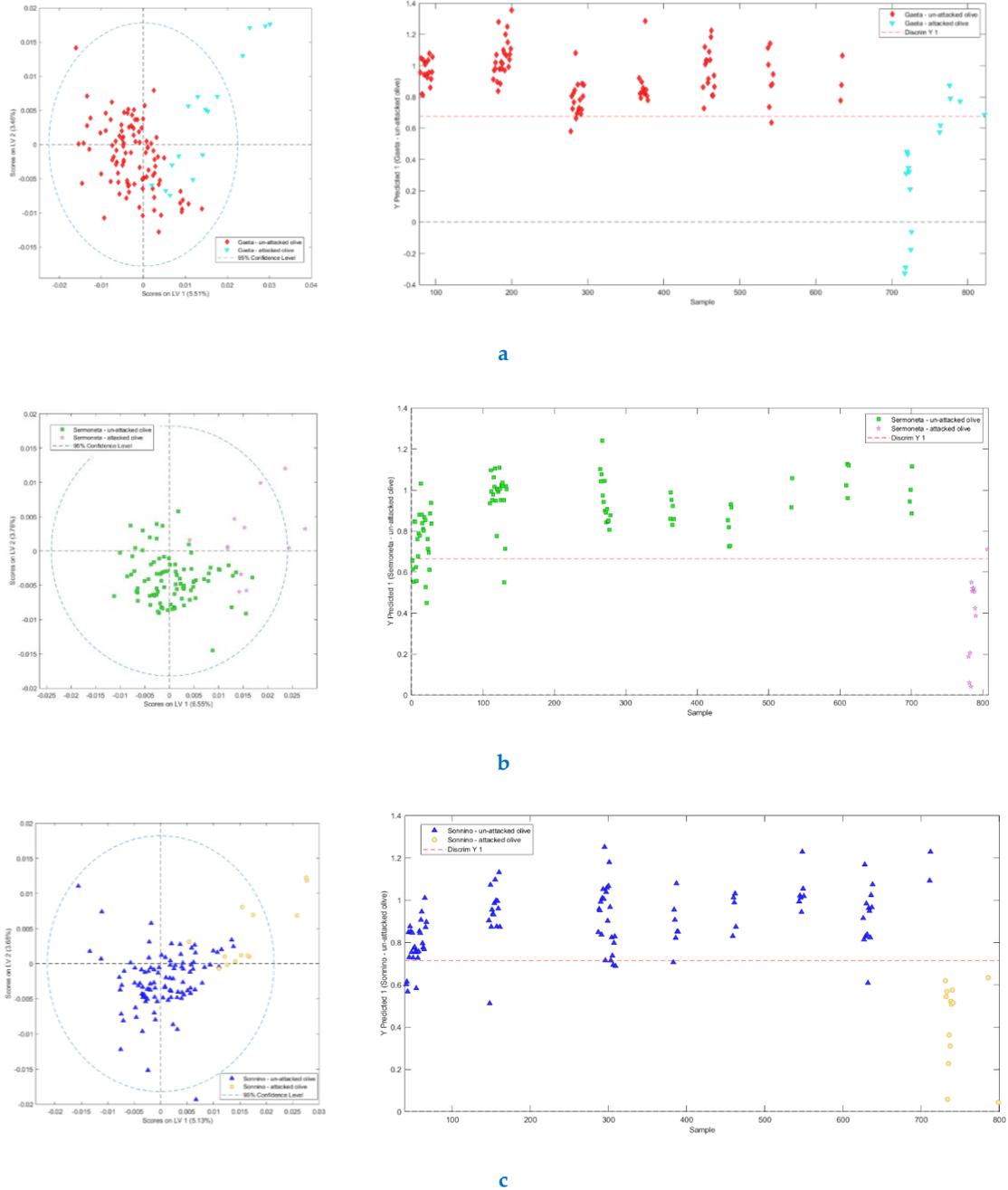


Figura 4.2.53. Scores plot delle prime due variabili latenti (a sinistra) e posizione del boundary di discriminazione (a destra; la linea tratteggiata rossa rappresenta la soglia di discriminazione), relativi alle classificazioni PLS-DA per le due classi (attacked olive fruits e un-attached olive fruits), relativamente ai set di dati: a) Gaeta, b) Sermoneta e c) Sonnino.

Casi di studio

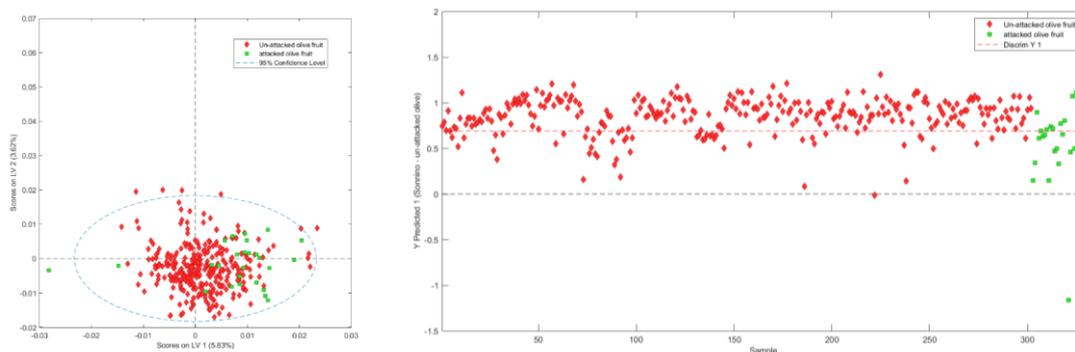


Figura 4.2.54. Scores plot delle prime due variabili latenti (a sinistra) e posizione del boundary di discriminazione (a destra; la linea tratteggiata rossa rappresenta la soglia di discriminazione), relativi alle classificazioni PLS-DA per le due classi (attacked olive fruits e un-attacked olive fruits), relativamente al set di dati che considera tutt'e tre le località insieme.

Tabella 4.2.17. Performance di classificazione in predizione delle classificazioni effettuate con i dataset del FieldSpec 4 (1000-2500 nm), relativamente ai set di dati: Gaeta, Sermoneta, Sonnino e tutt'e tre le località insieme.

| Dataset | Classe | Sensitivity (Pred) | Specificity (Pred) |
|---------------|-------------------------|--------------------|--------------------|
| GAETA | Un-attacked olive fruit | 0.970 | 0.750 |
| | Attacked olive fruit | 0.750 | 0.970 |
| SERMONETA | Un-attacked olive fruit | 0.899 | 0.917 |
| | Attacked olive fruit | 0.917 | 0.899 |
| SONNINO | Un-attacked olive fruit | 0.899 | 1.000 |
| | Attacked olive fruit | 1.000 | 0.899 |
| ALL THE ZONES | Un-attacked olive fruit | 0.951 | 0.778 |
| | Attacked olive fruit | 0.778 | 0.951 |

Anche nel caso dei dataset del SisuCHEMA XL si ottengono risultati simili a quanto visto per il FieldSpec 4 nell'intervallo 1000-2500 nm, come si può notare in Tabella 4.2.18. Le immagini dei dataset di validazione e le corrispondenti mappe di predizione, con riferimento a GAETA, SERMONETA, e SONNINO, sono riportate in Figura 4.2.55. Mentre, in Figura 4.2.56 è riportato il dataset di validazione e la mappa di predizione PLS-DA riferita al dataset ALL THE ZONES, in cui le drupe provenienti da ciascuna zona sono considerate tutte insieme. Come si può vedere dalle mappe di previsione (Figura 4.2.55 e Figura 4.2.56) e dalla Tabella 4.2.18, la classe prevista come "Attacked olive fruit (side A and side B)" risulta avere una certa percentuale di pixel erroneamente classificati, ciò è dovuto principalmente all'effetto bordo causato dalla forma delle drupe. Le prestazioni complessive dei modelli di classificazione delle singole zone utilizzando i dati del

Casi di studio

SisuCHEMA XL sono abbastanza buone, mentre il modello globale mostra un tasso di errata classificazione leggermente maggiore per i frutti d'oliva attaccati da mosca olearia.

Tabella 4.2.18. Performance di classificazione in predizione delle classificazioni effettuate con i dataset del FieldSpec 4 (1000-2500 nm), relativamente ai set di dati: Gaeta, Sermoneta, Sonnino e tutt'e tre le località insieme.

| <i>Dataset</i> | <i>Classe</i> | <i>Sensitivity (Pred)</i> | <i>Specificity (Pred)</i> |
|----------------|--------------------------------|---------------------------|---------------------------|
| GAETA | <i>Un-attacked olive fruit</i> | 0.878 | 0.526 |
| | <i>Attacked olive fruit</i> | 0.526 | 0.878 |
| SERMONETA | <i>Un-attacked olive fruit</i> | 0.903 | 0.543 |
| | <i>Attacked olive fruit</i> | 0.543 | 0.903 |
| SONNINO | <i>Un-attacked olive fruit</i> | 0.751 | 0.689 |
| | <i>Attacked olive fruit</i> | 0.689 | 0.751 |
| ALL THE ZONES | <i>Un-attacked olive fruit</i> | 0.752 | 0.480 |
| | <i>Attacked olive fruit</i> | 0.480 | 0.752 |

Confrontando i risultati di previsione dei modelli discriminanti di set-up per il set di dati FieldSpec 4 (1000 - 2500 nm) e SisuCHEMA XL (1000 - 2500 nm), i risultati migliori, in termini di *sensitivity* e *specificity*, si ottengono utilizzando i dati del FieldSpec 4 (come mostrato in [Tabella 4.2.19](#)). Se invece si considera la modellazione di tutte le zone, utilizzando i dati del FieldSpec 4, la classe “*Un-attacked olive fruit*” presenta una *sensitivity* pari a 0.951 ed una *specificity* pari a 0.778; mentre con i dati del SisuCHEMA XL si ottiene una *sensitivity* pari a 0.752 ed una *specificity* pari a 0.480: una situazione con falsi negativi bassi ed alti falsi positivi.

Casi di studio

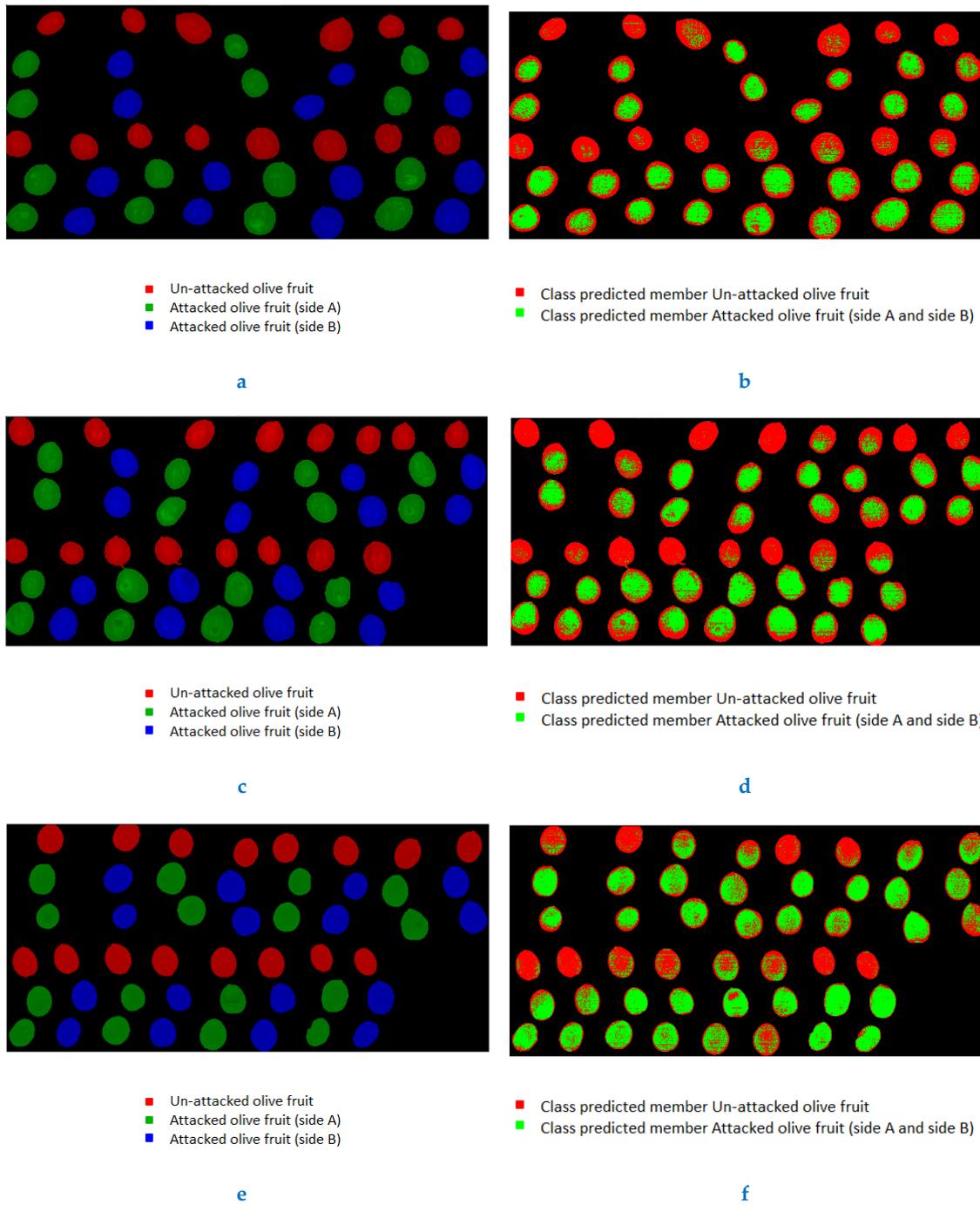


Figura 4.2.55. Set di validazione riferiti a: **a)** Gaeta, **c)** Sermoneta e **d)** Sonnino; mappe di predizione PLS-DA per le due classi (attacked olive fruits e un-attacked olive fruits), riferiti a: **b)** Gaeta, **b)** Sermoneta e **f)** Sonnino.

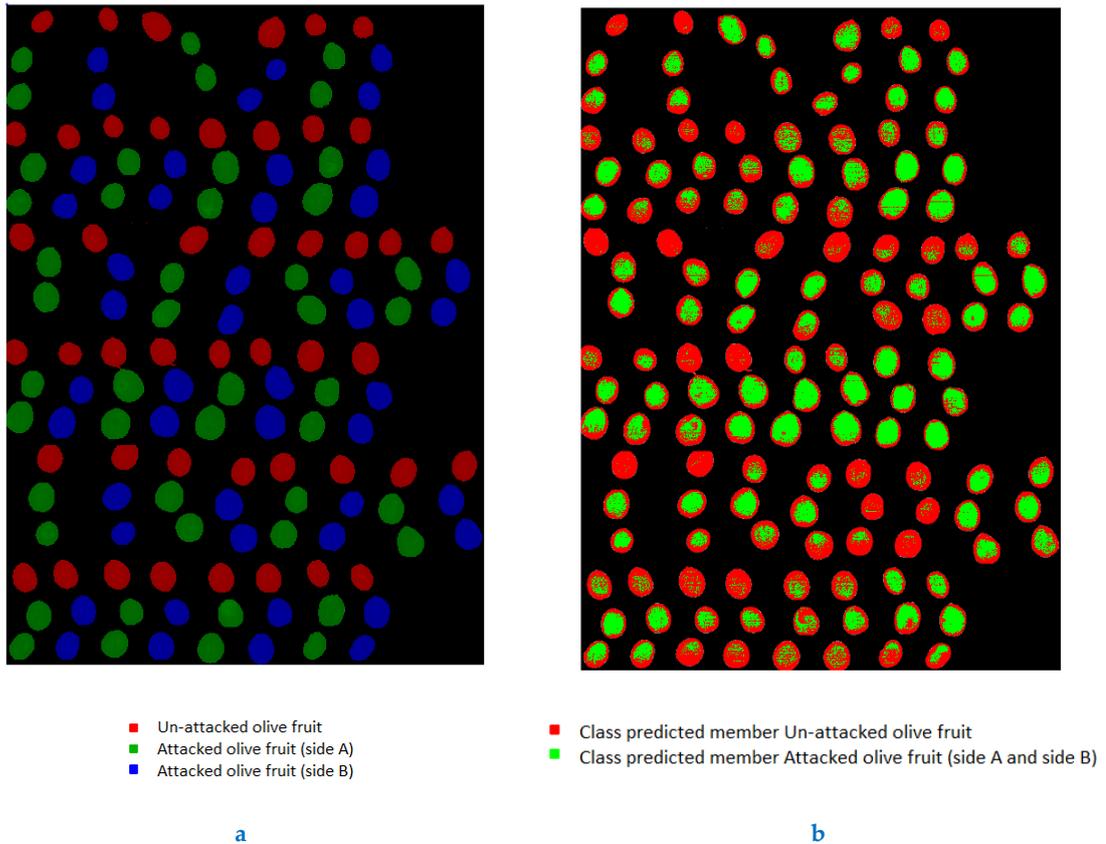


Figura 4.2.56. Set di validazione (a) e mappe di predizione PLS-DA (b) per le due classi (attacked olive fruits e un-attacked olive fruits), per il dataset concatenato riferito a tutt'e tre le località insieme.

Discussione dei risultati

In riferimento alla presenza o meno di segni di attacco della mosca olearia, gli spettri in riflettanza (1000 - 2500 nm) delle drupe di ulivo della cultivar itrana per diversi tempi di raccolta mostrano una certa variabilità. I modelli di classificazione sono stati calibrati sfruttando le informazioni spettrali delle drupe a diversi tempi di maturazione, al fine di garantire una maggiore robustezza nelle performance. I risultati di classificazione delle PLS-DA dimostrano che è possibile rilevare i frutti di ulivo attaccati dalla mosca delle olive sia con un sistema iperspettrale da laboratorio, che con uno spettrofotoradiometro portatile da campo.

Grazie all'utilizzo di uno spettrofotoradiometro da campo si può, quindi, effettuare una diagnosi precoce dell'attacco della mosca dell'oliva nelle drupe intatte, anche qualora le olive si trovassero ancora attaccate all'albero. Inoltre, l'utilizzo di tecniche iperspettrali se scalate in campo industriale, possono fornire preziose informazioni sullo stato di salute delle drupe. Informazioni, queste, che potranno poi essere sfruttate nel processo decisionale relativo alla raccolta e/o alla macinazione o come *core* analitico per macchine selezionatrici.

Casi di studio

Tabella 4.2.19. Performance di classificazione in predizione delle classificazioni effettuate con i dataset del FieldSpec 4 e SisuCHEMA XL (1000-2500 nm), relativamente ai set di dati: Gaeta, Sermoneta, Sonnino e tutt'e tre le località insieme.

| Dataset | Classe | FieldSpec 4 | | SisuCHEMA XL | |
|---------------|--------------------------------|--------------------|--------------------|--------------------|--------------------|
| | | Sensitivity (Pred) | Specificity (Pred) | Sensitivity (Pred) | Specificity (Pred) |
| GAETA | <i>Un-attacked olive fruit</i> | 0.970 | 0.750 | 0.878 | 0.526 |
| | <i>Attacked olive fruit</i> | 0.750 | 0.970 | 0.526 | 0.878 |
| SERMONETA | <i>Un-attacked olive fruit</i> | 0.899 | 0.917 | 0.903 | 0.543 |
| | <i>Attacked olive fruit</i> | 0.917 | 0.899 | 0.543 | 0.903 |
| SONNINO | <i>Un-attacked olive fruit</i> | 0.899 | 1.000 | 0.751 | 0.689 |
| | <i>Attacked olive fruit</i> | 1.000 | 0.899 | 0.689 | 0.751 |
| ALL THE ZONES | <i>Un-attacked olive fruit</i> | 0.951 | 0.778 | 0.752 | 0.480 |
| | <i>Attacked olive fruit</i> | 0.778 | 0.951 | 0.480 | 0.752 |

4.2.8 Valutazione del processo di maturazione del kiwi giallo mediante l'utilizzo di spettrofotometri NIR portatili

Problematica e stato dell'arte

Il consumo del kiwi, frutto dell'*Actinidia chinensis*, è molto diffuso in tutto il mondo per via delle proprie importanti qualità nutrizionali; questi frutti contengono una grande quantità di vitamine, soprattutto la vitamina C, che si attesta intorno agli 85 mg / 100 g di parte edibile. Il sapore, la forma e la stessa tessitura del kiwi possono variare con il tempo e le condizioni ambientali a cui sono esposti, perché si tratta di frutti climaterici, i quali maturano anche successivamente alla separazione della pianta. Proprio per questo motivo, è necessario monitorarne la qualità interna in maniera rapida ed efficiente. Siccome il kiwi viene commercializzato in tutto il mondo, c'è una certa propensione alla ricerca di nuovi metodi d'indagine che siano veloci, non distruttivi, non invasivi ed affidabili. È il sapore percepito dal gusto del consumatore che in qualche modo ne identifica la qualità. Gli indicatori che vengono comunemente utilizzati per definire la qualità di un kiwi, i quali influiscono sul gradimento dei consumatori sono: il contenuto di zuccheri, l'acidità, il contenuto di sostanza secca e la fermezza ([Wang et al., 2015](#)). Alcuni studi hanno dimostrato che il sapore di un kiwi maturo è prevalentemente determinato dal contenuto dei solidi solubili totali (TSSC, *Total Soluble Solid Content*) del frutto stesso, il quale, a sua volta, è correlato al contenuto di sostanza secca (DM, *Dry Matter*). Proprio per questo, sia il TSSC e il DM sono due parametri che vengono regolarmente utilizzati come riferimento per valutare la qualità globale del kiwi. Per valutare i parametri TSSC e DM vengono in genere utilizzati metodi distruttivi, cioè che portano alla distruzione stessa del frutto analizzato.

Ad ogni modo, esistono diversi metodi non-distruttivi di analisi di questi due parametri ([Crisosto et al., 2009](#); [Cantin et al., 2011](#)), come ad esempio l'analisi densimetrica e la spettroscopia NIR. In letteratura, sono presenti diversi studi riguardanti l'utilizzo della spettroscopia NIR applicata per la valutazione di parametri qualitativi di frutti di *Actinidia*. Una breve lista di tali studi è riportata in [Tabella 4.2.20](#).

Obiettivi dello studio

L'obiettivo principale di questo studio è stato quello di valutare l'utilizzo di due differenti spettrofotometri portatili per il monitoraggio dello stato di maturazione dei kiwi gialli, provenienti da coltivazioni site in Cisterna di Latina (Lazio, Italia), una località nota per le estese coltivazioni di *actinidia* contingenti le zone rurali. In particolare, gli indicatori valutati sono stati il contenuto delle sostanze allo stato solido dissolte (TSSC, *Total Soluble Solid Content*) misurato in gradi Brix (°Bx) ed il contenuto di sostanza secca (DM, *Dry Matter*), per diverse dimensioni e condizioni del frutto, selezionati a seguito di un'ispezione visiva.

Casi di studio

Tabella 4.2.20. Breve lista di pubblicazioni scientifiche nelle quali si fa riferimento all'utilizzo di spettroscopia ottica per determinare il contenuto totale dei solidi solubili totali (TSSC, Total Soluble Solid Content) ed il contenuto di sostanza secca (DM, Dry Matter); adattata da Wang et al., 2015 (*) AD: *Actinidia deliciosa*, HW: Hayward e ZH: Zhonghua.

| Attributo qualitativo | Cultivar* | Metodo | Intervallo spettrale [nm] | Metodo di regressione | r_p | RMSEP | Referenze bibliografiche |
|-----------------------|-----------|-----------------------|---------------------------|-----------------------|-------|------------|---------------------------|
| TSSC / Sugar Content | AD | Spettroscopia Imaging | 650–1100 | PLS | 0.92 | 1.18 | Martinsen & Schaare, 1998 |
| TSSC / Sugar Content | HW | Spettroscopia | 800–2500 | PLS | - | 0.68 (SEP) | Arazuri et al., 2005 |
| TSSC / Sugar Content | AD | Spettroscopia | 408–2492 | PLS | 0.99 | 0.49 | Lee et al., 2012 |
| Dry Matter | ZH | Spettroscopia | 1000–2500 | siPLS | 0.9 | 0.53% | Lue et al., 2010 |

È stato eseguito uno studio approfondito sulle correlazioni, con l'ausilio di tecniche chemiometriche, tra gli spettri di riflettanza *Near InfraRed* (NIR) acquisiti con strumenti portatili e questi parametri significativi del kiwi. L'analisi in componenti principali è stata utilizzata come metodologia di analisi esplorativa per gli spettri in riflettanza acquisiti. Mentre il metodo di regressione *Partial Least Squares* (PLS) è stato utilizzato al fine di prevedere quantitativamente il TSSC e il DM. La panoramica globale delle analisi effettuate è mostrata in [Figura 4.2.57](#).

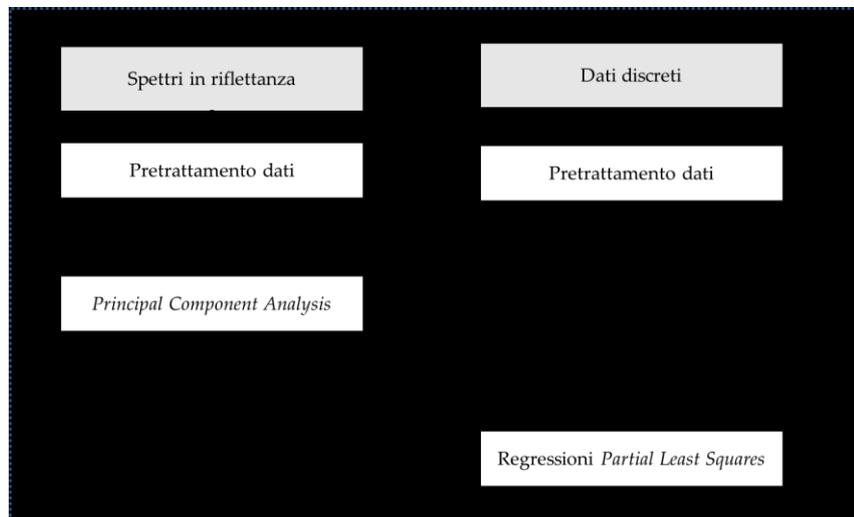


Figura 4.2.57. Diagramma di flusso delle analisi effettuate.

Procedura

Preparazione dei campioni

I frutti di *Actinidia chinensis*, della varietà “gold” (più comunemente detto kiwi giallo) analizzati in questo studio sono stati raccolti in una coltivazione situata nel comune di Cisterna di Latina (Lazio, Italia), una località ben conosciuta per le estese coltivazioni di

kiwi. I frutti raccolti sono stati successivamente immagazzinati in una camera a temperatura controllata (20°C). Sono stati analizzati 10 kiwi a settimana per 6 settimane, per un totale di circa 60 kiwi. In [Figura 4.2.58](#) è mostrato il primo set di kiwi analizzati al Tempo 0. Per ciascun frutto sono stati misurati i solidi solubili totali mediante il metodo rifrattometrico (in grado Brix) ed il peso secco. Inoltre, su ciascun frutto sono stati acquisiti spettri in riflettanza sul tessuto del frutto con due differenti tipologie di spettrofotometri portatili.

Misura del contenuto di sostanze allo stato solido dissolte

Tra i solidi solubili, gli zuccheri sono quelli che si trovano in maggiori concentrazioni all'interno della polpa di un frutto. Insieme agli zuccheri, nella polpa di un frutto, sono presenti anche altre sostanze solubili (i.e. pectina e amminoacidi). Il contenuto di sostanze allo stato solido dissolte, o TSSC, viene tradizionalmente valutato tramite refrattometro e riportato in gradi Brix ([Garner et al., 2015](#)). Un refrattometro portatile classico consente di misurare l'indice di rifrazione di un liquido. Nel caso della polpa di un frutto, il refrattometro consente la determinazione dell'angolo critico di rifrazione ([ChemBuddy, 2011](#)). La luce che passa attraverso il liquido disposto sulla superficie del prisma del refrattometro viene "piegata" e cioè rifratta. La luce rifratta viene poi focalizzata su una scala graduata, ingrandita da una lente in maniera tale da essere ben visibile ai fini della lettura della stessa ([MISCO, 2013](#); [A.KRÜSS Optronic, 2013](#)). In questo modo si può misurare il grado zuccherino del kiwi analizzato; i.e. assumendo che la sua polpa abbia lo stesso indice di rifrazione di un 1 g di saccarosio in 100 g di soluzione acquosa, allora si avrà un TSSC pari a 1° Bx ([Bionutrient Food Association, 2017](#)).

Per effettuare una misura sistematica del grado Brix, ciascun kiwi è stato tagliato a metà (lungo l'equatore del frutto) e successivamente sono state ottenute diverse fettine con un'affettatrice. Le fettine sono state ottenute seguendo la direzione ortogonale al taglio equatoriale. Metà di ciascun kiwi è stato spremuto per ottenere il succo da disporre uniformemente sul prisma del refrattometro. In questa maniera si sono ottenute le misure del grado Brix.

Misura del contenuto di sostanza secca

La sostanza secca o *Dry Matter* (DM) è la componente solida di un frutto e consiste essenzialmente nella parte rimanente del frutto a seguito dell'essiccamento o disidratazione completa. Questa parte che rimane a seguito dell'essiccamento è costituita prevalentemente da carboidrati solubili, come il fruttosio, il saccarosio ed il glucosio ([The New Zealand Institute for Plant and Food Research Limited, 2013](#)).

Per ciascun kiwi, sono state pesate 3 fette spesse circa 5 mm ($W_{f,i}$). Le fette sono state successivamente lasciate ad essiccare in un essiccatore per frutta a circa 70 ° C per 12 ore ([Figura 4.2.59a](#)). Le fette, a seguito del processo di essiccamento, sono state ripesate ($W_{d,i}$), come mostrato in [Figura 4.2.59b](#).

Casi di studio

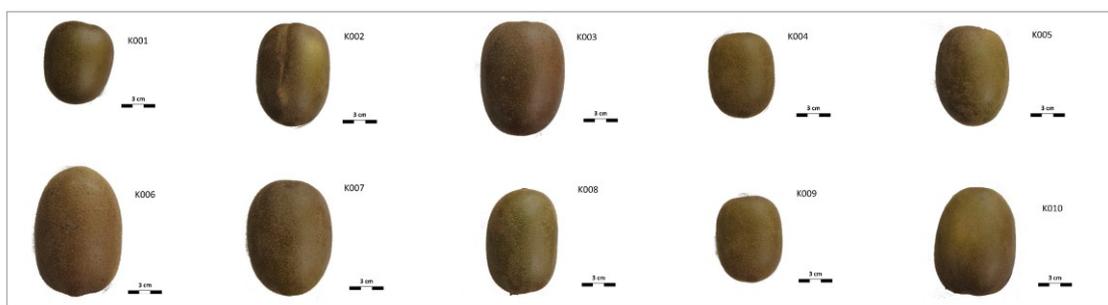


Figura 4.2.58. Il primo set di kiwi analizzati.



a



b

Figura 4.2.59. Disposizione delle fette di kiwi nell'essiccatore per frutta (a) e pesatura delle fette di kiwi a seguito del processo di essiccamento (b).

Per la pesatura delle fette di kiwi è stata utilizzata una bilancia di precisione della KERN & SOHN GmbH, avente una risoluzione di 0,0001 g. Per ciascun kiwi, il calcolo del DM è stato effettuato in accordo con l'equazione:

$$DM = \frac{\overline{W}_d}{\overline{W}_f} \times 100 \quad (4.2.3)$$

Dove \overline{W}_d è la media in grammi delle tre fette pesate prima del processo di essiccamento, mentre \overline{W}_f è la media in grammi delle tre fette pesate a seguito del processo di essiccamento.

Acquisizione di spettri in riflettanza, preparazione ed esportazione dati

Prima dell'analisi distruttiva della determinazione del grado zuccherino e la sostanza secca, sono stati acquisiti spettri in riflettanza con due unità spettrofotometriche portatili, opportunamente calibrati. Il primo set di acquisizioni è stato effettuato con lo spettrofotometro portatile JDSU MicroNIR™ (950-1650 nm): su ciascun kiwi sono stati acquisiti 6 spettri in riflettanza, grazie all'utilizzo dell'interfaccia software MicroNIR Pro. Il secondo set di acquisizioni è stato effettuato mediante lo spettrofotometri portatile ASD FieldSpec 4® Standard-Res (350 – 2500 nm). Anche in questo caso, sono stati acquisiti 6 spettri in riflettanza sull'esocarpo di ciascun frutto in maniera casuale.

Gli spettri acquisiti con MicroNIR™ sono stati salvati in formato `.csv` e successivamente importati in ambiente MATLAB® (MATLAB R2016b ver. 9.1.0.441655, The MathWorks, Inc., Natick, MA, USA), utilizzando lo script `micronir_import.m`. Mentre, gli spettri acquisiti con ASD FieldSpec 4® immagazzinati in singoli file in formato `.asd` sono stati convertiti in un singolo file con formato `.txt` (ASCII), mediante l'utilizzo del software ViewSpec Pro (Ver. 6.2.0.). Il file così ottenuto è stato importato anch'esso in ambiente MATLAB®, per le successive analisi. Tutti quanti i dataset importati in ambiente MATLAB® sono stati analizzati con il supporto del toolbox PLS_toolbox (ver. 8.2.1) della Eigenvector Research, Inc. Su ciascun dataset sono state impostate delle etichette (o *label*) sui campioni in riferimento a: la denominazione, il tempo di analisi, il contenuto di sostanza solubile disciolta (°Bx), la sostanza secca (DM) e gli intervalli maggiormente significativi di questi ultimi due. La definizione delle classi di TSSC e DM sono riportati rispettivamente in [Tabella 4.2.21](#) e [Tabella 4.2.22](#).

Tabella 4.2.21. Definizione delle classi di TSSC.

| °Bx class | Intervallo del TSSC [°Bx] | Classe |
|-----------|------------------------------|--------------------|
| 1 | 9.0-12.5 | <i>low</i> |
| 2 | 12.5-15.5 | <i>low-medium</i> |
| 3 | 15.5-18.5 | <i>medium-high</i> |
| 4 | 18.5-21.5 | <i>high</i> |

Pre-elaborazione dati ed analisi esplorativa

La pre-elaborazione dei dati è stata necessaria per esaltare le differenze spettrali presenti tra classi differenti e rimuovere eventuale rumore di *offset* e altri artefatti dovuti a fenomeni di *scattering* e/o ad altri fenomeni fisici ([Rinnan et al., 2009](#)). Il set di dati del FieldSpec 4 è stato ridotto da 350-2500 nm a 400-2500 nm per rimuovere la parte più rumorosa degli spettri.

Mentre, per il set di dati del MicroNIR è stato considerato l'intero range spettrale 950-1650 nm, in quanto non presenti zone di lunghezza d'onda particolarmente rumorose.

Per entrambi i set di dati provenienti dal FieldSpec 4 e dal MicroNIR è stata utilizzata la combinazione di algoritmi: EMSC (*Extended Multiplicative Scatter Correction*) per la correzione degli artefatti dovuti a fenomeni di *scattering* ([Eigenvector Research Inc. 2013b](#)), GLS-W (*Generalized Least Squares Weighting*, con $\alpha = 0.01$) per attenuare le differenze presenti tra gruppi di campioni simili ed esaltare quelle presenti tra le classi di campioni, *Smoothing* basato sull'algoritmo di S-G per attenuare eventuali *spikes* ([Savitzky & Golay, 1964](#); [Rinnan et al., 2009](#)) e *Mean Center* (MC).

Come tecnica di analisi esplorativa è stata scelta l'analisi in componenti principali ([Wold et al., 1987](#)). In particolare, per ciascun set di dati è stato analizzato il

raggruppamento degli *scores* in PCA, in accordo con le classi relative a: a) valori di TSSC, b) valori di DM, c) classi di TSSC e d) classi di DM. Per ciascuna PCA, sono state scelte le prime due componenti principali. Gli *outliers* sono stati identificati ed esclusi dai dataset per effettuare le successive analisi.

Tabella 4.2.22. Definizione delle classi di DM.

| DM class | Intervallo del DM [%] | classe |
|----------|--------------------------|-------------|
| 1 | 16-17 | low |
| 2 | 18-19 | low-medium |
| 3 | 20-21 | medium |
| 4 | 22-23 | medium-high |
| 5 | 24-25 | high |

Modelli di regressione

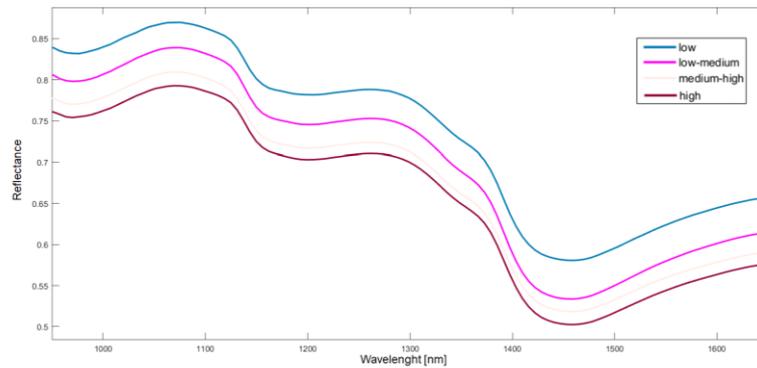
Per valutare la correlazione tra gli spettri acquisiti e lo stato di maturazione dei kiwi analizzati è stata utilizzata la tecnica di regressione ai minimi quadrati parziali (PLS). Per ciascun set di dati (il set di dati del FieldSpec 4 e quello contenente i dati del MicroNIR) sono stati computati due modelli PLS (*Geladi et al., 1986; Wold et al., 2001; Abdi, 2003*): il primo utilizzando i valori di TSSC (°Bx) come **Y** ed il secondo con i valori di DM. Per ciascun dataset, la scelta del set di addestramento del modello ed il set di validazione dello stesso è stato scelto applicando l'algoritmo di *Kennard-Stone* (*Kennard & Stone, 1969*): il 70% dei dati contenuti nel dataset iniziale sono stati utilizzati per la calibrazione e la convalida incrociata del modello, mentre il rimanente 30% è stato utilizzato come set di validazione. Come algoritmo di *cross-validation* è stato utilizzato il *random subset* con 20 iterazioni (*Eigenvector Research Inc., 2016*).

Analisi dei risultati

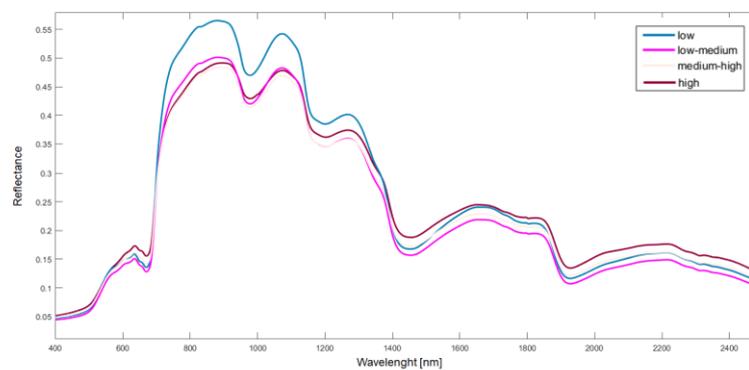
Spettri medi in riflettanza ed analisi in componenti principali

Gli spettri medi grezzi, acquisiti con FieldSpec 4 e MicroNIR, riferiti alle classi significative di TSSC e DM sono mostrati rispettivamente in [Figura 4.2.60](#) e [Figura 4.2.61](#). Gli spettri medi mostrano differenze in accordo con le variabili analizzate, quali: tempo di analisi, valori di DM e valori di TSSC (si veda anche [Allegato XVI- Case study – Kiwi: maturazione](#)).

Casi di studio



a



b

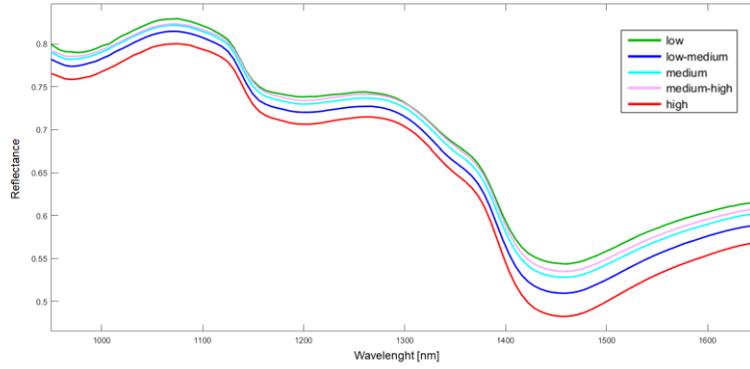
Figura 4.2.60. Variazioni in termini di riflettanza media per le classi di TSSC, rispettivamente riferite a: a) MicroNIR (950 – 1650 nm) e b) FieldSpec 4 (400 – 2500 nm).

Queste variazioni in termini di segnale sono maggiormente evidenti in PCA, come si può notare in Figura 4.2.62 e Figura 4.2.63. Effettuando un confronto tra queste ultime due figure, si può notare come gli *scores* della PCA effettuate con il set di dati del MicroNIR siano maggiormente sparpagliate rispetto a quelle effettuate con il set di dati del FieldSpec. La varianza spiegata dalla prima componente principale è circa 35-36 % per il set di dati del MicroNIR, mentre per il set di dati del FieldSpec è di circa 11-12 %.

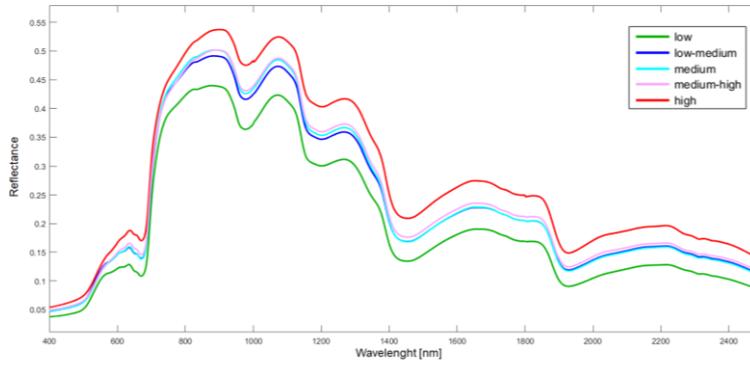
Regressioni *Partial Least Squares*

I risultati ottenuti dei modelli PLS con i dati del MicroNIR in riferimento al TSSC e al DM sono riportati rispettivamente in Figura 4.2.64a e Figura 4.2.64b. Mentre, le rette di regressione PLS con i dati del FieldSpec in riferimento al TSSC e al DM sono riportate rispettivamente in Figura 4.2.64c e Figura 4.2.64d. I risultati ottenuti dalle regressioni sono allineati con quanto ampiamente descritto in letteratura (si veda Tabella 4.2.20).

Casi di studio

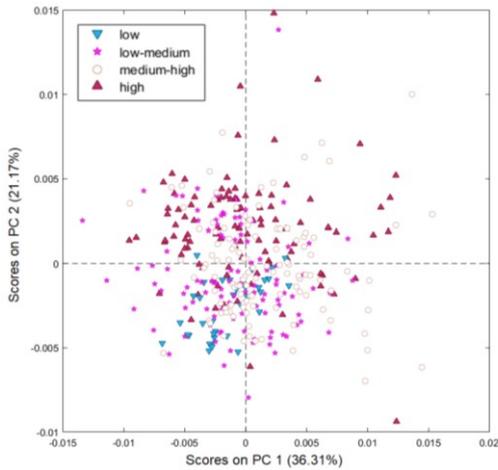


a

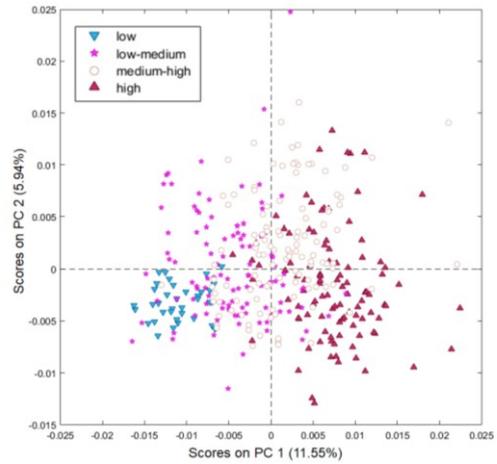


b

Figura 4.2.61. Variazioni in termini di riflettanza media per le classi di DM, rispettivamente riferite a: **a)** MicroNIR (950 – 1650 nm) e **b)** FieldSpec 4 (400 – 2500 nm).



a



b

Figura 4.2.62. PCA scores plot (Scores su PC 2 vs scores su PC 1) riferito alle analisi delle classi di TSCC per il set di dati del **a)** MicroNIR e del **b)** FieldSpec.

Casi di studio

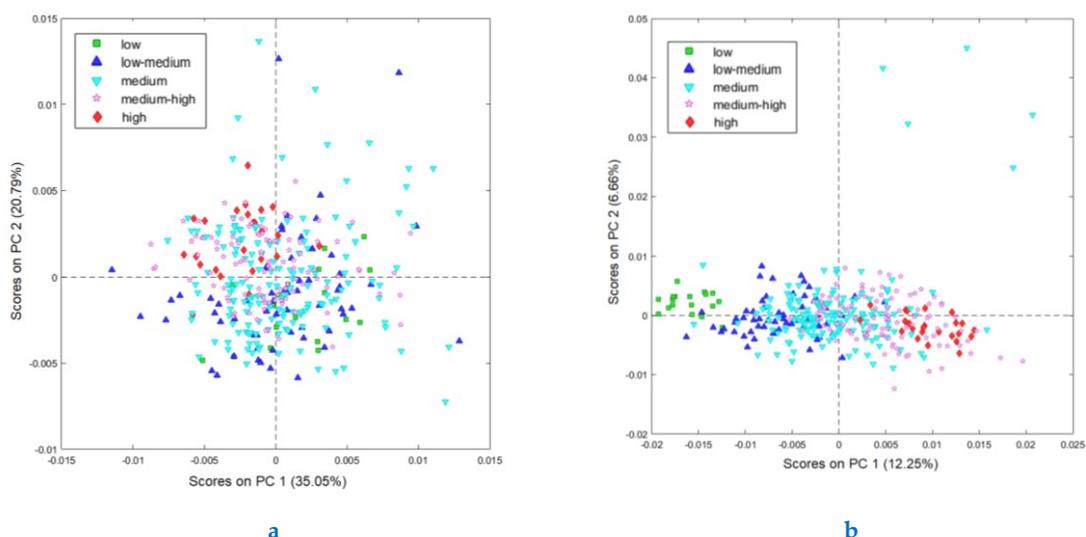


Figura 4.2.63. PCA scores plot (Scores su PC 2 vs scores su PC 1) riferito alle analisi delle classi di TSSC per il set di dati del a) MicroNIR e del b) FieldSpec.

Il TSSC viene in effetti predetto con un $R_p=0.990$ ed un $RMSEP=0.49$ °Bx nel range 408-2492 nm (come visto in Tabella 4.2.20), mentre nel caso della regressione PLS effettuata in questo studio per la previsione del grado zuccherino con il set di dati del FieldSpec (400-2500 nm), si ha un $R_p=0.922$ ed un $RMSEP=1.204$ °Bx (Tabella 4.2.23).

Tabella 4.2.23. Parametri delle regressioni PLS effettuate per la previsione del TSSC e del DM.

| Applicazione | TSSC | TSSC | DM | DM |
|------------------------------|----------------|-------------------------------|----------------|-------------------------------|
| Intervallo del set di dati Y | 9-21.5 °Bx | 9-21.5 °Bx | 16-25 % | 16-25 % |
| Strumento | JDSU MicroNIR™ | ASD FieldSpec 4® Standard-Res | JDSU MicroNIR™ | ASD FieldSpec 4® Standard-Res |
| Intervallo spettrale [nm] | 950 – 1650 | 400 - 2500 | 950 – 1650 | 400 - 2500 |
| LV | 8 | 8 | 7 | 8 |
| R^2_c | 0.580 | 0.869 | 0.582 | 0.816 |
| R^2_{cv} | 0.482 | 0.749 | 0.465 | 0.590 |
| R^2_p | 0.611 | 0.851 | 0.460 | 0.765 |
| R_p | 0.782 | 0.922 | 0.678 | 0.875 |
| RMSEC | 1.711 °Bx | 1.020 °Bx | 1.240 % | 0.815 % |
| RMSECV | 1.921 °Bx | 1.445 °Bx | 1.402 % | 1.274 % |
| RMSEP | 1.871 °Bx | 1.204 °Bx | 1.213 % | 0.791 % |

Casi di studio

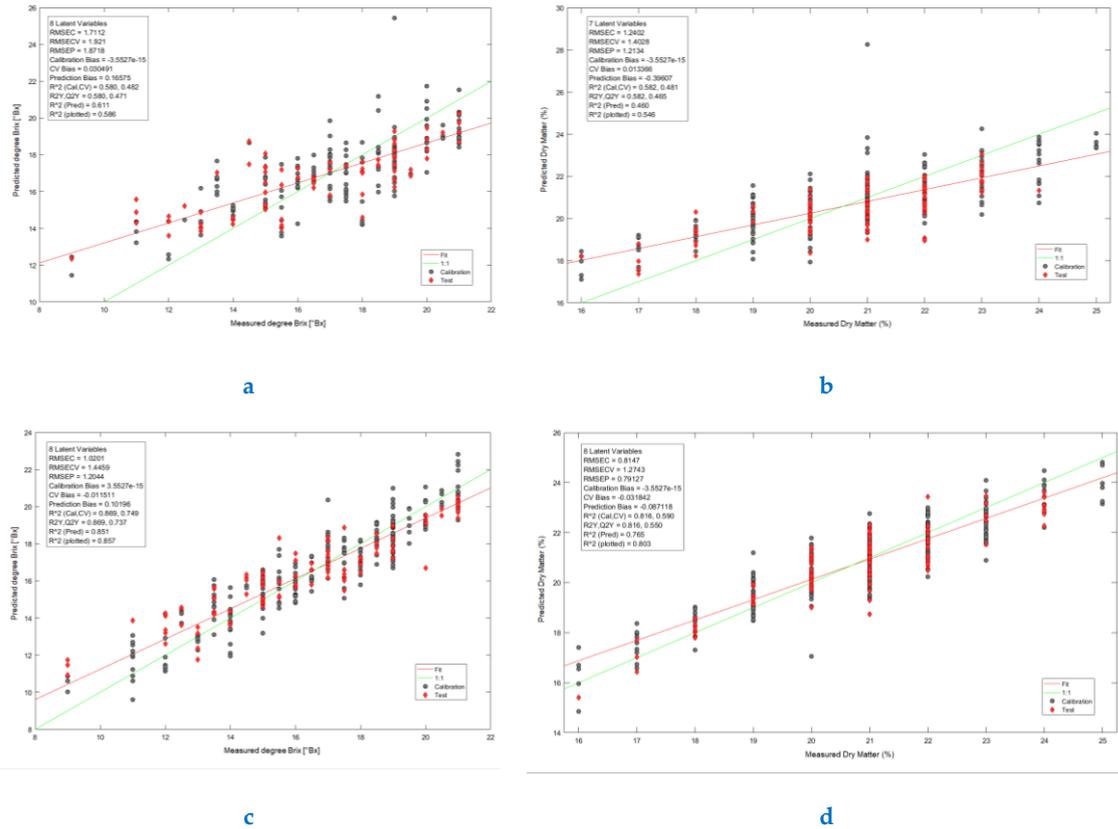


Figura 4.2.64. Valori predetti del TSSC dalla regressione PLS in funzione dei valori misurati per: a) MicroNIR™ (950-1650 nm) e b) FieldSpec 4® (400-2500 nm); valori predetti del DM dalla regressione PLS in funzione dei valori misurati per: c) MicroNIR™ (950-1650 nm) e d) FieldSpec 4® (400-2500 nm).

Discussione dei risultati

Sono stati testati due diversi spettrometri portatili al fine di dimostrare che è possibile monitorare in maniera rapida ed efficiente lo stato di maturazione di kiwi a polpa gialla intatti ("kiwi dorati").

Da quanto si è visto, gli spettrofotometri portatili potrebbero rappresentare una buona soluzione per valutare differenti parametri qualitativi del kiwi in modo rapido. Tale metodologia risulta, allo stesso tempo, affidabile come le tecniche tradizionali che vengono utilizzate; le quali, però, sono tecniche perlopiù invasive e distruttive.

I modelli utilizzati per prevedere quantitativamente il TSSC ed il DM hanno mostrato una buona capacità di prevedere i valori misurati con i metodi tradizionali e sono in linea con studi già effettuati da diversi gruppi di ricerca. La spettroscopia ottica nel visibile e nel vicino infrarosso potrebbe divenire uno strumento aggiuntivo da supporto ai produttori agricoli nel processo decisionale di raccolta, in quanto metodo affidabile e non distruttivo per la valutazione dello stato di maturazione del kiwi.

4.2.9 Valutazione di parametri qualitativi del grano tenero mediante spettrofotometria NIR: prima e dopo la macinazione

Problematica e stato dell'arte

Il frumento (*Triticum*) rappresenta uno degli alimenti di base di maggior consumo nel mondo. Questo cereale viene utilizzato come materia prima in differenti prodotti nell'industria agro-alimentari, come ad esempio la farina. La farina di grano tenero (*Triticum aestivum*) è uno dei principali ingredienti utilizzati per la panificazione. Questa può essere classificata in base al contenuto di glutine. Se il contenuto di glutine si attesta su percentuali basse (i.e. quantità minima di glutine secco del 7%), la farina di frumento viene detta farina di grano tenero. La farina di grano tenero è il principale ingrediente di svariati prodotti alimentari: i.e. torte, ciambelle, cracker, biscotti, ecc (Atwell, 2001; Chu, 2004). La qualità di tali prodotti è strettamente correlata alla qualità della farina utilizzata, la quale a sua volta è dipendente dalla qualità del grano stesso.

Le caratteristiche qualitative del grano possono essere suddivise in due gruppi: le caratteristiche fisiche e le caratteristiche chimiche. Le caratteristiche fisiche primarie sono: il contenuto di umidità, il colore e la durezza. Mentre, le caratteristiche chimiche più importanti sono: il contenuto proteico, il contenuto di glutine, il contenuto di amido ed il contenuto in ceneri (Ayman et al., 2018). La durezza del grano è uno dei parametri qualitativi più importanti ai fini della macinatura (Maghirang & Dowel, 2003). Mentre, i parametri principali che influenzano la qualità generale del cereale sono il contenuto di amido e quello di proteine. Dal punto di vista dell'imprenditore agricolo, la qualità del grano viene valutata in termini di resa, contenuto di umidità e resistenza agli stress climatici e/o idrici. Mentre, dal punto di vista del mugnaio e/o dell'addetto alla trasformazione del grano, la qualità viene valutata sulla base di parametri come: contenuto proteico, contenuto di umidità, resa in farina e performance reologiche e di panificazione. Ad ogni modo, la qualità della farina di grano tenero può variare da lotto a lotto, in quanto viene ottenuta in genere da differenti coltivazioni di frumento e a volte i rapporti nella stessa miscela di grano sono differenti (Varzakas, 2016). Ciascun produttore e/o addetto alla trasformazione del grano deve essere rifornito con frumento che rispetti degli standard qualitativi. Qualora questi standard non dovessero essere rispettati, ne potrebbe conseguire la perdita dell'intera produzione che genererebbe, a sua volta, una grande perdita in termini economici.

Ogni giorno, un impianto di macinazione del grano deve lavorare grandi quantitativi di grano. Mentre, gli impianti di panificazione industriale e di pasticceria devono interfacciarsi con più fornitori. Per questo motivo, da un lato è di necessaria importanza effettuare controlli qualitativi in maniera continuativa sui lotti di frumento che arrivano all'impianto di macinazione; mentre, dal lato dell'impianto di pianificazione è di grande importanza effettuare *quality check* su ciascun lotto di farina, per quanto riguarda parametri come: contenuto di umidità, contenuto proteico e caratteristiche reologiche.

Al giorno d'oggi, per effettuare analisi qualitative sul frumento vengono utilizzati metodi che sono prevalentemente distruttivi ed invasivi, come ad esempio metodi basati sulla chimica analitica tradizionale e/o ispezioni visive perlopiù soggettive. Questi metodi sono dispendiosi sia in termini di tempo che di costi. Da qui, la necessità dell'utilizzo di metodi che siano rapidi, economici, non distruttivi, non invasivi ed automatizzabili per effettuare stime sulla qualità del frumento (*Singh et al., 2006; Campolongo & Trenka, 2018*).

Metodi e procedure basate sulla spettroscopia NIR sono stati ampiamente utilizzati per valutare differenti parametri qualitativi su prodotti agri-food (*Bart et al., 2007; Lorente et al., 2011*). L'utilizzo della NIRS per applicazioni di controllo qualità sul frumento è stato studiato da diversi ricercatori al fine di: prevedere il contenuto proteico, il contenuto di amido ed altri parametri qualitativi (*Delwiche & Hruschka, 2000; Millar, 2003; Pojić et al., 2012; Cozzolino et al., 2014*) e per identificare la presenza di insetti e/o altri parassiti (*Dowell et al., 1998*). In letteratura, inoltre, sono riportati anche studi effettuati sulla farina di frumento al fine di prevedere: il contenuto di umidità, il contenuto in ceneri, il contenuto proteico, il contenuto di glutine ed altre caratteristiche reologiche, i.e. l'assorbimento di acqua e la stabilità dell'impasto ottenibile (*Hrušková et al., 2001; Başlar & Ertugay, 2011; Moroi et al., 2011; Dong & Sun, 2013*). Appare evidente, come il ruolo della spettroscopia NIR stia divenendo sempre più centrale nel settore cerealicolo e, più in generale, nell'industria agroalimentare. Infatti, la NIRS rappresenta una metodologia di analisi semplice, rapida ed accurata per effettuare misurazioni non-invasive e non-distruttive per molteplici attributi del frumento. Tant'è che la metodologia basata sulla NIRS è stata accettata anche dalla *International Organization for Standardization* (ISO) e da altri enti (i.e. AACC, AOAC and ICC) come standard per la valutazione di parametri qualitativi sui cereali e sui prodotti di comminazione derivati, i.e. contenuto di umidità, contenuto di grassi e contenuto di amido (*Pojić et al., 2012; International Organization for Standardization, 2017*).

In questo scenario, la spettroscopia NIR può essere vista come un metodo alternativo alle tecniche convenzionali per la valutazione di parametri qualitativi del grano e della farina di frumento.

Obiettivi dello studio

Lo scopo principale di questo studio è stato quindi quello di utilizzare la spettroscopia NIR come metodo di indagine non distruttivo per la valutazione di differenti parametri qualitativi del frumento in grani ed in farina. Per raggiungere questo scopo sono stati acquisiti spettri in riflettanza con uno spettrofotoradiometro portatile operante nel Vis-SWIR (350-2500 nm) con i quali sono stati costruiti modelli chemiometrici al fine di prevedere diversi parametri qualitativi. In particolare, sono stati messi a punto modelli *Partial least Squares* tra spettri e parametri di riferimento, al fine di prevedere: il contenuto di umidità, il contenuto proteico ed il contenuto di glutine umido.

La panoramica globale delle analisi effettuate è mostrata in [Figura 4.2.65](#).

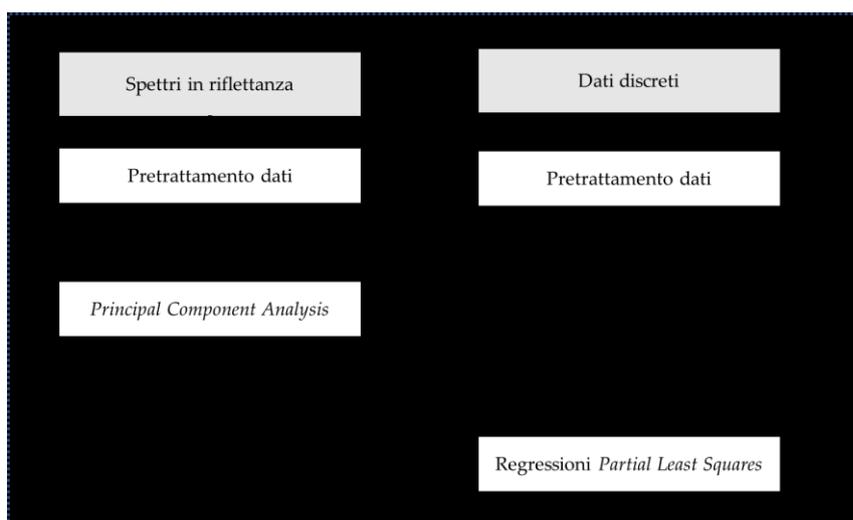


Figura 4.2.65. Diagramma di flusso delle analisi effettuate.

Procedura

Preparazione dei campioni

In questo studio sono stati analizzati campioni di frumento tenero in grani e macinato sotto forma di farina. In particolare, i campioni in grani analizzati (Figura 4.2.66) sono: a) “Grano tenero Francese”, b) “Grano tenero Romeno”, c) “Grano tenero Northern Spring”, e d) “Grano tenero Ucraino proteico”. Mentre, i corrispondenti campioni di farina (Figura 4.2.67) sono: a) “Farina Di Grano Francese”, b) “Farina Di Grano Romeno”, c) “Farina Di Grano Northern Spring” e d) “Farina Di Grano Ucraino Proteico”.



Figura 4.2.66. Campioni di frumento tenero in grani analizzati: a) “Grano tenero Francese”, b) Grano tenero Romeno”, c) “Grano tenero Northern Spring”, e d) “Grano tenero Ucraino proteico”.

Casi di studio



Figura 4.2.67. Campioni di farina di grano tenero: a) “Farina Di Grano Francese”, b) “Farina Di Grano Romeno”, c) “Farina Di Grano Northern Spring” e d) “Farina Di Grano Ucraino Proteico”.

Per ciascun campione sono stati forniti i parametri di riferimento relativi al: contenuto di umidità (12.4 ± 0.4 %), contenuto proteico (13.4 ± 1.8 %) e contenuto di glutine umido (28.8 ± 4.3 %). In Tabella 4.2.24 sono riportati i valori di riferimento per il frumento tenero in grani, mentre in Tabella 4.2.25 sono riportati i valori di riferimento per la farina di grano tenero.

Tabella 4.2.24. Parametri di riferimento per i campioni di grano tenero.

| | Grano tenero Francese | Grano tenero Romeno | Grano tenero Northern Spring | Grano tenero Ucraino proteico |
|---------------------------|--------------------------|------------------------|---------------------------------|----------------------------------|
| Moisture content (%) | 12.4 | 12.1 | 13.1 | 12.7 |
| Protein content (%) | 12.4 | 12.4 | 15 | 16 |
| Wet gluten content (%) | 24.6 | 24.5 | 30.6 | 32.8 |

Acquisizione di spettri in riflettanza, preparazione ed esportazione dati

Per ciascun campione di frumento tenero in grani e farina, sono stati acquisiti 20 spettri in riflettanza casualmente sulla superficie della capsula Petri contenente il materiale in esame (per un totale di 160 spettri). Lo strumento utilizzato per le acquisizioni è lo spettrofotoradiometro portatile ASD FieldSpec 4[®] Standard-Res (350 – 2500 nm), con l’ausilio della sonda a contatto.

Tabella 4.2.25. Parametri di riferimento per i campioni di farina di grano tenero.

| | Farina di grano Francese | Farina di grano Romano | Farina di grano Northern Spring | Farina di grano Ucraino proteico |
|---------------------------|-----------------------------|---------------------------|------------------------------------|-------------------------------------|
| Moisture content (%) | 12.1 | 11.9 | 12.5 | 12.6 |
| Protein content (%) | 11.3 | 11.3 | 14.5 | 14.5 |
| Wet gluten content (%) | 25.3 | 25.4 | 33.7 | 33.8 |

Mentre, gli spettri acquisiti con ASD FieldSpec 4[®] immagazzinati in singoli file in formato “.asd” sono stati convertiti in un singolo file con formato ASCII “.txt”, mediante l’utilizzo del software ViewSpec Pro (Ver. 6.2.0.). Il file così ottenuto è stato importato in ambiente MATLAB[®] (MATLAB R2018a ver. 9.4.0.813654), per le successive analisi.

I dati importati in ambiente MATLAB[®] sono stati analizzati con il supporto del toolbox PLS_toolbox (ver. 8.2.1) della Eigenvector Research, Inc. Su ciascun dataset sono state impostate delle etichette (o *label*) sui campioni di riferimento.

Pretrattamento dati ed analisi esplorativa

La pre-elaborazione dei dati è stata necessaria per esaltare le differenze spettrali presenti tra classi differenti e rimuovere eventuale rumore, fenomeni di *offset* e altri artefatti dovuti a fenomeni di *scattering* e/o ad altri fenomeni fisici (Wise et al., 2006; Wise et al., 2008; Rinnan et al., 2009).

La combinazione di algoritmi di pre-elaborazione utilizzati per i dati spettrali, a seguito della *Splice correction*, è stata la seguente: *Norris-Williams (NW) Gap Segment 2nd Derivative*, *Standard Normal Variate (SNV)* e *Mean Center (MC)*.

Come tecnica di analisi esplorativa è stata scelta l’analisi in componenti principali (Wold et al., 1987). In particolare, per ciascun set di dati è stato analizzato il raggruppamento degli *scores* in PCA, in accordo con le classi relative alla tipologia di campione analizzato e per l’identificazione di eventuali *outliers*. Gli *outliers* sono stati identificati ed esclusi dai dataset per effettuare le successive analisi.

Modelli di regressione

Per valutare la correlazione tra gli spettri acquisiti (X) ed i parametri di riferimento considerati (Y) è stata utilizzata la tecnica di regressione ai minimi quadrati parziali (PLS). È stata calibrata una retta di regressione per ciascuno dei parametri considerati. I modelli PLS (Geladi et al., 1986; Wold et al., 2001; Abdi, 2003) sono stati effettuati per:

- i) Spettri di riflettanza e parametri di riferimento del frumento tenero in grani (80 spettri);
- ii) Spettri in riflettanza e parametri di riferimento della farina di grano tenero (80 spettri);
- iii) Spettri in riflettanza e parametri di riferimento del frumento tenero in grani e della farina (160 spettri).

Per ciascun dataset, la scelta del set di addestramento del modello ed il set di validazione dello stesso è stato scelto applicando l'algoritmo di *Kennard-Stone* (*Kennard & Stone, 1969*): il 70% dei dati contenuti nel dataset iniziale sono stati utilizzati per la calibrazione e la convalida incrociata del modello, mentre il rimanente 30% è stato utilizzato come set di validazione. Come algoritmo di *cross-validation* è stato utilizzato *il Venetian Blinds* (*Eigenvector Research Inc. 2016b*), al fine di scegliere la complessità ottimale di ciascun modello e scegliere il giusto numero di variabili latenti.

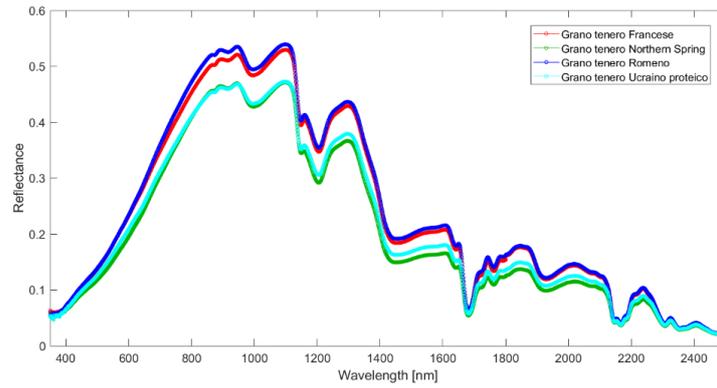
Analisi dei risultati

Spettri medi in riflettanza ed analisi in componenti principali

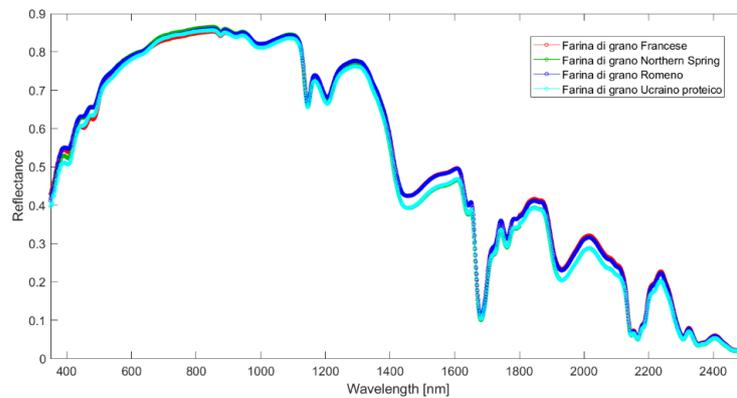
Il confronto tra la variazione degli spettri medi in riflettanza del frumento tenero in grani e quella della farina di grano tenero secondo la propria tipologia è riportata in [Figura 4.2.68](#). Come si può notare in [Figura 4.2.69](#), le principali somiglianze tra gli spettri medi del frumento tenero in grani e quella della farina di grano tenero sono riconducibili ad i minimi locali (assorbimenti) negli intorni di: 872 nm, 919 nm, 998 nm, 1148 nm, 1205 nm, 1454 nm, 1683 nm, 1762 nm, 1931 nm, 2161 nm, 2306 nm, 2357 nm e 2473 nm. Gli spettri medi pre-elaborati sono riportati nella sezione "Allegati" (*Allegato XVII- Case study – Grano e farine*).

I grafici dei punteggi delle PCA eseguite di chicchi di grano tenero e farina di grano tenero sono mostrati, rispettivamente, in [Figura 4.2.70a](#) e [Figura 4.2.70b](#). Come si può vedere dai grafici degli *scores*, la PCA è in grado di raggruppare automaticamente le diverse tipologie di frumento tenero in grani disposto in capsula di Petri (come mostrato nella [Figura 4.2.70a](#)), sebbene i diversi punteggi siano più sparpagliati rispetto a quanto accade, invece, per i punteggi della PCA effettuata sulle farine ([Figura 4.2.70b](#)). La scomposizione dei dati degli spettri in riflettanza della farina di grano tenero ha mostrato, in effetti, una migliore capacità di auto-raggruppamento. La PCA applicata sui set di dati uniti ([Figura 4.2.71](#)) mostra, invece, la presenza di due grandi *cluster* riferibili alla dimensione delle particelle analizzate (grani e farine). Pertanto, i punteggi della stessa varietà non si sovrappongono.

Casi di studio



a



b

Figura 4.2.68. Variazione degli spettri medi in riflettanza del frumento tenero in grani (a) e della farina di grano tenero (b), secondo la propria tipologia.

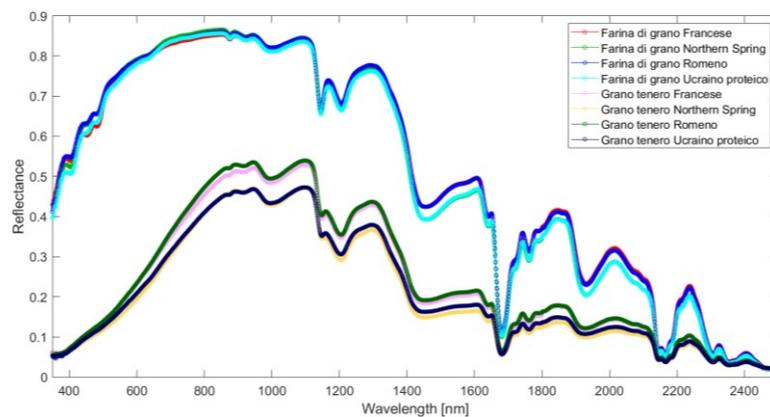


Figura 4.2.69. Confronto tra la variazione degli spettri medi in riflettanza del frumento tenero in grani e della farina di grano tenero, secondo la propria tipologia.

Casi di studio

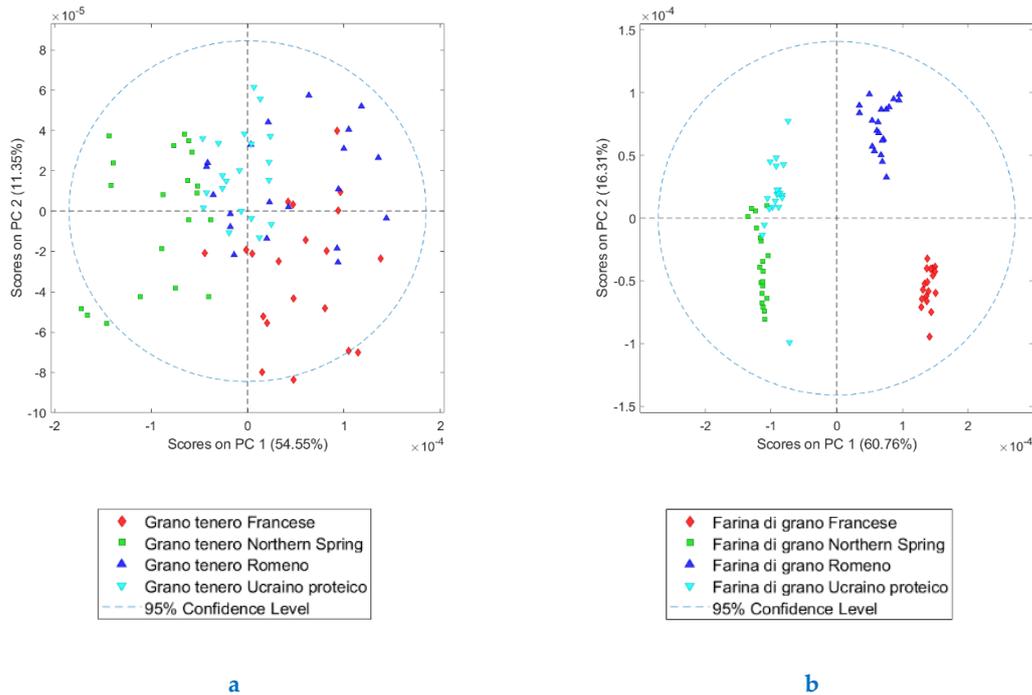


Figura 4.2.70. Confronto tra gli scores plot (scores su PC2 vs. scores su PC1) degli spettri di riflettanza del frumento tenero in grani (a) e delle rispettive farine (b).

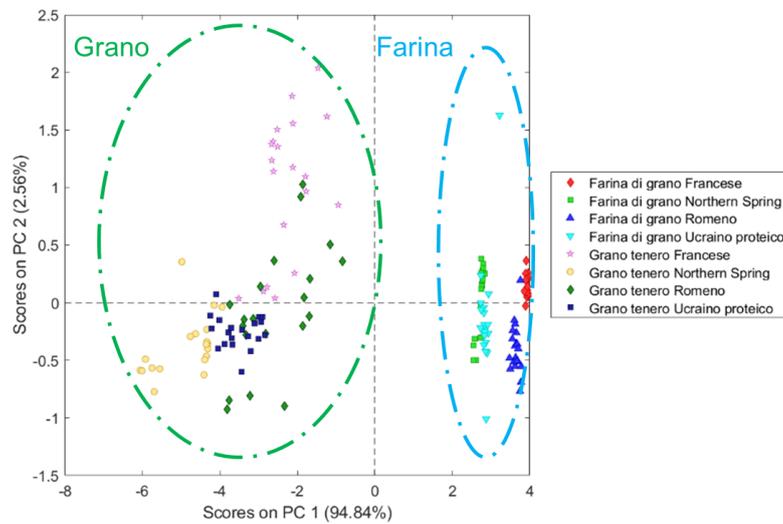


Figura 4.2.71. Scores plot (scores su PC2 vs. scores su PC1) degli spettri di riflettanza del frumento tenero in grani insieme a quelli delle rispettive farine. In evidenza la presenza di due grandi cluster riferibili alla dimensione delle particelle analizzate (grano e farina).

Regressioni Partial Least Squares

In Tabella 4.2.26 sono riassunti i parametri di performance ed i risultati delle regressioni PLS. In Figura 4.2.72 e Figura 4.2.73 sono riportate le restituzioni grafiche delle rette di regressione, rispettivamente per il frumento tenero in grani ed in farina. Mentre,

le restituzioni grafiche dei modelli di regressione calibrati utilizzando gli spettri del frumento tenero in grani ed in farina insieme sono riportati in [Figura 4.2.74](#).

Si può dire che le performance dei modelli di regressione del frumento tenero in grani e quello delle corrispettive farine sono altresì comparabili (si veda anche [Allegato XVII-Case study – Grano e farine](#)), sia in termini di RMSE che di R^2 . Globalmente, le regressioni effettuate per il frumento tenero in grani sono leggermente meno performanti rispetto a quelle effettuate per le farine. Ad ogni modo, utilizzando gli spettri sia dei chicchi che delle farine si sono ottenuti ottimi risultati in previsione per: contenuto di umidità (R^2 in previsione = 0.928; RMSEP = 0.2 %), contenuto proteico (R^2 in previsione = 0.908; RMSEP = 0.6 %) e contenuto di glutine umido (R^2 in previsione = 0.888; RMSEP = 1.3 %).

Discussione dei risultati

Dall'analisi in componenti principali è stato possibile valutare la tipologia di grano/farina considerata e la granulometria del prodotto. Inoltre, i risultati delle regressioni di PLS hanno mostrato prestazioni promettenti sia per la previsione degli attributi di qualità del grano dagli spettri di riflettanza acquisiti sul grano in chicchi che sulla farina. Da quanto visto, si è evinto come sia possibile prevedere il contenuto d'umidità, le proteine e il contenuto di glutine umido mediante l'utilizzo della spettrofotometria, operando nell'intervallo 350-2500 nm.

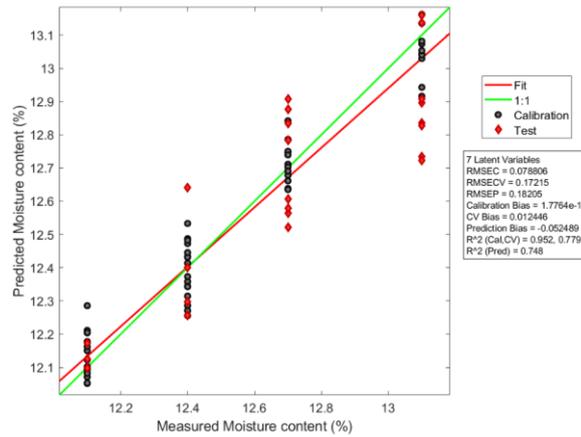
Seguendo questo approccio, potrebbe essere possibile definire strategie innovative per valutare diversi parametri del grano tenero da poter inserire nelle procedure processo di controllo qualità e per applicazioni di tracciabilità del prodotto. I.e. dal punto di vista dell'imprenditore agricolo per monitorare i lotti di grano da inviare a trasformazione e, dal lato dell'impianto di macinazione / trasformazione e/o impianto di panificazione per effettuare monitoraggi in tempo reale sui lotti forniti e valutarne la qualità di lavorazione. I risultati confermano che, utilizzando uno spettrofotometro portatile Vis-SWIR, i parametri di qualità del grano tenero possono essere valutati in modo più rapido rispetto alle tecniche tradizionali.

Casi di studio

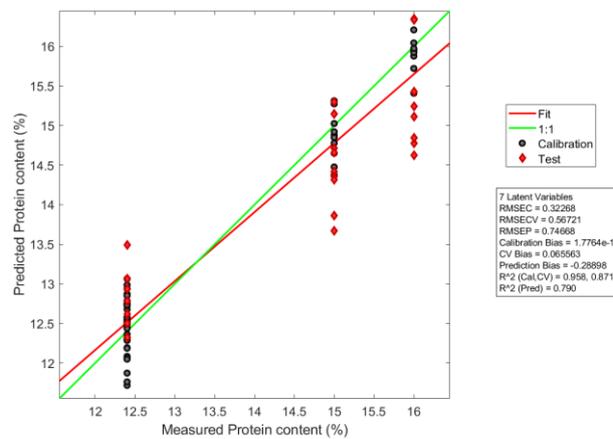
Tabella 4.2.26. Parametri di riferimento e parametri relativi alle regressioni PLS per: (*) frumento tenero in grani, (**) farina di grano tenero e (***) frumento tenero in grani e rispettive farine. LV = Latent Variables; RMSEC = Root Mean Squared Error Calibration; RMSECV = Root Mean Squares Error Cross Validation; RMSEP = Root Mean Squared Error Prediction; R² (Cal) = coefficiente di determinazione in calibrazione; R² (CV) = coefficiente di determinazione in cross-validation; R² (Pred) = coefficiente di determinazione in predizione.

| | | Contenuto di umidità (%) | Contenuto di proteine (%) | Contenuto di glutine umido (%) |
|---|------------------------|-----------------------------|------------------------------|-----------------------------------|
| <i>Parametri di riferimento</i> | Media | 12.4 | 13.4 | 28.8 |
| | Deviazione standard | 0.4 | 1.8 | 4.3 |
| <i>Parametri dei modelli di regressione *</i> | LV | 7 | 7 | 8 |
| | RMSEC | 0.1 | 0.3 | 0.8 |
| | RMSECV | 0.2 | 0.6 | 1.8 |
| | RMSEP | 0.2 | 0.7 | 1.7 |
| | R ² (Cal) | 0.952 | 0.958 | 0.945 |
| | R ² (CV) | 0.779 | 0.871 | 0.762 |
| | R ² (Pred) | 0.748 | 0.790 | 0.792 |
| <i>Parametri dei modelli di regressione **</i> | LV | 6 | 3 | 5 |
| | RMSEC | 0.0 | 0.2 | 0.4 |
| | RMSECV | 0.1 | 0.2 | 0.5 |
| | RMSEP | 0.0 | 0.2 | 0.4 |
| | R ² (Cal) | 0.988 | 0.986 | 0.991 |
| | R ² (CV) | 0.969 | 0.982 | 0.985 |
| | R ² (Pred) | 0.986 | 0.990 | 0.991 |
| <i>Parametri dei modelli di regressione ***</i> | LV | 7 | 6 | 6 |
| | RMSEC | 0.1 | 0.3 | 0.9 |
| | RMSECV | 0.1 | 0.5 | 1.2 |
| | RMSEP | 0.1 | 0.6 | 1.3 |
| | R ² (Cal) | 0.908 | 0.956 | 0.955 |
| | R ² (CV) | 0.843 | 0.909 | 0.909 |
| | R ² (Pred) | 0.928 | 0.908 | 0.888 |

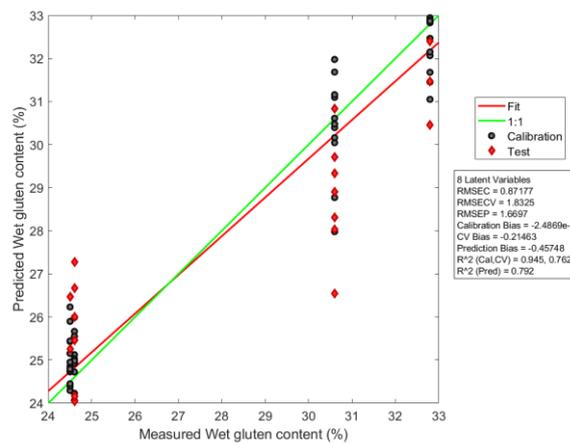
Casi di studio



a



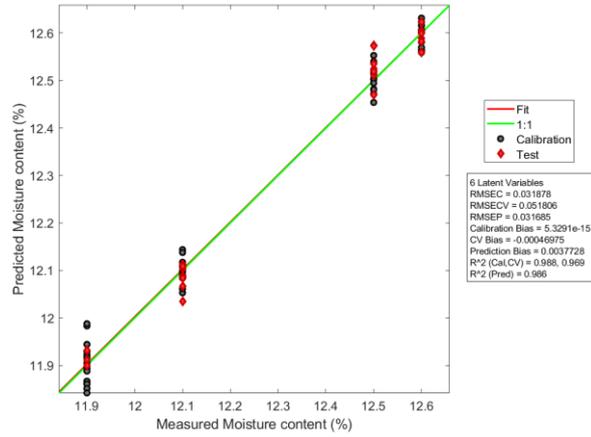
b



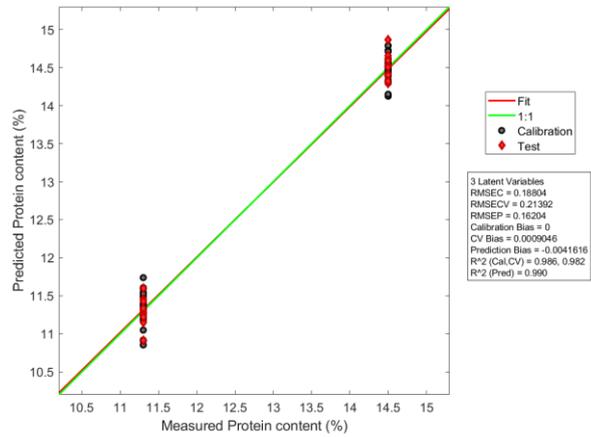
c

Figura 4.2.72. Risultati dei modelli di regressione riferiti al frumento tenero in grani per: a) contenuto di umidità, b) contenuto proteico e c) contenuto di glutine umido.

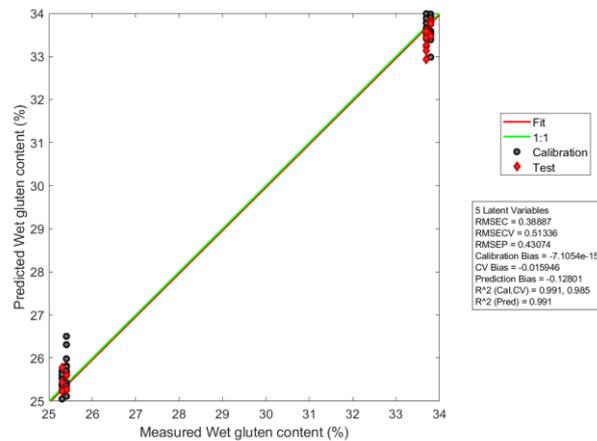
Casi di studio



a



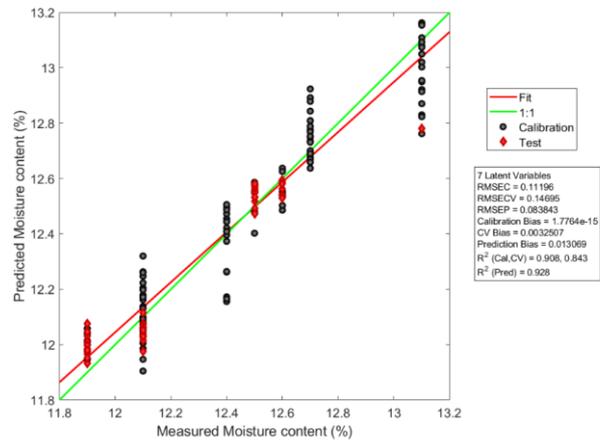
b



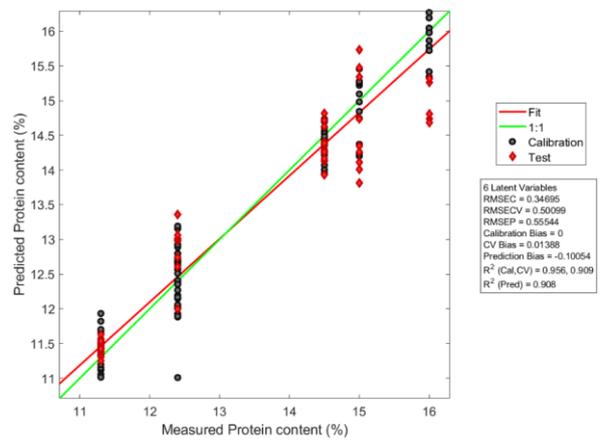
c

Figura 4.2.73. Risultati dei modelli di regressione riferiti alla farina di grano tenero per: a) contenuto di umidità, b) contenuto proteico e c) contenuto di glutine umido.

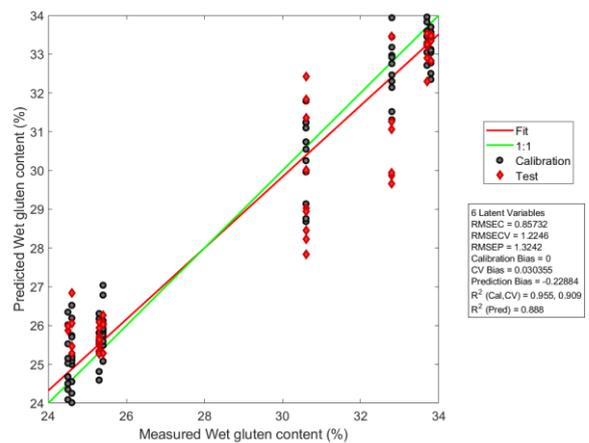
Casi di studio



a



b



c

Figura 4.2.74. Risultati dei modelli di regressione riferiti al frumento tenero in grani ed in farina per: a) contenuto di umidità, b) contenuto proteico e c) contenuto di glutine umido.

5 CONSIDERAZIONI FINALI E PROSPETTIVE FUTURE

L'attività di ricerca presentata in questa tesi dimostra come sia possibile ottenere importanti informazioni circa la caratterizzazione di materie prime secondarie e prodotti agroalimentari, mediante tecniche spettroscopiche supportate da indagini chemiometriche. Particolare attenzione è stata rivolta all'utilizzo di procedure innovative e allo sviluppo di appropriati algoritmi di statistica multivariata, *machine learning* e *computer vision* a complemento dell'utilizzo di tecniche di spettroscopia classica ed iperspettrale.

Le metodiche analitiche impiegate hanno riguardato l'uso della spettroscopia, quale tecnologia non distruttiva potenzialmente impiegabile nella selezione on-line e come tecnica di indagine speditiva. L'analisi di immagine iper-spetttrale (*HyperSpectral Imaging: HSI*) e la spettroscopia puntuale, in combinazione con altre tecniche d'indagine, consentono l'implementazione di una metodologia di caratterizzazione rapida, non distruttiva e non invasiva. I dati ottenuti sono stati sottoposti ad analisi chemiometrica con l'obiettivo di sviluppare ed ottimizzare la capacità discriminante delle funzioni di classificazione testate durante la sperimentazione e la capacità predittiva delle risposte spettrali.

Tali tecniche consentono l'individuazione delle procedure applicabili più consone per il controllo qualità e/o trattamento da utilizzare per le materie prime ed i prodotti presi in esame.

Dai casi applicativi analizzati e da una revisione sistematica della letteratura, si è visto come le tecniche basate sull'HSI siano estremamente versatili e scalabili per applicazioni di controllo qualità e classificazione *on-line* (i.e. come tecnologia di *sensing* per macchine selezionatrici), sia per materie prime che per prodotti agro-alimentari. L'HSI può essere considerato uno strumento valido per caratterizzare e monitorare differenti tipologie di materiali. Il grande potenziale di questa metodologia di analisi viene supportata anche dalle continue innovazioni in campo della sensoristica, nell'evoluzione tecnologica degli strumenti di acquisizione e di calcolatori con capacità sempre più prestanti. In questo contesto, l'utilizzo di approcci chemiometrici e di *image processing* per queste tipologie di dati diventa essenziale ai fini dell'analisi delle variabili che entrano in gioco e per estrarre il segnale utile (i.e. la "*fingerprint*") dal rumore.

Si è visto, inoltre, come la spettroscopia Vis-NIR, presenta numerosi vantaggi rispetto all'utilizzo di tecniche analitiche tradizionali (i.e. costi inferiori, dimensioni ridotte della strumentazione, rapidità di analisi, possibilità di effettuare analisi non distruttive e non invasive anche in situ e possibilità di ottenimento di risultati multi-parametrici), soprattutto per quanto riguarda la valutazione di parametri qualitativi di prodotti agro-alimentari.

Ad ogni modo, nell'industria agro-alimentare, vi sono numerosi fattori che ad oggi limitano la diffusione dell'utilizzo della spettroscopia Vis-NIR a tutti i livelli della catena di produzione, trasformazione e distribuzione. Ad esempio, il processo di sviluppo di

modelli di previsionali e/o classificazione richiede numerose risorse ed un'analisi molto approfondita. Inoltre, per applicazioni di tipo industriale è richiesto un set di dati di calibrazione robusto e sempre aggiornato. Senza considerare che la tecnologia degli spettrometri NIR è limitata nella rilevazione di composti presenti in basse concentrazioni (i.e. 0,1 %) e soffre delle interferenze dovute alle condizioni ambientali differenti (i.e. luce e umidità), le quali il ricercatore deve in qualche modo compensare in situ. A questi problemi, si affianca, quindi, la necessità di personale altamente specializzato, adeguatamente formato all'utilizzo degli strumenti e all'interpretazione dei risultati. Non per ultima di importanza, vi è inoltre la questione legata ai costi, che sono una limitazione soprattutto nella diffusione delle strumentazioni *in-line* nell'ambito industriale.

Il numero di studi pubblicati in questo settore è notevolmente aumentato negli ultimi due decenni; tuttavia, è necessario ancora molto lavoro per portare questa tecnologia allo stato dell'arte. Sebbene i casi di studio descritti in questa tesi e la letteratura consultata esaltino i benefici strumenti portatili e delle tecnologie HSI, è assolutamente chiaro che c'è ancora molto lavoro da fare per rendere questi strumenti facilmente utilizzabili sul campo in condizioni di produzione ed in tutti gli step della filiera, fino alla vendita, per ottenere garanzie di qualità del prodotto sempre maggiori. In questo contesto, andranno portate avanti altre sperimentazioni per approfondire le performances degli strumenti.

Le prospettive di sviluppo di questa attività di ricerca risiedono nell'ottimizzazione delle informazioni ottenibili dall'applicazione di tecniche spettroscopiche, nonché di correlare i risultati delle analisi distruttive e non distruttive con i parametri caratteristici del materiale analizzato.

RIFERIMENTI BIBLIOGRAFICI

- A.KRÜSS Optronic. (2013). *kruess*. Tratto da "Measuring ripeness in the fruit and vegetable industry - Measuring Brix using refractometry – Application report":
http://www.kruess.com/documents/Applikationsberichte/Brixmessung_Refraktometrie_en.pdf
- Abdelbasir, S. M., Hassan, S. S., H. Ayman, K. & El-Nasr, R. S. (2018). Status of electronic waste recycling techniques: a review. *Environmental Science and Pollution Research*, 25(17), 16533–16547. doi:<https://doi.org/10.1007/s11356-018-2136-6>
- Abdi, H. (2003). Partial Least Squares (PLS) Regression. In Lewis-Beck, M. & Bryman, A., *Encyclopedia of Social Sciences Research Methods*. Thousand Oaks (CA): Sage.
- Agelet, L., Ellis, D. D., Duvick, S., Goggi, A. S., Hurburgh, C. R. & Gardner, C. A. (2012). Feasibility of near infrared spectroscopy for analyzing corn kernel damage and viability of soybean and corn kernels. *Journal of Cereal Science*, 55, 160–165.
- Agresti, G., Bonifazi, G., Calienno, L., Capobianco, G., Lo Monaco, A., Pelosi, C., . . . Serranti, S. (2013). Surface Investigation of Photo-Degraded Wood by Colour Monitoring, Infrared Spectroscopy, and Hyperspectral Imaging. *Journal of Spectroscopy*. doi:<http://dx.doi.org/10.1155/2013/380536>
- Al-Amery, M., Geneve, R., Sanches, M., Armstrong, P., Maghirang, E., Lee, C., . . . Hildebrand, D. (2018). Near infrared spectroscopy used to predict soybean seed germination and vigour. *Seed Science Research*, 1-8. doi:<https://doi.org/10.1017/S0960258518000119>
- Alander, J. T., Bochko, V., Martinkauppi, B., Saranwong, S., & Mantere, T. (2013). A Review of Optical Nondestructive Visual and Near-Infrared Methods for Food Quality and Safety. *International Journal of Spectroscopy*, 2013. doi:<https://doi.org/10.1155/2013/341402>
- Amigo, J. M., Martí, I. & Gowen, A. (2013). Chapter 9 - Hyperspectral Imaging and Chemometrics: A Perfect Combination for the Analysis of Food Structure, Composition and Quality. *Data Handling in Science and Technology*, 28, 343-370. doi:<https://doi.org/10.1016/B978-0-444-59528-7.00009-0>
- Andersen, C. & Bro, R. (2010). Variable selection in regression - a tutorial. *Journal of chemometrics*, 24, 728–737. doi: <https://doi.org/10.1002/cem.1360>
- Arazuri, S., Jarén, C., & Arana, J. I. (2005). Selection of the Temperature in the Sugar Content Determination of Kiwi Fruit. *International Journal of Infrared and Millimeter Waves*, 26(4), 607–616. doi:10.1007/s10762-005-4076-8
- ASD Inc. (2008). *RS3™ User Manual, ASD Document 600545, Rev. E*.
- ASD Inc. (2015a). *FieldSpec® 4 Quick Start Guide, ASD Document 600978, Rev. C*.
- ASD Inc. (2015b). *FieldSpec® 4 User Manual, ASD Document 600979, Rev. D*.
- Atwell, W. A. (2001). CHAPTER 7: Products from Soft Wheat Flour: Problems, Causes, and Resolutions. *Wheat Flour*, 97–113. doi:10.1094/189112725x.007
- Ayman, I., Adrienne, C. V., Márton, J. & Ferenc, S. (2018). Applying Infrared Technique as a Nondestructive Method To Assess Wheat Grain Hardness. *International Journal of Science and Qualitative Analysis*, 4(3), 100-107. doi:10.11648/j.ij.sqa.20180403.16
- Bae, M., Han, E. & Hong, S. (1998). Use of near infrared spectroscopy in quality control of red pepper powder. *Near Infrared Spectroscopy*, 6, A333–A335.

Riferimenti bibliografici

- Ballabio, D. & Consonni, V. (2013). Classification tools in chemistry. Part 1: linear models. PLS-DA. *Analytical Methods*, 5, 3790–3798.
- Barker, M. & Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics*, 17, 166-173. doi: DOI: 10.1002/cem.785
- Barnes, R. J., Dhanoa, M. S. & Lister, S. J. (1989). Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy*, 43(5), 772–777. doi:https://doi.org/10.1366/0003702894202201
- Bart, M. N., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I. & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology*, 46(2), 99–118. doi:https://doi.org/10.1016/j.postharvbio.2007.06.024
- Başlar, M. & Ertugay, M. F. (2011). Determination of protein and gluten quality-related parameters of wheat flour using near-infrared reflectance spectroscopy (NIRS). *TÜBİTAK, Turk J Agric For*, 35, 139-144. doi:10.3906/tar-0912-507
- Beebe, K. R., Pell, R. J. & Seasholtz, M. B. (1998). *Chemometrics: A Practical Guide*. Wiley-Interscience.
- Berzaghi, P., Serva, L., Piombino, M., Mirisola, M. & Benozzo, F. (2005). Prediction Performances of Portable Near Infrared Instruments for at Farm Forage Analysis. *Ital. J. Anim. Sci.*, 4(Suppl. 3), 145-147.
- Bevilacqua, M., Bucci, R., Materazzi, S. & Marini, F. (2013). Application of near infrared (NIR) spectroscopy coupled to chemometrics for dried egg-pasta characterization and egg content quantification. *Food Chemistry*, 140, 726–734. doi: https://doi.org/10.1016/j.foodchem.2012.11.018
- Bionutrient Food Association. (2017). *bionutrient.org*. Tratto da Brix: <http://bionutrient.org/site/bionutrient-rich-food/brix>
- Blanco, M., & Villaroya, I. (2002). NIR spectroscopy: a rapid-response analytical tool. *TrAC Trends in Analytical Chemistry*, 21(4), 240-250. doi:10.1016/s0165-9936(02)00404-1
- Bonifazi, G. & Serranti, S. (2006a). Hyperspectral imaging based techniques in fluff characterization. *Proceedings of SPIE Optic East 2006, Advanced Environmental, Chemical, and Biological Sensing Technologies IV(6377)*. doi:10.1117/12.684661
- Bonifazi, G. & Serranti, S. (2006b). Imaging spectroscopy based strategies for ceramic glass contaminants removal in glass recycling. *Waste Management*, 26, 627-639.
- Bonifazi, G. & Serranti, S. (2007). Hyperspectral imaging based procedures applied to bottom ash characterization. *Optic East 2007, Proceedings of SPIE, Advanced Environmental, Chemical and Biological Sensing Technologies V(6755)*.
- Bonifazi, G. & Serranti, S. (2010). Hyperspectral Sensing Techniques Applied to Bio-masses Characterization: The Olive Husk Case. In D. Li, Y. Liu, & C. Y., *Computer and Computing Technologies in Agriculture IV : 4th IFIP TC 12 Conference, CCTA 2010* (Vol. Selected Papers, Part IV , p. 751-764). Nanchang, China. doi:10.1007/978-3-642-18369-0_89
- Bonifazi, G., Capobianco, G. & Serranti, S. (2018a). Asbestos containing materials detection and classification by the use of hyperspectral imaging. *Journal of hazardous materials*, 344, 981-993.
- Bonifazi, G., Gasbarrone, R. & Serranti, S. (2018b). Characterization of printed circuit boards from e-waste byproducts for copper beneficiation. *Proceeding of ECOMONDO 2018*, 65-69.
- Bonifazi, G., Gasbarrone, R. & Serranti, S. (2018c). Il riciclo meccanico dei rifiuti di apparecchiature elettriche ed elettroniche: una sfida tecnologica. In A. Budoni, M. Martone, & S. Zerunian, *La Bioregione Pontina:*

Riferimenti bibliografici

- esperienze, problemi, linee di ricerca per scenari di futuro*. Edizioni SdT, Firenze. Tratto da <http://www.societadeiterritorialisti.it/2016/01/17/collana-qricerche-e-studi-territorialistiq/>
- Bonifazi, G., Palmieri, R. & Serranti, S. (2015a). Short wave infrared hyperspectral imaging for recovered postconsumer single and mixed polymers characterization. In R. Widenhorn, & A. Dupret, *Image Sensors and Imaging Systems 2015, Proc. of SPIE-IS&T Electronic Imaging*, 9403. doi:10.1117/12.2081362
- Bonifazi, G., Palmieri, R. & Serranti, S. (2015b). Hyperspectral imaging applied to end-of-life (EOL) concrete recycling. *tm-Technisches Messen*, 82(12), 616–624.
- Bonifazi, G., Serranti, S. & Menesatti, P. (2005). Hyperspectral imaging based techniques in ornamental stone characterization. *Optical Sensors and Sensing Systems for Natural Resources and Food Safety and Quality*.
- Bonifazi, G., Serranti, S., Bonoli, A. & Dall'Ara, A. (2009). Innovative recognition-sorting procedures applied to solid waste: the hyperspectral approach. *Sustainable Development and Planning, IV(2)*, 885-889.
- Brown, S. D., Tauler, R. & Walczak, B. (2009). *Comprehensive Chemometrics - Volume 1: Chemical and Biochemical Data Analysis*. Elsevier Science.
- Bruker Corporation. (2009). *Physical Principles of Micro-X-ray Fluorescence*.
- Bruker Corporation. (2013). *M4 TORNADO - User manual*.
- Bruker Corporation. (2015). *M4 TORNADO - High performance Micro-XRF spectrometer*. Tratto da: https://www.bruker.com/fileadmin/user_upload/8-PDF-Docs/X-rayDiffraction_ElementalAnalysis/mXRF/Brochures/Bro_m4_tornado_8p_en_rev3_3_lores.pdf
- Bruni, G. (1964). *Chimica generale e inorganica*. Milano, Tamburini.
- Burger, J. (2009). Replacement of Hyperspectral Image Bad Pixels. *NIR news*, 20(7), 19-21. doi:<https://doi.org/10.1255/nim.1151>
- Burgers, J. & Gowen, A. (2011). Data handling in hyperspectral image analysis. *Chemometrics and Intelligent Laboratory Systems*, 108(1), 13-23. doi:10.1016/j.chemolab.2011.04.001
- Burse, M. M. (2017). A Brief History of Spectroscopy. *AccessScience*. doi:10.1036/1097-8542.BR0213171.
- Campolongo, G. & Trenka, E. (2018). *Application note: Rapid assessment of soft wheat flour - Quality parameters by Near Infrared Spectroscopy*. Tratto da BÜCHI Labortechnik AG, Switzerland: https://static1.buchi.com/sites/default/files/Barilla_White_Paper_Flour_.pdf
- Camps, C., Simone, C. & Gilli, C. (2012). Assessment of Tomato Quality Using Portable NIR Spectroscopy and PLSR with Wavelengths Selection. *Acta Hort.*, 936, 437-442.
- Cantin, C. M., Soto, A., Crisosto, G. M. & Crisosto, C. H. (2011). Evaluation of a non-destructive dry matter sensor for kiwifruit. *Acta Horticulturae(913)*, 627–632. doi:10.17660/actahortic.2011.913.86
- Capobianco, G., Bracciale, M. P., Sali, D., Sbardella, F., Belloni, P., Bonifazi, G., . . . Mariangela. (2017). Chemometrics approach to FT-IR hyperspectral imaging analysis of degradation products in artwork cross-section. *Microchemical Journal*, 132, 69-76.
- Capobianco, G., Prestileo, F., Serranti, S. & Bonifazi, G. (2015). Hyperspectral imaging-based approach for the in-situ characterization of ancient Roman wall paintings. *Periodico di Mineralogia, Special issue (3A)*, 407-418.
- Cavicchi, L. (2002). *Ortofrutta, per la qualità è efficace l'analisi sensoriale*. Tratto da Centro ricerche produzioni vegetali: <http://www.crpv.it/doc/5169/Ortofrutta.pdf>
- Cayuela, J. & Weiland, C. (2010a). Intact Orange Quality Prediction with Two Portable NIR Spectrometers. *Postharvest Biol. Technol.*, 58(2), 113-120.

Riferimenti bibliografici

- Cayuela, J., Del Carmen, M. & Camino, P. (2010b). Prediction of Quality of Intact Olives by Near Infrared Spectroscopy. *Eur. J. Lipid Sci. Tech.*, 112(11), 1209-1217.
- Cen, H. & He, Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science & Technology*, 18(2), 72-83.
doi:<https://doi.org/10.1016/j.tifs.2006.09.003>
- Chancerel, P. & Rotter, V. S. (2009). Recycling-oriented characterization of small waste electrical and electronic equipment. *Waste Management*, 29, 2336–2352.
- ChemBuddy. (2011). *Abbé refractometer*. Tratto da Refractometer.pl.: <http://www.refractometer.pl/Abbe-refractometer>
- Christen, D., Camps, C., Summermatter, A., Gabioud Rebeaud, S. & Baumgartner, D. (2012). Prediction of the Pre- and Postharvest Apricot Quality with Different VIS/NIRS Devices. *Acta Hortic.*, 966, 149-153.
- Chu, M. (2004). *Wheat Flour*. Tratto da Cooking for Engineers :
<http://www.cookingforengineers.com/article/63/Wheat-Flour>
- Cocchi, M., Seeber, R. & Ulrici, A. (2001). WPTER: wavelet packet transform for efficient pattern recognition of signals. *Chemometrics and Intelligent Laboratory System*, 57, 97–119.
- Cortes, C. & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.
- Costa, G., Noferini, M., Fiori, G. & Torrigani, P. (2009). Use of Vis/NIR Spectroscopy to assess fruit ripening stage and improve management in post-harvest chain. *Fresh Produce*, 3(1), 35–41. Tratto da:
[http://globalsciencebooks.info/Online/GSBOnline/images/0906/FP_3\(SI1\)/FP_3\(SI1\)35-41o.pdf](http://globalsciencebooks.info/Online/GSBOnline/images/0906/FP_3(SI1)/FP_3(SI1)35-41o.pdf)
- Costa, G., Noferini, M., Fiori, G., Montefiori, M., Miserochi, O. & Andreotti, C. (2001). La determinazione non-distruttiva di alcuni parametri di qualità della frutta: risultati delle esperienze condotte con il sistema NIRs (Near InfraRed spectroscopy). *III Workshop nazionale del Gruppo di Lavoro Postraccolta SOI. Postraccolta dei prodotti ortoflorofrutticoli*. Pisa.
- Cozzolino, D., Degner, S. & Eglinton, J. K. (2014). A Review on the Role of Vibrational Spectroscopy as An Analytical Method to Measure Starch Biochemical and Biophysical Properties in Cereals and Starchy Foods. *Foods*, 3(4), 605-621. doi:<https://doi.org/10.3390/foods3040605>
- Crisosto, C., Valero, C. & Slaughter, D. (2009). Evaluation of a kiwifruit nondestructive Sensor. *Proc. 8th Fruit, Nut and Vegetable Production Engineering Symposium, Progap-INIA's National Precision Agriculture Program*, (p. 443-448). Concepción, Octava Región del Bío Bío, Chile.
- Daniel, F., Mounier, A., Pérez-Arantegui, J., Pardos, C., Prieto-Taboada, N., Fdez-Ortiz de Vallejuelo, S. & Castro, K. (2016). Hyperspectral imaging applied to the analysis of Goya paintings in the Museum of Zaragoza (Spain). *Microchemical Journal*, 126, 113-120.
- Danner, M., Locherer, M., Hank, T. & Richter, K. (2015). *Spectral Sampling with the ASD FieldSpec 4 –Theory, Measurement, Problems, Interpretation*. *EnMAP Field Guides Technical Report, GFZ Data Services*. doi:<http://doi.org/10.2312/enmap.2015.008>.
- Daszykowski, M., Kaczmarek, K., Vander Heyden, Y., & Walczak, B. (2007). Robust statistics in data analysis – A review. *Chemom. Intell. Lab. Syst.*, 85(2), 203–219.
doi:<https://doi.org/10.1016/j.chemolab.2006.06.016>
- de Jong, S. (1993). SIMPLS: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 18(3), 251-263.
- Del Fiore, A., Reverberi, M., Ricelli, A., Pinzari, F., Serranti, S., Fabbri, A. & Fanelli, C. (2010). Early detection of toxigenic Fungi on maize by hyperspectral imaging analysis. *International Journal of Food Microbiology*, 144, 64-71.

Riferimenti bibliografici

- del Moral, F., Guillén, A., del Moral, L., O'Valle, F., Martínez, L. & del Moral, R. (2009). Duroc and Iberian Pork Neural Network Classification by Visible and Near Infrared Reflectance Spectroscopy. *J. Food Eng.*, 540-547.
- Dell'Aquila, A. (2006). Red-Green-Blue (RGB) colour density as a nondestructive marker in sorting deteriorated lentil (*Lens culinaris* Medik.) seeds. *Seed Science and Technology*, 34, 609–619.
- Delwiche, S. R., & Hruschka, W. R. (2000). Protein Content of Bulk Wheat from Near-Infrared Reflectance of Individual Kernels. *Cereal Chemistry*, 77, 86-88. doi:10.1094/CCHEM.2000.77.1.86
- Dong, X., & Sun, X. (2013). A case study of characteristic bands selection in near-infrared spectroscopy: nondestructive detection of ash and moisture in wheat flour. *Food Measure*, 7, 141–148. doi:10.1007/s11694-013-9149-0
- dos Santos, C. A., Lopo, M., Páscoa, R. N. & Lopes, J. A. (2013). A Review on the Applications of Portable Near-Infrared Spectrometers in the Agro-Food Industry. *Applied Spectroscopy*, 67(11), 1215–1233. doi:https://doi.org/10.1366/13-07228
- Dowell, F. E., Throne, J. E. & Baker, J. E. (1998). Automated Nondestructive Detection of Internal Insect Infestation of Wheat Kernels by Using Near-Infrared Reflectance Spectroscopy. *Journal of Economic Entomology*, 91(4), 899–904. doi:https://doi.org/10.1093/jee/91.4.899
- DV s.r.l. (2013). *dvopti*. Tratto da <http://www.dvopti.com/>
- Edelman, G., Gaston, E., van Leeuwen, T., Cullen, P. & Aalders, M. (2012). Hyperspectral imaging for non-contact analysis of forensic traces. *Forensic Science International*, 223(1-3), 28-39.
- Eigenvector Research Inc. (2012). *Advanced Preprocessing: Variable Centering*. Tratto da: [wiki.eigenvector: http://wiki.eigenvector.com/index.php?title=Advanced_Preprocessing:_Variable_Centering](http://wiki.eigenvector.com/index.php?title=Advanced_Preprocessing:_Variable_Centering)
- Eigenvector Research Inc. (2013a). *Advanced Preprocessing: Noise, Offset, and Baseline Filtering*. Tratto da: [wiki.eigenvector: http://wiki.eigenvector.com/index.php?title=Advanced_Preprocessing:_Noise,_Offset,_and_Baseline_Filtering](http://wiki.eigenvector.com/index.php?title=Advanced_Preprocessing:_Noise,_Offset,_and_Baseline_Filtering)
- Eigenvector Research Inc. (2013b). *Advanced Preprocessing: Multivariate Filtering*. Tratto da: [wiki.eigenvector: http://wiki.eigenvector.com/index.php?title=Advanced_Preprocessing:_Multivariate_Filtering#GL_SW_Algorithm](http://wiki.eigenvector.com/index.php?title=Advanced_Preprocessing:_Multivariate_Filtering#GL_SW_Algorithm)
- Eigenvector Research Inc. (2013c). *Kennardstone*. Tratto da: [wiki.eigenvector: http://wiki.eigenvector.com/index.php?title=Kennardstone](http://wiki.eigenvector.com/index.php?title=Kennardstone)
- Eigenvector Research Inc. (2014). *Model Building: Preprocessing Methods*. Tratto da: [wiki.eigenvector: http://wiki.eigenvector.com/index.php?title=Model_Building:_Preprocessing_Methods](http://wiki.eigenvector.com/index.php?title=Model_Building:_Preprocessing_Methods)
- Eigenvector Research Inc. (2015). *Pqnorm*. Tratto da: [wiki.eigenvector: http://wiki.eigenvector.com/index.php?title=Pqnorm](http://wiki.eigenvector.com/index.php?title=Pqnorm)
- Eigenvector Research Inc. (2016a). *Emscorr*. Tratto da: [wiki.eigenvector: http://wiki.eigenvector.com/index.php?title=Emscorr](http://wiki.eigenvector.com/index.php?title=Emscorr)
- Eigenvector Research Inc. (2016b). *Using Cross-Validation*. Tratto da [wiki.eigenvector: http://wiki.eigenvector.com/index.php?title=Using_Cross-Validation](http://wiki.eigenvector.com/index.php?title=Using_Cross-Validation)
- Eigenvector Research Inc. (2017a). *Advanced Preprocessing: Variable Scaling*. Tratto da [wiki.eigenvector](http://wiki.eigenvector.com/).
- Eigenvector Research Inc. (2017b). *Advanced Preprocessing: Sample Normalization*. Tratto da [wiki.eigenvector: http://wiki.eigenvector.com/index.php?title=Advanced_Preprocessing:_Sample_Normalization](http://wiki.eigenvector.com/index.php?title=Advanced_Preprocessing:_Sample_Normalization)

Riferimenti bibliografici

- Eigenvector Research Inc. (2018). *Interval PLS (IPLS) for Variable Selection*. Tratto da [wiki.eigenvector.com](http://wiki.eigenvector.com/index.php?title=Interval_PLS_(IPLS)_for_Variable_Selection): [http://wiki.eigenvector.com/index.php?title=Interval_PLS_\(IPLS\)_for_Variable_Selection](http://wiki.eigenvector.com/index.php?title=Interval_PLS_(IPLS)_for_Variable_Selection)
- Engel, J., Gerretzen, J., Szymańska, E., Jansen, J. J., Downey, G., Blanchet, L. & Buydens, L. M. (2013). Breaking with trends in pre-processing? *Trends in Analytical Chemistry*, 50, 96-106.
- European Commission. (2019). *Report From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions on the implementation of the Circular Economy Action Plan*. Tratto da https://ec.europa.eu/commission/sites/beta-political/files/report_implementation_circular_economy_action_plan.pdf
- Fasolato, L., Balzan, S., Riovanto, R., Berzaghi, P., Mirisola, M., Ferlito, J., . . . Novelli, E. (2012). Comparison of Visible and Near-Infrared Reflectance Spectroscopy to Authenticate Fresh and Frozen-Thawed Swordfish (*Xiphias gladius* L). *Prod. T.*, 21(5), 493-507.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters* , 27, 861-874.
- Ferguson, L. & Sibett, G. S. (1994). *Olive Production Manual*. California, USA: University of California, Division of Agriculture and Natural Resources.
- Fernández de la Ossa, M. Á., Amigo, J. M. & García-Ruiz, C. (2014). Detection of residues from explosive manipulation by near infrared hyperspectral imaging: A promising forensic tool. *Forensic Science International*, 242, 228-235. doi:10.1016/j.forsciint.2014.06.023
- Garner, D., Crisosto, C., Wiley, P. & Crisosto, G. (2015). *Measurement of Soluble Solids Content*. Tratto da <http://fruitandnuteducation.ucdavis.edu/files/162033.pdf>
- Geladi, P. & Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Analytica Chimica Acta*, 185, 1-17. doi:[https://doi.org/10.1016/0003-2670\(86\)80028-9](https://doi.org/10.1016/0003-2670(86)80028-9)
- Gemperline, P. (2006). *Practical Guide to Chemometrics* (2 ed.). CRC/Taylor & Francis.
- Giovanelli, G., Sinelli, N., Beghi, R., Guidetti, R. & Casiraghi, E. (2014). NIR spectroscopy for the optimization of postharvest apple management. *Postharvest Biology and Technology*, 13-20. doi:<https://doi.org/10.1016/j.postharvbio.2013.07.041>
- Glasstone, S. (1943). *Textbook of physical chemistry*. D. Van Nostrand Company.
- Gosselin, R., Rodrigue, D. & Duchesne, C. (2010). A Bootstrap-VIP approach for selecting wavelength intervals in spectral imaging applications. *Chemometrics and Intelligent Laboratory Systems*, 100(1), 12-21. doi:<https://doi.org/10.1016/j.chemolab.2009.09.005>
- Gosselin, R., Rodrigue, D. & Duchesne, C. (2011). A hyperspectral imaging sensor for on-line quality control of extruded polymer composite products. *Computers and Chemical Engineering*, 35(2), 296-306.
- Gowen, A. A., Feng, Y., Gaston, E. & Valdramidis, V. (2015). Recent applications of hyperspectral imaging in microbiology. *Talanta*, 137, 43-54. doi:<https://doi.org/10.1016/j.talanta.2015.01.012>
- Gowen, A. A., O'Donnell, C. P., Cullen, P. J., Downey, G. & Frias, J. (2007). Hyperspectral imaging – an emerging process analytical tool for food quality and safety control. *Trends in Food Science & Technology*, 18(12), 590-598. doi:10.1016/j.tifs.2007.06.001
- Gowen, A., O'Donnell, C., Cullen, P. & Bellc, S. (2008). Recent applications of chemical imaging to pharmaceutical process monitoring and quality control. *European Journal of Pharmaceutics and Biopharmaceutics*, 69(1), 10-22.
- Groß, F., Benning, R., Bindrich, U., Franke, K. & Heinz, V. (2011). Optical online measurement technique used for process control of the drying step during pasta production. *11th International Congress on Engineering and Food (ICEF11), Procedia Food Science*, 1, 1301 – 1308. doi:10.1016/j.profoo.2011.09.19.

Riferimenti bibliografici

- Guidetti, R., Beghi, R. & Bodria, L. (2010). Evaluation of Grape Quality Parameters by a Simple VIS/NIR System. *Trans. Asabe*, 53(2), 477-484.
- Guzmán, E., Baeten, V., Fernández Pierna, J. A. & García-Mesa, J. A. (2015). Determination of the olive maturity index of intact fruits using image analysis. *J Food Sci Technol*, 52(3), 1462-1470.
- Guzmán, I. & Bosland, P. W. (2017). Sensory properties of chile pepper heat – and its importance to food quality and cultural preference. *Appetite*, 117, 186-190. doi:10.1016/j.appet.2017.06.026
- Hans, G. & Geladi, P. (2007). *Techniques and applications of hyperspectral image analysis*. J. Wiley.
- Hecht, E. (2002). *Optics*. Addison-Wesley.
- Hollstein, F., Wohllebe, M. & Arnaiz, S. (2015). Identification and Sorting of Plastics Film Waste by NIR-Hyperspectral-Imaging. *Proceedings of the ICNIRS 2015*. doi:10.17648/nir-2015-34127
- Hopkins, D. (2001). What is a Norris Derivative? *NIR news*, 12(3), 3-5. doi: 10.1255/nirn.611
- Hrušková, M., Bednářová, M. & Novotný, F. (2001). Wheat flour dough rheological characteristics predicted by NIRSystems 6500. *Czech J. Food Sci.*, 19(6), 213-218. doi:10.17221/6610-CJFS
- Hu, B., Giacometti, L., Di Maio, F. & Rem, P. (2011). Recycling of WEEE by Magnetic Density Separation. *Volume Selected Proceedings of the Sixth International Conference on Waste Management and Technology (ICWMT 6)*.
- Hu, B., Serranti, S., Fraunolcz, N., Di Maio, F. & Bonifazi, G. (2013). Recycling-oriented characterization of polyolefin packaging waste. *Waste Management*, 33(3), 574-584. doi:https://doi.org/10.1016/j.wasman.2012.11.018.
- International Organization for Standardization. (1994). ISO 8402:1994 - Quality management and quality assurance – Vocabulary.
- International Organization for Standardization. (2005). ISO 9000:2005 - Quality management systems – Fundamentals and vocabulary.
- International Organization for Standardization. (2017). ISO 12099:2017 - Animal feeding stuffs, cereals and milled cereal products – Guidelines for the application of near infrared spectrometry.
- IOC. (2009). *Web site of International Olive Council*. Tratto da <http://www.internationaloliveoil.org/>.
- Jansen, M., van Velzen, T. & Pretz, T. (2015). *Handbook for sorting of plastic packaging waste concentrates*. Wageningen UR Food & Biobased Research. Tratto da: <https://www.kidv.nl/6226/wur-handbook-for-sorting-of-plastic-packaging-waste-concentrates.pdf>
- JDS Uniphase Corporation. (2013). *MicroNIR™ 1700 Spectrometer: Enabling handheld, portable, and embedded on-line sensor applications*. Tratto da: [diessechem: http://www.diessechem.com/upload/schede/1347279426-micronir1700-spectrometer-ds-co-ae.pdf](http://www.diessechem.com/upload/schede/1347279426-micronir1700-spectrometer-ds-co-ae.pdf)
- JDS Uniphase Corporation. (2014). *MicroNIR™ Pro User Guide: ED000124-01 Rev001*.
- Jia, S., Yang, L., An, D., Liu, Z., Yan, Y., Li, S., . . . Gu, J. (2016). Feasibility of analyzing frost-damaged and non-viable maize kernels based on near infrared spectroscopy and chemometrics. *Journal of Cereal Science*, 69, 145-150.
- Jolivot, R., Vabres, P. & Marzani, F. (2011). Reconstruction of hyperspectral cutaneous data from an artificial neural network-based multispectral imaging system. *Computerized Medical Imaging and Graphics*, 35(2), 85-88.

Riferimenti bibliografici

- Jürgens, M., Mayerhöfer, T., Popp, J., Lee, G., Matthews, D. L. & Wilson, B. C. (2013). Introduction to Biophotonics. In Handbook of Biophotonics. In J. Popp, V. V. Tuchin, C. A., & S. H. Heinemann, *Handbook of Biophotonics*. doi:10.1002/9783527643981.bphot001
- Kamruzzaman, M., Makino, Y. & Oshita, S. (2016). Online monitoring of red meat color using hyperspectral imaging. *Meat Science*, 116, 110–117.
- Kennard, R. & Stone, L. (1969). Computer Aided Design of Experiments. *Technometrics*, 11(1), 137-148. doi:http://dx.doi.org/10.1080/00401706.1969.10490666
- Kubik, M. (2007). Hyperspectral imaging: a new technique for the non-invasive study of artworks. In D. Creagh, & D. Bradley, *Physical Techniques in the Study of Art, Archaeology and Cultural Heritage* (Vol. 2, pp. 199–259). Elsevier.
- Kumagai, M., Matsuura, N., Li, H., Ohisa, N., Amano, T. & Ogawa, N. (2004). Application of a Portable Near Infrared Spectrometer for the Manufacturing of Noodle Products. *Infrared Spectroscopy*, 12(2), 127-131.
- Kumaravelu, C. & Gopal, A. (2015). A Review on the applications of Near-Infrared Spectrometer and Chemometrics for the Agro-Food Processing Industries. *2015 IEEE International Conference on Technological Innovations in ICT for Agriculture and Rural Development (TIAR 2015)*, (pp. 8-12). doi:10.1109/TIAR.2015.7358523
- Lamine, S., Petropoulos, G. P., Brewer, P. A., Bachari, N.-E.-I., Srivastava, P. K., Manevski, K., . . . Macklin, M. G. (2019). Heavy Metal Soil Contamination Detection Using Combined Geochemistry and Field Spectoradiometry in the United Kingdom. *Sensors*, 19(4), 1424-8220. doi:10.3390/s19040762
- Larrechi, M. & Callao, M. (2003). Strategy for introducing NIR spectroscopy and multivariate calibration techniques in industry. *Trends in Analytical Chemistry*, 22(10), 634-640. doi:https://doi.org/10.1016/S0165-9936(03)01005-7
- Leardi, R. & González, A. (1998). Genetic algorithms applied to feature selection in PLS regression: how and when to use them. *Chemometrics and Intelligent Laboratory Systems*, 41(2), 195-207. doi:https://doi.org/10.1016/S0169-7439(98)00051-3
- Leardi, R. & Nørgaard, L. (2005). Sequential application of backward interval partial least squares and genetic algorithms for the selection of relevant spectral regions. *Journal of Chemometrics*, 18, 486–497. doi: https://doi.org/10.1002/cem.893
- Lee, J., Kim, S., Seong, K., Kim, C., Um, Y. & Lee S.K. (2012). Quality prediction of kiwifruit based on near infrared spectroscopy. *Korean J. Hort. Sci.*, 30, 709–717 .
- Leitner, R., Mairer, H. & Kercek, A. (2003). Real-time classification of polymers with NIR spectral imaging and blob analysis. *Real-time Imaging* 9, 245–251.
- Li, W.-S., & Shaw, J.-T. (1997). Determination of Rough Rice Quality by a Portable Near-Infrared Spectroscopy. *1997 ISAMA Annual International Meeting*. Taipei, Taiwan.
- Lim, J., Kim, G., Mo, C. & Kim, M. S. (2015). Design and Fabrication of a Real-Time Measurement System for the Capsaicinoid Content of Korean Red Pepper (*Capsicum annum* L.) Powder by Visible and Near-Infrared Spectroscopy. *Sensors* , 27420-27435.
- Lim, J., Mo, C., Noh, S. H., Kang, S., Lee, K. & Kim, M. S. (2012). Capsaicinoids Content Prediction Model Development for Korean Red-pepper Powder Using a Visible and Near-infrared Spectroscopy. *Proceedings Volume 8369, Sensing for Agriculture and Food Quality and Safety IV*. doi: https://doi.org/10.1117/12.923656

Riferimenti bibliografici

- Liu, Y., Gao, R., Hao, Y., Sun, X. & Ouyang, A. (2012). Improvement of Near-Infrared Spectral Calibration Models for Brix Prediction in 'Gannan' Navel Oranges by a Portable Near-Infrared Device. *Food Bioprocess Technol.*, 5(3), 1106-1112.
- Liu, Y., Gao, R., Sun, X., Ouyang, A., Pan, Y. & Dong, X. (2010). Predicting Brix of Intact Pears by a Portable NIR Spectrometry with LS-SVM. *2010 Sixth International Conference on Natural Computation (ICNC 10)*, 2, pp. 909-913. New York, Stati Uniti.
- Liu, Z., Zhang, D., Yan, J., Li, Q. & Tang, Q. (2007). Classification of hyperspectral medical tongue images for tongue diagnosis. *Computerized Medical Imaging and Graphics*, 31, 672-678.
- Lorente, D., Aleixos Borrás, M., Gómez Sanchís, J., Cubero, S., García Navarrete, O. & Blasco Ivars, J. (2011). Recent advances and applications of hyperspectral imaging for fruit and vegetable quality assessment. *Food and Bioprocess Technology*, 5(4), 1121-1142. doi:10.1007/s11947-011-0725-1
- Lue, Q., Tang, M., Cai, J. & Lu, H. (2010). Long-term prediction of zhonghua kiwifruit dry matter by near infrared spectroscopy. *Scienceasia*, 36, 210-215.
- Maghirang, E. B. & Dowel, F. E. (2003). Hardness measurement of bulk wheat by single-kernel visible and Near-Infrared Reflectance Spectroscopy. *Cereal Chem.*, 80(3), 316-322.
- Malvern Panalytical Ltd. (2019). *ASD FieldSpec 4 Standard-Res Spectroradiometer*. Tratto da: <https://www.malvernpanalytical.com/en/products/product-range/asd-range/fieldspec-range/fieldspec-4-standard-res-spectroradiometer/>
- Marini, F., Bucci, R., Magrì, A. L., & Magrì, A. D. (2008). Artificial neural networks in chemometrics: History, examples and perspectives. *Microchemical Journal*, 88, 178-185.
- Martens, H., & Stark, E. (1991). Extended multiplicative signal correction and spectral interference subtraction: New preprocessing methods for near infrared spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis*, 9(8), 625-635. doi:[https://doi.org/10.1016/0731-7085\(91\)80188-F](https://doi.org/10.1016/0731-7085(91)80188-F)
- Martens, H. & Naes, T. (1989). *Multivariate Calibration*. Chichester: John Wiley & Sons Ltd.
- Martinsen, P. & Schaare, P. (1998). Measuring soluble solids distribution in kiwifruit using near-infrared imaging spectroscopy. *Postharvest Biology and Technology*, 14(3), 271-281. doi:10.1016/s0925-5214(98)00051-9
- Masoud, T., Gowen, A. A. & O'Donnell, C. P. (2011). Comparison of hyperspectral imaging with conventional RGB imaging for quality evaluation of *Agaricus bisporus* mushrooms. *Biosystems Engineering*, 108(2), 191-194. doi:<https://doi.org/10.1016/j.biosystemseng.2010.10.005>
- Masoumi, H., Safavi, S. M. & Khani, Z. (2012). Identification and Classification of Plastic Resins using Near Infrared Reflectance Spectroscopy. *International Journal of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering*, 6(5), 877-884.
- Massart, D., Vandeginste, B., Buydens, L., De Jong, S., Lewi, P. & Smeyers-Verbeke, J. (1998). *Handbook of Chemometrics and Qualimetrics: Part A*. Amsterdam: Elsevier Science.
- Mehta, N., Shaik, S., Devireddy, R. & Gartia, M. R. (2018). Single-Cell Analysis Using Hyperspectral Imaging Modalities. *Journal of Biomechanical Engineering*, 140(2), 020802-020802-16. doi: 10.1115/1.4038638
- Middleton Spectral Vision . (2004). *SisuCHEMA*. Tratto da: middletonspectral: <http://www.middletonspectral.com/products/hyperspectral-components-systems/systems/sisuchema/>
- Millar, S. J. (2003). *The Development Of Near Infrared (Nir) Spectroscopy Calibrations For The Prediction Of Wheat And Flour Quality*. Project Report No. 310, Home-Grown Cereals Authority (HGCA) . doi:https://cereals.ahdb.org.uk/media/376552/310_complete_final_report.pdf

Riferimenti bibliografici

- Misal, R. M., & Deshmukh, R. R. (2016). Application of Near-Infrared Spectrometer in Agro-Food Analysis: A Review. *International Journal of Computer Applications (0975 – 8887)*, 141(7).
- MISCO. (2013). *How to Use a Traditional Analog Handheld Refractometer*. Tratto da: MISCO Refractometer: <http://www.misco.com/training/traditional-handheld>
- Moroi, A., Vartolomei, N., Aruş, A., Nistor, I. & Lazăr, I. (2011). Prediction of the ash content of wheat flours using spectral and chemometric methods. *The Annals of the University Dunarea de Jos of Galati. Fascicle VI – Food Technology*, 35(2), 34-45.
- Moroni, M., Lupo, E., Marra, E. & Cenedese, A. (2013). Hyperspectral Image Analysis in Environmental Monitoring: Setup of a New Tunable Filter Platform. *Procedia Environmental Sciences*, 885-894.
- Moscetti, P., R., Haff, R., Stella, E., Contini, M., Monarca, D., . . . Massantini, R. (2015). Feasibility of NIR spectroscopy to detect olive fruit infested by *Bactrocera oleae*. *Postharvest Biology and Technology*, 99, 58–62.
- Moscetti, R., Haff, R. P., Monarca, D., Cecchini, M. & Massantini, R. (2016). Near-infrared spectroscopy for detection of hailstorm damage on olive fruit. *Postharvest Biology and Technology*, 120, 204–212. doi:10.1016/j.postharvbio.2016.06.011
- Naes, T. (2002). *User Friendly Guide to Multivariate Calibration and Classification*. NIR Publications.
- Nicolai, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology*, 46(2), 99-118.
- Noordam, J. C., van den Broek, W. H., & Buydens, L. M. (2005). Detection and classification of latent defects and diseases on raw French fries with multispectral imaging. *Journal of the Science of Food and Agriculture*, 85(13), 2249–2259. doi:https://doi.org/10.1002/jsfa.2226
- Nørgaard, L., Saudland, A., Agner, J., Nielsen, J., Munck, L., & Engelsen, S. (2000). Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. *Applied Spectroscopy*, 54(3), 413–419. doi:10.1366/0003702001949500
- Norris, K. (1983). Extracting information from spectrophotometric curves - Predicting chemical composition from visible and near infrared spectra. In H. Martens, & H. Russwurm Jr. (A cura di), *Food Research and Data Analysis—Proc. IUFOST Symposium* (p. 95-113). Londra, Regno Unito: Applied Science Publishers.
- Olesen, M., Carstensen, J. & Boelt, B. (2011a). Multispectral imaging as a potential tool for seed health testing of spinach (*Spinacia oleracea* L.). *Seed Science and Technology*, 39, 140–150.
- Olesen, M., Shetty, N., Gislum, R. & Boelt, B. (2011b). Classification of viable and non-viable spinach (*Spinacia oleracea* L.) seeds by single seed near infrared spectroscopy and extended canonical variates analysis. *Journal of Near Infrared Spectroscopy*, 19, 171–180.
- Olive Oil Source. (2017b). *Olive Fly Control*. Tratto da <https://www.oliveoilsource.com/page/olive-fly-control>
- Oliveira, C., Bernardes, A. M. & Gerbase, A. (2012). Collection and recycling of electronic scrap: A worldwide overview and comparison with the Brazilian situation. *Waste Management*, 32, 1592-1610.
- Oliveri, P., Casolino, M. C. & Forina, M. (2010). Chemometric Brains for Artificial Tongues. *Advances in Food and Nutrition Research*, 61, 57-117. doi:https://doi.org/10.1016/B978-0-12-374468-5.00002-7.
- Ornberg, R. L., Woerner, B. M. & Edwards, D. A. (1999). Analysis of Stained Objects in Histological Sections by Spectral Imaging and Differential Absorption. *The Journal of Histochemistry & Cytochemistry*, 47(10), 1307–1313. doi:10.1177/002215549904701010

Riferimenti bibliografici

- Ortiz, C. & Torregrosa, A. (2014). Mechanical properties of citrus and impact damage under different storage conditions. *Transactions of the ASABE*, 57(2), 593-598. doi:10.13031/trans.57.10479
- Othman, Z. A., Ahmed, Y. B., Habila, M. A. & Ghafar, A. A. (2011). Determination of Capsaicin and Dihydrocapsaicin in Capsicum Fruit Samples using High Performance Liquid Chromatography. *Molecules*, 16(10), 8919-8929. doi:10.3390/molecules16108919
- Ozdemir, F. & Akinci, I. (2004). Physical and Nutritional Properties of Four Major Commercial Turkish Hazelnut Varieties. *Journal of Food Engineering*, 63, 341-347. doi:10.1016/j.jfoodeng.2003.08.006
- Palmieri, R., Bonifazi, G. & Serranti, S. (2014a). Recycling-oriented characterization of plastic frames and printed circuit boards from mobile phones by electronic and chemical imaging. *Waste Management*, 34, 2120-2130. doi:http://dx.doi.org/10.1016/j.wasman.2014.06.003
- Palmieri, R., Bonifazi, G. & Serranti, S. (2014b). Automatic detection and classification of EOL-concrete and resulting recovered products by hyperspectral imaging. 1-14.
- Panda, S. K. & Suresh, M. (2018). Estimation of Pungency of Green Chillies using UV Spectrophotometer. *international Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, 7(2).
- Pariante, E. S., Cancilla, J. C., Wierzechos, K. & Torrecilla, J. S. (2018). On-site images taken and processed to classify olives according to quality – The foundation of a high-grade olive oil. *Postharvest Biology and Technology*, 140, 60–66. doi:10.1016/j.postharvbio.2018.02.012
- Park, T. S., Sim, M. J., E., K. D. & In Cho, S. (2008). Analysis of Capsaicinoids from Hot Red Pepper Powder by Near-Infrared Spectroscopy. *2008 ASABE Annual International Meeting 2008*. Rhode Island, Stati Uniti.
- Pasikatan, M. C. & Dowell, F. E. (2001). Sorting systems based on optical methods for detecting and removing seeds infested internally by insects or fungi: a review. *Applied Spectroscopy Reviews*, 36, 399–416.
- Pasquini, C. (2003). Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. *J. Braz. Chem. Soc.*, 14(2), 198-219. doi:http://dx.doi.org/10.1590/S0103-50532003000200006
- Pérez-Marín, D., Paz, P., Guerrero, J., Garrido-Varo, A. & Sánchez, M. (2010). Miniature Handheld NIR Sensor for the On-Site Non-Destructive Assessment of Post-Harvest Quality and Refrigerated Storage Behavior in Plums. *J. Food Eng.*, 99(3), 294-302.
- Picon, A., Ghita, O., Iriondo, P. M., Bereciartua, A. & Whelan, P. F. (2010). Automation of waste recycling using hyperspectral image analysis. *Proceedings of 15th IEEE International Conference on Emerging Technologies and Factory Automation, ETFA 2010*, (p. 1-4). Bilbao, Spagna.
- Pisani, M., Zucco, M., Caricato, V. & Egidi, A. (2013). Hyperspectral imaging: a tool for biological measurements. *16th International Congress of Metrology*. doi:10.1051/metrology/201314007
- Pojić, M., Mastilović, J. & Majcen, N. (2012). The Application of Near Infrared Spectroscopy in Wheat Quality Control. In T. Theophile, *Infrared Spectroscopy - Life and Biomedical Sciences*. IntechOpen. doi:10.5772/34676
- Polder, G., van der Heijden, G. W. & Young, I. T. (2002). Spectral image analysis for measuring ripeness of tomatoes. *Transactions of the ASAE*, 45(4). doi:10.13031/2013.9924
- Rantanen, J. & Khinast, J. (2015). The Future of Pharmaceutical Manufacturing Sciences. *Journal of Pharmaceutical Sciences*, 104(11), 3612-3638. doi:https://doi.org/10.1002/jps.24594.
- Rinnan, Å., van den Berg, F. & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry*, 28(10), 1201-1222. doi:https://doi.org/10.1016/j.trac.2009.07.007

Riferimenti bibliografici

- Robinson, B. H. (2009). E-waste: An assessment of global production and environmental impacts. *Science of The Total Environment*, 408(2), 183-191.
- Rocha, W., Sabin, G., Março, P. & Poppi, R. (2010). Quantitative analysis of piroxicam polymorphs pharmaceutical mixtures by hyperspectral imaging and chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 106(2), 198-204.
- Roggo, Y., Chalus, P., Maurer, L., Lema-Martinez, C., Edmond, A. & Jent, N. (2007). A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *Journal of Pharmaceutical and Biomedical Analysis*, 44, 683–700. doi:https://doi.org/10.1016/j.jpba.2007.03.023
- Salhofer, S., Spitzbart, M. & Maurer, K. (2011). Recycling of LCD Screens in Europe - State of the Art and Challenges. In J. Hesselbach & C. Hermann, *Glocalized Solutions for Sustainability in Manufacturing* (p. 454-458).
- Saranwong, S., Sornsrivichai, J. & Kawano, S. (2003). Performance of a Portable Near Infrared Instrument for Brix Value Determination of Intact Mango Fruit. *J. Near Infrared Spectrosc.*, 11(3), 175-181.
- Savitzky, A. & Golay, M. J. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36(8), 1627–1639. doi:10.1021/ac60214a047
- Schropp, C., Raulf, K. & Robben, M. (2014). *Sensor Technologies: Impulses for the Raw Materials Industry*. Aachen.
- Serranti, S. & Bonifazi, G. (2010). Post-consumer polyolefins (PP-PE) recognition by combined spectroscopic sensing techniques. *The Open Waste Management Journal* 3, 36–46.
- Serranti, S., Bonifazi, G. & Gasbarrone, R. (2018a). Kiwifruits ripening assessment by portable hyperspectral devices. *Sensing for Agriculture and Food Quality and Safety X, Proc. SPIE 10665*. doi:https://doi.org/10.1117/12.2297353
- Serranti, S., Bonifazi, G. & Gasbarrone, R. (2018b). Olive fruit ripening evaluation and quality assessment by hyperspectral sensing devices. *Sensing for Agriculture and Food Quality and Safety X, Proc. SPIE 10665*. doi:https://doi.org/10.1117/12.2297352
- Serranti, S., Bonifazi, G., Bonoli, A. & Dall'Ara, A. (2009). Composting products quality assessment and monitoring by hyperspectral imaging based logics. *Proceedings of Waste-to-Resources, III International Symposium MBT&MRF*, (pp. 584–597). Hanover, Germany.
- Serranti, S., Gargiulo, A. & Bonifazi, G. (2010). The utilization of hyperspectral imaging for impurities detection in secondary plastics. *The Open Waste Management Journal*, 3, 57-71.
- Serranti, S., Gargiulo, A. & Bonifazi, G. (2012). Dried fruits quality assessment by hyperspectral imaging. *Proc. SPIE 8369, Sensing for Agriculture and Food Quality and Safety IV*, 83690U. doi:10.1117/12.918561
- Serranti, S., Palmieri, R. & Bonifazi G. (2013). Opto-electronic combined characterization of end-of-life mobile phone wastes for a sustainable processing. *Proceedings Sardinia 2013, Fourteenth International Waste Management and Landfill Symposium, S. Margherita di Pula, Cagliari, Italy*.
- Serranti, S., Palmieri, R. & Bonifazi, G. (2015). Hyperspectral imaging applied to demolition waste recycling: an innovative approach for quality control. *Journal of Electronic Imaging*, 24(4), 043003-1_043003-9.
- Serranti, S., Palmieri, R., Bonifazi, G. & Cózar, A. (2018). Characterization of microplastic litter from oceans by an innovative approach based on hyperspectral imaging. *Waste Management*, 76, 117-125. doi:https://doi.org/10.1016/j.wasman.2018.03.003
- Shimamoto, J., Hasegawa, K., Hattori, S., Hattori, Y. & Mizuno, T. (2003). Non-Destructive Determination of the Fat Content in Glazed Bigeye Tuna by Portable Near Infrared Spectrophotometer. *Fisheries Sci.*, 69(6), 1247-1256.

Riferimenti bibliografici

- Silva, E. J., Britto, A. S., Oliveira, L. S., Enembreck, F., Sabourin, R. & Koerich, A. L. (2017). A two-step cascade classification method. *2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK*, (p. 573-580).
- Singh, C., Paliwal, J., Jayas, D. & White, N. (2006). Near-infrared spectroscopy: Applications in the grain industry. *CSBE/SCGAB 2006*, (p. Paper No. 06-189). Edmonton, Alberta, U.S.A.
- Smith, G., Gravett, I., Edwards, C., Curtis, J. & Buwalda, J. (1994). Spatial analysis of the canopy of kiwifruit vines as it relates to the physical, chemical and postharvest attributes of the fruit. *Annals of Botany*, 73(1), 99-111.
- Specim, Spectral Imaging Ltd. (2016). <http://www.specim.fi/downloads/VIS-VNIR-ImSpectors-ver1-2016.pdf>. Tratto da specim.
- Specim, Spectral Imaging Ltd. (2019). *SisuCHEMA - Chemical Imaging Analyzer*. Tratto da specim: <http://www.specim.fi/products/sisuchema/>.
- Spreafico, L. (1978). *L'analisi delle sementi. Quaderno n. 34 dell'E.N.S.E. (Aggiornamento e revisione del quaderno n.4)*. Milano: Ente Nazionale Sementi Elette.
- Stella, E., Moschetti, R., Haff, R. P., Monarca, D., Cecchini, M., Contini, M. & Massantini, R. (2015). Review: Recent Advances in the Use of Non-Destructive near Infrared Spectroscopy for Intact Olive Fruits. *Journal of Near Infrared Spectroscopy*, 23(4), 197-208. doi:10.1255/jnirs.1169.
- Stuart, M. B., McGonigle, A. J. & Willmott, J. R. (2019). Hyperspectral Imaging in Environmental Monitoring: A Review of Recent Developments and Technological Advances in Compact Field Deployable Systems. *Sensors*, 19(3071). doi:http://dx.doi.org/10.3390/s19143071
- Sun, D.-W. (2010). *Hyperspectral Imaging for Food Quality Analysis and Control*. San Diego, CA (U.S.A.): Academic Press/Elsevier.
- Sun, X., Zhang, H. & Liu, Y. (2009). Nondestructive Assessment of Quality of Nanfeng Mandarin Fruit by a Portable Near Infrared Spectroscopy. *Int. J. Agric. Biol. Eng.*, 2(1), 65-71.
- Suresh, S. S., Mohanty, S. & Nayak, S. K. (2017). Preparation and characterization of recycled blends using poly(vinyl chloride) and poly(methyl methacrylate) recovered from waste electrical and electronic equipments. *Journal of Cleaner Production*, 149, 863-873.
- Suresh, S. S., Mohanty, S. & Nayak, S. K. (2018). Preparation of poly(vinyl chloride)/poly(methyl methacrylate) recycled blends: effect of varied concentration of PVC and PMMA in stability of PVC phase on recycled blends. *Proceedings of Materialstoday*, 5(2), 8899-8907.
- Tamburini, E., Battaglia, M., Suma, L. & G., M. M. (2016). La tecnica NIR applicata al controllo qualità per la produzione di pasta di semola, pasta all'uovo. *Atti del congresso NIR ITALIA 2016*.
- Tarantili, P., Mitsakaki, A. & Petoussi, M. (2010). Processing and properties of engineering plastics recycled from waste electrical and electronic equipment (WEEE). *Polymer Degradation and Stability*, 95, 405-410. doi:10.1016/j.polyimdegradstab.2009.11.029.
- Testoni, A. & Lovati, F. (2010). Shelf-life e qualità dei frutti. In *Metodi innovativi di gestione dei frutti nella fase post-raccolta* (p. 39-51).
- The MathWorks, Inc. (2019). *MathWorks*. Tratto da mathworks: <https://it.mathworks.com/>
- The New Zealand Institute for Plant and Food Research Limited. (2013). *Plant & Food RESEARCH*. Tratto da What is dry matter?: <http://www.plantandfood.co.nz/growingfutures/case-studies/delivering-superior-apples/what-is-dry-matter>.
- The Olive Oil Source. (2017a). *Harvest Timing*. Tratto da <https://www.oliveoilsource.com/page/harvest-timing#index>.

Riferimenti bibliografici

- Tianqi, C. & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (p. 785-794). San Francisco, California, USA.
- UC Cooperative Extension. (2006). *Olive Maturity Index*. Tratto da Sonoma County: <http://ucanr.edu/datastoreFiles/391-406.pdf>.
- Ulrici, A., Serranti, S., Ferrari, C., Cesare, D., Foca, G. & Bonifazi, G. (2013). Efficient chemometric strategies for PET-PLA discrimination in recycling plants using hyperspectral imaging. *Chemometrics and Intelligent Laboratory Systems*, 122, 31-39. doi:<https://doi.org/10.1016/j.chemolab.2013.01.001>
- UNEP. (2009). *Recycling - from e-waste to resources*. United Nations Environment Programme & United Nations University.
- Varzakas, T. (2016). Quality and Safety Aspects of Cereals (Wheat) and Their Products. *Critical Reviews in Food Science and Nutrition*, 56(15), 2495-2510. doi:10.1080/10408398.2013.866070.
- Vasefi, F., MacKinnon, N. & Farkas, D. (2016). Chapter 16 - Hyperspectral and Multispectral Imaging in Dermatology. In M. R. Hamblin, P. Avci & G. K. Gupta, *Imaging in Dermatology* (p. 187-201). Academic Press.
- Veit, H. M. & Moura, B. A. (2015). Electronic Waste: Generation and management. In H. M. Veit & B. A. Moura, *Electronic Waste: recycling techniques* (p. 3-10).
- Wang, D., Ram, M. & Dowell, F. (2002). Classification of damaged soybean seeds using near-infrared spectroscopy. *Transactions of the ASAE*, 45, 1943-1948.
- Wang, H., Peng, J., Xie, C., Bao, Y. & He, Y. (2015). Fruit Quality Evaluation Using Spectroscopy Technology: A Review. *Sensors*, 15(5), 11889-11927. doi:10.3390/s150511889
- Wang, T. & Wang, X. (2013). Study on Germination of Tomato Seed Based on Near-Infrared Spectroscopy. *The 11th IEEE International Conference on Electronic Measurement & Instruments*.
- Wirth, K. (2019). *X-Ray Fluorescence (XRF)*. Tratto da: [serc.carleton.edu](https://serc.carleton.edu/research_education/geochemsheets/techniques/XRF.html): https://serc.carleton.edu/research_education/geochemsheets/techniques/XRF.html
- Wise, B. M., Gallagher, N. B., Bro, R., Shaver, J. M., Windig, W. & Koch, R. S. (2008). *PLS_Toolbox 4.2 Reference Manual for use with MATLAB™*. Eigenvector Research, Inc.
- Wise, B., Gallagher, N., Bro, R., Shaver, J., Windig, W. & Koch, R. (2006). *Chemometrics Tutorial for PLS_Toolbox and Solo*. Eigenvector Research Inc.
- Wold, S. (1976). Pattern recognition by means of disjoint principal components models. 8(3), 127-139. doi:[https://doi.org/10.1016/0031-3203\(76\)90014-5](https://doi.org/10.1016/0031-3203(76)90014-5)
- Wold, S., Esbensen, K. & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1-3), 37-52. doi:[https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9)
- Wold, S., Sjöström, M. & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2), 109-130. doi:[https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1).
- World, S. (1995). Chemometrics; what do we mean with it, and what do we want from it? *Chemometric and Intelligent Laboratory Systems*, 30, 109-115.
- Wright, K. (2017). *The NIPALS algorithm*. Tratto da [cran.r-project](https://cran.r-project.org/web/packages/nipals/vignettes/nipals_algorithm.pdf): https://cran.r-project.org/web/packages/nipals/vignettes/nipals_algorithm.pdf
- Wu, D. & Sun, D.-W. (2013). Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review – Part II: Applications. *Innovative Food Science and Emerging Technologies*, 19, 15-28. doi:<https://doi.org/10.1016/j.ifset.2013.04.016>

Riferimenti bibliografici

- Wu, W., Mallet, Y., Walczak, B., Penninckx, W., Massart, D., Heuerding, S. & Erni, F. (1996). Comparison of regularized discriminant analysis linear discriminant analysis and quadratic discriminant analysis applied to NIR data. *Analytica Chimica Acta*, 329(3), 257-265. doi:[https://doi.org/10.1016/0003-2670\(96\)00142-0](https://doi.org/10.1016/0003-2670(96)00142-0).
- Yağci, S. & Göğüş, F. (2008). Response Surface Methodology for Evaluation of Physical and Functional Properties of Extruded Snack Foods Developed from Food-by-products. *Journal of Food Engineering*, 86, 122-132 .
- Yale University. (1997). *Multiple Linear Regression*. Tratto da stat.yale.:
<http://www.stat.yale.edu/Courses/1997-98/101/linmult.htm>
- Yang, F., Li, Y., Gu, X., Ma, J., Fan, X., Wang, X. & Zhang, Z. (2011). Determination of Acidity and Vitamin C in Apples Using Portable NIR Analyzer. *Spectroscopy and Spectral Analysis*, 31(9), 2386-2389.
- Zardetto, S. (2005). Potential applications of near infrared spectroscopy for evaluating thermal treatments of fresh egg pasta. *Food Control*, 249–256. doi:10.1016/j.foodcont.2004.03.002
- Zeng, X., Mathewes, J. A. & Li, J. (2018). Urban Mining of E-Waste is Becoming More Cost-Effective Than Virgin Mining. *Environmental Science & Technology*, 52(8), 4835-4841. doi:10.1021/acs.est.7b04909
- Zhang, C., Guo, C., Liu, F., Kong, W. & He, Y. (2016). Hyperspectral imaging analysis for ripeness evaluation of strawberry with support vector machine. *Journal of Food Engineering*, 179, 11-18.
- Zhang, S. & Forsberg, E. (1997). Mechanical separation-oriented characterization of electronic scrap. *Resources, Conservation and Recycling*, 21(4), 247-269. doi:[https://doi.org/10.1016/S0921-3449\(97\)00039-6](https://doi.org/10.1016/S0921-3449(97)00039-6)

ALLEGATO I - MATLAB SCRIPT - SISUCHEMA_SWIR_IMPORT.M

Lo script `Sisuchema_SWIR_import.m` consente di aprire i tre file `.hdr` (uno contenente l'informazione del bianco di riferimento, uno quella della *dark current* ed uno contenente l'immagine iperspettrale grezza) ottenuti dal software ChemaDAQ™ con lo strumento SisUCHEMA™ XL Chemical Imaging Workstation (3.1 *Sistema iperspettrale SWIR*), applicare la correzione dell'immagine (equazione 3.1.1) e ricavare il dataset corretto dell'immagine iperspettrale in riflettanza.

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Sisuchema SWIR importer %%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Riccardo Gasbarrone and Roberta Palmieri 2016 %%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% This script allows to import raw and calibration data into MATLAB
% and calculates the spectra reflectance correction.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%info
h = msgbox('1.  Open .hdr raw data, 2.  Open .hdr dark reference,
3.  Open .hdr white.','INFO','help')
pause (10)
%select first raw data
Raw = opotekenvirdr
%select dark reference (black)
Dark = opotekenvirdr
%select white
White = opotekenvirdr
%select first column of Raw (pixels positions)
%A=Raw(:,1);
%A2=double(A);
%select raw data
%B2=double(Raw);
B3=Raw;
n=size(Raw);
%B3=dataset(B2)
%select dark values
D1=double(Dark);
%Applies bicubic interpolation smoothing & create dataset
D2=imresize (D1, n, 'bicubic');
D3=dataset(D2)
%select white values
W1=double(White);
%Applies bicubic interpolation smoothing & create dataset
W2=imresize(W1, n, 'bicubic' );
W3=dataset(W2)
%Calculate real reflectance
C=(B3-D3);
F=(W3-D3);
REFL=C./F;
clearvars -except REFL
%create dataset
%REFL=dataset(REFL);
%add wavelenghts
wn=[942.170000000000;948.500000000000;...;2542.140000000000];
REFL.axisScale{2}=wn;

```

ALLEGATO II - MATLAB SCRIPT - FIELDSPEC_IMPORT2.M

Lo script `fieldspec_import2.m` consente di importare in ambiente MATLAB gli spettri acquisiti con ASD FieldSpec 4® (3.3 *Spettrofotometri portatile VIS-SWIR Fieldspec® 4*) convertiti in file `.txt`. La routine applica la correzione degli spettri (equazione 3.1.1) ed effettua, inoltre, il *pre-processing* SC (si veda il paragrafo 2.3.1.3 *Altri metodi di pre-elaborazione*).

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% FieldSpec 4 Importer %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Riccardo Gasbarrone 2016 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% VER 2.1 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% This works on. ascii data. After the conversion of .ASD files in
% ViewSpecPro, this script allows the importation of the ascii file
% into Matlab for processing within PLS_Toolbox.
% REMEMBER: this script replaces commas with dots using comma2dot.m script
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%replace commas with dots
[filename,pathname] = uigetfile('*.txt', 'pick text file');
wholefile = fileread(filename); % read in entire file
newfiledata = strrep(wholefile,',','.');%replace commas with full stops
fileid = fopen(filename,'w'); %open file to write
fprintf(fileid,'%s',newfiledata); %print to file
fclose(fileid);
%Import data to ascii
[stmfile, stmpath] = uigetfile('*.txt', 'pick text file');
Data = importdata(fullfile(stmpath, stmfile), '\t');
%select wavelenghts
w=Data.data(:,1);
%transpose wavelenghts
w=w';
%select only spectra(select all the other columns
cubespectra=Data.data(:,2:end);
%transpose spectra matrix
cubespectra=cubespectra';
%select labels
labels=Data.textdata(2:end);
%PLS_toolbox create new dataset
A=dataset(cubespectra);
A.axisScale(2)=w;
A.label{1}=labels;
%DEVIDE THE CHANNELS: VNIR,SWIR1 AND SWIR2.
%VNIR: 350-1000 nm wavelength.
%SWIR 1 for the wavelength range of 1001 nm to 1800 nm.
%SWIR 2 for the wavelength range of 1801 nm to 2500 nm.
%comment those strings below if %you don't want to operate this process.
VNIR=A(:,1:651);%VNIR: 350-1000 nm wavelength.
SWIR1=A(:,652:1451);%1001 nm to 1800 nm.
SWIR2=A(:,1452:end);%1801 nm to 2500 nm.
sizeSWIR1=length(SWIR1);
sizeSWIR2=length(SWIR2);
% Apply splice correction pre-process
%http://gfzpublic.gfzpotdam.de/pubman/item/escidoc:1388298/component/escidoc:
1388299/EnMAP_FieldGuide_ASD_2015_008.pdf
R1001=A(:, 652);
R1000=A(:, 651);
R999=A(:, 650);
R1801=A(:, 1452);
R1800=A(:, 1451);
R1799=A(:, 1450);
f1000=R1001-(2.*R1000-R999);
f1000=f1000';
f1000_r= repmat(f1000, sizeSWIR1);
f1000_r=f1000_r';

```

Allegati

```
f1800=R1801-(2.*R1800-R1799);
f1800=f1800';
f1800_r= repmat(f1800, sizeSWIR2);
f1800_r=f1800_r';
SWIR1=SWIR1-f1000_r;
SWIR2=SWIR2-f1800_r;
%create final dataset
Afinale=cat(2, VNIR, SWIR1, SWIR2);
%assign file name
Name=strrep(stmfile, '..txt', '');
assignin('base', Name, Afinale);
%clear
clear('A', 'Afinale', 'cubespectra', 'Data', 'f1000', 'f1000_r', 'f1800', 'f1800_r',
'labels', 'Name', 'R1000', 'R1001', 'R1799', 'R1800', 'R1801', 'R999', 'sizeSWIR1',
'sizeSWIR2', 'stmfile', 'stmpath', 'SWIR1', 'SWIR2', 'VNIR', 'w', 'ans', 'fileid',
'filename', 'newfiledata', 'pathname', 'wholefile' );
```

ALLEGATO III - MATLAB SCRIPT - MICRONIR_IMPORT.M

Lo script `micronir_import.m` consente di esportare in ambiente MATLAB® gli spettri acquisiti con lo spettrometro JDSU MicroNIR™ 1700 ([3.4 Spettrometro portatile NIR MicroNIR™](#)) e salvati in formato `.csv`. Inoltre, questo script applica la correzione degli spettri in riflettanza (equazione 3.1.1).

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% MicroNir JDSU Importer %%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Riccardo Gasbarrone 2016 %%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%This script allows the importation of MicroNir JDSU raw data .csv in
%MatLab Environment and calculate the real spectra. REMEMBER:
%The last rows must be black and white
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Initialize importfile2 function
data=importfile2;
%data=csvimport
%csvopen = importfile(filename, startRow, endRow)
load ('filename.mat', 'FileName')
%Extract wavelength%convert to matrix
w=data(1,2:end-3);
w=table2array(w);
%Extract spectra names
label=data(2:end-2, 1);
label=table2array(label);
%Extract spectra
raw=data(2:end-2,2:end-3);
raw=table2array(raw);
%extract black
bl=data(end-1,2:end-3);
bl=table2array(bl);
n=size(raw,1);
black= repmat(bl,n,1);
%extract white
wh=data(end,2:end-3);
wh=table2array(wh);
white= repmat(wh,n,1);
%Calculate real spectra
A=(raw-black);
B=(white-black);
REFL=A./B;
REFL=dataset(REFL);
REFL.axisScale{2}=w;
%clear variables
%clearvars -except REFL label w
REFL.label{1}=label;
%%%clearvars -except FileName REFL label w
%Erase .csv in char filename
FileName=strrep(FileName, '.csv', '')
%renaming REFL with the real data file name
assignin('base',FileName,REFL);
%clear old file
clear ('REFL')
%to concatenate all the loaded file, try this:
%ALL=cat(1,k0001, k0002, k0003,... )

```

ALLEGATO IV - MATLAB SCRIPT - IMPORTXRFFMAP.M

La funzione `importxrffmap.m` consente di importare in ambiente MATLAB® la mappa XRF (scansionata con lo spettrofotometro Bruker M4 TORNADO; si veda [3.5 Spettrometro Micro-XRF M4 TORNADO](#)), salvata in formato LISPIX `.raw`.

```
function my_image = importxrffmap(filename, options)

if nargin == 0 | isempty(filename)
    [file,pathname] = evriuigetfile('*.raw','Load LISPIX RAW file');
    if file==0
        %User cancel.
        imagedso = [];
        return
    end
    filename = [pathname,file];
end

%parse inputs
switch nargin
    case 1
        options = [];
end
options = reconopts(options,'lispixrawreadr');

[pth,file,ext] = fileparts(filename);

if ~exist(fullfile(pth,[file '.rpl'])) | ~exist(fullfile(pth,[file '.raw']))
    error('The .rpl and .raw files must both exist in the same folder (is this an actual
Lispix RAW file?')
end

%-----
%read RPL file and parse results
info = parseripple(fullfile(pth,[file '.rpl']));

%-----
%type is based on information from .rpl file
% data-type: signed, unsigned, or float
% data-length: # bytes per pixel      1, 2, 4, or 8
switch lower(info.type)
    case 'signed'
        type = ['int' num2str(str2double(info.length)*8)];
    case 'unsigned'
        type = ['uint' num2str(str2double(info.length)*8)];
    case 'float'
        switch info.length
            case {1 2}
                error('unable to read images using 1 or 2 byte floating point format')
            case 4
                type = 'single';
            case 8
                type = 'double';
        end
    end
    otherwise
        error('unrecognized data type %s',info.type)
end

%-----
%open and read RAW data file

%Set MACHINEFORMAT using Byte-Order field:
%  big-endian, little-endian, or dont-care
myprecision = '';
switch info.byte_order
    case 'big-endian'
        myprecision = 'b';
    case 'little-endian'
        myprecision = 'l';
    case 'dont-care'
```

Allegati

```
    myprecision = 'n';
end

[fid,message] = fopen(fullfile(pth,[file '.raw']),'r',myprecision);
if fid<1
    error('Unable to open RAW file: %s',message);
end
if str2double(info.offset)>0;
    junk = fread(fid,'uint8',str2double(info.offset));
end

% load the requested pixels
my_image = loaddata(fid, type, info, options);
fclose(fid);

%-----
function imagedso = addaxisscale(imagedso, info)
% Add depth axis scale and name. Apply binning to scale if necessary.
depthscaleorigin = str2double(info.depthscaleorigin);
depthscaleincrement = str2double(info.depthscaleincrement);
zlen = str2double(info.depth);
binsize = str2double(info.depthbinsize);
zscale = depthscaleorigin + (1:zlen)*depthscaleincrement;
depthscale = coadd(zscale, binsize, 2);

imagedso.axisscale{2} = depthscale;
imagedso.axisscalename{2} = info.depthscaleunits;

%-----
function out = loaddata(f, type, info, options)
% Load data from the .raw file, binning in depth dimension, and extracting
% a subset rectangle with corners specified by (w1,h1) and (w2,h2),
% inclusive.

opts = coadd('options');
opts.maxitems = options.maxitems;

height = str2double(info.height);
width = str2double(info.width);
depth = str2double(info.depth);
bytesperpixel = str2double(info.length);

% Subset width
w2 = min([str2double(info.width2) width]);
w1 = min([str2double(info.width1) w2]);
if ~(w1>0 & w2>0)
    w1 = 1;
    w2 = width;
end
w1 = max([w1 1]);
w2 = max([w2 w1]);
dwidth = w2-w1+1;
% Subset height
h2 = min([str2double(info.height2) height]);
h1 = min([str2double(info.height1) h2]);
if ~(h1>0 & h2>0)
    h1 = 1;
    h2 = height;
end
h1 = max([h1 1]);
h2 = max([h2 h1]);
dheight = h2-h1+1;

npixels = dwidth*dheight;
if strcmp(options.waitbar,'on') | (strcmp(options.waitbar,'auto') & dheight>250)
    wbhandle = waitbar(0,'Loading LISPIX File. (Close to abort)');
else
    wbhandle = [];
end
% limit how small the waitbar increment can be to waitfract
waitfract = 1/50;
nwaitsteps = ceil(dheight*waitfract);
nwaitsteps = max(10, nwaitsteps);

% Binning in wavelength
binsize = str2double(info.depthbinsize); %30;
```

Allegati

```

depth2 = ceil(depth/binsize);
out = nan(dheight, dwidth, depth2);
opts.dim = 1; % spec is read as depth x 1. Do coadd on depth dimension.

offset = (h1-1)*width*depth*bytesperpixel;
offsetw1 = (w1-1)*depth*bytesperpixel;
offsetw2 = (width-w2)*depth*bytesperpixel;
status = fseek(f,offset,'bof'); % skip (h1-1)*width spectra
for ih = 1:dheight
    if ~isempty(wbhandle);
        if ~ishandle(wbhandle); error('Aborted by user'); end
        if mod(ih, nwaitsteps)==0
            waitbar(ih/dheight);
        end
    end
end

spec = nan(depth,dwidth);
status1 = fseek(f, offsetw1, 'cof'); % skip (w1-1) spectra
for iw = 1:dwidth
    spec(:,iw) = fread(f,depth,type);
end
status2 = fseek(f, offsetw2, 'cof'); % skip (width-w2) spectra
spec = coadd(spec, binsize, opts);
out(ih, :, :) = spec';
end
if ~isempty(wbhandle) & ishandle(wbhandle); close(wbhandle); end

%-----
function info = parseripple(filename)

[fid,message] = fopen(filename);
if fid<1
    error('Unable to open RPL file: %s',message);
end
content = fread(fid,'char');
fclose(fid);
content = str2cell(char(content)',true);
%content is now a cell array of strings from the file. HOWEVER, if no
%linefeeds were present, we may still have to parse the single line.
% Nx2 array, first column is key name and second column is field name for
% info structure output.
myfields = {'Width' 'width';...
    'Height' 'height';
    'Depth' 'depth';
    'Offset' 'offset';
    'Data-type' 'type';
    'Data-length' 'length';
    'Byte-Order' 'byte_order';
    'Record-by' 'record_by';
    'depthbinsize' 'depthbinsize';
    'width1' 'width1';
    'width2' 'width2';
    'height1' 'height1';
    'height2' 'height2';
    'depthscaleorigin' 'depthscaleorigin';
    'depthscaleincrement' 'depthscaleincrement';
    'depthscaleunits' 'depthscaleunits'
};

%Make lower case for now, may need orginal case in future.
myfields = lower(myfields);

for i = 1:length(myfields)
    %Match regexp up to value then after value then cut it out.
    [s1 f1] = regexp(lower(content),[myfields{i,1} ' [\s]+:[\s]*']);
    [s2 f2] = regexp(lower(content),[myfields{i,1} ' [\s]+:[\s]*\d*\w*[-.]?\w*']);
    myloc = ~cellfun('isempty',f1);
    if ~any(myloc)
        switch myfields{i,1}
            case {'height' 'depth' 'offset' 'type' 'length' 'byte_order' 'record_by'}
                %Didn't find any matches so warn and keep going.
                warning(['Could not find value for field: ' myfields{i,1} ' in .rpl file. This
may cause problems reading the file.']);
                continue
            case 'depthbinsize' % Didn't find depthbinsize so set = 1.
                info.depthbinsize = '1';
            case 'width1' % didn't find width1, so flag using neg.

```

Allegati

```
    info.width1 = '-1';
case 'width2'
    info.width2 = '-1';
case 'height1'
    info.height1 = '-1';
case 'height2'
    info.height2 = '-1';
case 'depthscaleorigin'
    info.depthscaleorigin = '0';
case 'depthscaleincrement'
    info.depthscaleincrement = '1';
case 'depthscaleunits'
    info.depthscaleunits = 'Arbitrary';
end
else
    info.(myfields{i,2}) = content{myloc}(f1{myloc}+1:f2{myloc});
end
end
```

ALLEGATO V - MATLAB SCRIPT - CREATE_DATASETXRF.M

la funzione `create_datasetXRF.m` re-interpola la matrice di dati, riducendone le dimensioni, al fine di semplificare la gestione del set di dati e la procedura di analisi successiva, nonché di diminuire il tempo necessario di elaborazione dati.

```
function newvar=create_datasetXRF(datacube)
% creates a dataset object for image analysis with MIA toolbox
% starting from data imported from the instrument as MATLAB hypercube
% I/O: newvar=create_dataset(datacube)
% where datacube is the variable arranged as xpixel x ypixels x wavelengths
% and newvar is the dataset object

newvar=dataset(reshape(datacube,size(datacube,1)*size(datacube,2),size(datacube,3)));
newvar.type='image';
newvar.imagesize=[size(datacube,1) size(datacube,2)]
wn=[[-0.9591000000000000 -0.9491000000000000 (...) 39.99090000000000]];
newvar.axisscale{2}=wn;
newvar.include{2}=(135:4096);
```

ALLEGATO VI - CASE STUDY – ANALISI MICRO-XRF DI SOTTOPRODOTTI MDS PROVENIENTI DA RAEE

Nel seguente allegato sono riportati i risultati ottenuti (e le restituzioni grafiche degli stessi) riferibili alle analisi effettuate sull'analisi elementale quantitativa, mediante micro-XRF, effettuata su un sottoprodotto di separazione magnetica per densità (MDS) a partire da un flusso di RAEE in pezzatura (*4.1.1 Caratterizzazione di un flusso di RAEE risultante da comminuzione e separazione magnetica per densità*).

Distribuzione degli elementi all'interno del campione

In [Figura A - 1](#) sono mostrati gli elementi trovati nel campione, a partire dalla deconvoluzione dello spettro totale riferito alla ROI selezionata ([Figura A - 2](#)). La mappa XRF raffigurante tutti gli elementi trovati è riportata in [Figura A - 3](#).

Analisi quantitativa elementale

In [Tabella A - 1](#) è riportata l'analisi quantitativa elementale mediante XRF effettuata sul campione; in particolare è riportata la concentrazione in peso per ciascun elemento ritrovato. Dall'analisi quantitativa, si evince che i metalli di maggior interesse economico contenuti nel campione rappresentano circa il 20.8 % della concentrazione in massa totale degli elementi trovati [Figura A - 4](#). Mentre, la concentrazione in massa del rame presente nel campione rispetto al totale rappresenterebbe l'1.93 %.

Allegati

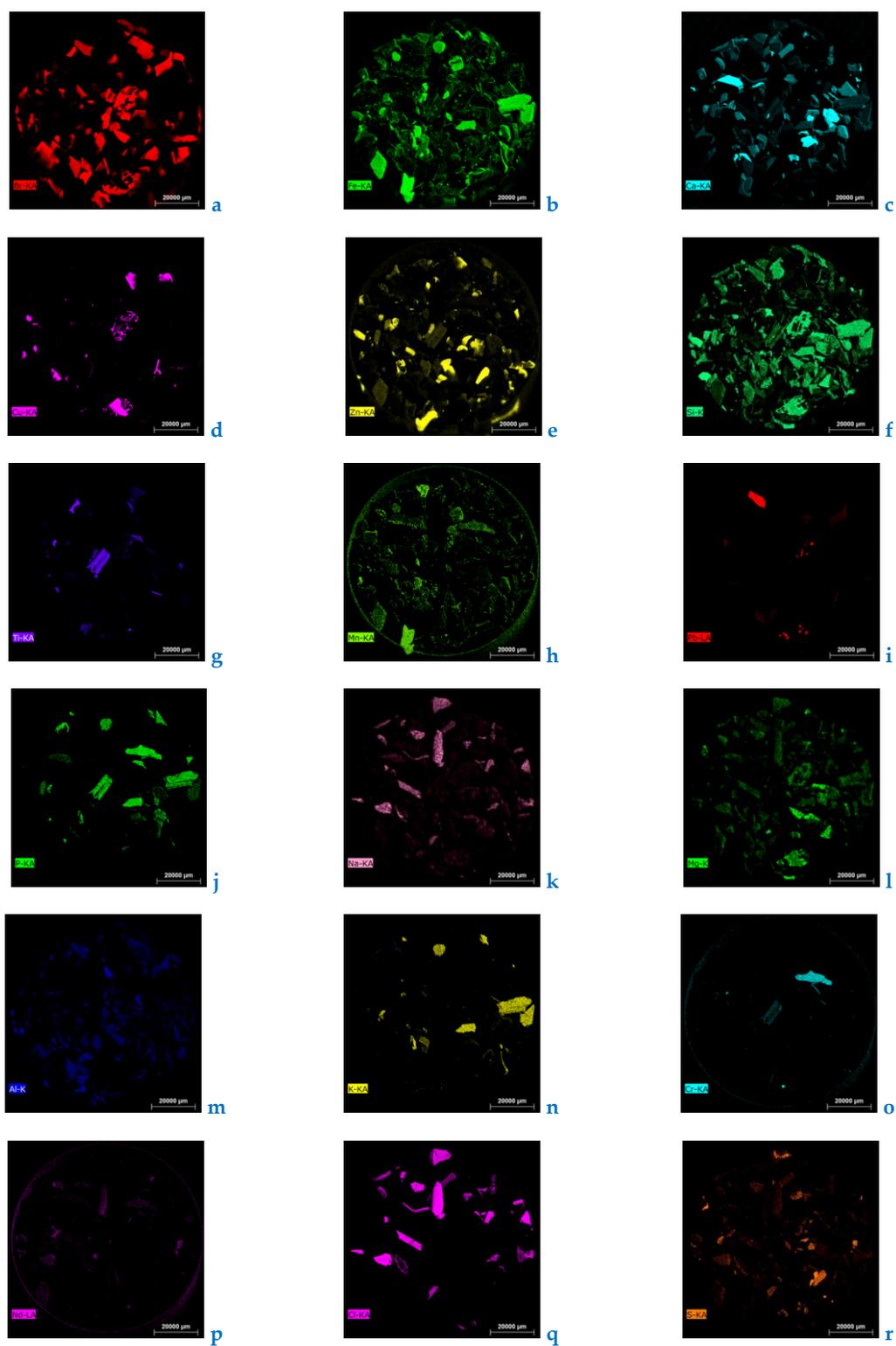


Figura A - 1. Mappe (XRF) elementali di: **a)** Br = non-metal; **b)** Fe = metal; **c)** Ca = alkaline earth metal; **d)** Cu = metal; **e)** Zn = metal; **f)** Si = non-metal; **g)** Ti = metal; **h)** Mn = alkaline earth metal; **i)** Pb = metal; **j)** P = non-metal; **k)** Na = alkali metal; **l)** Mg = metal; **m)** Al = metal; **n)** K = alkali metal; **o)** Cr = metal; **p)** Nd = rare earth; **q)** Cl = non-metal; **r)** S = non-metal.

Allegati

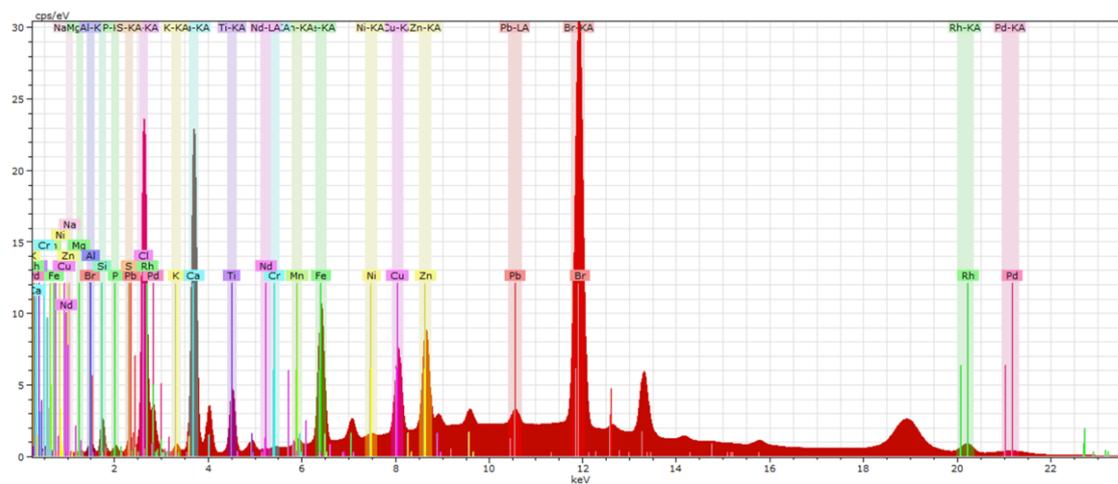


Figura A - 2. Spettro XRF totale della ROI selezionata; sono posti in evidenza i picchi riferibili agli elementi trovati.



Figura A - 3. Mappa XRF di tutti gli elementi trovati nel campione.

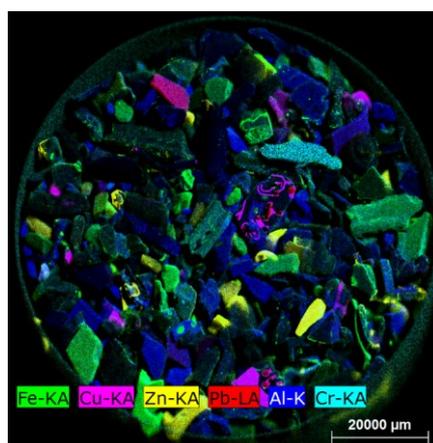


Figura A - 4. Mappa XRF dei metalli di maggior interesse economico.

Allegati

Tabella A - 1. Report dell'analisi quantitativa mediante XRF.

| <i>Element</i> | <i>AN</i> | <i>Series</i> | <i>[norm. at. %]</i> | <i>[norm. wt. %]</i> | <i>Error in wt. %</i> <i>(1 Sigma)</i> |
|-------------------|-----------|-----------------|----------------------|----------------------|---|
| <i>Sodium</i> | 11 | <i>K-series</i> | 4.436 | 2.634 | 1.95x10 ⁻⁰² |
| <i>Magnesium</i> | 12 | <i>K-series</i> | 5.222 | 3.277 | 9.50x10 ⁻⁰³ |
| <i>Aluminium</i> | 13 | <i>K-series</i> | 10.554 | 7.354 | 7.70x10 ⁻⁰³ |
| <i>Silicon</i> | 14 | <i>K-series</i> | 12.617 | 9.151 | 6.10x10 ⁻⁰³ |
| <i>Phosphorus</i> | 15 | <i>K-series</i> | 1.587 | 1.269 | 1.70x10 ⁻⁰³ |
| <i>Sulphur</i> | 16 | <i>K-series</i> | 0.978 | 0.810 | 1.00x10 ⁻⁰³ |
| <i>Chlorine</i> | 17 | <i>K-series</i> | 24.342 | 22.285 | 5.40x10 ⁻⁰³ |
| <i>Potassium</i> | 19 | <i>K-series</i> | 0.718 | 0.725 | 1.10x10 ⁻⁰³ |
| <i>Calcium</i> | 20 | <i>K-series</i> | 23.507 | 24.328 | 4.50x10 ⁻⁰³ |
| <i>Titanium</i> | 22 | <i>K-series</i> | 3.936 | 4.867 | 1.70x10 ⁻⁰³ |
| <i>Chromium</i> | 24 | <i>K-series</i> | 0.059 | 0.080 | 2.00x10 ⁻⁰⁴ |
| <i>Manganese</i> | 25 | <i>K-series</i> | 0.143 | 0.203 | 2.00x10 ⁻⁰⁴ |
| <i>Iron</i> | 26 | <i>K-series</i> | 2.286 | 3.297 | 9.00x10 ⁻⁰⁴ |
| <i>Nickel</i> | 28 | <i>K-series</i> | 0.051 | 0.077 | 1.00x10 ⁻⁰⁴ |
| <i>Copper</i> | 29 | <i>K-series</i> | 1.176 | 1.930 | 6.00x10 ⁻⁰⁴ |
| <i>Zinc</i> | 30 | <i>K-series</i> | 1.044 | 1.763 | 5.00x10 ⁻⁰⁴ |
| <i>Bromine</i> | 35 | <i>K-series</i> | 7.082 | 14.613 | 1.90x10 ⁻⁰³ |
| <i>Neodymium</i> | 60 | <i>L-series</i> | 0.016 | 0.062 | 0.00x10 ⁺⁰⁰ |
| <i>Lead</i> | 82 | <i>L-series</i> | 0.237 | 1.271 | 7.00x10 ⁻⁰⁴ |
| <i>Sum:</i> | | | 100 | 100 | |

ALLEGATO VII- CASE STUDY - RICONOSCIMENTO DELLA TIPOLOGIA DI POLIMERO DEI FRAMMENTI DI *LIGHT GUIDE PANEL* PROVENIENTI DA MONITOR A SCHERMO PIATTO GIUNTI A FINE VITA

In questo allegato è riportata la procedura effettuata per l'identificazione della tipologia di polimero dei frammenti di *Light Guide Panel* (LGP) studiati ([4.1.2 Identificazione a cascata di materiali provenienti da monitor a schermo piatto giunti a fine vita mediante HyperSpectral Imaging](#)).

Siccome, Il *Light Guide Panel* (LGP) dell'unità di retroilluminazione (BLU) dei monitor a schermo e degli schermi a cristallo liquido (LCD) può essere composto principalmente da due tipologie di polimeri: il policarbonato (PC) o dal polimetilmetacrilato (PMMA). Al fine di riconoscere la tipologia di polimero principale presente nei frammenti di LGP è stato quindi messo a punto un modello di classificazione PLS-DA, addestrato con spettri in riflettanza di pellet in PC e PMMA vergini ([Figura A - 5](#)).



Figura A - 5. Plastiche vergini in pellet utilizzate per l'addestramento del modello PLS-DA: a) PMMA e b) PC.

A seguito dell'identificazione dei frammenti di LGP, le particelle riconosciute come "Other (1)", sono state rimosse dall'immagine. L'immagine così ottenuta è stata utilizzata come set di validazione del modello calibrato con gli spettri dei polimeri vergini.

Spettri medi grezzi e pre-elaborati

Da un primo confronto visivo tra lo spettro medi dei frammenti di LGP e quelli dei polimeri vergini, ci si accorge che lo spettro dello LGP rassomiglia a quello del PMMA ([Figura A - 6](#)). Gli spettri medi grezzi dei polimeri vergini sono stati pre-elaborati con l'algoritmo MC ([Figura A - 7](#)), prima della calibrazione del modello PLS-DA.

Allegati

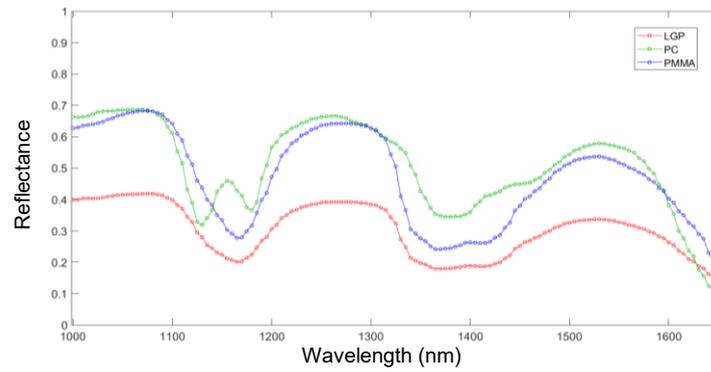
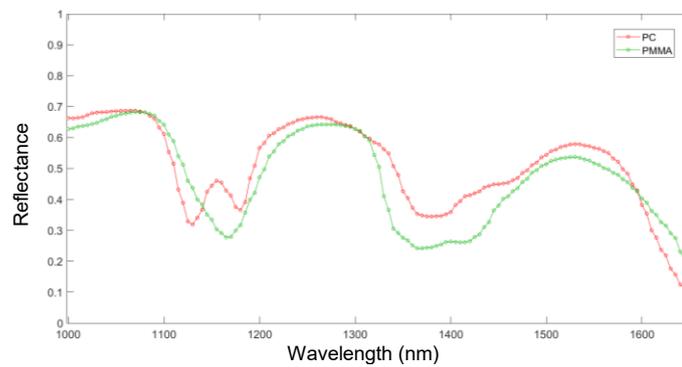
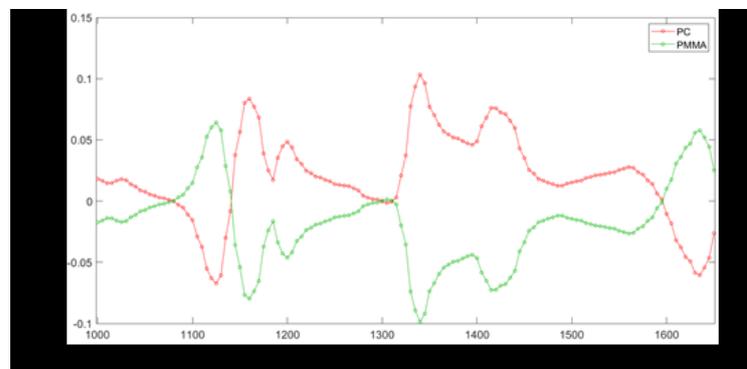


Figura A - 6. Confronto tra lo spettro medio dei frammenti di LGP e quelli dei polimeri vergini, PC e PMMA.



a



b

Figura A - 7. Spettri medi grezzi (a) e pre-elaborati (b) dei polimeri vergini PC e PMMA.

Classificazione PLS-DA finalizzata al riconoscimento della tipologia di polimero

I risultati in termini di mappa di predizione sono riportati in [Figura A - 8](#). Mentre la *Sensitivity* e la *Specificity* in calibrazione, *cross-validation* e validazione sono mostrati in [Tabella A - 2](#). Da quanto si evince dai risultati, il modello riconosce i frammenti di LGP come PMMA.

Allegati

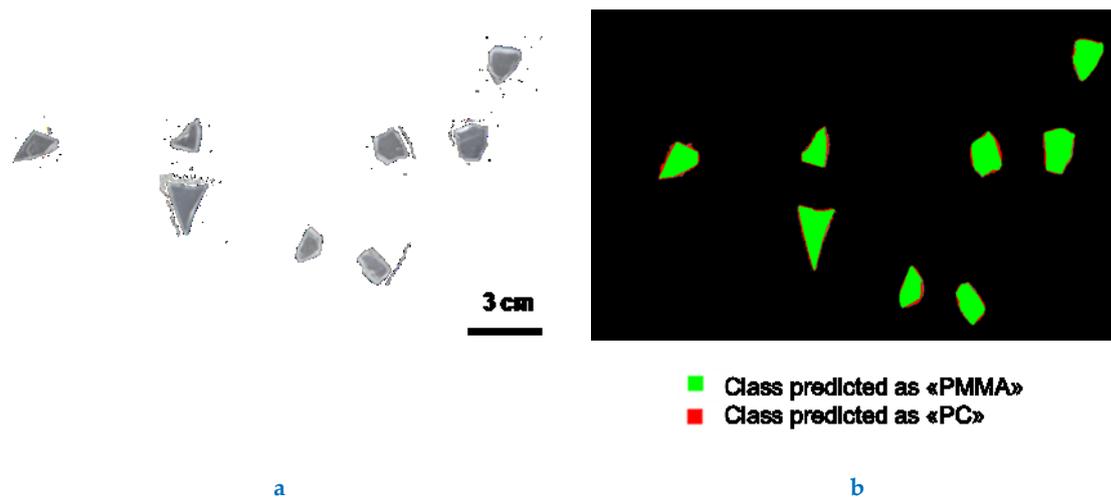


Figura A - 8. Immagine digitale del set di validazione (a) e mappa di predizione ottenuta dalla classificazione PLS-DA (b).

Tabella A - 2. Indicatori di performance del modello PLS-DA.

| | Classe | Sensitivity | Specificity |
|------------------|--------|-------------|-------------|
| Calibration | PC | 1.000 | 1.000 |
| | PMMA | 1.000 | 1.000 |
| Cross-validation | PC | 1.000 | 1.000 |
| | PMMA | 1.000 | 1.000 |
| Validation | PC | - | 0.133 |
| | PMMA | 0.866 | - |

ALLEGATO VIII- CASE STUDY – PACKAGING

Nel seguente allegato sono riportati alcuni dei risultati ottenuti (e le restituzioni grafiche degli stessi) riferibili alle analisi effettuate per l'impostazione di un processo di identificazione a cascata, basato su modelli PLS-DA per discriminare: la frazione plastica, la frazione cellulosica e i poliaccoppiati di carta e plastica in un flusso di rifiuti di imballaggi (4.1.3 *Identificazione a cascata di materiali provenienti da rifiuti di imballaggi mediante HyperSpectral Imaging*).

In particolare, sono riportati: gli spettri medi grezzi e pre-elaborati dei campioni analizzati e gli *scores plot* relativi alle analisi in componenti principali effettuate sugli spettri pre-processati dei campioni.

Spettri medi in riflettanza

In [Figura A - 9a](#) sono mostrati gli spettri medi estratti da ROI delle classi "Paper & cardboard" e "Polymer", mentre in [Figura A - 9b](#) sono mostrati i corrispettivi spettri pre-elaborati con la combinazione di algoritmi SNV + SMOOTHING + MC. Gli spettri medi grezzi e pre-elaborati per le classi "Paper & cardboard 2" e "Paper & cardboard (Laminated card)" sono mostrati rispettivamente in [Figura A - 9c](#) e [Figura A - 9d](#). Mentre, gli spettri medi e pre-elaborati per le classi "Polymer 2" and "Polymer (Laminated plastic)" sono mostrati rispettivamente in [Figura A - 9e](#) e [Figura A - 9f](#).

Principal Component Analysis

L'analisi in componenti principali per le classi "Paper & cardboard" e "Polymer" con 3 componenti principale risolve la maggior parte della varianza (circa 99.18% di varianza catturata). Gli *scores* appartenenti alle due classi si raggruppano in due differenti cluster ([Figura A - 10a](#)). Gli *scores* della classe "Paper & cardboard" sono maggiormente sparpagliati rispetto a quelli della classe "Polymer", ciò è probabilmente dovuto a problemi di *scattering* e riflessione sulla superficie bianca altamente riflettente di alcune particelle cellulosiche. Un comportamento simile può essere osservato per le classi "Paper & cardboard 2" - "Paper & cardboard (Laminated card)" ([Figura A - 10b](#)) e "Polymer" - "Polymer (Laminated plastic)" ([Figura A - 10c](#)). Come ci si aspettava, nel grafico degli *scores* delle prime due componenti principali riportato in [Figura A - 10b](#), si notano sovrapposizioni tra le due classi "Paper & cardboard 2" e "Paper & cardboard (Laminated card)", anche se la PC1 risolve la maggior parte della varianza (91.22%). Mentre, nel caso delle classi "Polymer 2" - "Polymer (Laminated plastic)", visualizzando il grafico degli *scores* delle prime due componenti principali riportato in [Figura A - 10c](#), si nota come il raggruppamento di *scores* "Polymer 2" è parzialmente sovrapposto a quello della classe "Polymer (Laminated plastic)" maggiormente sparpagliato rispetto al primo. Questo fatto è probabilmente dovuto alla rugosità superficiale dei frammenti di poliaccoppiati.

Allegati

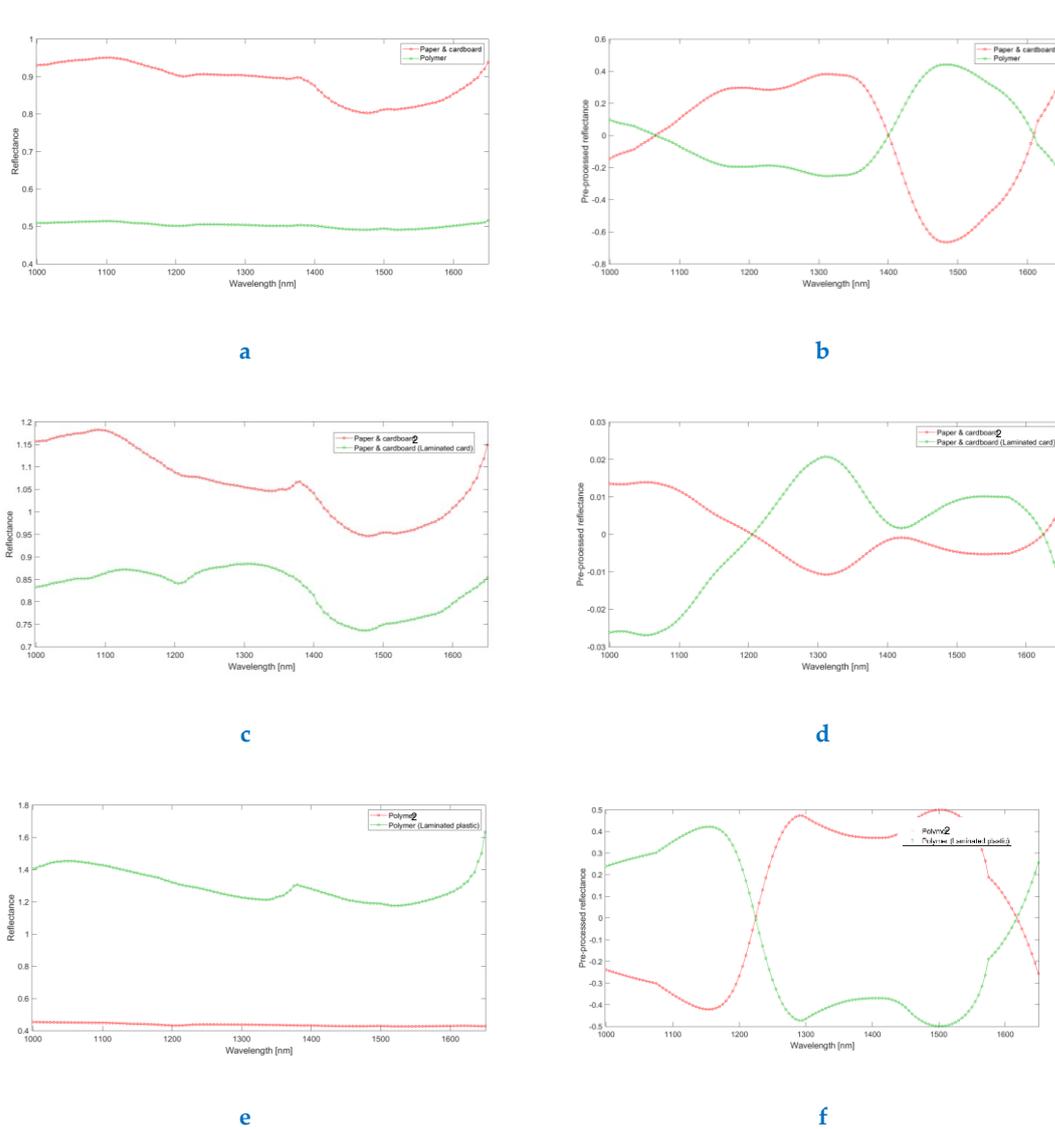
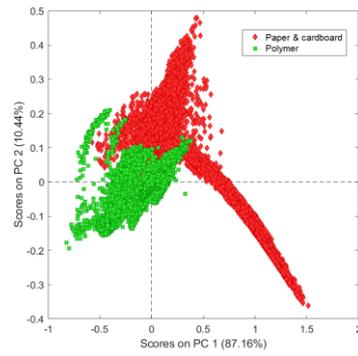
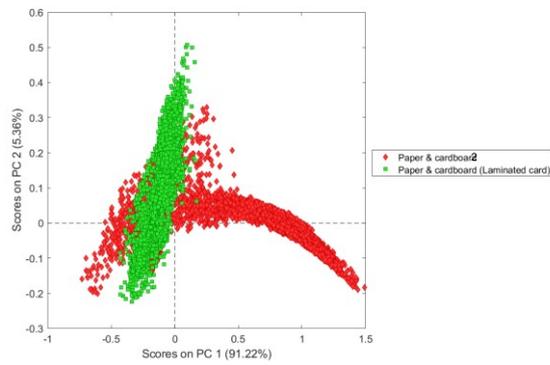


Figura A - 9. Set di calibrazione "Paper & cardboard" - "Polymer": spettri medi grezzi (a) e pre-elaborati (b); set di calibrazione "Paper & cardboard 2" - "Paper & cardboard (Laminated card)": spettri medi grezzi (c) e pre-elaborati (d); set di calibrazione "Polymer 2" - "Polymer (Laminated plastic)". La riflettanza è riportata in scala arbitraria non normalizzata.

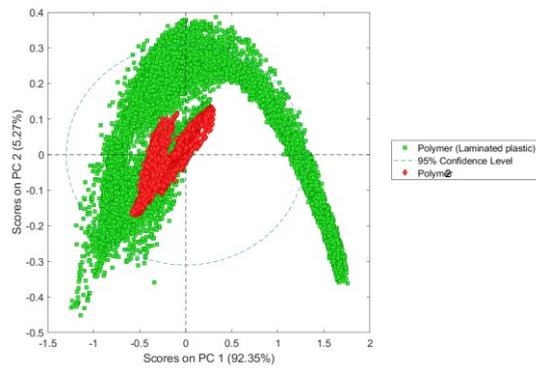
Allegati



a



b



c

Figura A - 10. Scores plot delle prime due componenti principali per le classi: a) "Paper & cardboard" - "Polymer", b) "Paper & cardboard" - "Paper & cardboard (Laminated card)" e c) "Polymer" and "Polymer (Laminated plastic)".

ALLEGATO IX- CASE STUDY - PEPERONCINO

Nel seguente allegato sono riportati i risultati ottenuti (e le restituzioni grafiche degli stessi) riferibili alle analisi effettuate sulla valutazione qualitativa della piccantezza e di altri attributi di campioni di peperoncino essiccato (4.2.1 *Valutazione qualitativa della piccantezza di campioni di peperoncino essiccato mediante spettrofotometria puntuale*). In particolare, sono riportati: gli spettri medi grezzi dei campioni analizzati, gli *scores plot* relativi alle analisi in componenti principali effettuate sugli spettri pre-processati dei campioni, le rette di regressione con riferimento ai modelli PLS per contenuto in ceneri (%) - *Ash content* (%), contenuto di umidità (%) - *Moisture content* (%) e piccantezza - *Pungency* (SHU). Infine, sono riportati i grafici delle posizioni della linea di *boundary*, ricavati dal modello PLS-DA, per le diverse classi di piccantezza analizzate.

Spettri medi in riflettanza

Gli spettri medi in riflettanza sono mostrati in [Figura A - 11](#). Nella [Figura A - 11a](#) e nella [Figura A - 11b](#) sono mostrati gli spettri di riflettanza di tutti i campioni analizzati, rispettivamente secondo la denominazione dei campioni e le classi di piccantezza. Mentre nella [Figura A - 11c](#) e nella [Figura A - 11d](#) sono mostrati gli spettri medi in riflettanza dei campioni di peperoncino in polvere, rispettivamente secondo la denominazione dei campioni e le classi di piccantezza. I campioni di peperoncino frantumati sono mostrati nella [Figura A - 11e](#) e nella [Figura A - 11f](#).

Principal Component Analysis

I grafici degli *scores* in [Figura A - 12](#) mostrano come a seconda dei valori di piccantezza dei campioni analizzati si vengono a generare diversi cluster. Dall'analisi esplorativa si evidenzia anche la granulometria del campione. Infatti, in [Figura A - 12a](#) (come anche in [Figura A - 12b](#)) si può notare come gli *scores* dei peperoncini frantumati (all'interno dell'ovale tratteggiato in blu), anche se maggiormente sparpagliati, risultino essere nettamente separati dagli *scores* dei peperoncini macinati (cerchiati con l'ovale tratteggiato in verde).

Regressioni Partial Least Squares

Gli *scatter plot* delle regressioni PLS, relativamente al contenuto in ceneri e al contenuto di umidità (valori predetti in funzione di quelli misurati), sono mostrati rispettivamente in [Figura A - 13a](#) ed in [Figura A - 13b](#).

Classificazione Partial Least Squares – Discriminant Analysis

Le posizioni del *boundary* discriminante per ciascuna classe prevista di piccantezza sono mostrate in [Figura A - 14](#).

Allegati

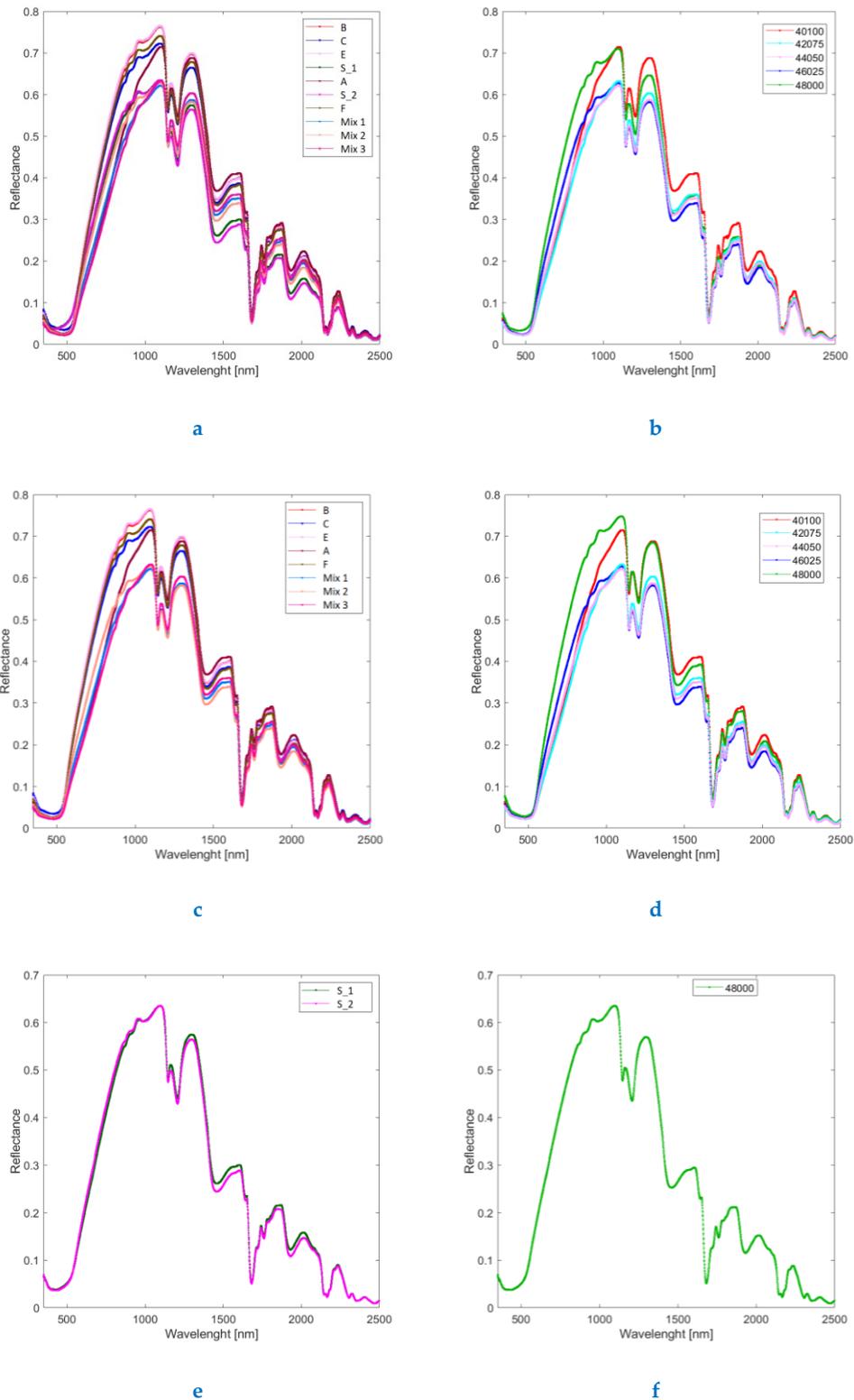


Figura A - 11. Spettri medi in riflettanza (450-2500 nm) di: tutti i campioni (peperoncino essiccato macinato e tritato) etichettati per denominazione del campione (a) e grado di piccantezza - SHU (b); campioni di peperoncino macinati etichettati per denominazione del campione (c) e grado di piccantezza - SHU (d); campioni di peperoncino essiccato tritato, etichettati per denominazione del campione (e) e grado di piccantezza - SHU (f).

Allegati

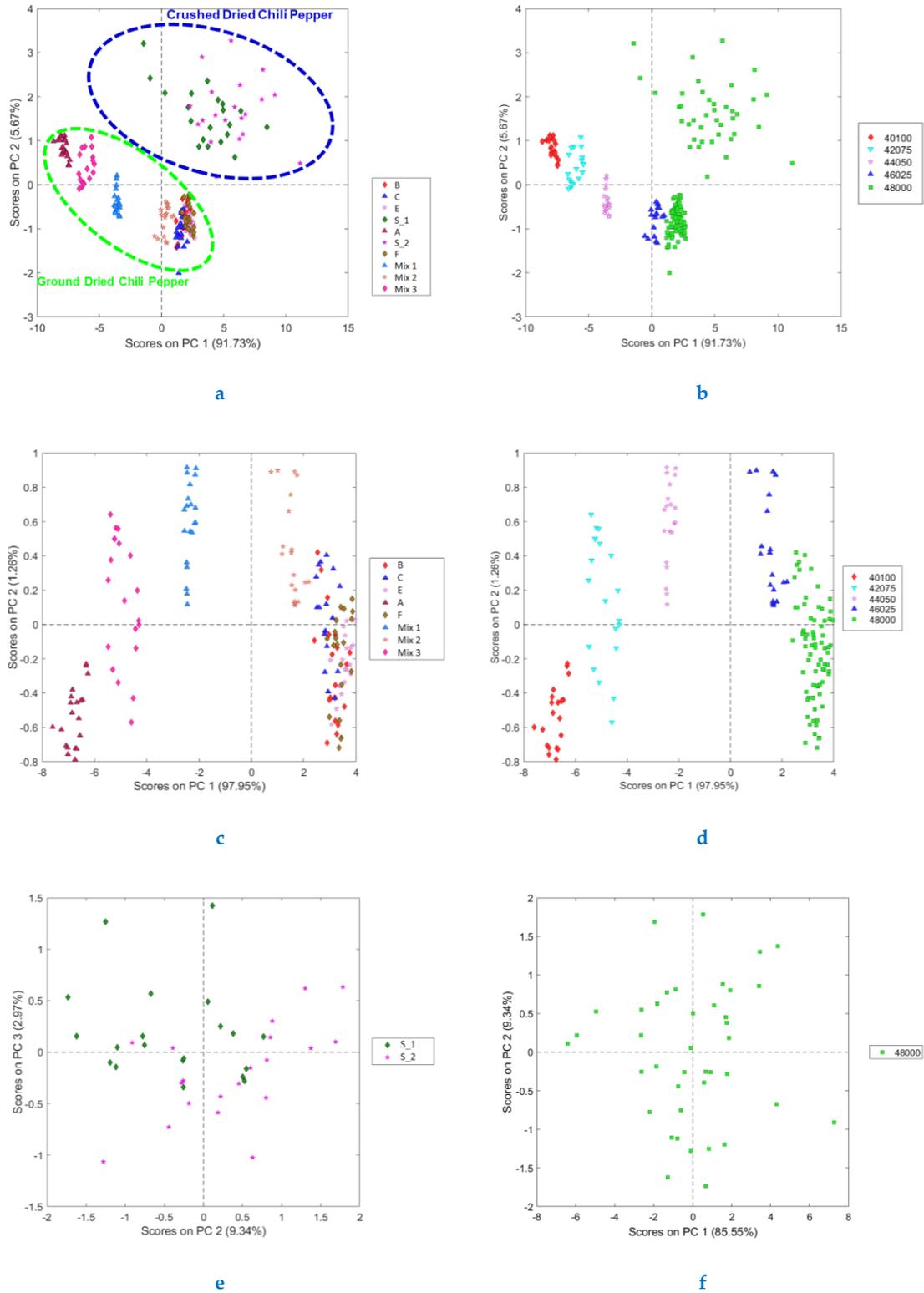
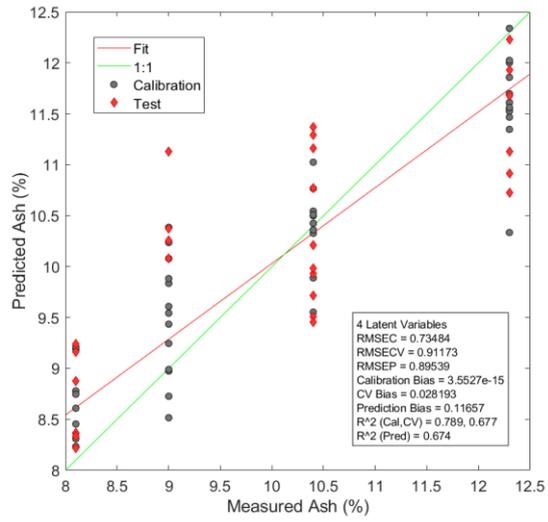
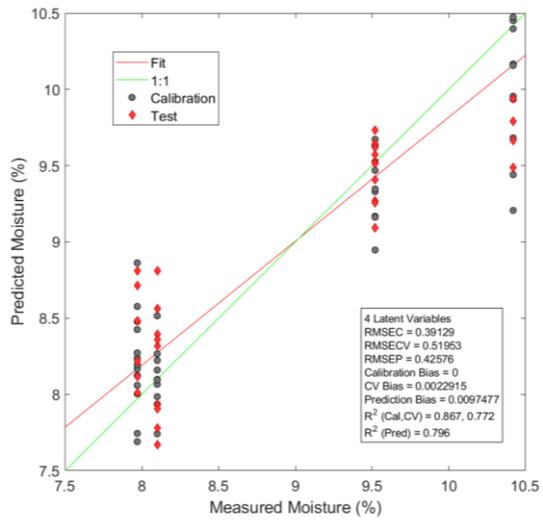


Figura A - 12. Scores su PC 1 vs scores su PC 2 (spettri di riflettanza Vis-SWIR: 450-2500 nm) per: tutti i campioni (peperoncino essiccato macinato e tritato), etichettati per denominazione del campione (a) e grado di piccantezza - SHU (b); campioni di peperoncino macinati, etichettati per denominazione del campione (c) e grado di piccantezza - SHU (d); campioni di peperoncino essiccato tritato, etichettati per denominazione del campione (e) e grado di piccantezza - SHU (f).

Allegati



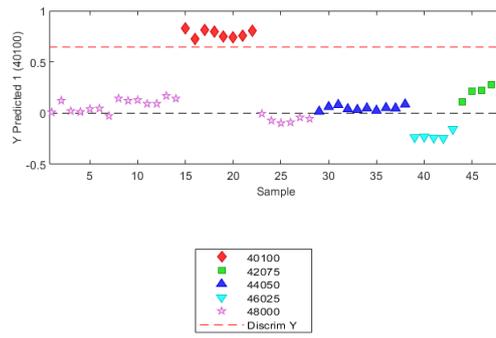
a



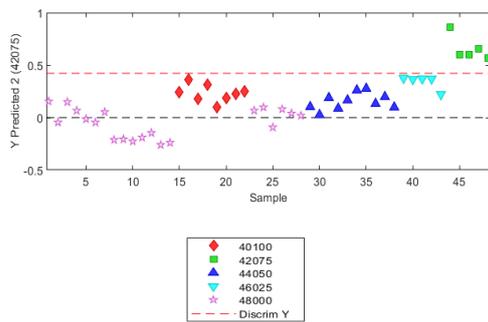
b

Figura A - 13. Diagrammi di regressione PLS effettuate con gli spettri Vis-SWIR (ASD FieldSpec 4; 450-2500 nm) per: contenuto in ceneri (a) e contenuto di umidità (b).

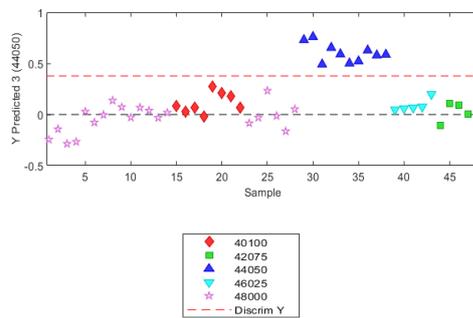
Allegati



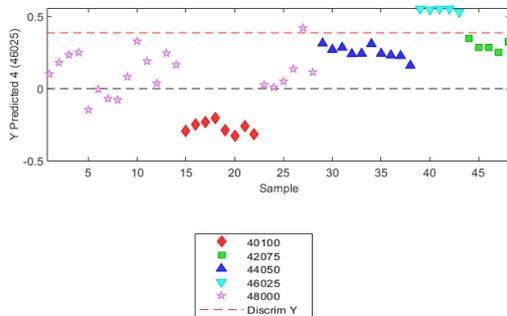
a



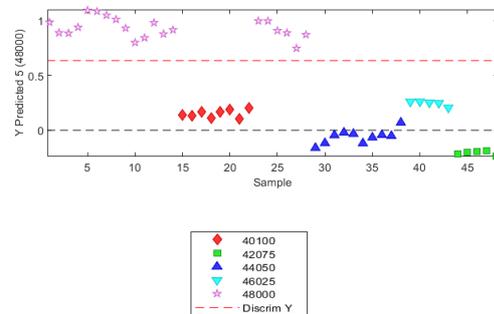
b



c



d



e

Figura A - 14. Posizioni del boundary di discriminazione (linea tratteggiata in rosso), determinato dal modello PLS-DA per le classi: 40100 SHU (a), 42075 SHU (b), 44050 SHU (c), 46025 SHU (d) e 4800 SHU (e).

ALLEGATO X- CASE STUDY – PASTA SURGELATA

Nel seguente allegato sono riportati i risultati ottenuti (e le restituzioni grafiche degli stessi) riferibili alle analisi effettuate sulla valutazione qualitativa della salatura e del contenuto in cenere di pasta surgelata ([4.2.2 Analisi qualitativa di pasta surgelata mediante spettrofotometria puntuale](#)). In particolare, sono riportati: gli spettri medi grezzi dei campioni analizzati, gli *scores plot* relativi alle analisi in componenti principali effettuate sugli spettri pre-processati dei campioni ed i grafici delle posizioni della linea di *boundary*, ricavati dai modelli PLS-DA, per le diverse classi di contenuto in ceneri analizzate.

Spettri medi in riflettanza

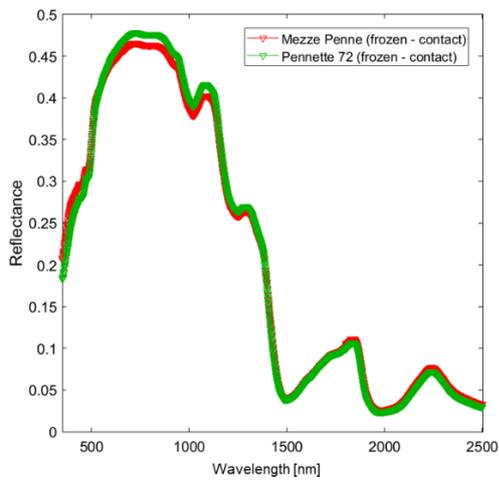
Gli spettri medi in riflettanza dei campioni di pasta analizzati congelati e scongelati, acquisiti con sonda a contatto e a distanza, sono mostrati rispettivamente in [Figura A - 15](#) e [Figura A - 16](#).

Principal Component Analysis

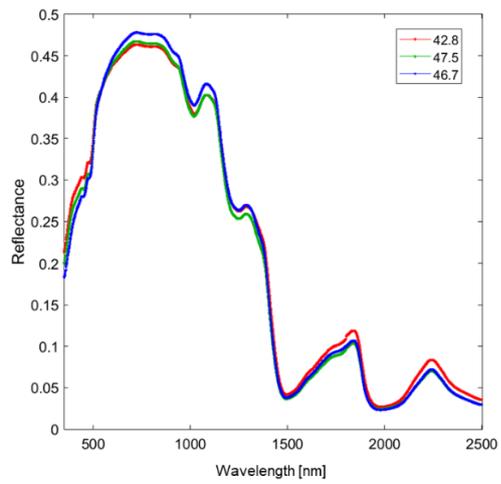
I grafici degli *scores* con riferimento agli spettri in riflettanza dei campioni di pasta analizzati congelati e scongelati, acquisiti con sonda a contatto e a distanza, sono mostrati rispettivamente in [Figura A - 17](#) ed in [Figura A - 18](#), mostrano come a seconda delle casistiche analizzate, si vengono a creare dei cluster. Come si può notare, nel caso dei campioni surgelati, grazie alla prime due componenti principali si può catturare la maggior parte della varianza presente.

Classificazione Partial Least Squares – Discriminant Analysis

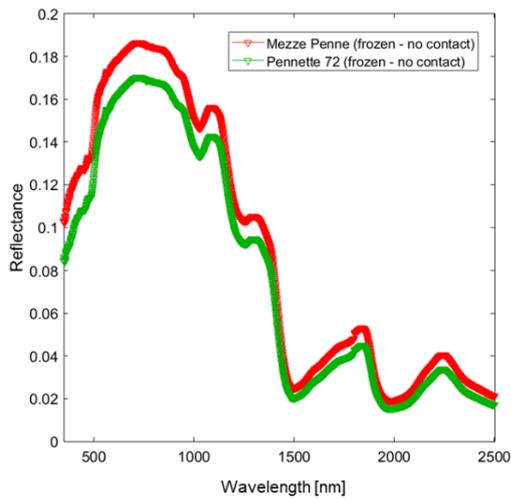
Le posizioni del *boundary* discriminante per ciascuna classe prevista di piccantezza sono mostrate in [Figura A - 19](#).



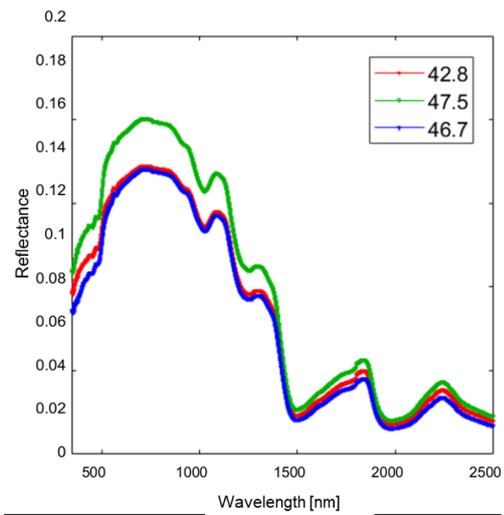
a



b



c



d

Figura A - 15. Spettri medi in riflettanza (350-2500 nm) di: campioni surgelati acquisiti con sonda a contatto, con classi relative alla tipologia del campione (a) e alle classi di residuo secco (b); campioni surgelati acquisiti con sonda a distanza, con classi relative alla tipologia del campione (c) e alle classi di residuo secco (d).

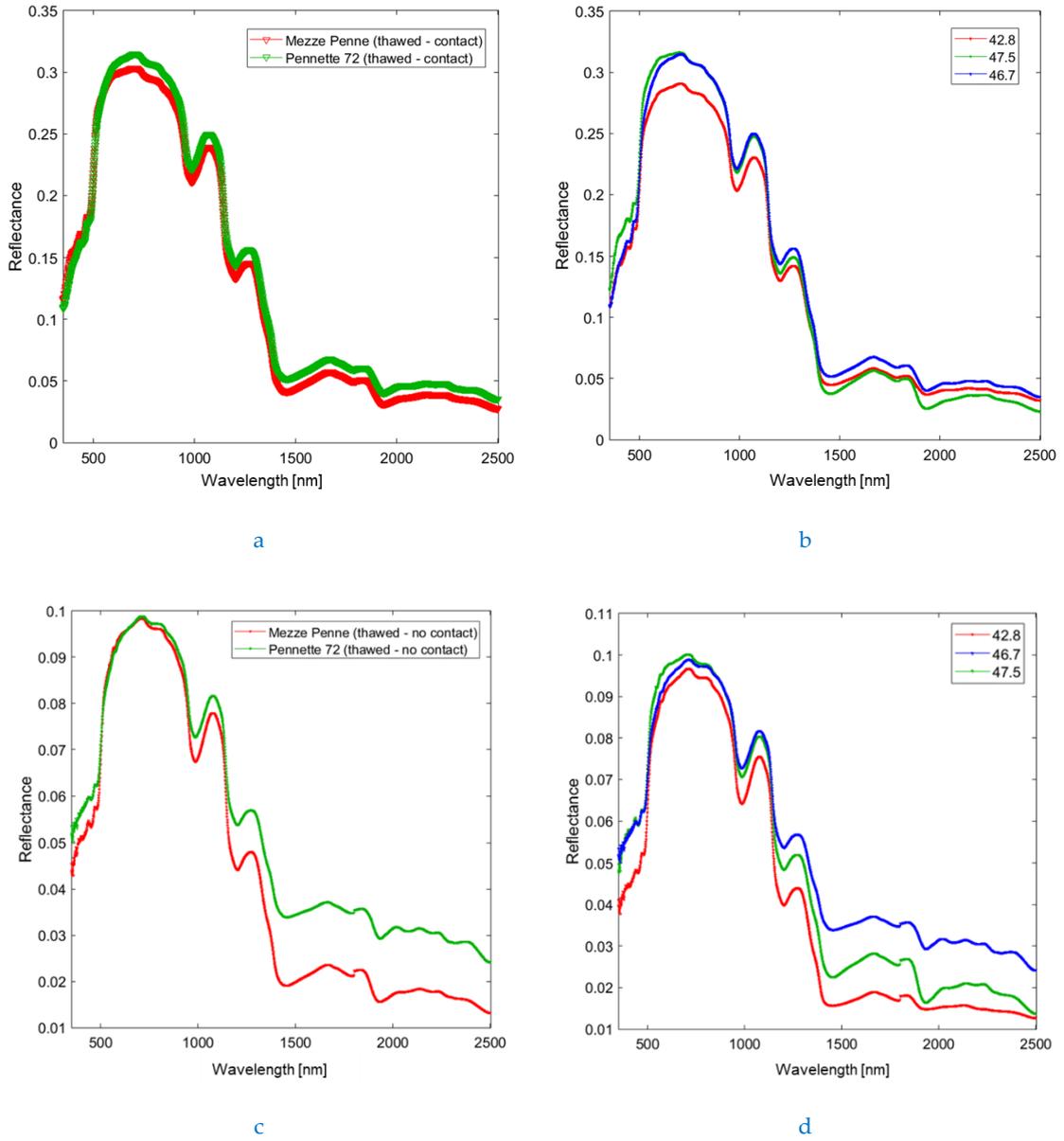
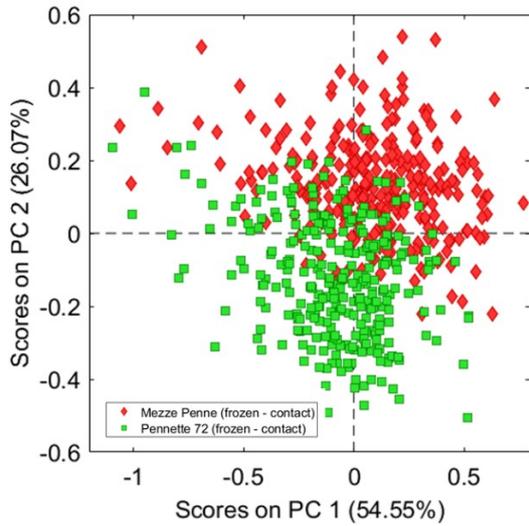
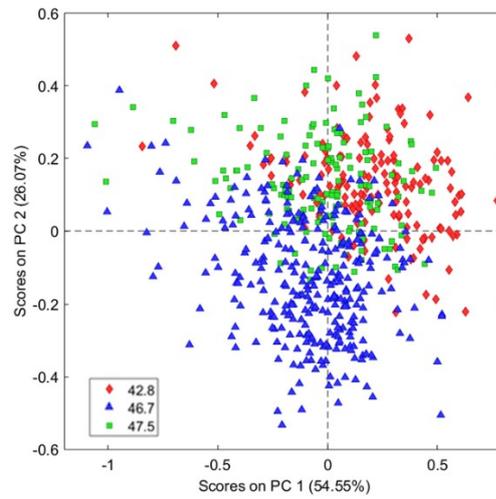


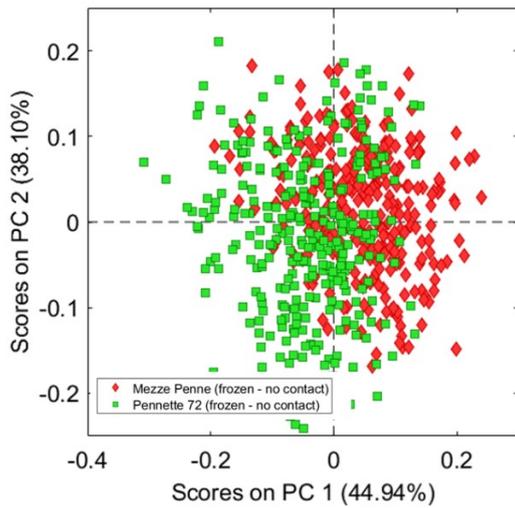
Figura A - 16. Spettri medi in riflettanza (350-2500 nm): campioni scongelati acquisiti con sonda a contatto, con classi relative alla tipologia del campione (a) e alle classi di residuo secco (b); campioni scongelati acquisiti con sonda a distanza, con classi relative alla tipologia del campione (c) e alle classi di residuo secco (d).



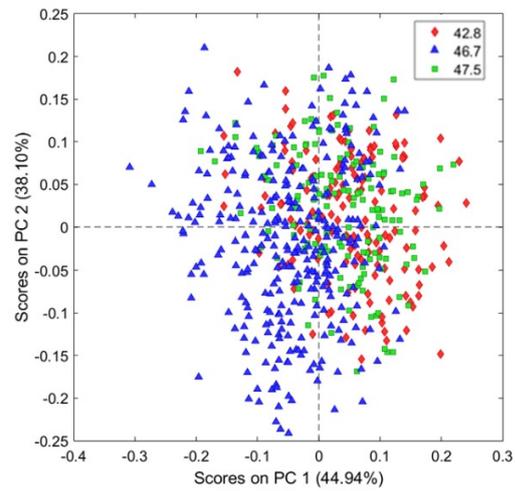
a



b



c



d

Figura A - 17. Scores su PC 1 vs scores su PC 2 (spettri di riflettanza Vis-SWIR: 350-2500 nm) per: campioni surgelati acquisiti con sonda a contatto, con classi relative alla tipologia del campione (a) e alle classi di residuo secco (b); campioni surgelati acquisiti con sonda a distanza, con classi relative alla tipologia del campione (c) e alle classi di residuo secco (d).

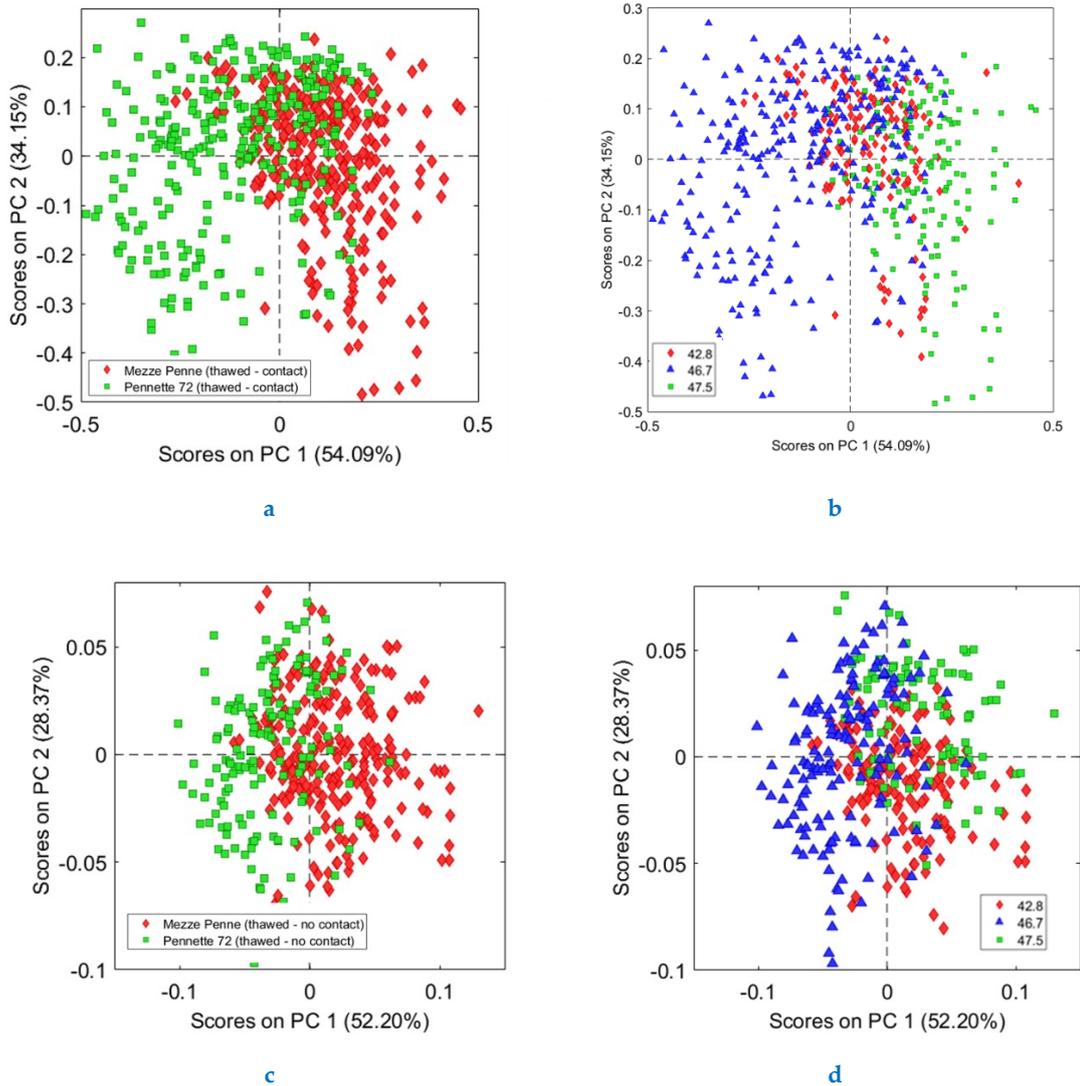
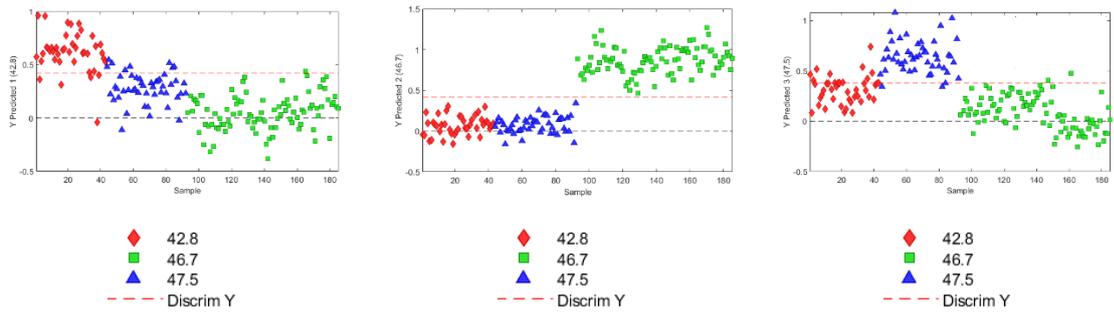
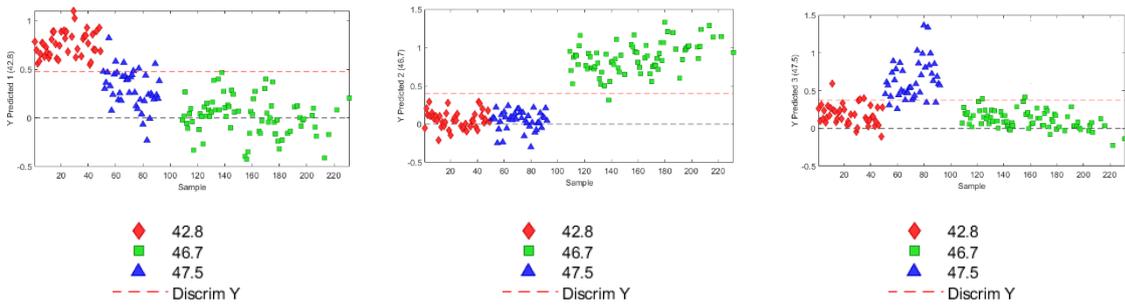


Figura A - 18. Scores su PC 1 vs scores su PC 2 (spettri di riflettanza Vis-SWIR: 350-2500 nm) per: campioni scongelati acquisiti con sonda a contatto, con classi relative alla tipologia del campione (a) e alle classi di residuo secco (b); campioni scongelati acquisiti con sonda a distanza, con classi relative alla tipologia del campione (c) e alle classi di residuo secco (d).

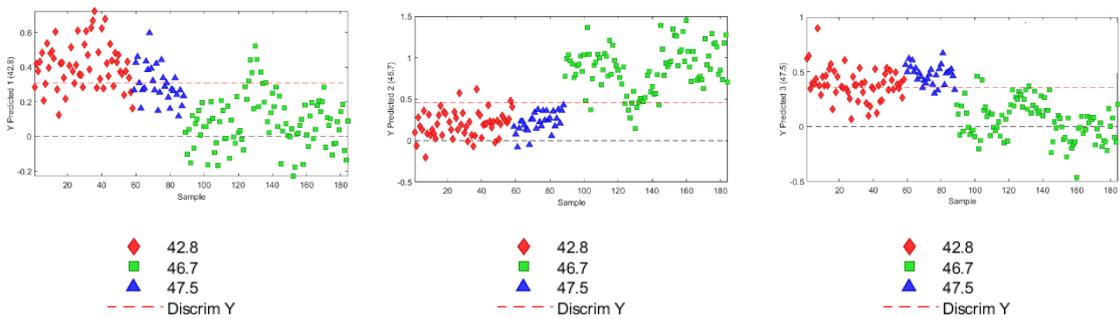
Allegati



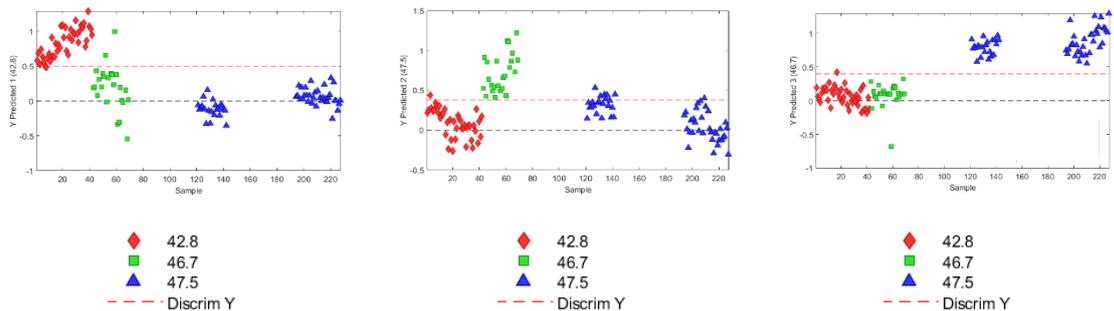
a



b



c



d

Figura A - 19. Posizioni del boundary di discriminazione di ciascuna delle tre classi di tenore in ceneri (linea tratteggiata in rosso), determinate dai modelli PLS-DA per: campioni surgelati acquisiti con sonda a contatto (a), campioni scongelati acquisiti con sonda a contatto (b), campioni surgelati acquisiti a distanza (c) e campioni scongelati acquisiti a distanza (d).

ALLEGATO XI- CASE STUDY – NOCCIOLE

Nel seguente allegato si riportano i risultati ottenuti dalle analisi effettuate per la discriminazione di nocciole buone dai relativi scarti (4.2.3 *Classificazione mediante hyperspectral imaging di nocciole buone e relativi scarti*), nei tre casi considerati:

- I. analisi effettuate utilizzando la combinazione di *pre-processing* SNV +MC;
- II. analisi effettuate utilizzando la combinazione di *pre-processing* DETREND +SNV +MC,
- III. analisi effettuate utilizzando la combinazione di *pre-processing* GLS-W +MC.

I - Caso SNV + MC

Spettri medi grezzi e pre-processati

Gli spettri medi grezzi e pre-processati (utilizzando la combinazione di algoritmi SNV + MC) del dataset Nocciole di calibrazione sono mostrati in [Figura A - 20](#).

Analisi in componenti principali (PCA)

Lo *scores plot* PC 1 vs PC 2 della PCA effettuata sul dataset di calibrazione Nocciole è mostrato in [Figura A - 21](#), mentre il *loadings plot* è mostrato in [Figura A - 22](#).

Classificazione Partial Least Squares – Discriminant Analysis (PLS-DA)

La mappa di predizione, risultante dalla classificazione PLS-DA, è mostrata in [Figura A - 23b](#). Mentre, in [Tabella A - 3](#) sono mostrati i parametri relativi alla matrice di confusione per la valutazione delle performance di classificazione del modello PLS-DA in predizione.

Selezione delle variabili mediante Interval-PLSDA

In [Figura A - 24](#) è mostrato il risultato della selezione delle variabili mediante I-PLSDA. Le variabili selezionate ricadono negli intervalli: 1339 nm-1390 nm, 1566 – 1616 nm, 1848-1899 nm e 2130-2180 nm.

Nuovo modello PLS-DA con variabili selezionate

La mappa di predizione, risultante dalla classificazione PLS-DA a seguito della selezione delle variabili, è mostrata in [Figura A - 25](#). Mentre, in [Tabella A - 4](#) sono mostrati i parametri relativi alla matrice di confusione per la valutazione delle performance di classificazione del modello PLS-DA in predizione, a seguito della selezione delle variabili.

Allegati

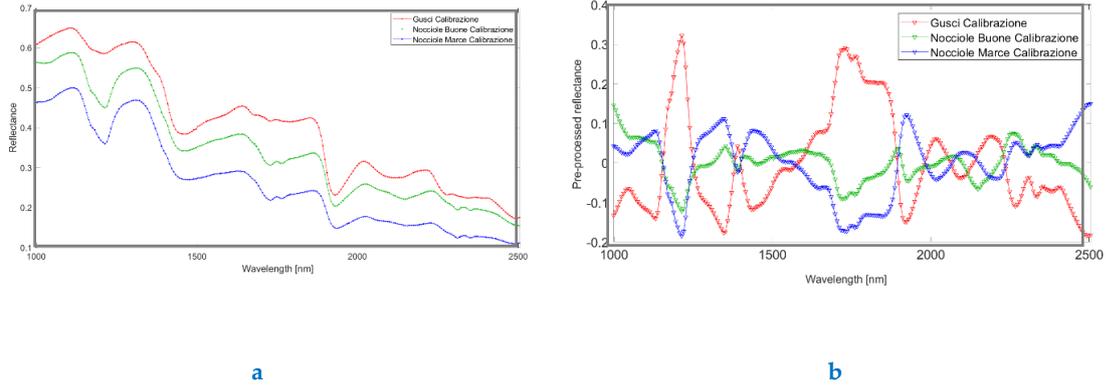


Figura A - 20. Spettri medi grezzi (a) e pre-processati (b) del dataset di calibrazione con algoritmo SNV + MC.

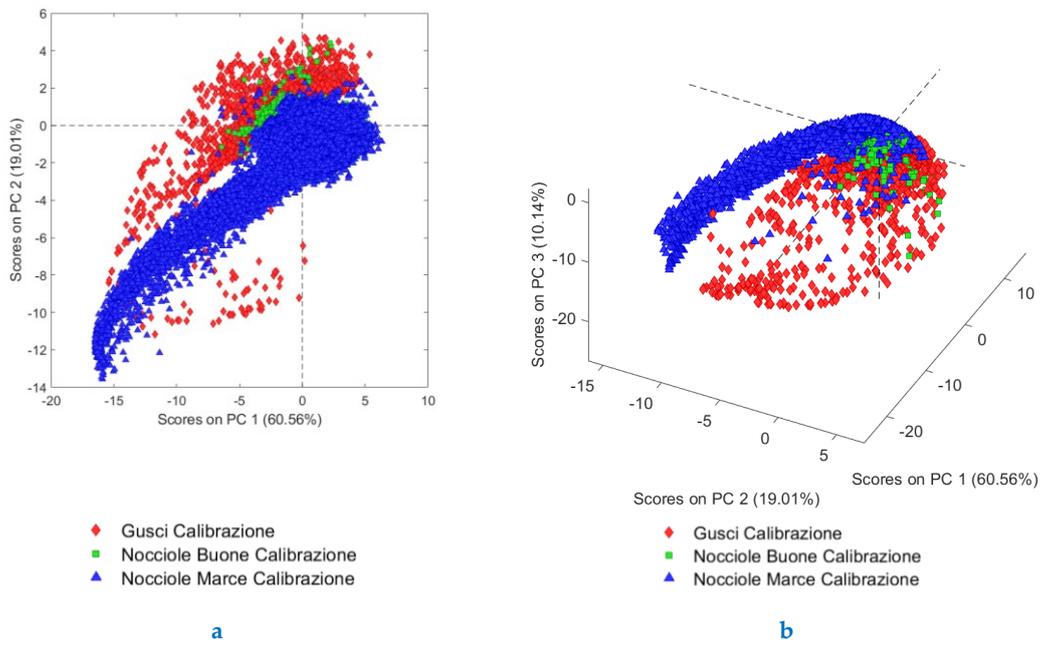


Figura A - 21. Scores plot PC 1-PC 2 (a) e scores on PC 1 vs scores on PC 2 vs scores on PC 3 (b).

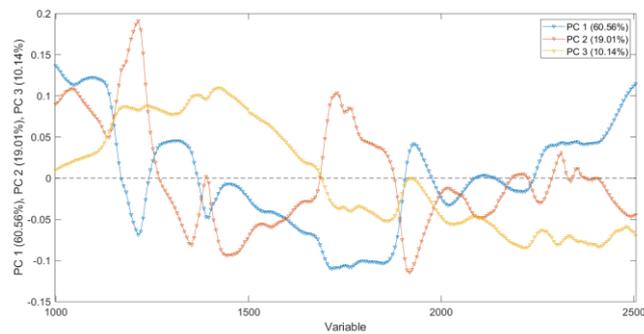


Figura A - 22. Loadings plot (PC 1, PC 2 e PC 3 in funzione della lunghezza d'onda).

Allegati

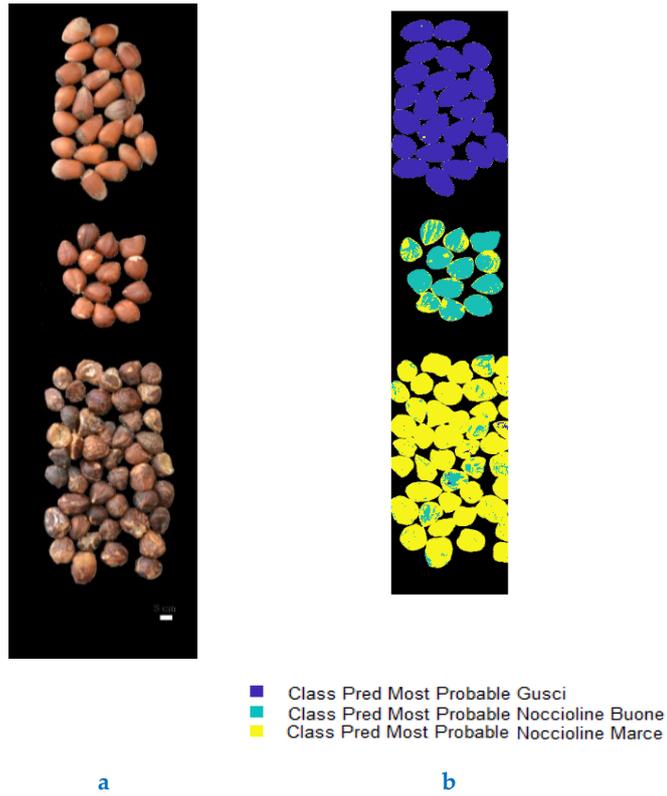
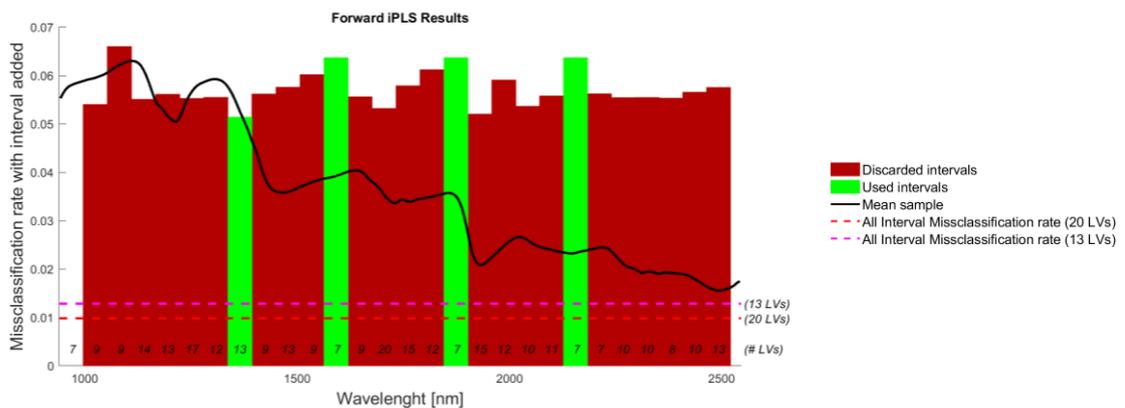


Figura A - 23. Immagine RGB (a) e risultato della classificazione PLS-DA: mappa - class predicted most probable (b).

Tabella A - 3. Parametri relativi alla matrice di confusione per la valutazione delle performance di classificazione del modello PLS-DA in predizione.

| Classe | Sensitivity | Specificity | Numero di spettri | Mis-classification Error | Precisione |
|------------------|-------------|-------------|-------------------|--------------------------|------------|
| Gusci | 0.997 | 1.000 | 30892 | 0.001 | 0.999 |
| Noccioline Buone | 0.658 | 0.972 | 13897 | 0.077 | 0.809 |
| Noccioline Marce | 0.953 | 0.893 | 45021 | 0.077 | 0.900 |



Allegati

Figura A - 24. Risultato della selezione delle variabili mediante I-PLSDA.

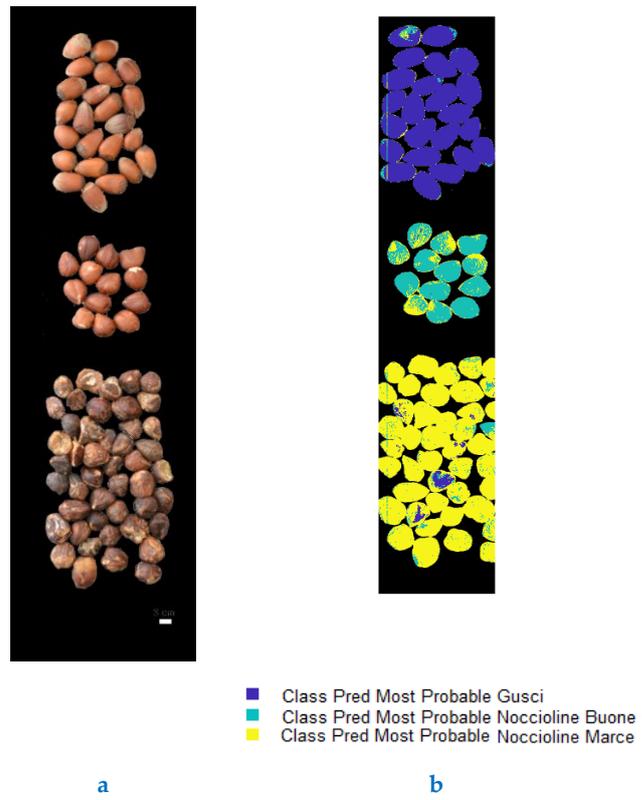


Figura A - 25. Immagine RGB (a) e risultato della classificazione PLS-DA: mappa - class predicted most probable (b).

Allegati

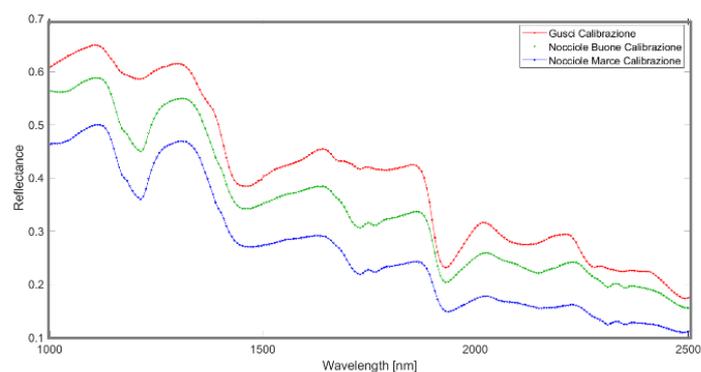
Tabella A - 4. Parametri relativi alla matrice di confusione per la valutazione delle performance di classificazione del modello PLS-DA in predizione, a seguito della selezione delle variabili.

| Classe | Sensitivity | Specificity | Numero di spettri | Mis-classification Error | Precisione |
|----------------|-------------|-------------|-------------------|--------------------------|------------|
| Gusci | 0.948 | 0.987 | 30892 | 0.026 | 0.975 |
| Nocciole Buone | 0.664 | 0.969 | 13897 | 0.079 | 0.794 |
| Nocciole Marce | 0.955 | 0.884 | 45021 | 0.080 | 0.893 |

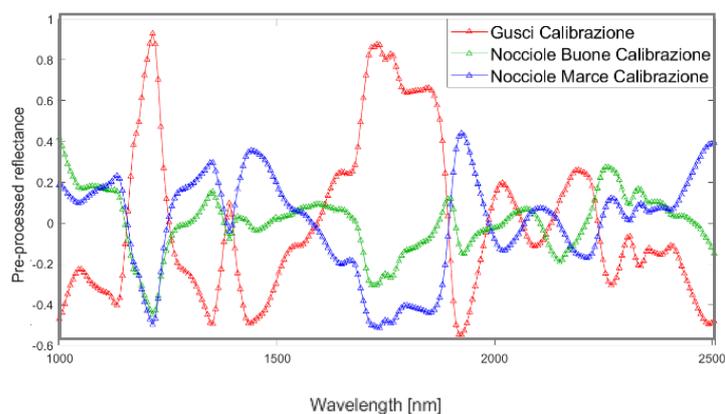
II - Caso DETREND + SNV + MC

Spettri medi pre-processati

Gli spettri medi grezzi e pre-processati (utilizzando la combinazione di algoritmi DETREND + SNV + MC) del dataset Nocciole di calibrazione sono mostrati in [Figura A - 26](#).



a



b

Figura A - 26. Spettri medi grezzi (a) e pre-processati (b) del dataset di calibrazione con algoritmo DETREND + SNV + MC.

Analisi in componenti principali (PCA)

Lo *scores plot* della PCA effettuata sul dataset di calibrazione Nocchie è mostrato in [Figura A - 27](#), mentre il *loadings plot* è mostrato in [Figura A - 28](#).

Classificazione Partial Least Squares – Discriminant Analysis (PLS-DA)

La mappa di predizione, risultante dalla classificazione PLS-DA, è mostrata in [Figura A - 29b](#). Mentre, in [Tabella A - 5](#) sono mostrati i parametri relativi alla matrice di confusione per la valutazione delle performance di classificazione del modello PLS-DA in predizione.

Selezione delle variabili mediante Interval-PLSDA

In [Figura A - 30](#) è mostrato il risultato della selezione delle variabili mediante I-PLSDA. Le variabili selezionate ricadono negli intervalli: 1000 – 1049 nm, 1339 – 1390 nm, 1622 – 1673 nm e 2074-2124 nm.

Nuovo modello PLS-DA con variabili selezionate

La mappa di predizione, risultante dalla classificazione PLS-DA a seguito della selezione delle variabili, è mostrata in [Figura A - 31b](#). Mentre, in [Tabella A - 6](#) sono mostrati i parametri relativi alla matrice di confusione per la valutazione delle performance di classificazione del modello PLS-DA in predizione, a seguito della selezione delle variabili.

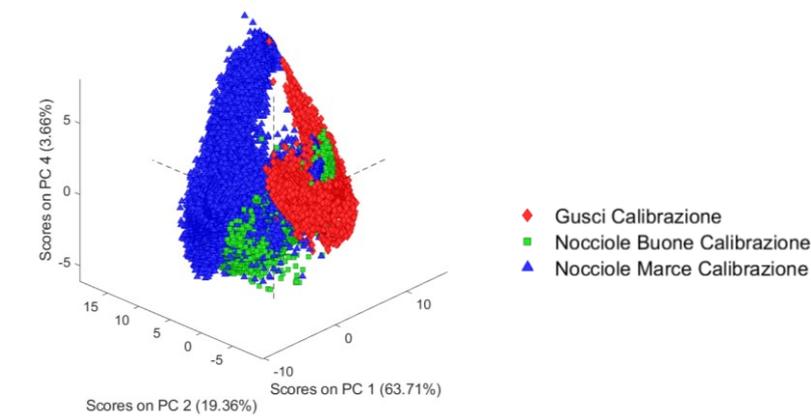


Figura A - 27. Scores plot (scores on PC 1 vs scores on PC 2 vs scores on PC 4).

Allegati

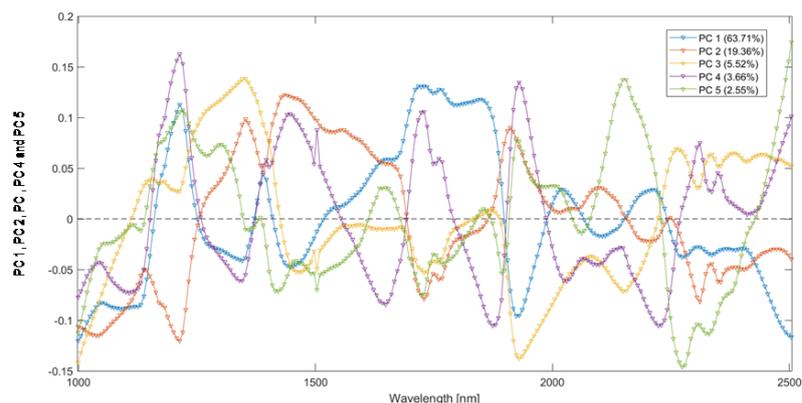


Figura A - 28. Loadings plot (PC 1, PC 2, PC 3, PC 4 e PC 5 in funzione della lunghezza d'onda.

Tabella A - 5. Parametri relativi alla matrice di confusione per la valutazione delle performance di classificazione del modello PLS-DA in predizione.

| Classe | Sensitivity | Specificity | Numero di spettri | Mis-classification Error | Precisione |
|----------------|-------------|-------------|-------------------|--------------------------|------------|
| Gusci | 0.999 | 0.999 | 30892 | 0.001 | 0.998 |
| Nocciole Buone | 0.695 | 0.977 | 13897 | 0.067 | 0.847 |
| Nocciole Marce | 0.960 | 0.905 | 45021 | 0.068 | 0.910 |

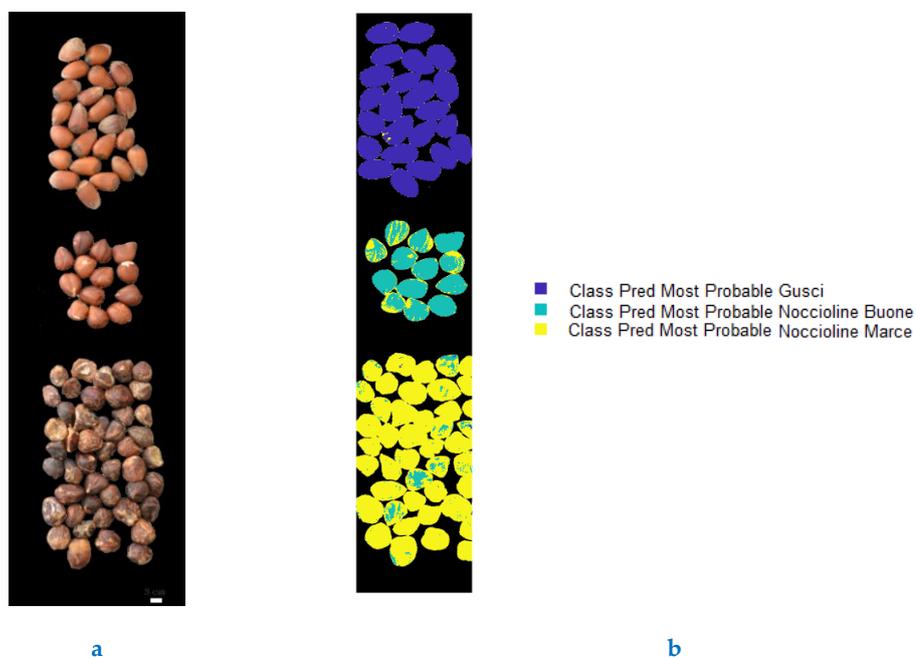


Figura A - 29. Immagine RGB (a) e risultato della classificazione PLS-DA: mappa - class predicted most probable (b).

Allegati

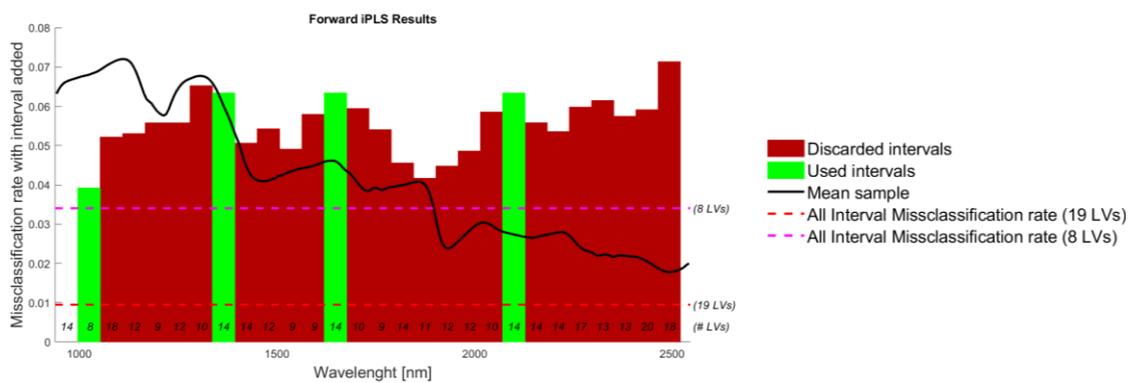


Figura A - 30. Risultato della selezione delle variabili mediante I-PLSDA.

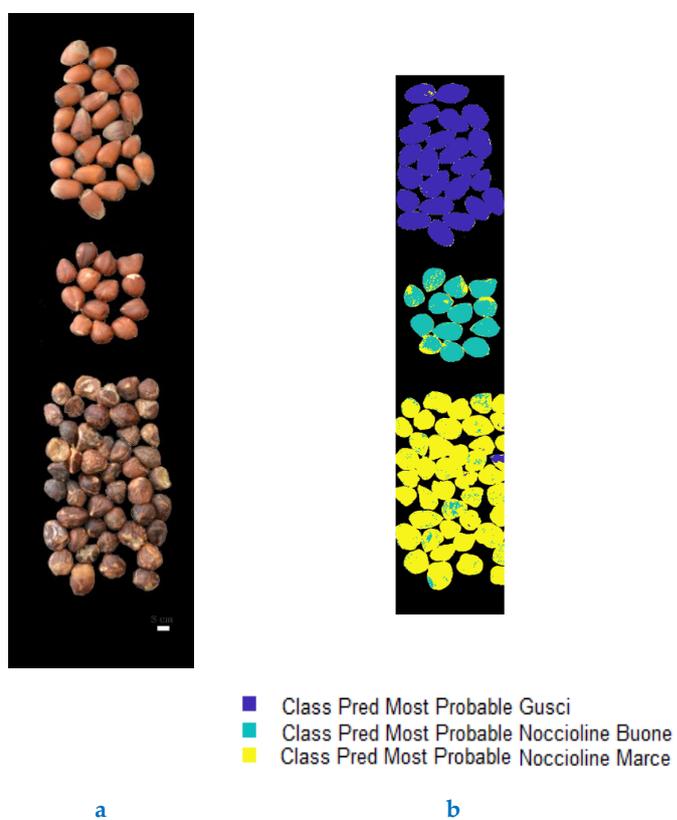


Figura A - 31. Immagine RGB (a) e risultato della classificazione PLS-DA: mappa - class predicted most probable (b).

Allegati

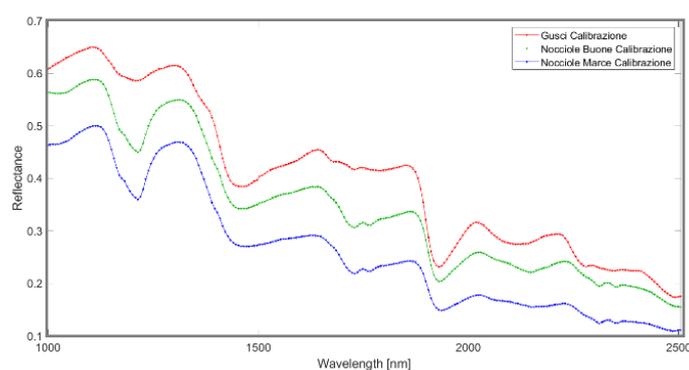
Tabella A - 6. Parametri relativi alla matrice di confusione per la valutazione delle performance di classificazione del modello PLS-DA in predizione, a seguito della selezione delle variabili.

| Classe | Sensitivity | Specificity | Numero di spettri | Mis-classification Error | Precisione |
|----------------|-------------|-------------|-------------------|--------------------------|------------|
| Gusci | 0.992 | 0.997 | 30892 | 0.005 | 0.994 |
| Nocciole Buone | 0.824 | 0.980 | 13897 | 0.044 | 0.885 |
| Nocciole Marce | 0.965 | 0.942 | 45021 | 0.047 | 0.943 |

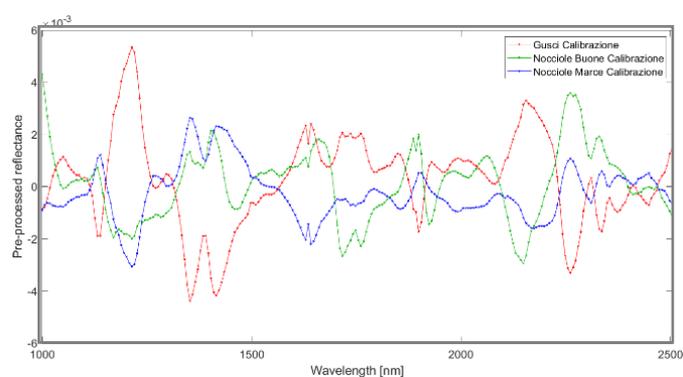
Allegato III - Caso GLS-W + MC

Spettri medi pre-processati

Gli spettri medi grezzi e pre-processati (utilizzando la combinazione di algoritmi GLS-W + MC) del dataset Nocciole di calibrazione sono mostrati in [Figura A - 32](#).



a



b

Figura A - 32. Spettri medi grezzi (a) e pre-processati (DETREND + SNV + MC) (b) del dataset di calibrazione.

Analisi in componenti principali (PCA)

Lo *scores plot* PC 1 vs PC 2 della PCA effettuata sul dataset di calibrazione Nocchie è mostrato in [Figura A - 33](#), mentre il *loadings plot* è mostrato in [Figura A - 34](#).

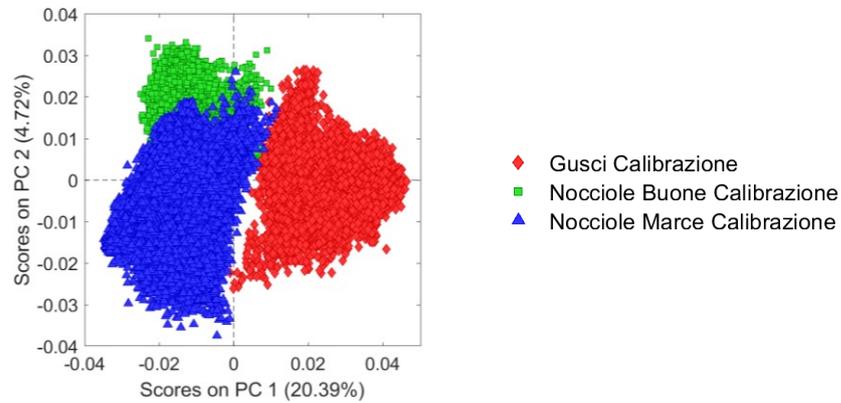


Figura A - 33. Scores plot (scores on PC 1 vs scores on PC 2).

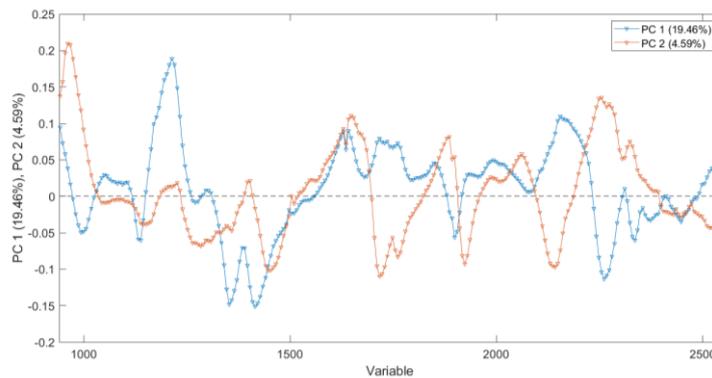


Figura A - 34. Loadings plot (PC 1 e PC 2 in funzione della lunghezza d'onda).

Classificazione Partial Least Squares – Discriminant Analysis (PLS-DA)

In [Tabella A - 7](#) sono mostrati i parametri relativi alla matrice di confusione per la valutazione delle performance di classificazione del modello PLS-DA in predizione. Mentre, la mappa di predizione, risultante dalla classificazione PLS-DA, è mostrata in [Figura A - 35b](#).

Allegati

Tabella A - 7. Parametri relativi alla matrice di confusione per la valutazione delle performance di classificazione del modello PLS-DA in predizione.

| Classe | Sensitivity | Specificity | Numero di spettri | Mis-classification Error | Precisione |
|------------------|-------------|-------------|-------------------|--------------------------|------------|
| Gusci | 0.999 | 0.998 | 30892 | 0.002 | 0.995 |
| Noccioline Buone | 0.949 | 0.976 | 13897 | 0.028 | 0.877 |
| Noccioline Marce | 0.956 | 0.983 | 45021 | 0.030 | 0.983 |

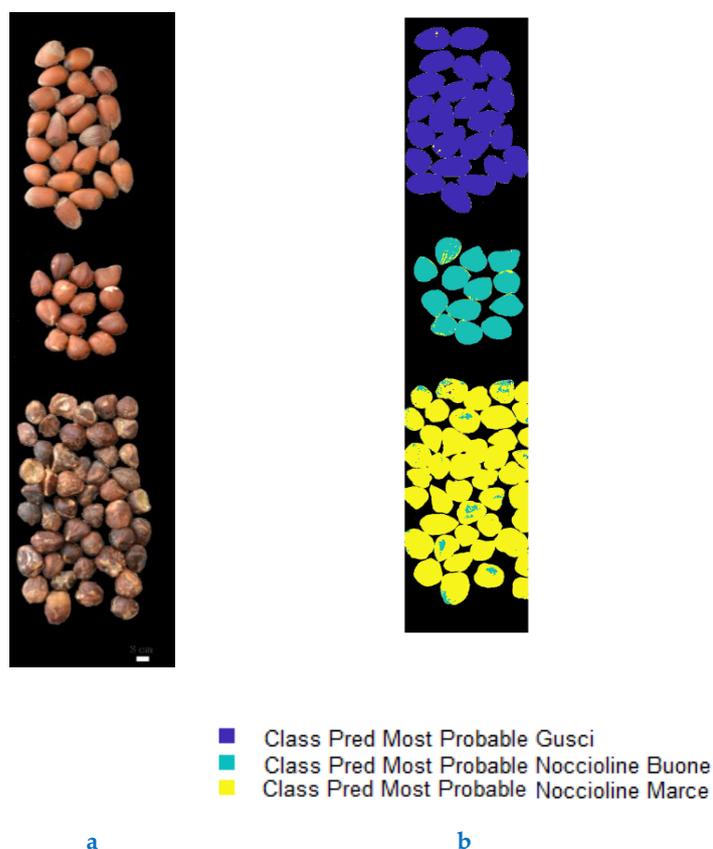


Figura A - 35. Immagine RGB (a) e risultato della classificazione PLS-DA: mappa - class predicted most probable (b).

Selezione delle variabili mediante Interval-PLSDA

In Figura A - 36 è mostrato il risultato della selezione delle variabili mediante I-PLSDA. Le variabili selezionate dall'algoritmo I-PLSDA ricadono nei seguenti intervalli: 999 – 1049 nm, 1112 – 1163 nm, 1848 - 1899 nm e 2130 - 2180 nm.

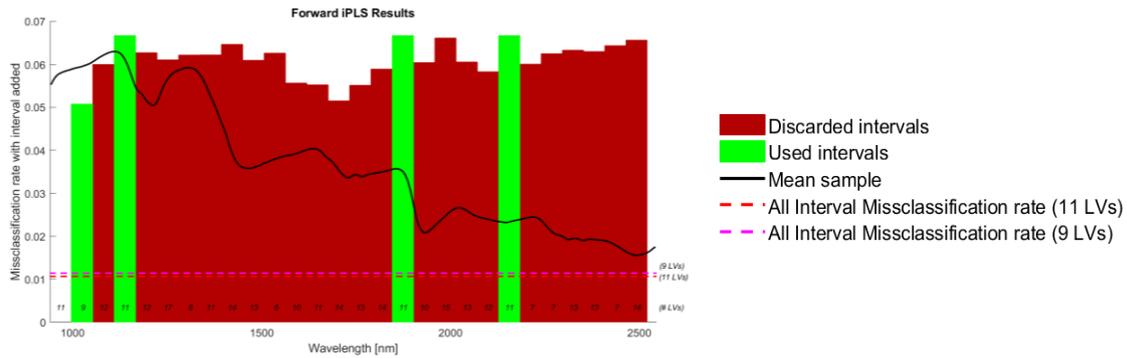


Figura A - 36. Risultato della selezione delle variabili mediante I-PLSDA.

Nuovo modello PLS-DA con variabili selezionate

La mappa di predizione, risultante dalla classificazione PLS-DA a seguito della selezione delle variabili, è mostrata in Figura A - 37. In Tabella A - 8 sono mostrati i parametri relativi alla matrice di confusione per la valutazione delle performance di classificazione del modello PLS-DA in predizione, a seguito della selezione delle variabili.

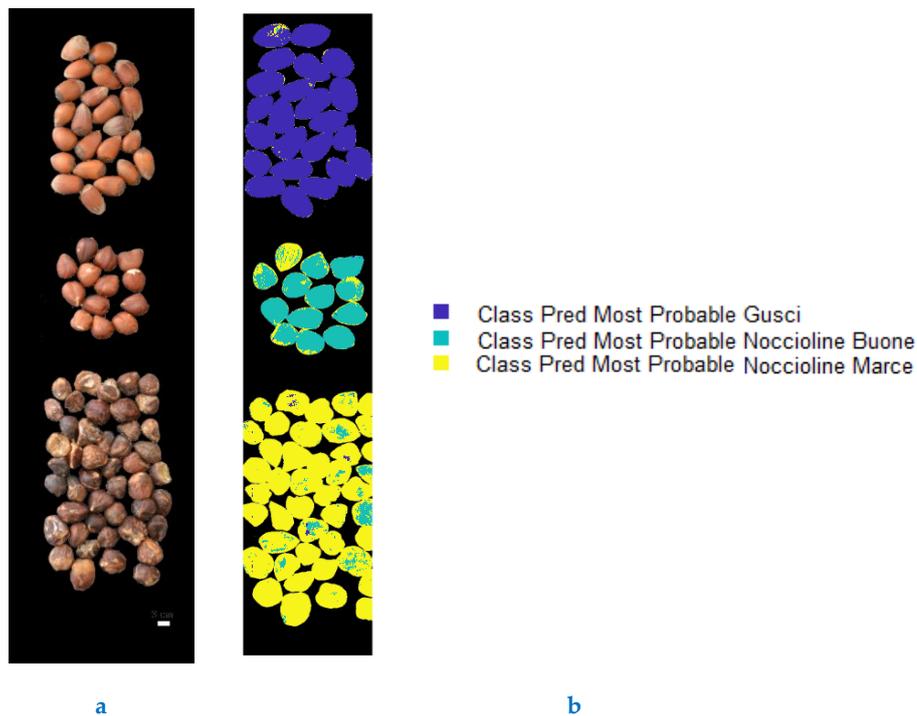


Figura A - 37. Immagine RGB (a) e risultato della classificazione PLS-DA; mappa - class predicted most probable (b).

Allegati

Tabella A - 8. Parametri relativi alla matrice di confusione per la valutazione delle performance di classificazione del modello PLS-DA in predizione, a seguito della selezione delle variabili.

| Classe | <i>Sensitivity</i> | <i>Specificity</i> | Numero di spettri | Mis-classification Error | Precisione |
|----------------|---------------------------|---------------------------|--------------------------|---------------------------------|-------------------|
| Gusci | 0.977 | 0.991 | 30892 | 0.014 | 0.983 |
| Nocciole Buone | 0.665 | 0.982 | 13897 | 0.067 | 0.871 |
| Nocciole Marce | 0.961 | 0.883 | 45021 | 0.078 | 0.892 |

ALLEGATO XII- CASE STUDY – PISTACCHI

Nel seguente allegato si riportano tutti i risultati ottenuti dalle analisi effettuate per la discriminazione di pistacchi buoni da relativi scarti ([4.2.4 Classificazione mediante hyperspectral imaging di pistacchi buoni e relativi scarti](#)), nei tre casi considerati:

- I. analisi effettuate utilizzando la combinazione di *pre-processing* SNV +MC;
- II. analisi effettuate utilizzando la combinazione di *pre-processing* DETREND +SNV +MC,
- III. analisi effettuate utilizzando la combinazione di *pre-processing* GLS-W +MC.

I - Caso SNV + MC

Spettri medi grezzi e pre-processati

Gli spettri medi grezzi e pre-processati (utilizzando la combinazione di algoritmi SNV + MC) del dataset Pistacchi di calibrazione sono mostrati in [Figura A - 38](#).

Analisi in componenti principali (PCA)

Gli *scores plot* della PCA effettuata sul dataset di calibrazione Pistacchi sono mostrati in [Figura A - 39](#), mentre il *loadings plot* è mostrato in [Figura A - 40](#).

Classificazione Partial Least Squares – Discriminant Analysis (PLS-DA)

La mappa di predizione, risultante dalla classificazione PLS-DA, è mostrata in [Figura A - 41b](#). Mentre, in [Tabella A - 9](#) sono mostrati i parametri relativi alla matrice di confusione per la valutazione delle performance di classificazione del modello PLS-DA in predizione.

Selezione delle variabili mediante Interval-PLSDA

In [Figura A - 42](#) è mostrato il risultato della selezione delle variabili mediante I-PLSDA. Le variabili selezionate ricadono negli intervalli: 1000-1049 nm, 1169-1220 nm, 1220-1277 nm e 1340-1390 nm.

Nuovo modello PLS-DA con variabili selezionate

In [Tabella A - 10](#) sono mostrati i parametri relativi alla matrice di confusione per la valutazione delle performance di classificazione del modello PLS-DA in predizione, a seguito della selezione delle variabili. La mappa di predizione, risultante dalla classificazione PLS-DA a seguito della selezione delle variabili, è mostrata in [Figura A - 43](#).

Allegati

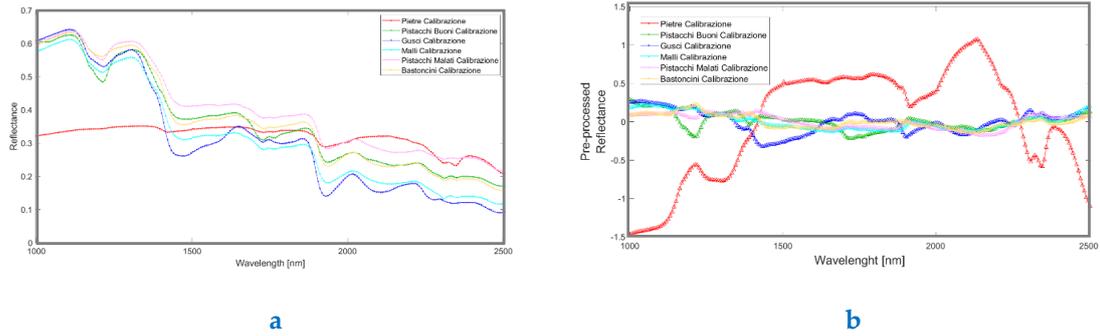


Figura A - 38. Spettri medi grezzi (a) e pre-processati (b) del dataset di calibrazione con algoritmo SNV + MC.

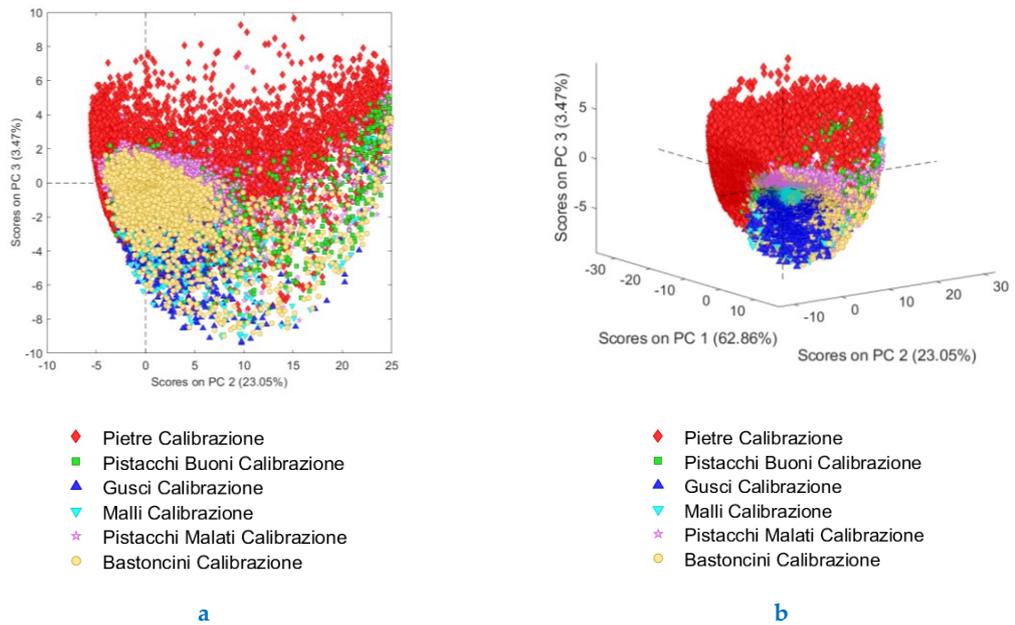


Figura A - 39. Scores plot: scores on PC 2 vs scores on PC 3 (a) e scores on PC 1 vs scores on PC 2 vs scores on PC 3 (b).

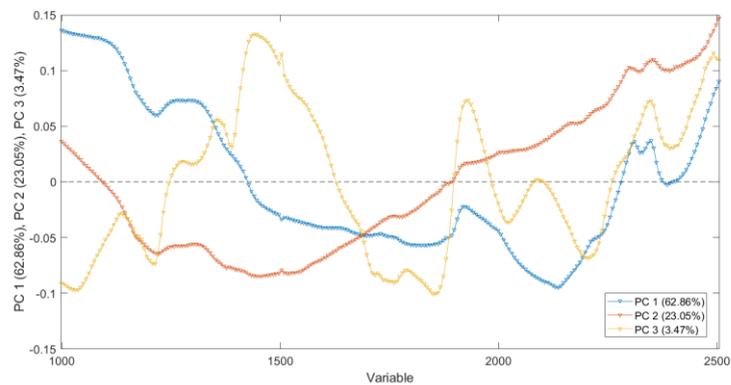


Figura A - 40. Loadings plot (PC 1, PC 2 e PC 3 in funzione della lunghezza d'onda).

Tabella A - 9. Parametri relativi alla matrice di confusione per la valutazione delle performance di classificazione del modello PLS-DA in predizione.

Allegati

| Classe | Sensitivity | Specificity | Numero di spettri | Mis-classification Error | Precisione |
|------------------|-------------|-------------|-------------------|--------------------------|------------|
| Pietre | 0.992 | 0.999 | 14565 | 0.002 | 0.996 |
| Pistacchi Buoni | 0.707 | 0.937 | 20527 | 0.114 | 0.761 |
| Gusci | 0.912 | 0.987 | 13460 | 0.024 | 0.921 |
| Malli | 0.776 | 0.964 | 20655 | 0.078 | 0.859 |
| Pistacchi Malati | 0.480 | 0.927 | 14635 | 0.143 | 0.552 |
| Bastoncini | 0.749 | 0.907 | 8960 | 0.108 | 0.462 |

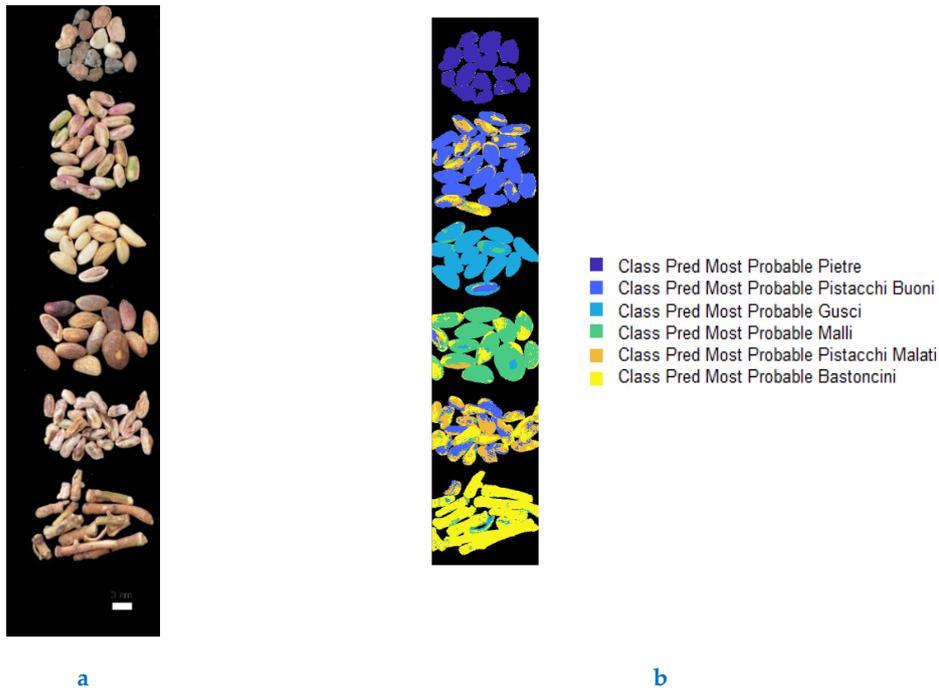


Figura A - 41. Immagine RGB (a) e risultato della classificazione PLS-DA (b).

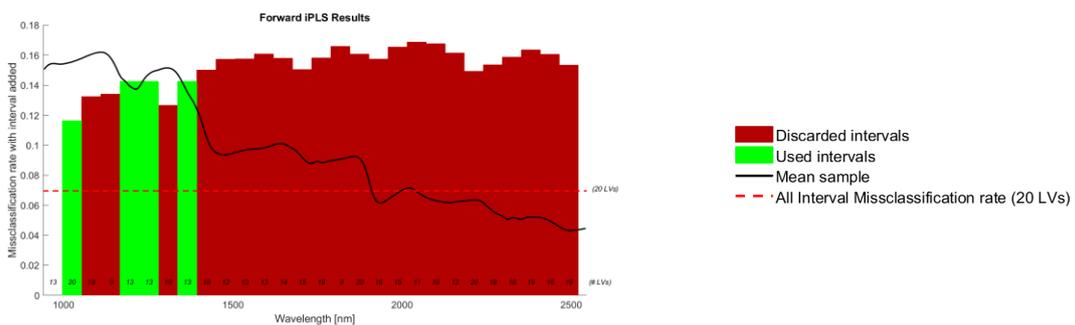


Figura A - 42. Risultato della selezione delle variabili mediante I-PLSDA.

Tabella A - 10. Parametri relativi alla matrice di confusione per la valutazione delle performance di classificazione del modello PLS-DA in predizione, a seguito della selezione delle variabili.

Allegati

| Classe | <i>Sensitivity</i> | <i>Specificity</i> | Numero di spettri | <i>Mis-classification Error</i> | Precisione |
|------------------|--------------------|--------------------|-------------------|---------------------------------|------------|
| Pietre | 0.995 | 0.989 | 14565 | 0.010 | 0.942 |
| Pistacchi Buoni | 0.878 | 0.915 | 20527 | 0.093 | 0.746 |
| Gusci | 0.904 | 0.976 | 13460 | 0.034 | 0.864 |
| Malli | 0.858 | 0.957 | 20655 | 0.065 | 0.852 |
| Pistacchi Malati | 0.374 | 0.964 | 14635 | 0.129 | 0.663 |
| Bastoncini | 0.687 | 0.953 | 8960 | 0.073 | 0.608 |

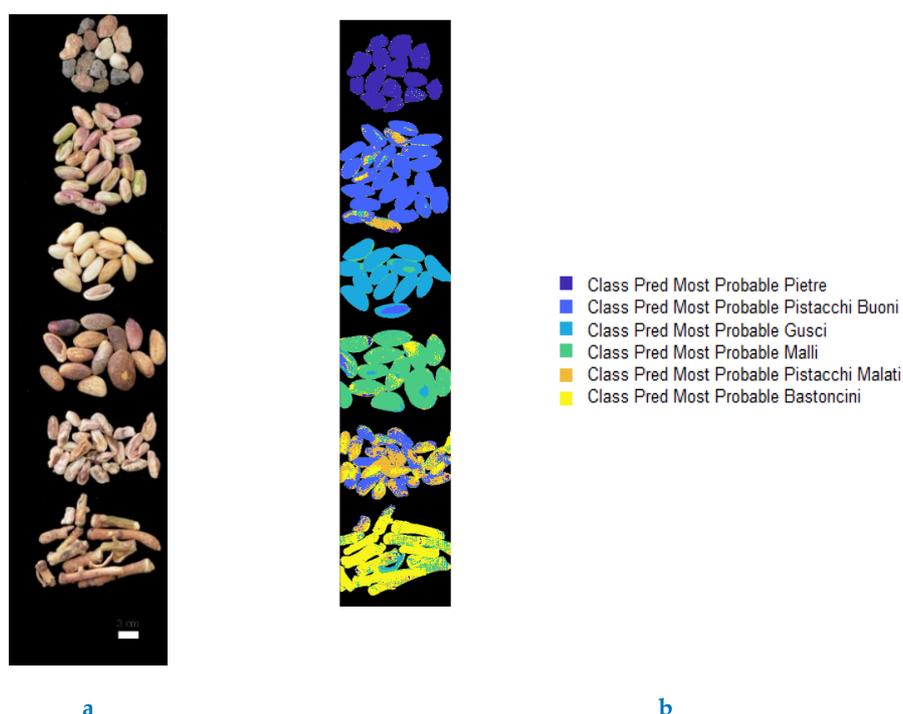


Figura A - 43. Immagine RGB (a) e risultato della classificazione PLS-DA: mappa - class predicted most probable (b).

II - Caso DETREND + SNV + MC

Spettri medi pre-processati

Gli spettri medi grezzi e pre-processati (utilizzando la combinazione di algoritmi DETREND + SNV + MC) del dataset Pistacchi di calibrazione sono mostrati in [Figura A - 44](#).

Analisi in componenti principali (PCA)

Gli *scores plot* della PCA effettuata sul dataset di calibrazione Pistacchi sono mostrati in [Figura A - 45](#), mentre il *loadings plot* è mostrato in [Figura A - 46](#).

Allegati

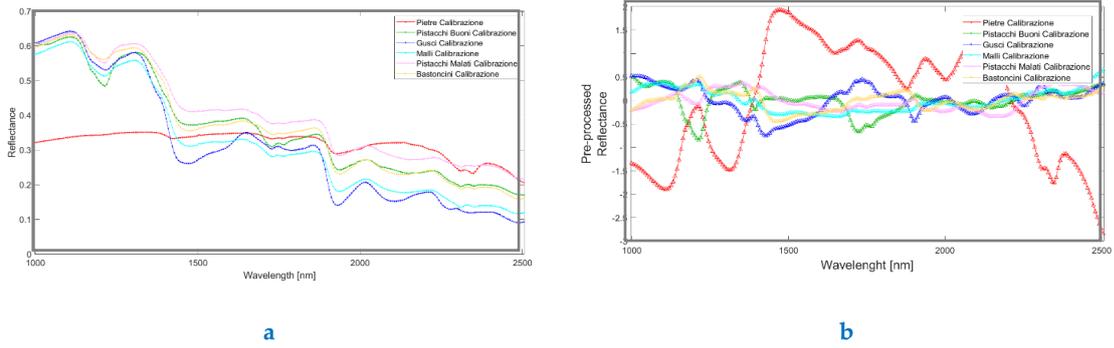


Figura A - 44. Spettri medi grezzi (a) e pre-processati (b) del dataset di calibrazione con algoritmo DETREND + SNV + MC.

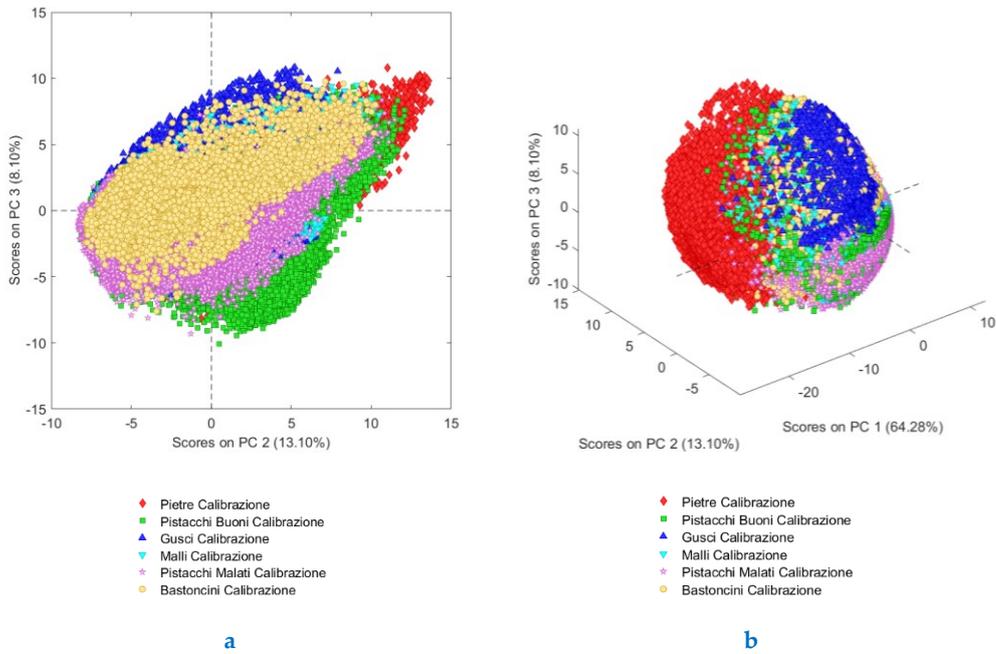


Figura A - 45. Scores plot: scores on PC 2 vs scores on PC 3 (a) e scores on PC 1 vs scores on PC 2 vs scores on PC 3 (b).

Allegati

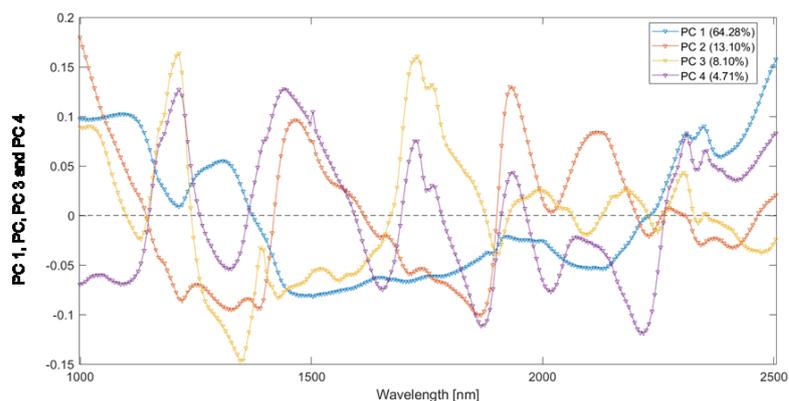


Figura A - 46. Loadings plot (PC 1, PC 2, PC 3 e PC 4 in funzione della lunghezza d'onda).

Classificazione Partial Least Squares – Discriminant Analysis (PLS-DA)

La mappa di predizione, risultante dalla classificazione PLS-DA, è mostrata in [Figura A - 47 b](#). Mentre, in [Tabella A - 11](#) sono mostrati i parametri relativi alla matrice di confusione per la valutazione delle performance di classificazione del modello PLS-DA in predizione.

Selezione delle variabili mediante Interval-PLSDA

In [Figura A - 48](#) è mostrato il risultato della selezione delle variabili mediante I-PLSDA. Le variabili selezionate ricadono negli intervalli: 1123-1163 nm, 1340-1390 nm, 1736-1786 nm, 2417-2461 nm.

Tabella A - 11. Parametri relativi alla matrice di confusione per la valutazione delle performance di classificazione del modello PLS-DA in predizione.

| Classe | Sensitivity | Specificity | Numero di spettri | Mis-classification Error | Precisione |
|------------------|-------------|-------------|-------------------|--------------------------|------------|
| Pietre | 1.000 | 0.993 | 14565 | 0.006 | 0.963 |
| Pistacchi Buoni | 0.779 | 0.936 | 20527 | 0.099 | 0.775 |
| Gusci | 0.900 | 0.971 | 13460 | 0.039 | 0.842 |
| Malli | 0.804 | 0.946 | 20655 | 0.086 | 0.810 |
| Pistacchi Malati | 0.566 | 0.939 | 14635 | 0.120 | 0.633 |
| Bastoncini | 0.592 | 0.955 | 8960 | 0.080 | 0.584 |

Allegati

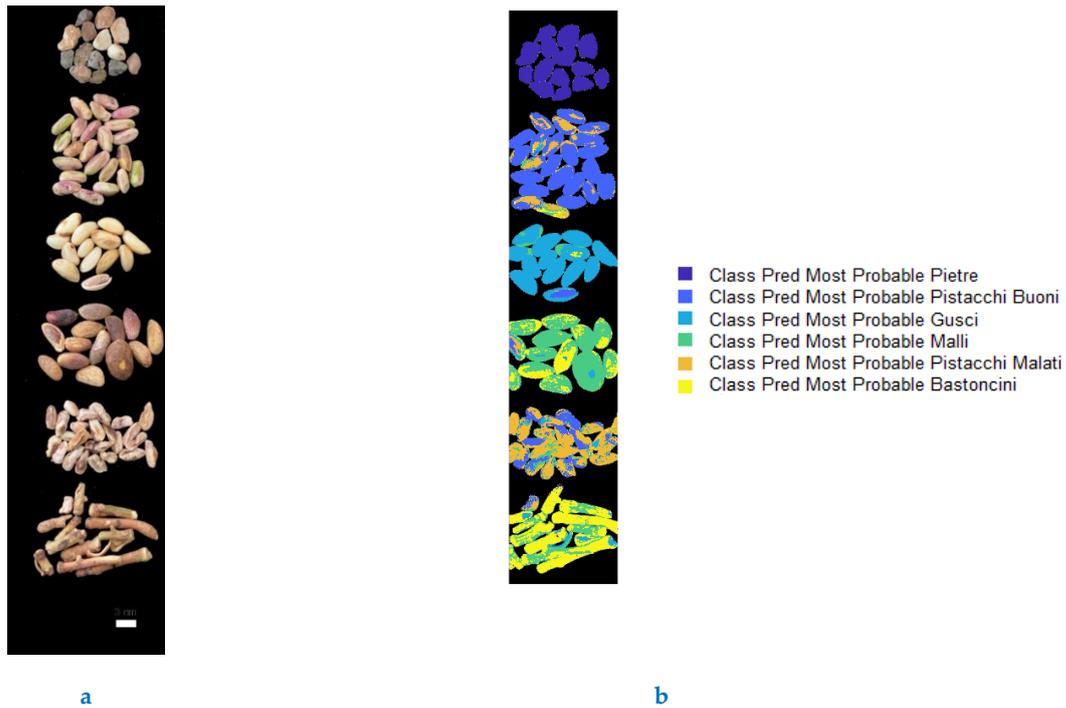


Figura A - 47. Immagine RGB (a) e risultato della classificazione PLS-DA: mappa - class predicted most probable (b).

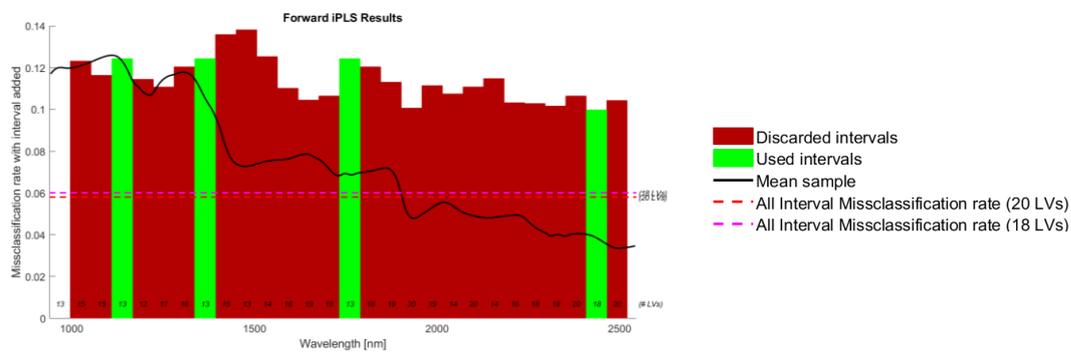


Figura A - 48. Risultato della selezione delle variabili mediante I-PLSDA.

Nuovo modello PLS-DA con variabili selezionate

In Tabella A - 12 sono mostrati i parametri relativi alla matrice di confusione per la valutazione delle performance di classificazione del modello PLS-DA in predizione, a seguito della selezione delle variabili. La mappa di predizione, risultante dalla classificazione PLS-DA a seguito della selezione delle variabili, è mostrata in Figura A - 49b.

Allegati

Tabella A - 12. Parametri relativi alla matrice di confusione per la valutazione delle performance di classificazione del modello PLS-DA in predizione, a seguito della selezione delle variabili.

| Classe | Sensitivity | Specificity | Numero di spettri | Mis-classification Error | Precisione |
|------------------|-------------|-------------|-------------------|--------------------------|------------|
| Pietre | 0.926 | 0.984 | 14565 | 0.025 | 0.914 |
| Pistacchi Buoni | 0.868 | 0.922 | 20527 | 0.090 | 0.760 |
| Gusci | 0.904 | 0.991 | 13460 | 0.022 | 0.942 |
| Malli | 0.868 | 0.922 | 20655 | 0.090 | 0.760 |
| Pistacchi Malati | 0.469 | 0.952 | 14635 | 0.124 | 0.647 |
| Bastoncini | 0.516 | 0.966 | 8960 | 0.078 | 0.616 |

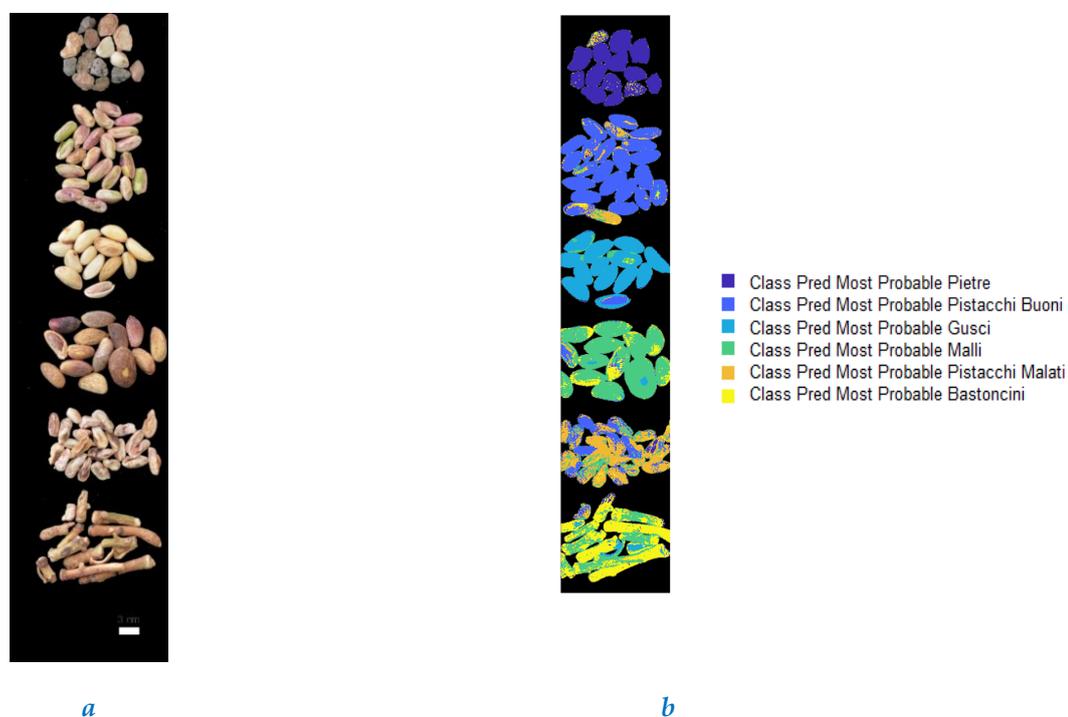


Figura A - 49. Immagine RGB (a) e risultato della classificazione PLS-DA: mappa - class predicted most probable (b).

III - Caso GLS-W + MC

Spettri medi pre-processati

Gli spettri medi grezzi e pre-processati (utilizzando la combinazione di algoritmi GLS-W + MC) del dataset Pistacchi di calibrazione sono mostrati in [Figura A - 50](#).

Analisi in componenti principali (PCA)

Allegati

Lo scores plot PC 1 vs PC 2 della PCA effettuata sul dataset di calibrazione Pistacchi è mostrato in [Figura A - 51](#), mentre il *loadings plot* è mostrato in [Figura A - 52](#).

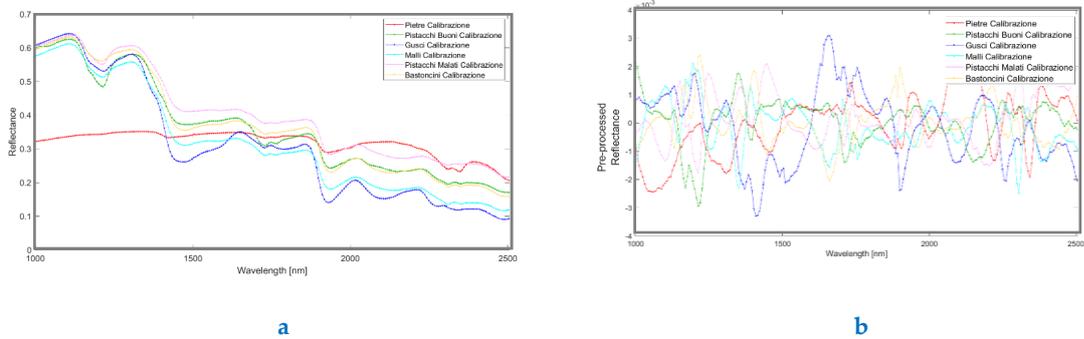


Figura A - 50. Spettri medi grezzi (a) e pre-processati (b) del dataset di calibrazione con algoritmo DETREND + SNV + MC.

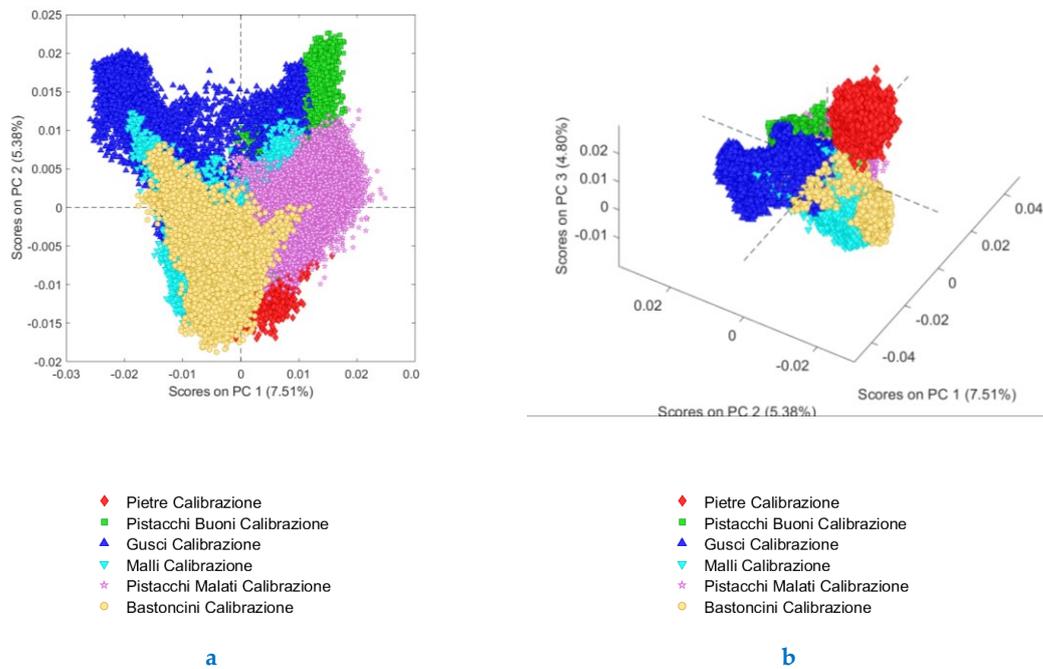


Figura A - 51. Scores plot: scores on PC 1 vs scores on PC 2(a) e scores on PC 1 vs scores on PC 2 vs Scores on PC 3 (b).

Allegati

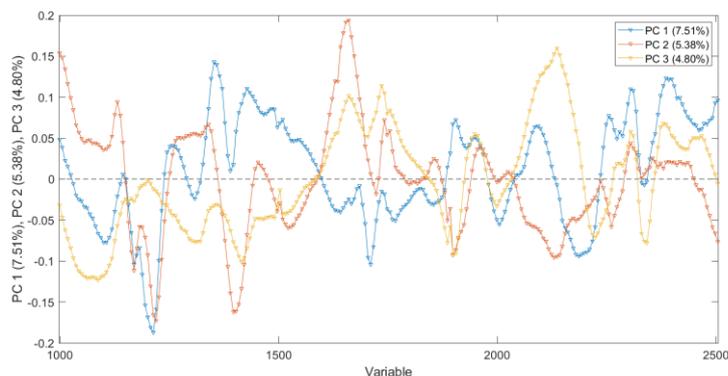


Figura A - 52. Loadings plot (PC 1, PC 2 e PC 3 in funzione della lunghezza d'onda).

Classificazione Partial Least Squares – Discriminant Analysis (PLS-DA)

La mappa di predizione, risultante dalla classificazione PLS-DA, è mostrata in [Figura A - 53b](#). Mentre, in [Tabella A - 13](#) sono mostrati i parametri relativi alla matrice di confusione per la valutazione delle performance di classificazione del modello PLS-DA in predizione.

Selezione delle variabili mediante Interval-PLSDA

In [Figura A - 54](#) è mostrato il risultato della selezione delle variabili mediante I-PLSDA. Le variabili selezionate dall'algoritmo I-PLSDA ricadono nei seguenti intervalli: 1123-1163 nm, 1226-1276 nm, 1340-1390 nm e 1736-1786 nm.

Tabella A - 13. Parametri relativi alla matrice di confusione per la valutazione delle performance di classificazione del modello PLS-DA in predizione.

| Classe | Sensitività | Specificity | Numero di spettri | Mis-classification Error | Precisione |
|------------------|-------------|-------------|-------------------|--------------------------|------------|
| Pietre | 1.000 | 0.964 | 14565 | 0.030 | 0.839 |
| Pistacchi Buoni | 0.871 | 0.983 | 20527 | 0.042 | 0.936 |
| Gusci | 0.908 | 0.992 | 13460 | 0.020 | 0.950 |
| Malli | 0.929 | 0.981 | 20655 | 0.030 | 0.934 |
| Pistacchi Malati | 0.853 | 0.962 | 14635 | 0.055 | 0.807 |
| Bastoncini | 0.768 | 0.993 | 8960 | 0.029 | 0.921 |

Allegati

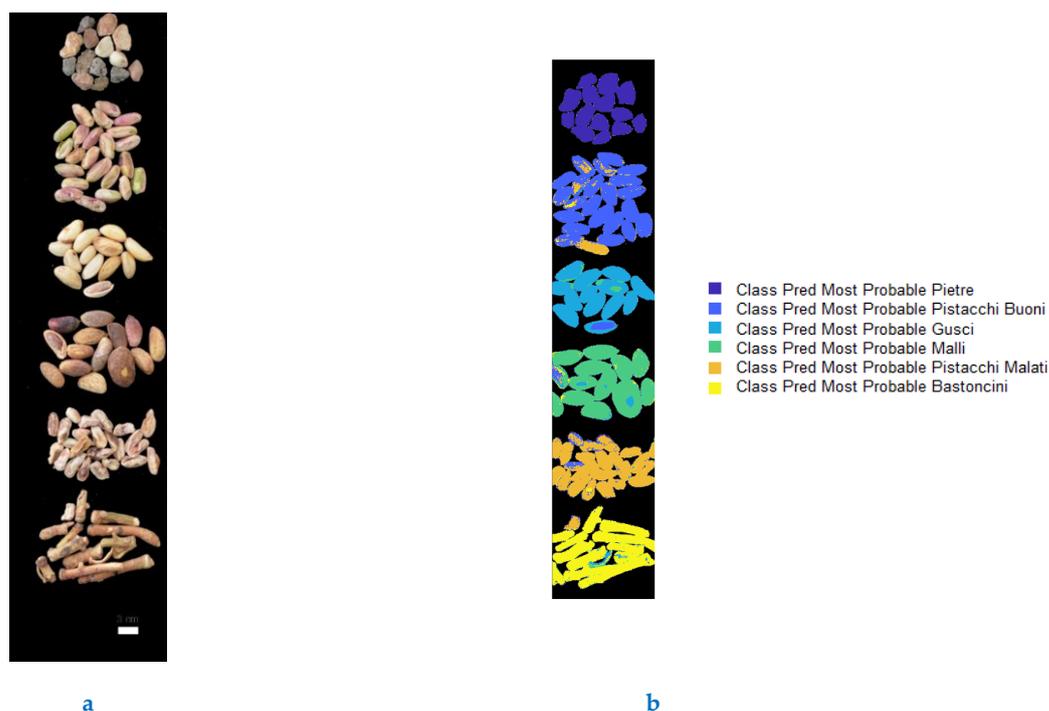


Figura A - 53. Immagine RGB (a) e risultato della classificazione PLS-DA: mappa - class predicted most probable (b).

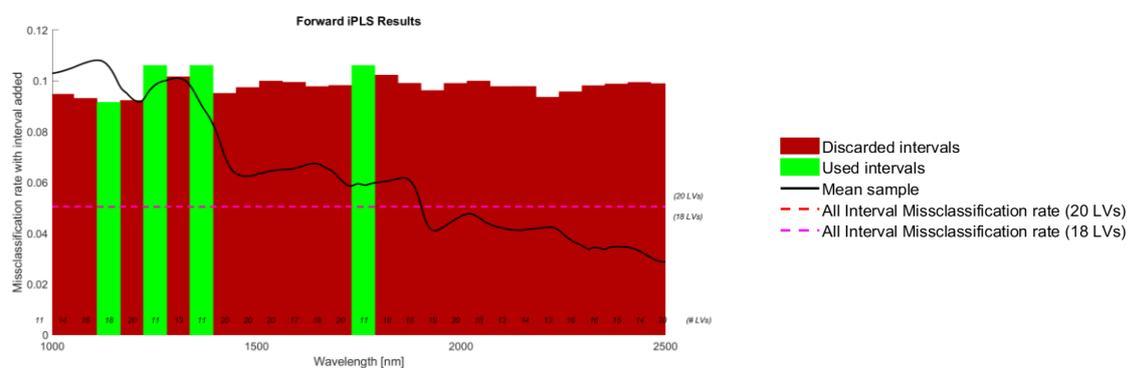


Figura A - 54. Risultato della selezione delle variabili mediante I-PLSDA.

Nuovo modello PLS-DA con variabili selezionate

In [Tabella A - 14](#) sono mostrati i parametri relativi alla matrice di confusione per la valutazione delle performance di classificazione del modello PLS-DA in predizione, a seguito della selezione delle variabili. La mappa di predizione, risultante dalla classificazione PLS-DA a seguito della selezione delle variabili, è mostrata in [Figura A - 55b](#).

Tabella A - 14. Parametri relativi alla matrice di confusione per la valutazione delle performance di classificazione del modello PLS-DA in predizione, a seguito della selezione delle variabili.

Allegati

| Classe | <i>Sensitivity</i> | <i>Specificity</i> | Numero di spettri | <i>Mis-classification Error</i> | Precisione |
|------------------|--------------------|--------------------|-------------------|---------------------------------|------------|
| Pietre | 0.974 | 0.971 | 14565 | 0.029 | 0.861 |
| Pistacchi Buoni | 0.767 | 0.964 | 20527 | 0.080 | 0.858 |
| Gusci | 0.885 | 0.987 | 13460 | 0.028 | 0.922 |
| Malli | 0.785 | 0.965 | 20655 | 0.075 | 0.865 |
| Pistacchi Malati | 0.728 | 0.934 | 14635 | 0.098 | 0.675 |
| Bastoncini | 0.658 | 0.945 | 8960 | 0.083 | 0.559 |

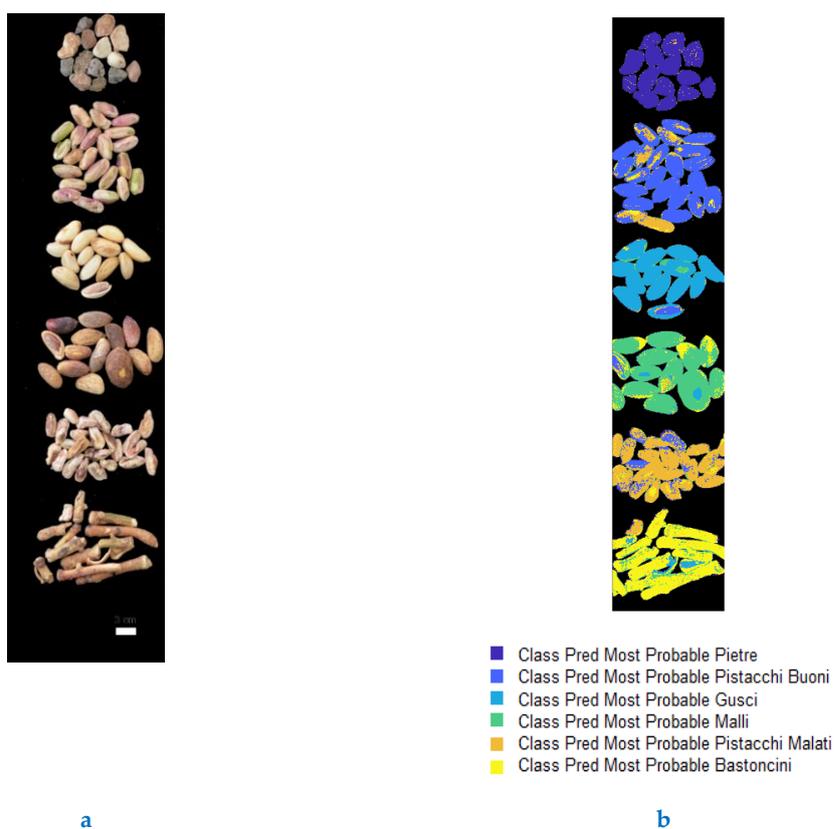


Figura A - 55. Immagine RGB (a) e risultato della classificazione PLS-DA: mappa - class predicted most probable (b).

ALLEGATO XIII- CASE STUDY - SEMENTI

Nel seguente allegato sono riportati i risultati ottenuti (e le restituzioni grafiche degli stessi) riferibili alle analisi effettuate sullo studio di germinabilità di sementi (4.2.5 *Studio sulla germinabilità di semi con l'uso di tecniche di hyperspectral imaging nell'infrarosso ad onde corte: Solanum Melongena, Cucurbita Pepo e Citrullus lanatus*). Più nel dettaglio alle analisi in componenti principali effettuate sui seguenti dataset di calibrazione:

- Dataset Melanzana - *Solanum Melongena*;
- Dataset Zucca - *Cucurbita Pepo*;
- Dataset Cocomero – *Citrullus Lanatus*.

Principal Component Analysis

In [Figura A - 56](#) sono riportati gli score plot risultanti dalle PCA effettuate per le tre combinazioni di *pre-processing* sul dataset di calibrazione. Mentre, in [Figura A - 57](#) sono riportati i *loadings plot* delle corrispettive analisi in componenti principali applicate sui dataset di calibrazione per Melanzana - *Solanum Melongena*, Zucca - *Cucurbita Pepo* e Cocomero – *Citrullus Lanatus*.

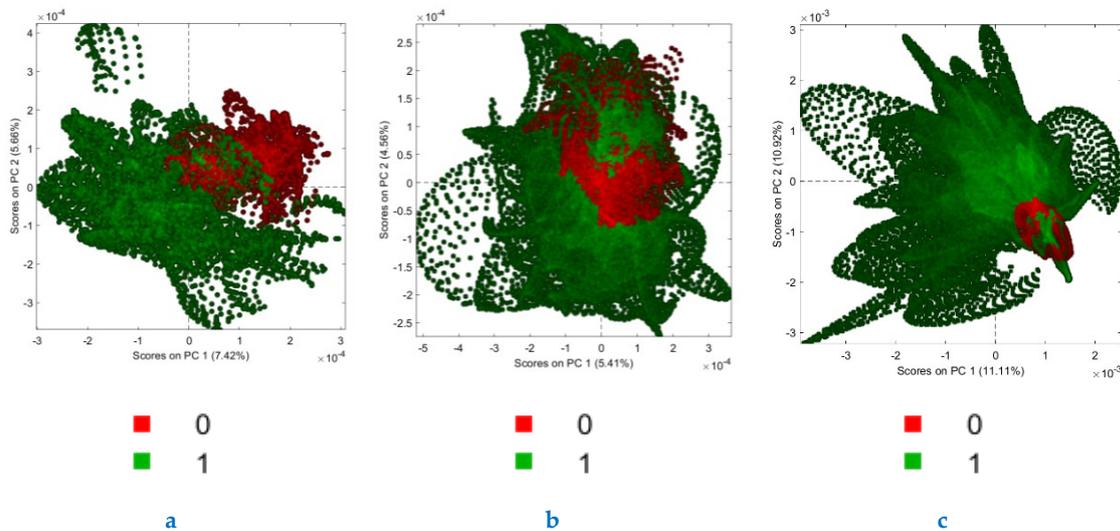
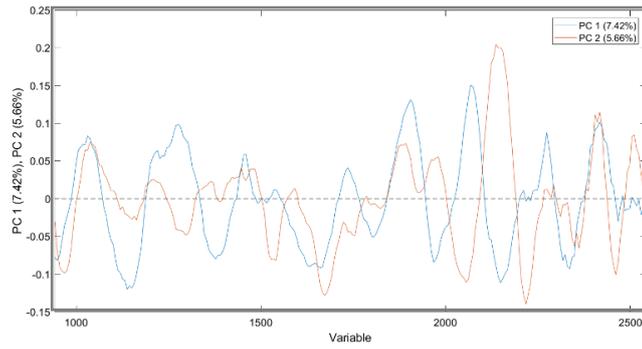
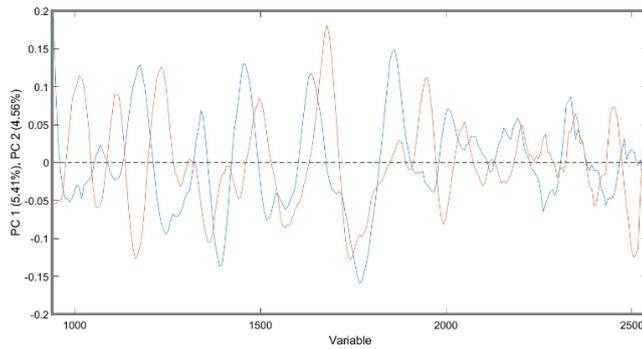


Figura A - 56. Scores su PC 1 vs scores su PC 2 (spettri in riflettanza SWIR: 1000-2500 nm) per: Melanzana - Solanum Melongena (a), Zucca - Cucurbita Pepo (b) e Cocomero – Citrullus Lanatus (c). Note della legenda: 0 = non germinato, 1= germinato.

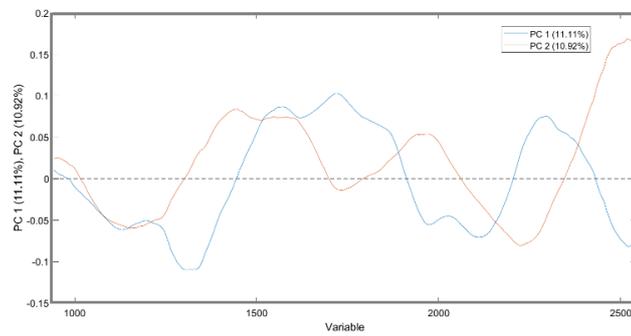
Allegati



a



b



c

Figura A - 57. Loadings plot: PC1 e PC 2 in funzione della lunghezza d'onda per Melanzana - Solanum Melongena (a), Zucca - Cucurbita Pepo (b) e Cocomero - Citrullus Lanatus (c).

ALLEGATO XIV- CASE STUDY – OLIVE: INVAIATURA

Nel seguente allegato sono riportati alcuni dei risultati ottenuti (e le restituzioni grafiche degli stessi) riferibili alle analisi di invaiatura dell'oliva "Itrana" mediante tecniche di HyperSpectral Imaging e spettrofotometria puntuale (4.2.6 *Valutazione del processo di maturazione dell'oliva itrana mediante l'utilizzo di un sistema iperspettrale da laboratorio ed uno spettroradiometro portatile*). In particolare, sono riportati: i risultati della regressione MLR effettuata per indice di maturazione e tempo di raccolta, gli spettri medi grezzi dei campioni analizzati e gli *scores plot* relativi alle analisi in componenti principali effettuate sugli spettri pre-processati dei campioni.

Correlazione tra indice di maturazione e tempo di raccolta

In **Tabella A - 15** sono riportati i punteggi relativi agli indici di maturazione, con riferimento all'analisi visiva effettuata sull'insieme di campioni per ciascun tempo di raccolta, a seconda delle località di provenienza. Dai risultati ottenuti dalla regressione MLR, effettuata per le variabili indice di maturazione (MI) e tempo di raccolta, si può evincere come l'MI sia strettamente correlato al tempo di raccolta: con $R^2 = 0.844$ in predizione e $RMSEP = 1.435$ MI. In **Figura A - 58** è mostrato l'indice di maturazione MI in funzione del tempo di raccolta, per ciascuna località di provenienza. La retta di regressione MLR (MI vs. *Time of Harvest*) ed i rispettivi parametri, risultanti dalla modellazione, sono mostrati in **Figura A - 59**.

Tabella A - 15. Indice di maturazione per ciascuna zona investigata.

| Tempi di raccolta (<i>Time of harvest</i>) | Maturity Index (MI) per Zona di provenienza | | |
|---|---|-----------|---------|
| | Gaeta | Sermoneta | Sonnino |
| T0 | 0 | 0 | 0 |
| T1 | 1 | 0 | 2 |
| T2 | 2 | 2 | 2 |
| T3 | 3 | 2 | 2 |
| T4 | 2 | 5 | 3 |
| T5 | 3 | 4 | 3 |
| T6 | 5 | 6 | 3 |
| T7 | 5 | 6 | 4 |

Principal Component Analysis

I grafici degli *scores* sulle prime due componenti principali, risultanti dalle analisi in componenti principali effettuate sulle ROI dell'ipercubo ricavato dalle immagini iperspettrali del sistema HSI (1000-2500 nm), in base delle classi riferite ai tempi di raccolta ed ai punteggi MI, sono mostrati in **Figura A - 60**.

Allegati

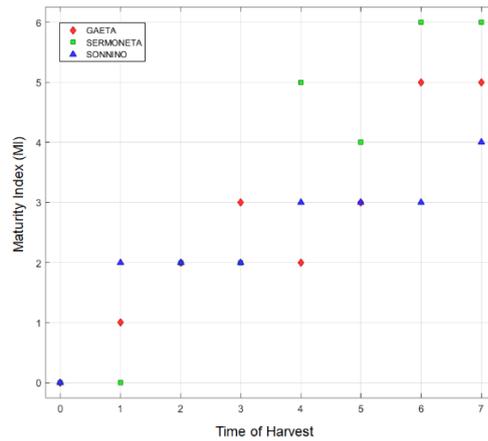


Figura A - 58. Indice di maturazione in funzione del tempo di raccolta dei campioni di oliva Itrana per ciascuna località di provenienza (Gaeta, Sermoneta e Sonnino).

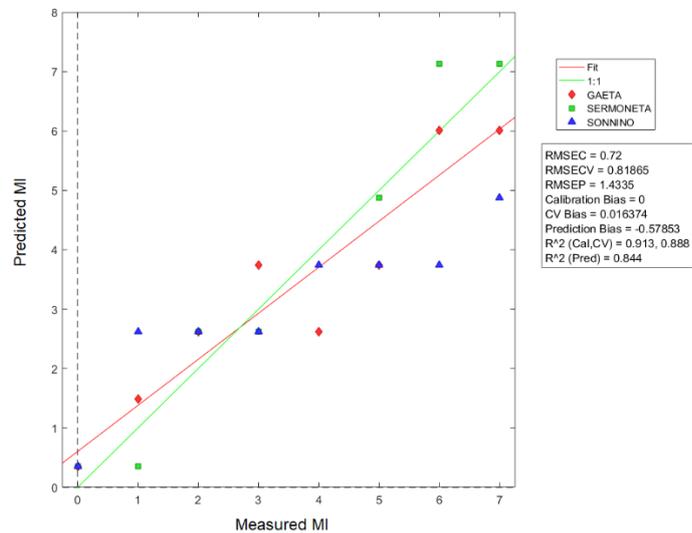
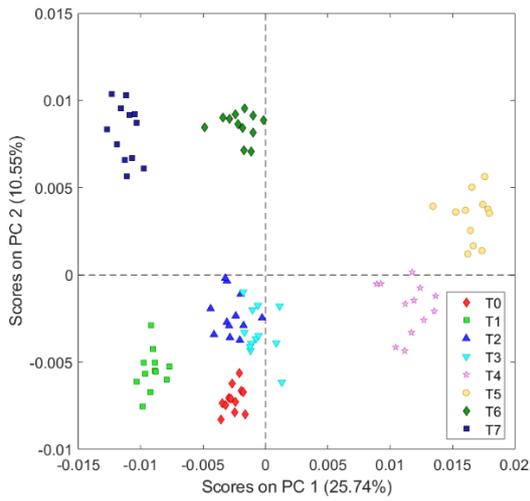


Figura A - 59. Punteggi dell'indice di maturazione predetti a partire dai tempi di raccolta in funzione dei punteggi MI valutati con metodo colorimetrico tradizionale.

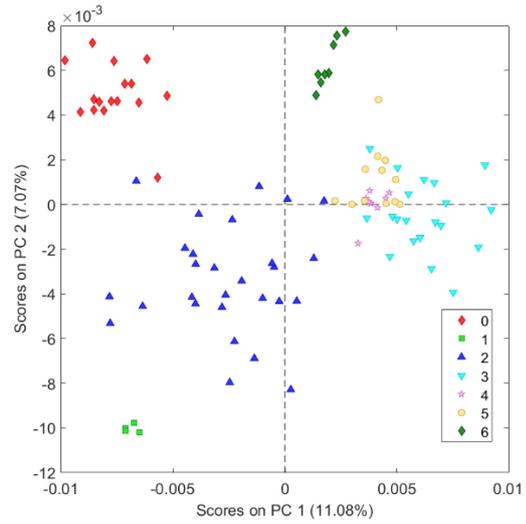
Come si può notare, i punti nelle nuvole che si vengono a formare sul grafico degli *scores*, relativamente alla PCA effettuata con le classi dei tempi di raccolta (Figura A - 60a), sono maggiormente raggruppate, rispetto a quelli dei punteggi MI (Figura A - 60b). I grafici degli *scores* su PC 1 VS *scores* su PC 2 relativi alle analisi in componenti principali effettuate sugli spettri Vis-SWIR (400-2500 nm) sono mostrati in Figura A - 61.

Differentemente da quanto avviene nel caso della PCA effettuata con i dati del SisucHEMA XL relativamente alle classi di punteggio MI, gli *scores* della PCA effettuata con i dati del FieldSpec (400-2500 nm) relativi alle classi di punteggio MI, appaiono maggiormente raggruppati (Figura A - 61b). Ciò può essere riferibile all'influenza del colore nella zona del visibile (400 - 800 nm).

Allegati

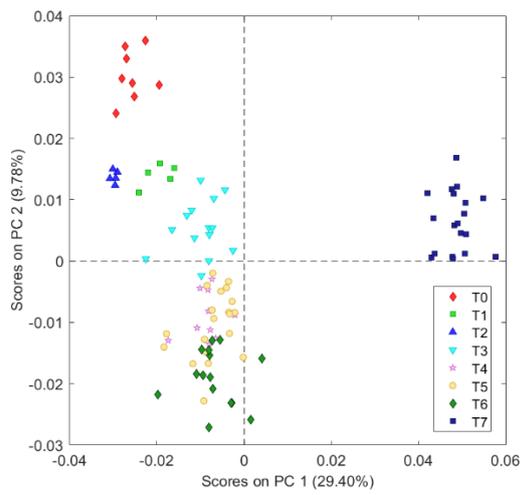


a

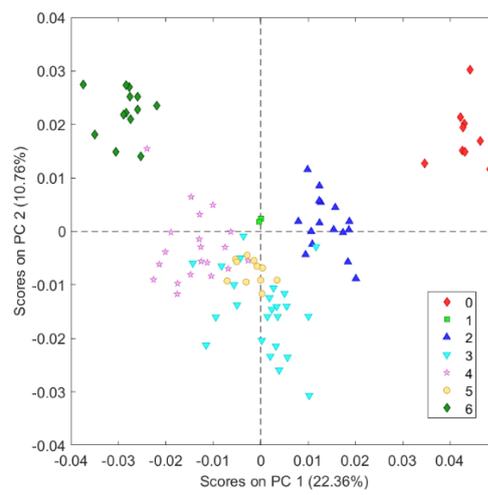


b

Figura A - 60. Grafici degli Scores sulle prime due componenti principali (PC1 vs. PC 2), relative alle PCA effettuate sugli spettri del sistema SisuCHEMA XL™ (1000 – 2500 nm), per le classi tempi di raccolta (a) e i punteggi MI (b).



a



b

Figura A - 61. Grafici degli Scores sulle prime due componenti principali (PC1 vs. PC 2), relative alle PCA effettuate sugli spettri del FieldSpec 4 (400-2500 nm), per le classi tempi di raccolta (a) e i punteggi MI (b).

ALLEGATO XV- CASE STUDY – OLIVE: MOSCA OLEARIA

Nel seguente allegato sono riportati alcuni dei risultati ottenuti (e le restituzioni grafiche degli stessi) riferibili alle analisi sulla presenza della mosca olearia in drupe della cultivar “Itrana” mediante spettrofotometria puntuale nel range 400-2500 nm (4.2.7 *Identificazione di olive attaccate da mosca olearia mediante l'utilizzo di uno spettroradiometro portatile*). In particolare, sono riportati: gli spettri medi grezzi dei campioni analizzati, gli *scores plot* relativi alle analisi in componenti principali effettuate sugli spettri pre-processati dei campioni ed i risultati dei modelli di classificazione PLS-DA.

Spettri medi in riflettanza

Comparando gli spettri medi in riflettanza, come riportato in [Figura A - 62](#), si nota chiaramente una certa variabilità tra le firme spettrali delle drupe attaccate da mosca olearie e quelle sane, non solo nell'intervallo del visibile.

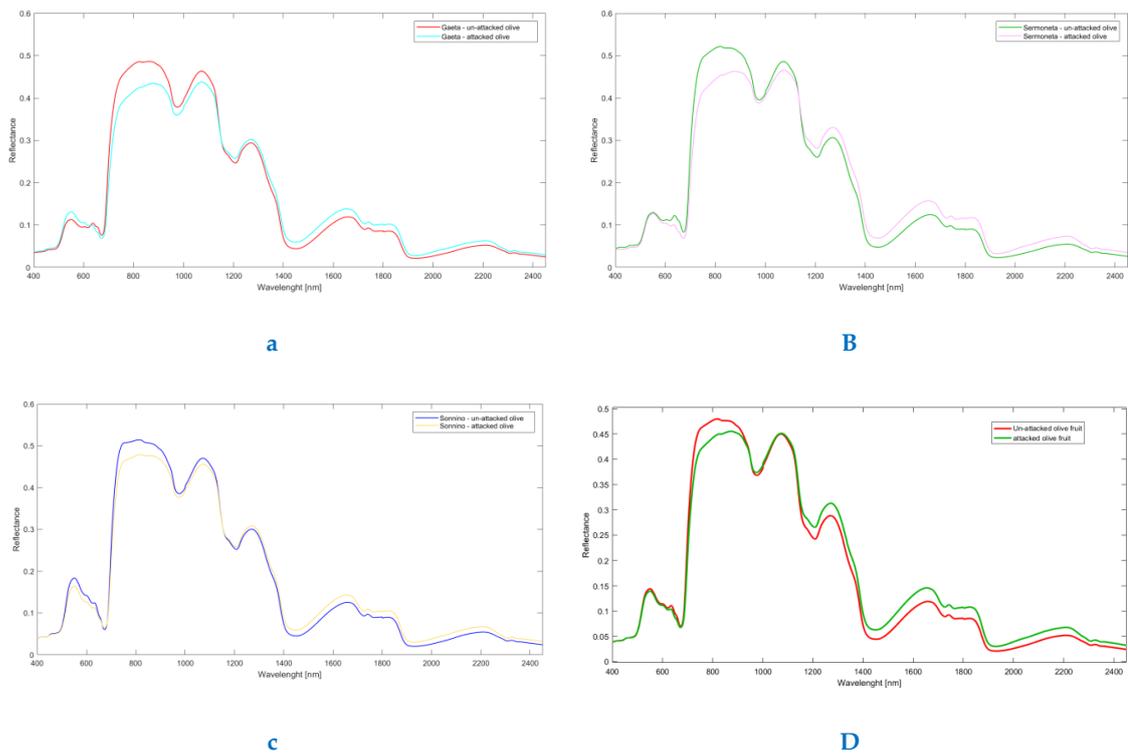


Figura A - 62. Spettri medi grezzi in riflettanza, acquisiti con spettroradiometro FieldSpec 4 (Vis-SWIR: 400 – 2500 nm), delle drupe attaccate da mosca (Attacked olive fruits) e drupe sane (un-attacked olive fruits) per: a) Gaeta, b) Sermoneta, c) Sonnino e d) tutt'e tre le località insieme.

Analisi esplorativa

La variabilità notata tra drupe attaccate da mosca olearie e quelle sane comparando gli spettri medi in riflettanza ([Figura A - 62](#)), appare maggiormente evidente nelle analisi PCA. Come si può notare negli *scores plot*, mostrati in [Figura A - 63](#), tale variabilità è ben

Allegati

spiegata con la prima componente principale (PC 1): appaiono evidenti due nuvole di scores distinte, relative alle due classi di olive.

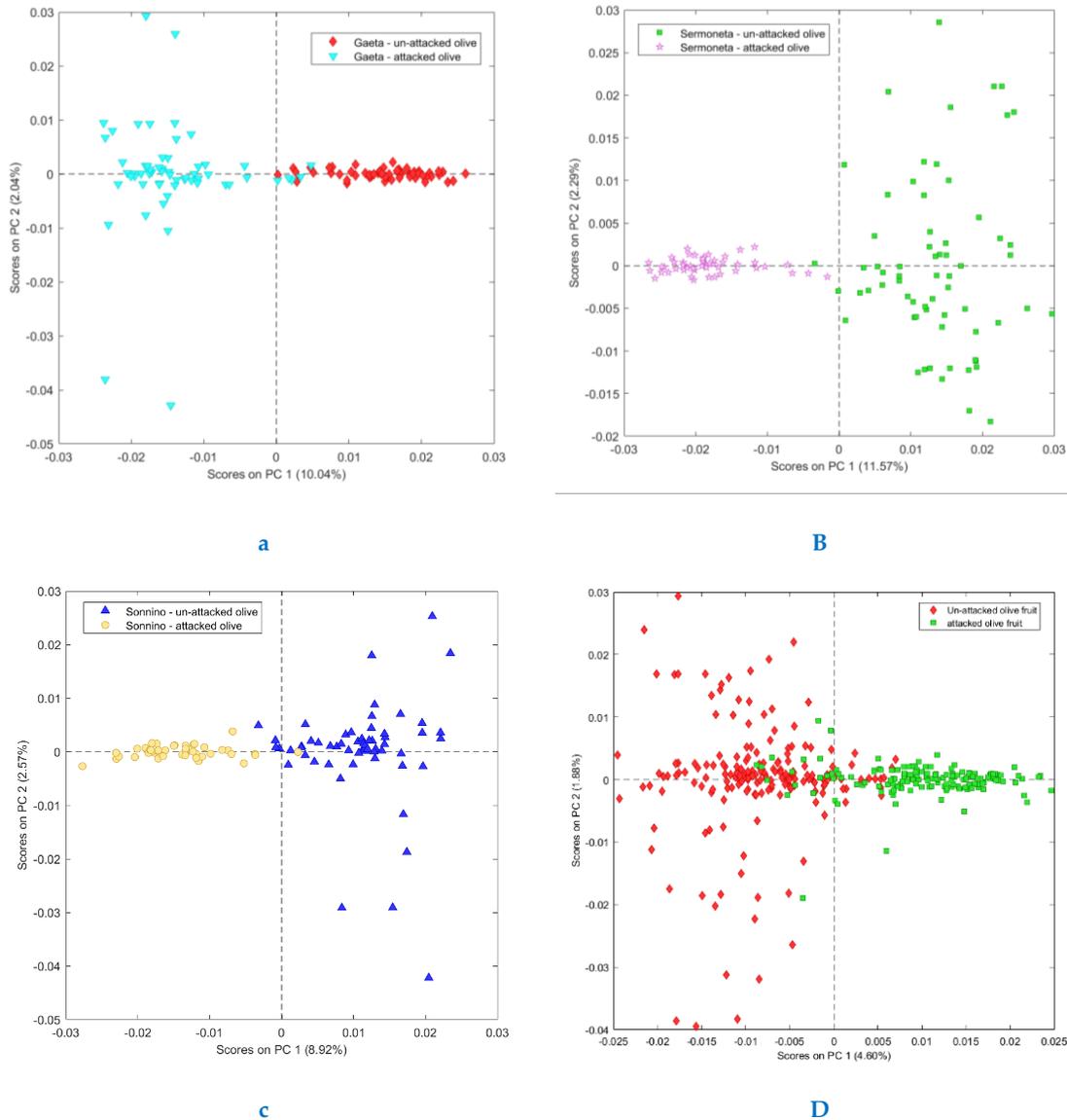


Figura A - 63. Grafici degli Scores sulle prime due componenti principali (PC1 vs. PC 2), relative alle PCA effettuate sugli spettri del FieldSpec 4® (400-2500 nm), delle drupe attaccate da mosca (Attacked olive fruits) e drupe sane (un-attacked olive fruits) per: a) Gaeta, b) Sermoneta, c) Sonnino e d) tutt'e tre le località insieme.

Classificazione Partial Least Squares – Discriminant Analysis

I modelli di classificazione, in tutti e quattro i casi considerati, riescono a discriminare con buona precisione le due classi di olive. Questo può essere notato sia dagli *scores plot* delle prime due variabili latenti, mostrati in Figura A - 64, che dai grafici relativi alla posizione del *boundary* dei classificatori (riportati in Figura A - 65).

Allegati

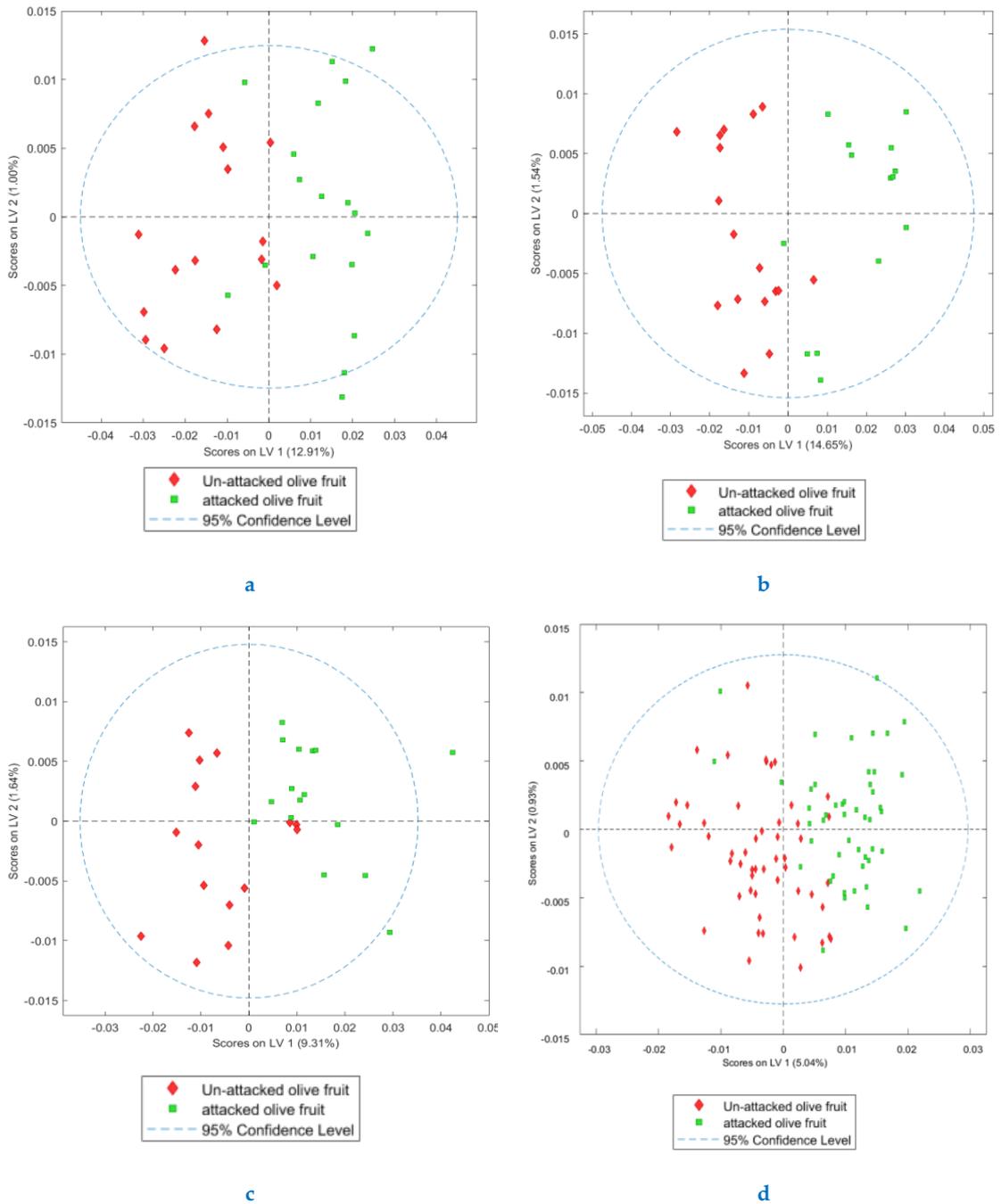


Figura A - 64. Scores plot delle prime due variabili latenti, relativi alle classificazioni PLS-DA per le due classi (attacked olive fruits e un-attacked olive fruits), relativamente ai set di dati: **a)** Gaeta, **b)** Sermoneta, **c)** Sonnino e **d)** tutt'e tre le località insieme.

Allegati

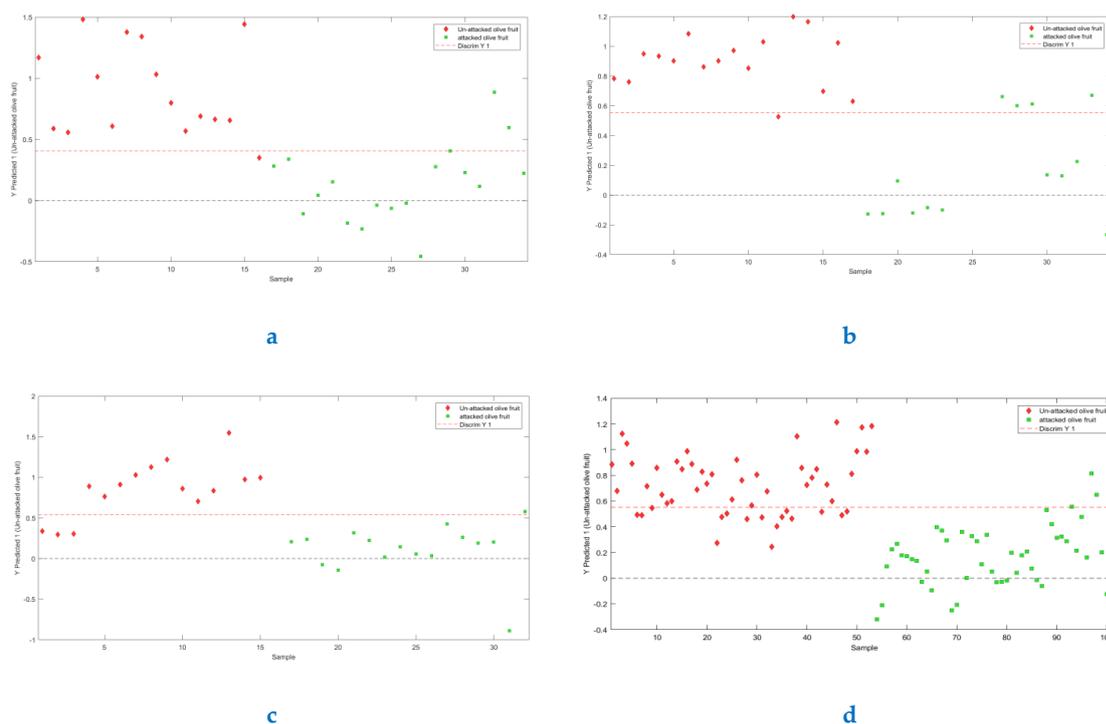


Figura A - 65. Posizioni dei boundary di discriminazione determinati dalle classificazioni PLS-DA per le due classi (attacked olive fruits e un-attacked olive fruits), relativamente ai set di dati: a) Gaeta, b) Sermoneta, c) Sonnino e d) tutt'e tre le località insieme.

Le metriche di performance dei classificatori, in termini di sensibilità e specificità in predizione sono riportati in [Tabella A - 16](#).

Tabella A - 16. Performance di classificazione in predizione, relativamente ai set di dati: Gaeta, Sermoneta, Sonnino e tutt'e tre le località insieme.

| Dataset | Classe | Sensitivity (Pred) | Specificity (Pred) |
|---------------|-------------------------|--------------------|--------------------|
| GAETA | Un-attacked olive fruit | 0.938 | 0.833 |
| | Attacked olive fruit | 0.833 | 0.938 |
| SERMONETA | Un-attacked olive fruit | 0.941 | 0.714 |
| | Attacked olive fruit | 0.714 | 0.941 |
| SONNINO | Un-attacked olive fruit | 0.800 | 0.938 |
| | Attacked olive fruit | 0.938 | 0.800 |
| ALL THE ZONES | Un-attacked olive fruit | 0.698 | 0.939 |
| | Attacked olive fruit | 0.939 | 0.698 |

ALLEGATO XVI- CASE STUDY – KIWI: MATURAZIONE

Nel seguente allegato sono riportati parte dei risultati ottenuti (e le restituzioni grafiche degli stessi) riferibili alle analisi effettuate sulle analisi di maturazione del kiwi giallo (4.2.8 *Valutazione del processo di maturazione del kiwi giallo mediante l'utilizzo di spettrofotometri NIR portatili*). In particolare, sono riportati: gli spettri medi grezzi dei campioni analizzati e gli *scores plot* relativi alle analisi in componenti principali effettuate sugli spettri pre-processati dei campioni.

Spettri medi in riflettanza

Gli spettri medi acquisiti con JDSU MicroNIR™ e ASD FieldSpec 4® Standard Res mostrano differenze in accordo con le variabili analizzate, quali: tempo di analisi (T0 a T5, ad intervallo di 7 giorni), valori di DM e valori di TSSC. In [Figura A - 66](#) è riportato il confronto tra la variazione degli spettri medi con il tempo di analisi per i due spettrofotometri utilizzati. Mentre, in [Figura A - 67](#) e [Figura A - 68](#) sono riportati rispettivamente gli spettri medi dei due strumenti identificati secondo il TSSC ed il DM.

Principal Component Analysis

In [Figura A - 69](#) è riportato il confronto tra gli *scores plot* relativamente alla variazione del tempo di analisi per i due spettrofotometri utilizzati. Mentre, in [Figura A - 67](#) e [Figura A - 68](#) sono riportati rispettivamente gli *scores plot* con la variazione del TSSC ed il DM per i due spettrofotometri utilizzati.

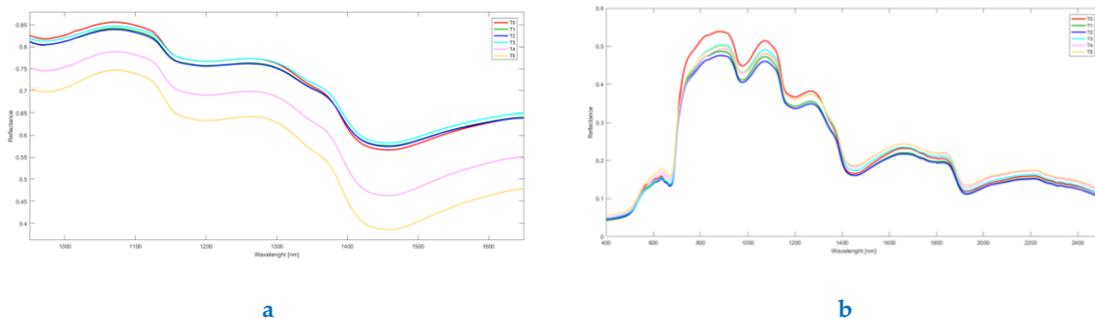


Figura A - 66. Variazione degli spettri medi con il tempo di analisi per: a) MicroNIR (950-1650 nm) e b) FieldSpec 4 (400-2500 nm).

Allegati

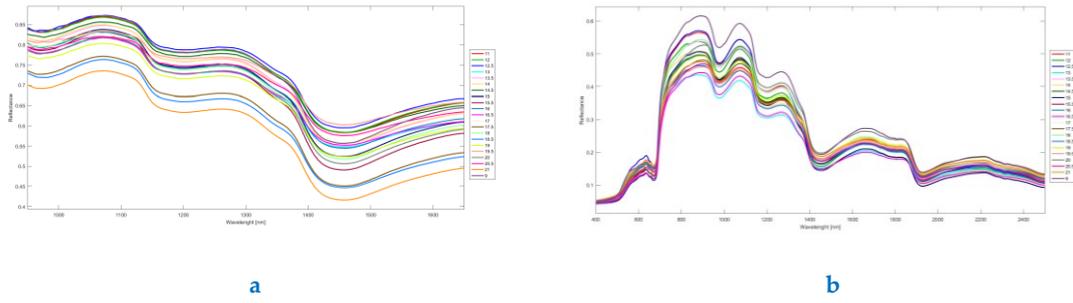


Figura A - 67. Variazione degli spettri medi con il TSS per: a) MicroNIR (950-1650 nm) e b) FieldSpec 4 (400-2500 nm).

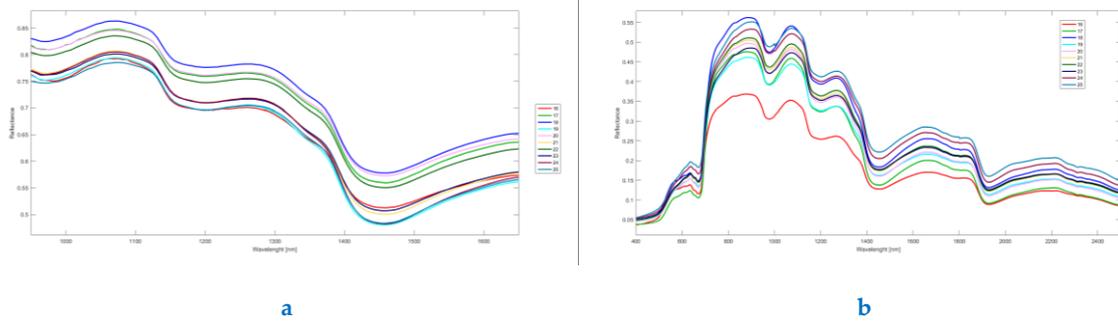


Figura A - 68. Variazione degli spettri medi con il DM per: a) MicroNIR (950-1650 nm) e b) FieldSpec 4 (400-2500 nm).

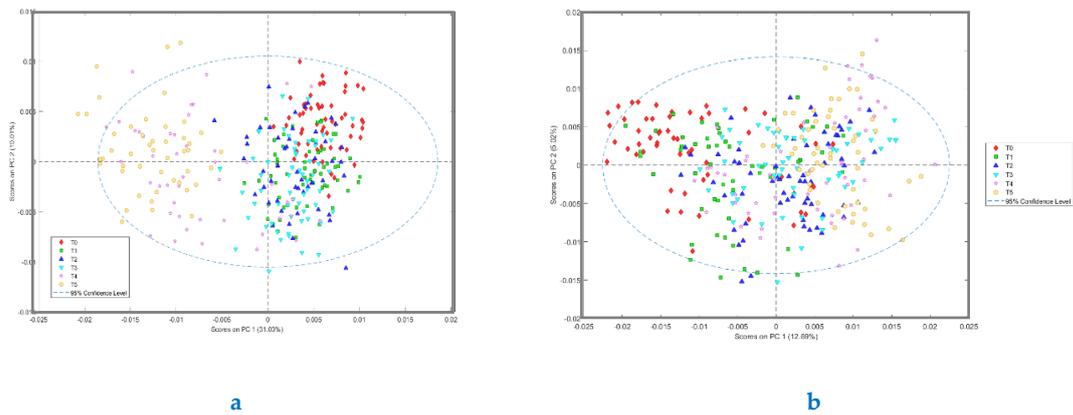


Figura A - 69. Scores plot (scores su PC2 vs scores su PC1) del tempo di analisi per: a) MicroNIR (950-1650 nm) e b) FieldSpec 4 (400-2500 nm).

Allegati

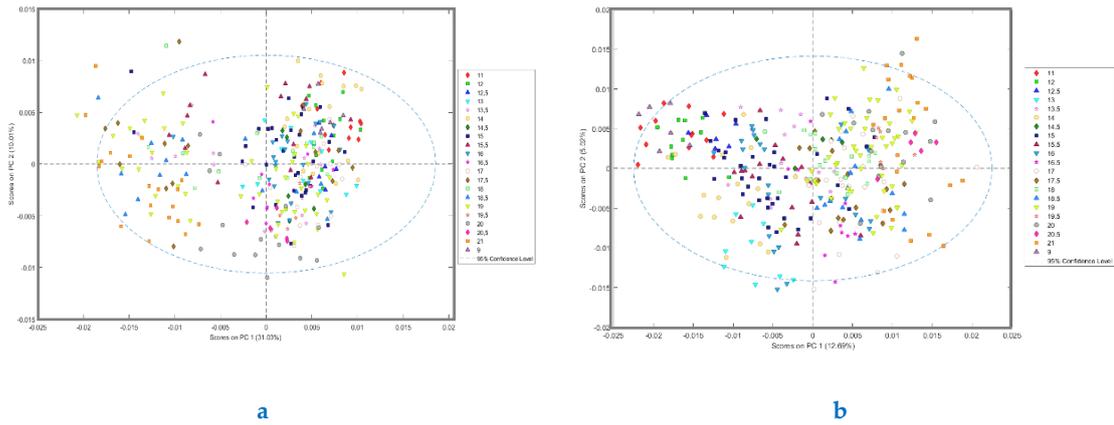


Figura A - 70. Scores plot (scores su PC2 vs scores su PC1) del TSSC per: **a)** MicroNIR (950-1650 nm) e **b)** FieldSpec 4 (400-2500 nm).

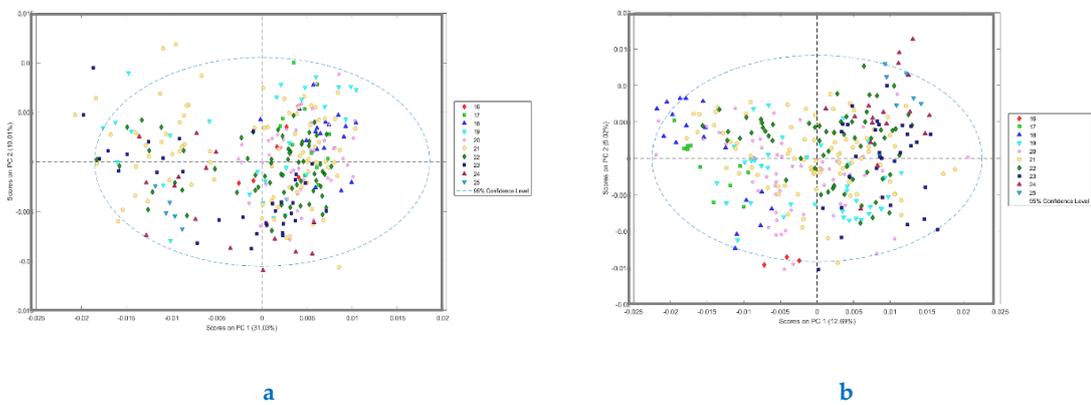


Figura A - 71. Scores plot (scores su PC2 vs scores su PC1) del DM per: **a)** MicroNIR (950-1650 nm) e **b)** FieldSpec 4 (400-2500 nm).

ALLEGATO XVII- CASE STUDY – GRANO E FARINE

Nel seguente allegato sono riportati i risultati ottenuti (e le restituzioni grafiche degli stessi) riferibili alle analisi effettuate sulla valutazione di alcuni parametri qualitativi di frumento tenero in grani (Tabella A - 17) e farina di grano tenero (Tabella A - 18), studiati (4.2.9 Valutazione di parametri qualitativi del grano tenero mediante spettrofotometria NIR: prima e dopo la macinazione). In particolare, sono riportati: gli spettri medi grezzi e pre-elaborati dei campioni analizzati e i risultati delle regressioni PLS effettuate con gli spettri acquisiti su frumento tenero in grani per i seguenti parametri: Umidità (%), Peso specifico (kg/hl), Proteine (%), Glutine Umido (%), W (%), Durezza (%). Sono Inoltre, riportati i risultati delle regressioni PLS effettuate con gli spettri acquisiti sulla farina di grano tenero per i seguenti parametri: Umidità (%), Proteine (%), Ceneri (%), Glutine Umido (%), W (%), Assorbimento (%) ed Amido (%).

Tabella A - 17. Parametri qualitativi dei campioni di frumento tenero in grani.

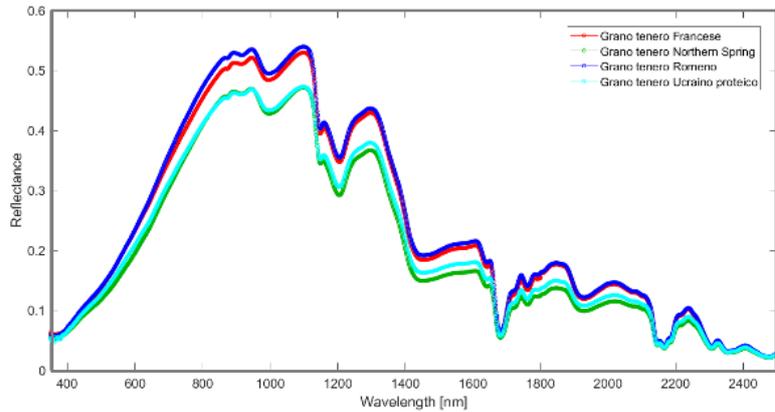
| | Grano tenero Francese | Grano tenero Romano | Grano tenero Northern Spring | Grano tenero Ucraino proteico |
|---------------------------|----------------------------------|--------------------------------|---|--|
| Umidità (%) | 12.4 | 12.1 | 13.1 | 12.7 |
| Peso specifico (kg/hl) | 77.9 | 78.8 | 83.4 | 77.4 |
| Proteine (%) | 12.4 | 12.4 | 15 | 16 |
| Glutine Umido (%) | 24.6 | 24.5 | 30.6 | 32.8 |
| W (%) | 226 | 157 | 321 | 361 |
| Durezza (%) | 65 | 32 | 62 | 82 |

Tabella A - 18. Parametri qualitativi dei campioni di farina di grano tenero.

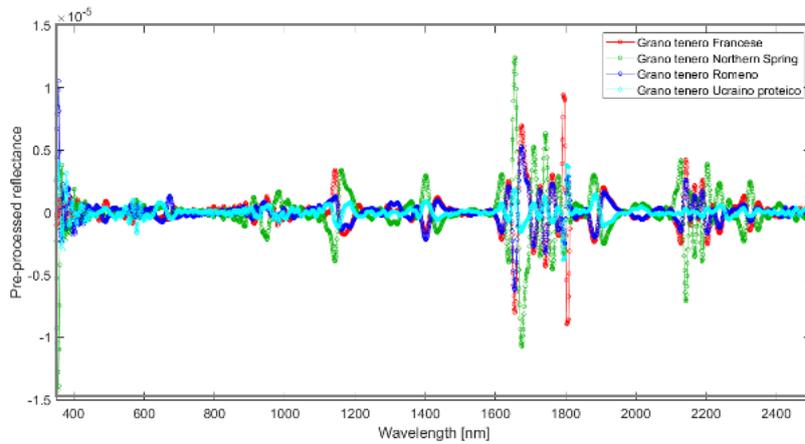
| | Farina di grano Francese | Farina di grano Romano | Farina di grano Northern Spring | Farina di grano Ucraino proteico |
|-------------------|-------------------------------------|-----------------------------------|--|---|
| Umidità (%) | 12.1 | 11.9 | 12.5 | 12.6 |
| Proteine (%) | 11.3 | 11.3 | 14.5 | 14.5 |
| Ceneri (%) | 0.6 | 0.6 | 0.7 | 0.6 |
| Glutine Umido (%) | 25.3 | 25.4 | 33.7 | 33.8 |
| W (%) | 253 | 254 | 370 | 359 |
| Assorbimento (%) | 57.1 | 56.9 | 61.6 | 60.2 |
| Amido (%) | 24.4 | 23 | 26.7 | 24.9 |

Spettri medi grezzi e pre-elaborati

In [Figura A - 72](#) sono riportati gli spettri medi grezzi in riflettanza e pre-elaborati dei campioni di frumento tenero in grani. Mentre, in [Figura A - 73](#) sono riportati gli spettri medi grezzi in riflettanza e pre-elaborati dei campioni di farina di grano tenero.



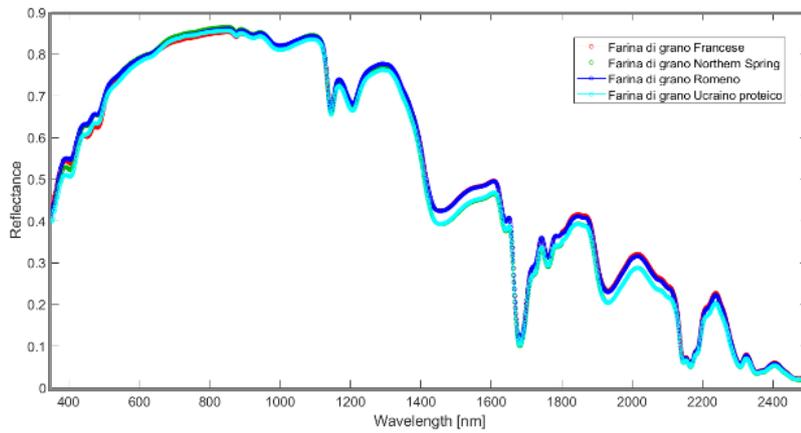
a



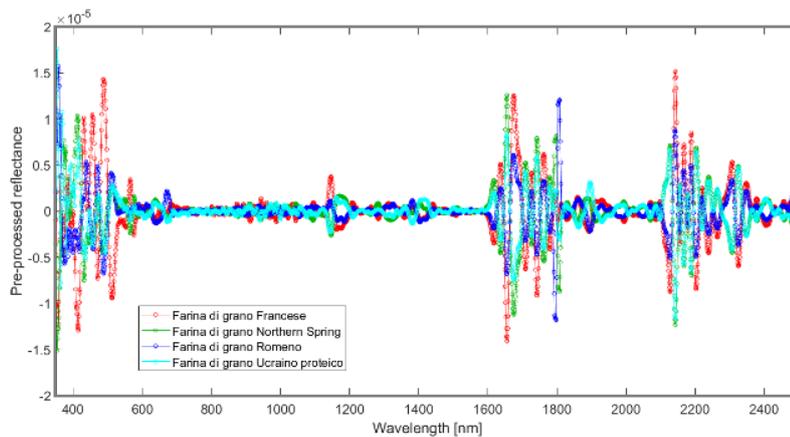
b

Figura A - 72. Spettri medi grezzi (a) in riflettanza e pre-elaborati (b) dei campioni di frumento tenero in grani.

Allegati



a

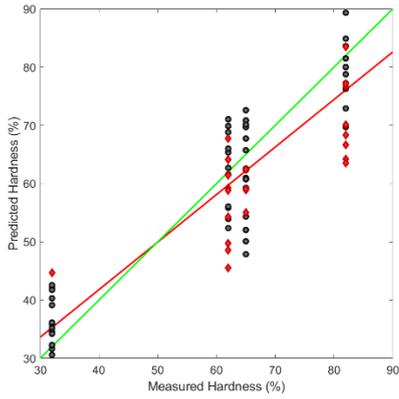


b

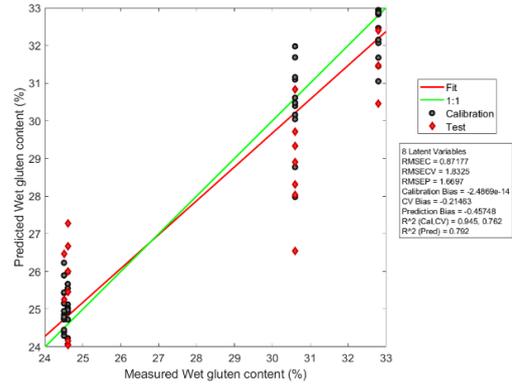
Figura A - 73. Spettri medi grezzi (a) in riflettanza e pre-elaborati (b) dei campioni di farina di grano tenero.

Regressioni Partial Least Squares

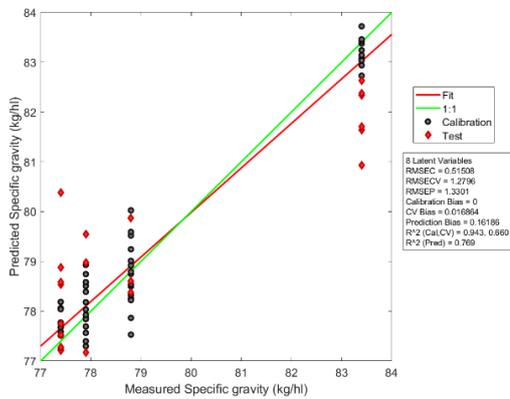
I risultati delle regressioni PLS e la restituzione grafiche delle rette di regressione effettuate con gli spettri acquisiti sui campioni di frumento tenero in grani per i parametri: Durezza (%), Glutine Umido (%), Peso specifico (kg/hl), Proteine (%), Umidità (%) e W (%) sono riportati in [Figura A - 74](#). Mentre, i risultati delle regressioni PLS e la restituzione grafiche delle rette di regressione effettuate con gli spettri acquisiti sui campioni di farina di grano tenero per i parametri: Amido (%), Assorbimento (%), Ceneri (%), Glutine Umido (%), Proteine (%), W (%) e Umidità (%) sono riportati in [Figura A - 75](#) e [Figura A - 76](#).



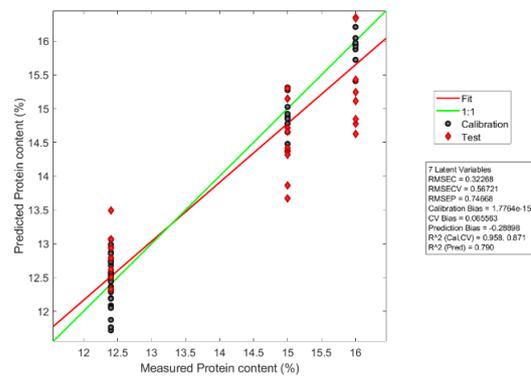
a



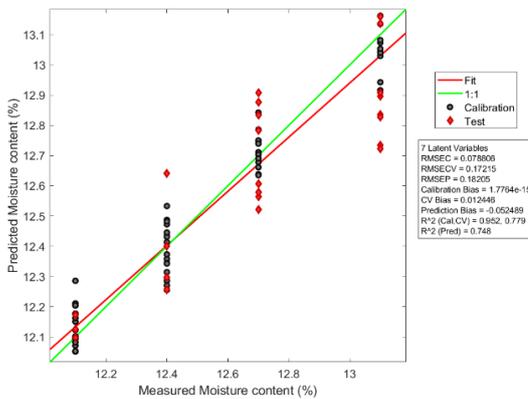
b



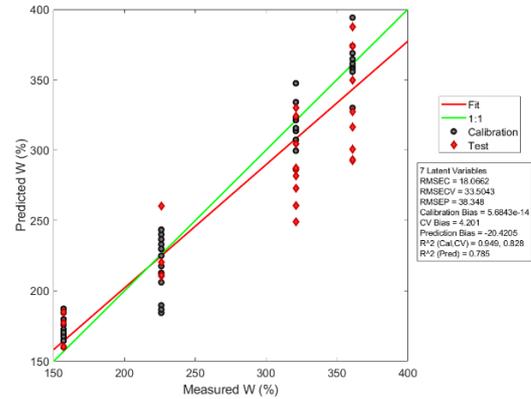
c



d



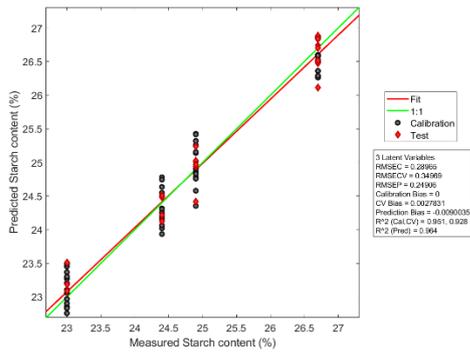
e



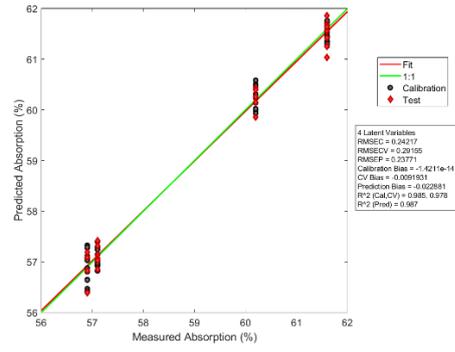
f

Figura A - 74. Rette di regressione PLS effettuate con gli spettri acquisiti sui campioni di frumento tenero in grani per i parametri: a) Durezza (%), b) Glutine Umido (%), c) Peso specifico (kg/hl), d) Proteine (%), e) Umidità (%) ed f) W (%).

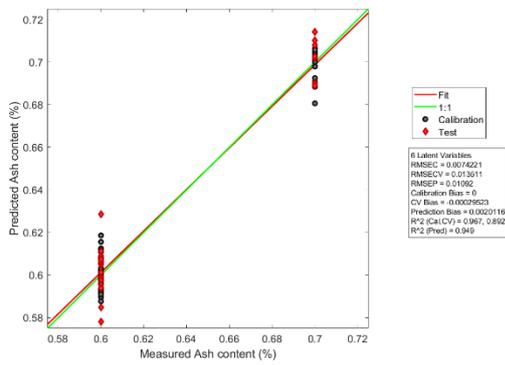
Allegati



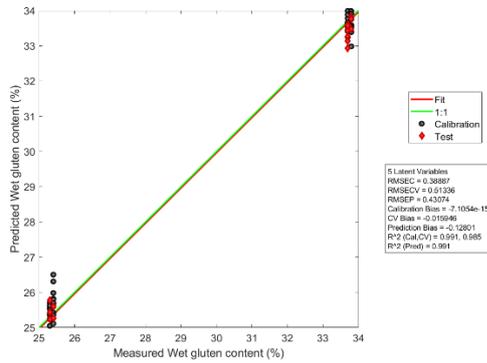
a



b



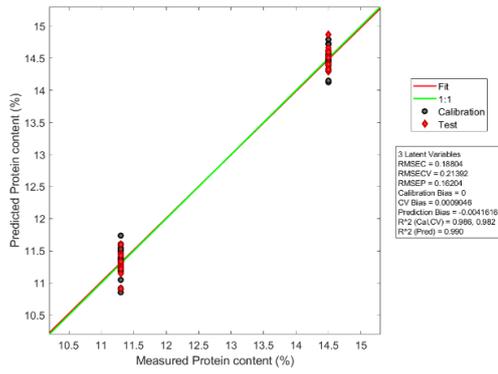
c



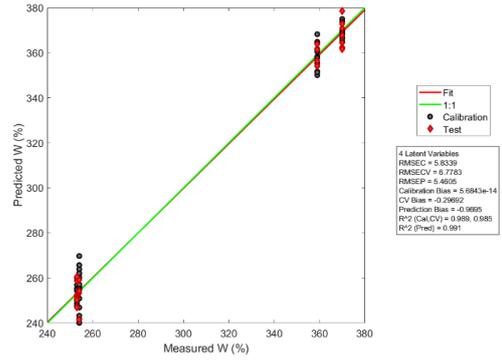
d

Figura A - 75. Rette di regressione PLS effettuate con gli spettri acquisiti sui campioni di farina di grano tenero per i parametri: a) Amido (%), b) Assorbimento (%), c) Ceneri (%) e d) Glutine Umido (%).

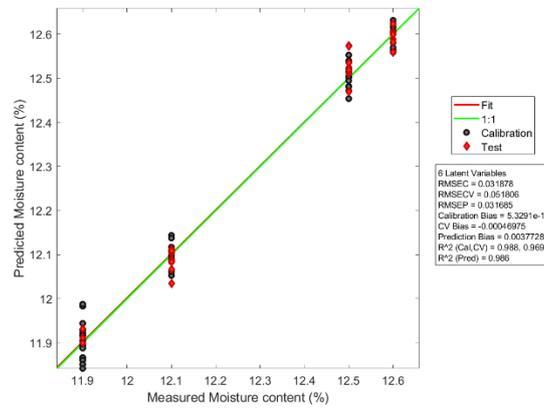
Allegati



a



b



c

Figura A - 76. Rette di regressione PLS effettuate con gli spettri acquisiti sui campioni di farina di grano tenero per i parametri: a) Proteine (%), b) W (%) e c) Umidità (%).

LISTA DELLE PUBBLICAZIONI

Materie prime

Giuseppe Bonifazi, Riccardo Gasbarrone and Silvia Serranti, "Characterization of printed circuit boards from e-waste byproducts for copper beneficiation". *Green and circular economy: ricerca, innovazione e nuove opportunità - Atti dei convegni aperti a call for papers a cura di Fabio Fava. Waste Management And Valorization (Ricerche applicate per il trattamento ed il recupero di residui solidi), ECOMONDO 2018 Green & Circular Economy, 6-9 Novembre 2018, pp. 65 - 69. Maggioli S.p.A.*

Giuseppe Bonifazi, Riccardo Gasbarrone e Silvia Serranti, "Il riciclo meccanico dei rifiuti di apparecchiature elettriche ed elettroniche: una sfida tecnologica" a cura di A. Budoni, M. Martone, S. Zerunian, 2018, "La Bioregione Pontina: esperienze, problemi, linee di ricerca per scenari di futuro" Edizioni SdT, Firenze, ebook open access: <http://www.societadeiterritorialisti.it/2016/01/17/collana-qricerche-e-studi-territorialistiq/>

Prodotti agro-alimentari

S. Serranti, G. Bonifazi, R. Gasbarrone, "Olive fruit ripening evaluation and quality assessment by hyperspectral sensing devices", *Proc. SPIE 10665, Sensing for Agriculture and Food Quality and Safety X, 106650R (15 May 2018); doi: 10.1117/12.2297352; https://doi.org/10.1117/12.2297352.*

S. Serranti, G. Bonifazi, R. Gasbarrone, "Kiwifruits ripening assessment by portable hyperspectral devices", *Proc. SPIE 10665, Sensing for Agriculture and Food Quality and Safety X, 106650S (15 May 2018); doi: 10.1117/12.2297353; https://doi.org/10.1117/12.2297353.*

Giuseppe Bonifazi, Riccardo Gasbarrone, and Silvia Serranti "Dried red chili peppers pungency assessment by visible and near infrared spectroscopy ", *Proc. SPIE 10986, Algorithms, Technologies, and Applications for Multispectral and Hyperspectral Imagery XXV, 109861S (14 May 2019); doi: 10.1117/12.2517069; https://doi.org/10.1117/12.2517069.*

Altro

R. Gasbarrone, A. Currà, A. Cardillo, G. Bonifazi, S. Serranti, "Near infrared spectroscopy of human muscles", *Proc. SPIE 10489, Optical Biopsy XVI: Toward Real-Time Spectroscopic Imaging and Diagnosis, 1048914 (21 February 2018); doi: 10.1117/12.2287814; https://doi.org/10.1117/12.2287814*

G. Bonifazi, A. Cardillo, A. Currà, R. Gasbarrone, S. Serranti, "Near infrared spectroscopy as a tool for in vivo analysis of human muscles ", *Proc. SPIE 10662, Smart Biomedical and Physiological Sensor Technology XV, 106620O (14 May 2018); doi: 10.1117/12.2304311; https://doi.org/10.1117/12.2304311*

Lista delle pubblicazioni

Currà Antonio, Gasbarrone Riccardo, Cardillo Alessandra, Trompetto Carlo, Fattapposta Francesco, Pierelli Francesco, Missori Paolo, Bonifazi Giuseppe, Serranti Silvia. Near infrared spectroscopy as a tool for in vivo analysis of human muscles. Scientific Reports 9, Article number: 8623 (2019). <https://doi.org/10.1038/s41598-019-44896-8>

Currà A, Pierelli F, Gasbarrone R, Mannarelli D, Nofroni I, Matone V, Marinelli L, Trompetto C, Fattapposta F and Missori P (2019) The Ventricular System Enlarges Abnormally in the Seventies, Earlier in Men, and First in the Frontal Horn: A Study Based on More Than 3,000 Scans. Front. Aging Neurosci. 11:294. doi: 10.3389/fnagi.2019.00294

RINGRAZIAMENTI

Desidero ringraziare tutti coloro che mi hanno aiutato nella ricerca e nella realizzazione finale di questo elaborato. Tutte le persone che verranno citate in questi ringraziamenti hanno svolto un ruolo fondamentale nella stesura della tesi, ma desidero precisare che ogni errore o imprecisione è imputabile soltanto a me.

Ringrazio anzitutto il mio tutor Professor Giuseppe Bonifazi e la Professoressa Silvia Serranti, senza il loro supporto e la loro guida sapiente questa tesi non esisterebbe.

Intendo poi ringraziare la mia “collega” Ing. Roberta Palmieri che mi ha accompagnato durante questa strada tortuosa e ne ha allietato la percorrenza. Ringrazio anche il Dott. Giuseppe Capobianco che, insieme all’Ingegnere Palmieri, hanno fornito un prezioso supporto per le analisi effettuate e per la strumentazione utilizzata presso il laboratorio.

Desidero poi ringraziare il Professor Antonio Currà ed il Professor Eugenio Lendaro per avermi fatto spaziare in diversi ambiti di ricerca con approccio multidisciplinare.

Ringrazio, inoltre, tutti i miei colleghi e compagni di ricerca che mi hanno supportato e “sopportato” durante questi anni.

Infine, ho desiderio di ringraziare con affetto la mia famiglia ed in particolare i miei genitori per il sostegno ed il grande aiuto che mi hanno dato e per essermi stati vicini ogni momento durante questi anni di lavoro.

Dedico questo mio lavoro da ingegnere e da aspirante ricercatore ai miei nipotini, il vero “futuro del Bel Paese”, con l’augurio che anche loro un giorno riescano a coronare tutti i propri sogni contribuendo a rendere il mondo un posto migliore.

“Ogni tecnologia sufficientemente avanzata è indistinguibile dalla magia”

Sir Arthur Charles Clarke

