



Università di Genova

Dipartimento di Farmacia - DIFAR

Corso di Laurea in Chimica e Tecnologia Farmaceutiche

STUDIO DELLA PENETRAZIONE DELLA RADIAZIONE NIR ATTRAVERSO MATERIALI POLIMERICI: UNA VALUTAZIONE QUANTITATIVA MEDIANTE REGRESSIONE MULTIVARIATA

Relatore

Prof.ssa Cristina Malegori

Correlatori

Prof. Paolo Oliveri

Dott.ssa Sara Gariglio

Candidata

Giulia Ferrari

Anno accademico 2025-2026

Riassunto

La radiazione nel vicino infrarosso (NIR) sta riscontrando un maggiore successo per la capacità di penetrare oltre agli strati superficiali e per la possibilità di combinare tecniche spettroscopiche con tecniche di imaging (HSI - hyperspectral imaging).

Si sono analizzati campioni polimerici stratificati tramite HSI nella regione NIR (HSI-NIR). Applicando una strategia *image based* si sono costruiti dei modelli qualitativi mediante l'algoritmo PLS (regressione ai minimi quadrati parziali) per valutare la profondità di penetrazione della radiazione.

Si sono analizzate due tipologie di campioni in PETG (polietilene tereftalato glicole) e PLA (acido polilattico), alti 3 e 5 cm. Ciascuna di esse è costituita da un training set e un test set interno, con campioni parallelepipedi, e da un test set esterno cilindrico, per analizzare l'effetto di diversi *infill*.

Con i dati ottenuti si sono costruiti dei modelli di regressione multivariata utilizzando pretrattamenti con effetti diversi. La combinazione SNV-centraggio sulla media ha permesso di correggere i fenomeni legati allo scattering della radiazione. La combinazione Log₁₀-centraggio sulla media mitiga la non linearità osservata. Visti tali effetti si è deciso di utilizzarli in combinazione, Log₁₀-SNV-centraggio sulla media, per sfruttarli entrambi, ma non si sono migliorate le prestazioni del modello.

I risultati evidenziano una differente penetrazione della radiazione NIR attraverso i due materiali, raggiungendo maggiori profondità nel PLA rispetto al PETG, in cui la radiazione è penetrata fino a 12 mm. Nei campioni con il PLA sopra si è riscontrata una differente penetrazione nel set da 3 cm, arrivando a 16 mm, e in quello da 5 cm, in cui sembra di penetrare fino a 40 mm.

Il lavoro conferma la capacità di penetrazione della radiazione NIR, sottolineando le potenzialità del suo utilizzo. Inoltre, la strategia di elaborazione delle immagini potrà fungere da esempio per determinare la profondità di penetrazione in altri campioni.

Abstract

Near-infrared (NIR) radiation is gaining increasing interest due to its ability to penetrate beyond superficial layers and the possibility of combining spectroscopic techniques with imaging techniques (hyperspectral imaging, HSI).

Stratified polymeric samples were analysed using HSI in the NIR region (NIR-HSI). Applying an image-based strategy, qualitative models were constructed using the Partial Least Squares (PLS) algorithm to evaluate the penetration depth of the radiation.

Two types of samples, made with polyethylene terephthalate glycol (PETG) and polylactic acid (PLA), with thicknesses of 3 and 5 cm, were analysed. Each set consists of a training set and an internal test set, composed of parallelepiped-shaped samples, and an external test set composed of cylindrical samples, used to analyse the effect of different infills.

Using the obtained data, multivariate regression models were constructed using pretreatments with different effects. The combination of Standard Normal Variate (SNV) and mean centering allowed the correction of phenomena related to radiation scattering. The combination of Log_{10} and mean centering mitigates the observed non-linearity. Given these effects, it was decided to use them in combination, Log_{10} -SNV-mean centering, to exploit both advantages; however, this did not improve the model's performance.

The results highlight a difference in the penetration of NIR radiation through the two materials, reaching greater depths in PLA compared to PETG, where the radiation penetrated up to 12 mm. In samples with PLA on top, different penetration depths were found: reaching 16 mm in the 3 cm set, and appearing to penetrate up to 40 mm in the 5 cm set.

This work confirms the penetration capability of NIR radiation, highlighting the potential of its application. Furthermore, the image processing strategy can serve as a reference for determining the penetration depth in other samples.

Indice

1	INTRODUZIONE	1
1.1	Spettroscopia nel vicino infrarosso	1
1.1.1	Cenni storici	1
1.1.2	Spettroscopia vibrazionale	3
1.1.3	Spettri NIR	7
1.1.4	Applicazioni	9
1.1.5	Imaging iperspettrale	10
1.2	Chemimetria	12
1.2.1	Generalità	12
1.2.2	Pretrattamenti	13
1.2.3	Metodi supervisionati e non supervisionati	15
1.3	Penetrazione della radiazione NIR	22
1.3.1	Fattori che influenzano la penetrazione	23
1.3.2	Stato dell'arte	25
2	SCOPO	27
3	MATERIALI E METODI	29
3.1	Campioni	29
3.2	Acquisizione delle immagini	31
3.2.1	Strumento e specifiche di acquisizione	31
3.2.2	Esecuzione dell'analisi	31
3.3	Trattamento dei dati	34
3.3.1	Dataset	35
3.3.2	Identificazione delle regioni di interesse e analisi object based	36
3.3.3	Pretrattamenti	37
3.3.4	Spettri grezzi e pretrattati	39
3.3.5	Analisi delle componenti principali	39

3.3.6	Regressione ai minimi quadrati parziali	40
4	RISULTATI E DISCUSSIONE	42
4.1	Campioni da 3 cm	42
4.1.1	Spettri grezzi e pretrattati	42
4.1.2	Analisi delle componenti principali	46
4.1.3	Regressione ai minimi quadrati parziali	50
4.2	Campioni 5 cm	68
4.2.1	Spettri grezzi e pretrattati	68
4.2.2	Analisi delle componenti principali	71
4.2.3	Regressione ai minimi quadrati parziali	74
5	CONCLUSIONI	91
	BIBLIOGRAFIA	94

Indice delle immagini

Figura 1-1:Tipologie di vibrazione [9]	4
Figura 1-2: Oscillatore armonico e anarmonico [10]	5
Figura 1-3: Modalità di acquisizione degli spettri NIR. [11]	8
Figura 1-4: Rappresentazione schematica di una immagine iperspettrale e di una immagine RGB [19]	11
Figura 1-5 Scattering di Rayleigh e Scattering di Mie [33]	24
Figura 3-1: Set up delle acquisizioni	33
Figura 3-2: Valori dell'intensità globale lungo PC2 e individuazione delle ROI all'interno delle maschere	37
Figura 4-1: Spettri grezzi del PLA e del PETG	43
Figura 4-2: Spettri dei campioni con altezza 3 cm in configurazione PETG-PLA.	44
Figura 4-3:Spettri dei campioni con altezza 3 cm in configurazione PLA-PETG.	45
Figura 4-4: PCA dei campioni di altezza 3 cm in configurazione PETG-PLA.	48
Figura 4-5: PCA dei campioni di altezza 3 cm in configurazione PLA-PETG.	50
Figura 4-6: Modello PLS, utilizzando come pretrattamento SNV-centraggio sulla media, dei campioni di altezza 3 cm in configurazione PETG-PLA.	53
Figura 4-7: Modello PLS, utilizzando come pretrattamento SNV-centraggio sulla media, dei campioni di altezza 3 cm in configurazione PLA-PETG.	56
Figura 4-8: Modello PLS, utilizzando come pretrattamento Log10-centraggio sulla media, dei campioni di altezza 3 cm in configurazione PETG-PLA.	59
Figura 4-9: Modello PLS, utilizzando come pretrattamento Log10-centraggio sulla media, dei campioni di altezza 3 cm in configurazione PLA-PETG.	61
Figura 4-10: Modello PLS utilizzando come pretrattamento Log10-SNV-centraggio sulla media dei campioni di altezza 3 cm in configurazione PETG-PLA.	64
Figura 4-11: Modello PLS utilizzando come pretrattamento Log10-SNV-centraggio sulla media dei campioni di altezza 3 cm in configurazione PLA-PETG.	66
Figura 4-12: Spettri dei campioni con altezza 5 cm in configurazione PETG-PLA.	69
Figura 4-13: : Spettri dei campioni con altezza 5 cm in configurazione PLA-PETG.	70
Figura 4-14: PCA dei campioni di altezza 5 cm in configurazione PETG-PLA.	72
Figura 4-15: PCA dei campioni di altezza 5 cm in configurazione PLA-PETG.	74

Figura 4-16: Modello PLS, utilizzando come pretrattamento SNV-centraggio sulla media, dei campioni di altezza 5 cm in configurazione PETG-PLA.	76
Figura 4-17: Modello PLS, utilizzando come pretrattamento SNV-centraggio sulla media dei campioni, di altezza 5 cm in configurazione PLA-PETG.	79
Figura 4-18: Modello PLS utilizzando come pretrattamento Log ₁₀ -centraggio sulla media dei campioni di altezza 5 cm in configurazione PETG-PLA.	81
Figura 4-19: Modello PLS utilizzando come pretrattamento Log ₁₀ -centraggio sulla media dei campioni di altezza 5 cm in configurazione PLA-PETG.	83
Figura 4-20: Modello PLS, utilizzando come pretrattamento Log ₁₀ -SNV-centraggio sulla media, dei campioni di altezza 5 cm in configurazione PETG-PLA.	86
Figura 4-21: Modello PLS utilizzando come pretrattamento Log ₁₀ -SNV-centraggio sulla media dei campioni di altezza 5 cm in configurazione PLA-PETG.	88

Indice tabelle

Tabella 3-1: Altezze e geometrie di riempimento dei cilindri da 3 cm	30
Tabella 3-2: Altezza e geometria di riempimento dei cilindri da 5 cm	30
Tabella 3-3: Altezze dei regoli contenuti nel test set interno per i campioni da 3 cm	35
Tabella 3-4: Altezze dei regoli contenuti nel test set interno per i campioni da 5 cm	36
Tabella 4-1: Riassunto dei valori di RMSE e di bias per ciascuna configurazione e ciascun pretrattamento per i campioni di altezza 3 cm	67
Tabella 4-2: Riassunto dei valori di RMSE e di bias per ciascuna configurazione e ciascun pretrattamento per i campioni di altezza 3 cm	90

1 INTRODUZIONE

1.1 SPETTROSCOPIA NEL VICINO INFRAROSSO

La spettroscopia nel vicino infrarosso (NIR - Near Infrared) si basa sull'utilizzo della radiazione per determinare un'interazione con la materia nell'intervallo spettrale che include lunghezze d'onda comprese tra i 780-2500 nm. La regione spettrale del NIR si colloca, all'interno dello spettro elettromagnetico, a valori di energia intermedi rispetto a quelli del visibile e del medio infrarosso (MIR – Mid Infrared).

1.1.1 Cenni storici

Nel 1800 William Herschel scoprì la regione del NIR tramite uno studio sul riscaldamento causato dalla radiazione solare [1]. Utilizzando un prisma di vetro per disperdere la radiazione su tre termometri, notò che, appena oltre la regione del rosso nel visibile, la temperatura era più alta, nonostante non fosse più presente luce visibile. Per questo motivo Herschel chiamò questa radiazione infrarosso – oltre il rosso. In realtà, Herschel non osservò tutto l'infrarosso, ma solo una sua porzione. Infatti il vetro, utilizzato come materiale disperdente durante l'esperimento, non trasmette molto il MIR e quindi quello che si osservò realmente era il NIR [1].

Il primo spettro NIR è stato registrato diversi anni dopo la scoperta della radiazione. Questo passo importante, compiuto nel 1881, è dovuto a William de Wiveleslie Abney e Edward Robert Festing [1]. Il loro lavoro è stato importante non solo perché ha consentito di avere la prima analisi NIR utilizzando una lastra fotografica che si è scoperta essere sensibile a questa radiazione, ma ha anche rappresentato la prima volta in cui si è cercato di interpretare uno spettro NIR. I lavori in questa regione spettrale sono continuati con William Weber Coblentz [1], che ha costruito per primo uno spettrometro che lavorasse alle lunghezze d'onda del NIR. I suoi lavori hanno avuto una rilevanza notevole. Registrando gli spettri di centinaia di composti ha scoperto che sostanze con gli stessi sostituenti hanno profili diversi. Nonostante l'unicità

degli spettri, Coblenz è stato in grado di riconoscere delle caratteristiche comuni come l'assorbimento di certi gruppi funzionali alle stesse lunghezze d'onda.

Nonostante alcuni lavori [2,3] sia in campo teorico che in progettazione strumentale, l'impatto che il NIR ha avuto è stato limitato soprattutto se paragonato a quello del MIR, scoperto successivamente. La poca risonanza avuta dal NIR è evidente nella definizione data da Wetzel nel 1983 [4], che descrive questa tecnica come "*sleeper among spectroscopic techniques*" (dormiente tra le tecniche spettroscopiche).

La difficoltà nello sviluppo della spettroscopia NIR è dovuta alle caratteristiche delle bande spettrali che caratterizzano questa regione. Segnali di debole intensità, insieme alla presenza di bande di *overtone* e di combinazione non hanno riscontrato grande favore da parte degli spettroscopisti classici, che ritenevano questi aspetti di difficile interpretazione.

Nonostante le difficoltà iniziali nell'uso della spettroscopia NIR, a partire dagli anni 40 del '900 si sono sviluppati diversi lavori che dimostravano l'utilità di questa tecnica analitica [5]. Una figura fondamentale è stata quella di Karl Norris che, nel 1968, studiando il contenuto di umidità in prodotti agricoli, abbandonò l'approccio legato alla legge di Lambert-Beer per l'analisi quantitativa [6]. Sviluppò invece un modello basato su due lunghezze d'onda per correlare l'informazione spettrale con la concentrazione dell'analita. Inoltre, Norris introdusse l'utilizzo della riflettanza diffusa come metodo di analisi non distruttivo [6].

Da questo punto in poi, la spettroscopia NIR subì un'evoluzione notevole, sia in termini di strumentazione, sia in termini di acquisizione dei dati e del loro trattamento. In questo sviluppo svolse un ruolo fondamentale la chemiometria, che permise di ottenere le informazioni che, per le caratteristiche spettrali del NIR, un tempo erano nascoste. Grazie a questi progressi la spettroscopia NIR è passata da "*sleeping technique to the morning star of spectroscopy*" (tecnica dormiente all'astro nascente della spettroscopia) [7].

Questa evoluzione che la spettroscopia NIR ha subito nel tempo è legata ai grandi vantaggi che presenta. Infatti, questa tecnica permette di acquisire spettri di campioni solidi o liquidi senza necessità di pretrattamento [8]. Il metodo di acquisizione, quindi, è non distruttivo e rapido, in quanto l'acquisizione degli spettri richiede tipicamente pochi secondi[6]. Inoltre, con lo sviluppo tecnologico, si sono prodotti strumenti sempre più piccoli e portatili, che possono essere usati per svolgere delle rapide analisi in loco. Un altro vantaggio della spettroscopia NIR è che, attraverso una singola analisi spettrale, si possono ricavare sia

informazioni fisiche che chimiche [6]. Questo aspetto è di grande importanza quando si ha a che fare con la caratterizzazione dei campioni.

Nel prossimo paragrafo si fornirà una descrizione della spettroscopia vibrazionale, nella quale rientra anche la spettroscopia NIR, al fine di spiegare le caratteristiche particolari derivanti dall'interazione della radiazione in questo range spettrale con la materia.

1.1.2 Spettroscopia vibrazionale

Come detto in precedenza, le tecniche spettroscopiche sono basate sull'interazione tra la radiazione e la materia. Considerando la duplice natura della radiazione, ondulatoria e corpuscolare, si comprende come essa, sottoforma di fotoni, sia in grado di cedere alla materia dei quanti di energia [6]. Se l'energia trasferita è sufficiente, si ha la possibilità di far passare l'analita dallo stato fondamentale, in cui si trova naturalmente, a uno stato eccitato.

Mentre questa tesi si concentrerà per lo più sulle spettroscopie vibrazionali, tra cui rientra la tecnica NIR, ne esistono altre due tipologie: quella elettronica e quella rotazionale. La classificazione in una tipologia piuttosto che in un'altra è funzione delle transizioni che vengono promosse dalla radiazione incidente sulla materia. Questo aspetto è dipendente dall'energia trasportata dalla luce che è pari a:

$$E = h\nu = \frac{hc}{\lambda} \quad (1.1)$$

dove h è la costante di Planck (pari a $6,626 \cdot 10^{-34}$ Js), ν la frequenza della radiazione, c la velocità della luce (pari a $3 \cdot 10^8$ m/s) e λ la lunghezza d'onda della luce.

Le diverse tipologie di transizioni che possono avvenire richiedono quantità di energia diverse e in particolare si ha che: $E_{\text{Trans_elettroniche}} > E_{\text{Trans_Vibrazionali}} > E_{\text{Trans_rotazionali}}$ [5].

In questa scala, l'energia che viene trasferita dalla radiazione NIR consente di effettuare delle transizioni vibrazionali.

Allo stato fondamentale, gli atomi che costituiscono le molecole vibrano l'uno rispetto all'altro. I meccanismi di vibrazione sono di vario tipo e vengono illustrati in Figura 1-1 [5]:

- *Stretching* (stiramento): variazione della lunghezza del legame. Può essere simmetrico o asimmetrico.

- *Bending* (piegamento): variazione dell'ampiezza dell'angolo di legame. I movimenti che possono portare a questi cambiamenti sono: *scissoring*, *rocking*, *wagging* e *twisting*.

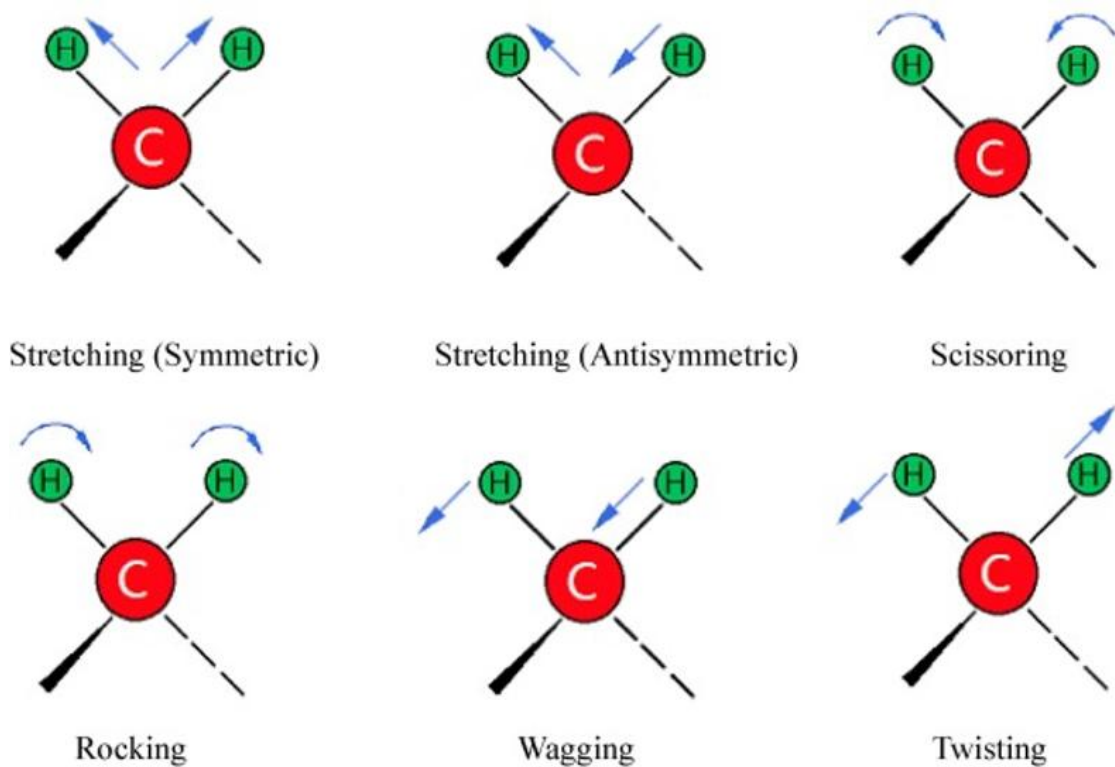


Figura 1-1: Tipologie di vibrazione [9]

La vibrazione di una molecola e le transizioni tra i suoi livelli energetici possono essere descritte per mezzo del modello dell'oscillatore armonico [5]. In questo sistema si compiono varie approssimazioni. Si considera una molecola diatomica in cui gli atomi sono approssimati a due sfere e il legame che li unisce è assimilabile ad una molla. Gli atomi che formano la molecola, anche nello stato fondamentale, vibrano determinando una variazione della lunghezza di legame. Questo comportamento può essere descritto dalla legge di Hooke [5], in cui il valore di energia potenziale è pari a:

$$V = \frac{1}{2}k(r - r_e)^2 \quad (1.2)$$

dove r è la distanza internucleare durante la vibrazione, r_e è la distanza internucleare all'equilibrio e k è la costante di forza del legame.

L'energia potenziale di questo sistema può essere descritta da una curva che ha la caratteristica di essere parabolica e simmetrica rispetto alla distanza internucleare di equilibrio, come si può vedere dalla curva in verde in Figura 1-2.

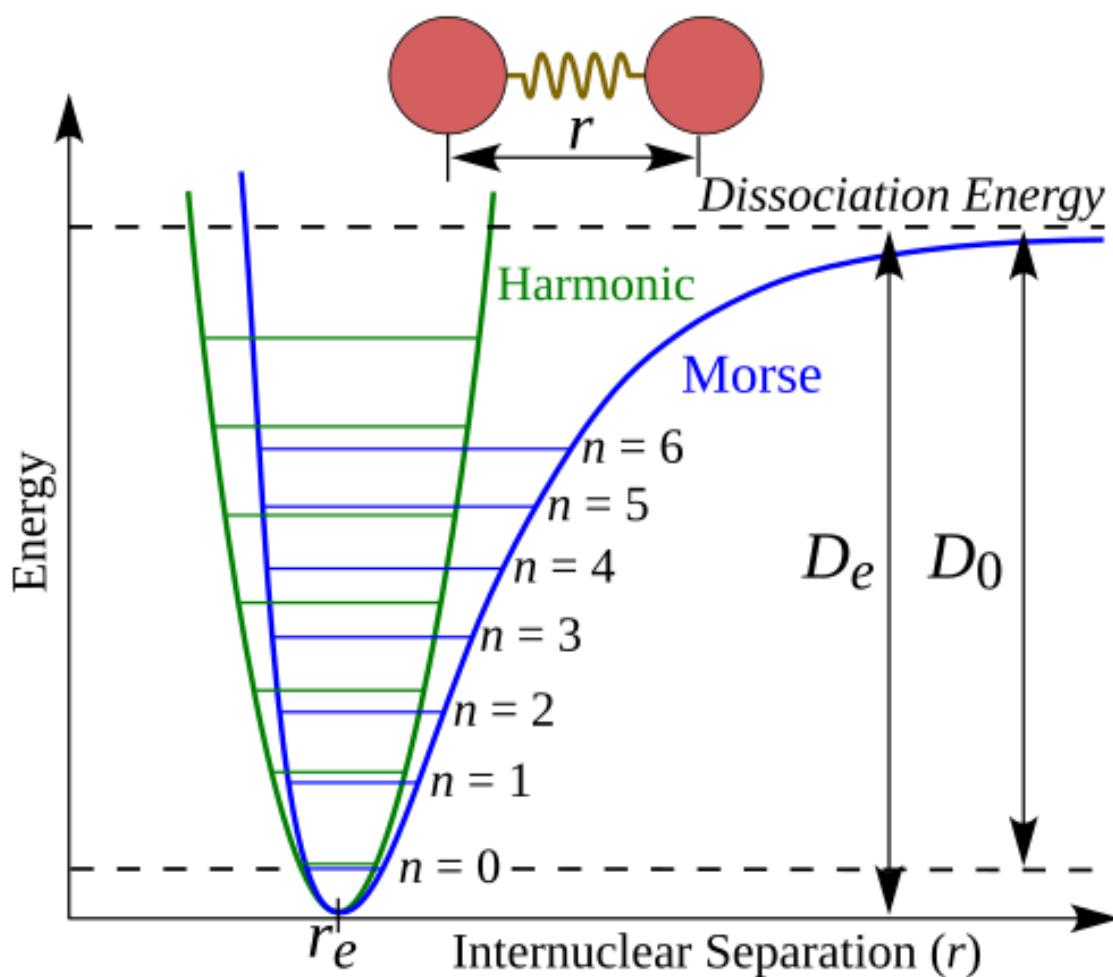


Figura 1-2: Oscillatore armonico, curva verde, e anarmonico, curva blu [10]

L'utilizzo di questo modello consente anche di determinare la frequenza vibrazionale:

$$\nu_0 = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}} \quad (1.3)$$

dove μ è la massa ridotta del sistema ed è pari a:

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \quad (1.4)$$

dove m_1 e m_2 sono le masse dei due atomi coinvolti nel legame. Dall'equazione (1.3) si ricava che le frequenze di vibrazione sono molto influenzate dalla struttura del composto di

interesse, in particolare dalla massa degli atomi che costituiscono la molecola e dalla forza di legame. Questa caratteristica è di fondamentale importanza per comprendere le caratteristiche della spettroscopia NIR.

Al modello descritto bisogna apportare delle modifiche affinché sia adatto per la descrizione di molecole [5]. Infatti, dall'equazione (1.2) si può ottenere qualsiasi valore di energia. Questo è in contrasto con il fatto che, da un punto di vista quantistico, i livelli energetici sono quantizzati. Per tenere conto di questo aspetto l'equazione diventa:

$$E_v = \left(n + \frac{1}{2}\right) h\nu_0 \quad (1.5)$$

dove n è il numero quantico vibrazionale, che può assumere solo valori interi, e ν_0 è la frequenza vibrazionale fondamentale.

Affinché si abbia interazione tra la radiazione infrarossa e la molecola, oltre all'energia sufficiente per far avvenire una transizione vibrazionale, è necessaria un'altra caratteristica: il vettore elettrico della radiazione deve oscillare alla stessa frequenza del momento di dipolo della molecola. Infatti, una molecola è attiva all'infrarosso solo se il dipolo cambia in funzione della vibrazione della molecola. Questo determina il fatto che la spettroscopia infrarossa è molto sensibile a gruppi funzionali polari.

Il modello dell'oscillatore armonico è tuttavia solo una prima approssimazione di quanto accade durante le transizioni vibrazionali. Questo modello è infatti caratterizzato da livelli energetici discreti ed equidistanti e dal fatto che le transizioni consentite avvengono solo tra livelli energetici adiacenti. Queste due caratteristiche non permettono di spiegare alcuni fenomeni fondamentali che avvengono durante la spettroscopia NIR.

Per questo motivo, il modello migliore per descrivere questi fenomeni è l'oscillatore anarmonico [5]. In questo caso i livelli energetici sono pari a:

$$E_n = h\nu_0 \left(n + \frac{1}{2}\right) - \chi h\nu_0 \left(n + \frac{1}{2}\right)^2 \quad (1.6)$$

con χ costante di anarmonicità.

In questo caso la distanza tra i livelli energetici non è costante e diminuisce all'aumentare del livello energetico. Questo modello permette di spiegare le transizioni tra livelli energetici non adiacenti, che vengono chiamate *overtone* e sono uno dei fenomeni caratteristici della spettroscopia NIR.

Nella spettroscopia NIR sono presenti anche delle bande derivanti da fenomeni detti di combinazione, che avvengono quando l'energia fornita alle molecole consente di far avvenire più transizioni vibrazionali contemporaneamente. Questo determina la presenza di bande a frequenze pari alla somma o alla differenza della frequenza delle singole transizioni.

1.1.3 Spettri NIR

Le caratteristiche dell'oscillatore anarmonico e le transizioni promosse dalla radiazione NIR determinano le proprietà stesse degli spettri NIR, che sono caratterizzati da bande generalmente di bassa intensità e ampie. L'intensità è influenzata da χ , la costante di anarmonicità: minore è il suo valore, minore è l'intensità della banda di *overtone* [5]. Questo spiega perché le vibrazioni dei gruppi X-H sono quelle più rappresentate nella regione NIR, dal momento che hanno χ maggiore [5]. Al contrario, gruppi funzionali come C=O hanno χ bassa e di conseguenza l'intensità delle bande è inferiore nelle lunghezze d'onda del NIR, rispetto a quanto accade nel MIR [5]. Inoltre, l'intensità delle bande di *overtone* è influenzata dalla tipologia di transizione. Se la transizione comporta una variazione di $n = \pm 2$, avremo delle bande di maggiore intensità rispetto a quelle che si ottengono quando la transizione comporta una variazione pari a $n = \pm 3$ o superiore. Le transizioni con $n = \pm 1$, dette fondamentali, non sono rappresentate nella spettroscopia NIR perché l'energia che viene fornita in questa zona dello spettro elettromagnetico è più elevata. Per questo motivo le transizioni fondamentali sono rappresentate nel MIR.

L'ampiezza dei segnali, invece, è legata al fatto che le numerose bande di *overtone* e di combinazione spesso si sovrappongono tra di loro [5]. Questi aspetti rendono l'interpretazione spettrale complessa. Di conseguenza l'identificazione di gruppi funzionali specifici è meno immediata rispetto alla spettroscopia MIR.

Gli spettri NIR possono essere acquisiti in modalità differenti [5], rappresentate schematicamente in Figura 1-3:

- Trasmissione: la radiazione emessa dalla sorgente incontra il campione e la quantità che non viene assorbita raggiunge il rilevatore che è posto dalla parte opposta rispetto alla sorgente (Figura 1-3 a). Misura l'intensità della radiazione che non è stata assorbita dal campione e la sua unità di misura è la trasmittanza (T).

- Assorbimento: la misura viene effettuata in maniera analoga alla trasmissione (Figura 1-3 a). La differenza è rappresentata dall'elaborazione matematica dell'unità di misura, l'assorbanza (A), che deriva dalla trasmittanza alla quale è legata secondo la relazione:

$$A = -\log(T) \quad (1.7)$$

- Transflessione: la radiazione, dopo aver attraversato il campione, incontra una superficie riflettente, che la spinge a riattraversare l'analita e quindi raggiunge il detector, che si trova dalla stessa parte della sorgente (Figura 1-3 b). Viene misurata in transflettanza.
- Riflessione diffusa: la radiazione emessa dalla sorgente colpisce il campione che riflette la radiazione in tutte le direzioni. Parte di questa luce verrà registrata dal rivelatore, posto dalla stessa parte della sorgente (Figura 1-3 c). La riflessione viene misurata in riflettanza diffusa (R).

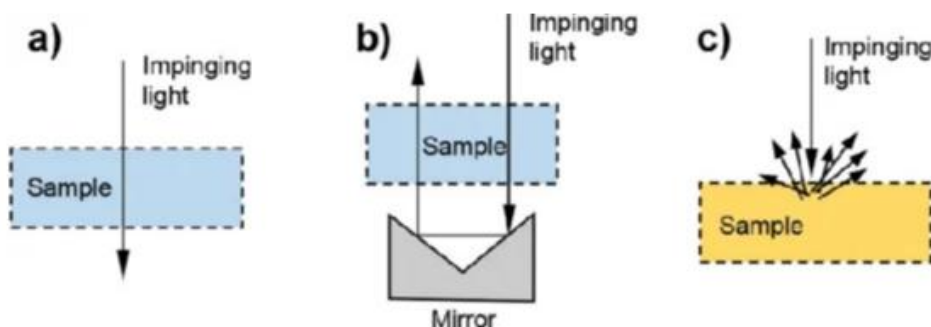


Figura 1-3: Modalità di acquisizione degli spettri NIR. a) Trasmittanza b) Transflettanza c) Riflettanza diffusa [11]

Indipendentemente dal metodo in cui si acquisiscono gli spettri, per ottenere degli spettri di buona qualità, prima dell'analisi è necessario effettuare due scansioni di riferimento dette "bianco" e "nero" [6]. Il nero, che consiste nell'acquisizione di uno spettro impedendo alla luce di raggiungere il rivelatore, consente di rimuovere il rumore elettrico di fondo dello strumento, definendo la linea di base dello spettro. Al contrario il bianco, ovvero l'acquisizione di uno spettro misurato utilizzando un materiale che riflette completamente la radiazione in tutto il range di lunghezze d'onda di interesse, consente di andare a identificare il limite superiore dello spettro. Se l'analisi del bianco non è effettuata correttamente, si possono "danneggiare" gli spettri. In particolare, questo si può tradurre in una riduzione della risoluzione dei picchi. Può altresì verificarsi il fenomeno della saturazione, in cui i picchi dell'analita hanno intensità maggiore del bianco, risultando tagliati. Tutti questi difetti causano perdita di informazione.

1.1.4 Applicazioni

Le metodiche analitiche che sfruttano la radiazione nel range del NIR hanno moltissimi campi di applicazione. L'impiego molto elevato è dovuto al fatto che è una tecnica che presenta molti aspetti positivi. Infatti, la spettroscopia NIR non necessita di preparazione del campione e si basa sull'interazione tra la materia e una radiazione non sufficientemente energetica da rovinare il campione. Questo la rende un metodo non distruttivo, che quindi consente di riutilizzare l'analita anche dopo il campionamento [6].

Un altro vantaggio è che lo svolgimento dell'analisi richiede poco tempo e questo permette di utilizzare questa metodica analitica anche per il monitoraggio di processo. In base alle necessità della linea di produzione si possono sfruttare analisi:

- *Off-line*: campionamento dalla linea di processo e analisi in un laboratorio dedicato. Questo tipo di analisi può introdurre ritardi nella produzione e non cattura cambiamenti transienti nel processo.
- *At-line*: campionamento automatico o manuale dalla linea di processo e trasporto a uno strumento analitico vicino.
- *On-line*: estrazione di un campione dalla linea di processo e introduzione nella zona di analisi con caratteristiche simili a quelle della zona di processo.
- *In-line*: lo strumento analitico è direttamente integrato nella linea di processo; l'analisi è effettuata nelle condizioni di processo.

La versatilità di utilizzo che la spettroscopia NIR fa sì che trovi molti campi di applicazione. In ambito farmaceutico può essere utilizzata lungo tutta la catena di processo: a partire dal controllo di qualità delle materie prime [12], al monitoraggio di processi specifici, come ad esempio l'essiccazione, oppure installata direttamente in-line [13] fino al controllo dell'uniformità del dosaggio nella forma farmaceutica [14]. Un altro campo in cui si hanno varie applicazioni è quello alimentare. La spettroscopia NIR può essere utilizzata per verificare il contenuto proteico di vari prodotti e per verificare la presenza di adulteranti, sempre di origine proteica, all'interno di essi. Questo può contribuire a determinare la sicurezza alimentare per individui con necessità particolari [15]. Anche in campo alimentare questa tecnica analitica si trova impiegata per controllare processi di essiccamento ed è particolarmente utile per i casi in cui un controllo in-line tramite metodi tradizionali non risulta essere possibile [16]. Sono poi

presenti diverse applicazioni in campo medico, a testimonianza della versatilità di queste metodiche. In questo ambito la spettroscopia NIR può essere utilizzata come mezzo diagnostico di supporto alle tecniche classiche. La sua utilità è dovuta al fatto che tessuti sani e tumorali presentano delle differenze nell'assorbimento, nello scattering e nel contenuto di acqua, lipidi e ossigeno [17].

1.1.5 Imaging iperspettrale

Un aspetto che ha ampliato ulteriormente l'utilizzo della spettroscopia NIR è lo sviluppo di nuove metodiche come l'imaging iperspettrale (*hyperspectral imaging* HSI), una tecnica analitica che combina l'imaging e la spettroscopia in modo da sfruttare i vantaggi di entrambe.

Le tecniche di imaging classiche consentono di osservare le caratteristiche superficiali e prevalentemente fisiche dei campioni, tra cui: colore, dimensioni, difetti esterni. Le immagini ottenute sono costituite da pixel, per ognuno dei quali viene registrata una risposta. Le immagini RGB, quelle più diffuse, si basano sull'analisi della risposta dei canali del rosso, del verde e del blu per andare a definire la combinazione dei tre colori primari che definisce ogni pixel [18]. Il grande aspetto positivo di queste tecniche è che rendono possibile la localizzazione nello spazio delle proprietà evidenziate, ma hanno lo svantaggio di non consentire l'osservazione delle caratteristiche chimiche, esplorabili solo per mezzo di metodi spettroscopici.

Combinando l'imaging e la spettroscopia, grazie all'HSI, si possono quindi ottenere sia le informazioni fisiche che quelle chimiche e la loro disposizione spaziale. Le immagini iperspettrali, anch'esse costituite da pixel, riportano per ognuno di essi centinaia di canali, ognuno dei quali corrisponde a una lunghezza d'onda. Lo spettro del campione, misurato in riflettanza, viene tradotto in un numero che rappresenta un valore in scala di grigio per ogni lunghezza d'onda. Si ottengono quindi tante immagini quante sono le lunghezze d'onda caratterizzate da pixel, ciascuno dei quali è codificato da un certo valore all'interno della scala di grigio. La sovrapposizione delle immagini ottenute per ciascuna lunghezza d'onda analizzata fornisce una matrice tridimensionale, come schematizzato in Figura 1-4. Queste matrici sono dette ipercubi, dei cubi di dati in cui si hanno due dimensioni, x e y, che portano l'informazione spaziale, e una terza dimensione z in cui è contenuta l'informazione spettrale [18]. Questa struttura ci permette di andare a osservare le immagini in maniera differente. Si ha la possibilità

di studiare lo spettro di ogni pixel, osservando la sua risposta lungo l'asse z, oppure si può osservare l'intera immagine registrata a una specifica lunghezza d'onda.

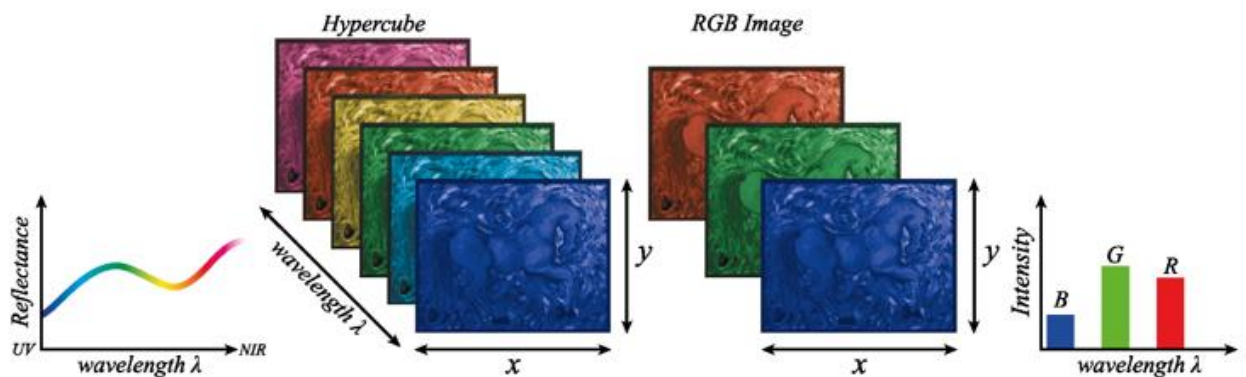


Figura 1-4: a sinistra rappresentazione schematica di una immagine iperspettrale. A destra rappresentazione di una immagine RGB [19]

Le immagini iperspettrali possono essere acquisite in tre modalità differenti [20]:

- *Whisk broom* (scansione puntuale): il campione o il rivelatore vengono mossi lungo le due dimensioni spaziali (x e y), acquisendo un pixel alla volta. Consente di avere risoluzione elevata, ma necessita di un tempo di analisi relativamente lungo.
- *Push broom* (scansione lineare): il campione o il rivelatore vengono mossi lungo una direzione (x o y), registrando lo spettro per una linea di pixel alla volta. È caratterizzato da una risoluzione peggiore rispetto al metodo precedente ma ha un tempo di analisi più contenuto.
- *Snapshot* (scansione planare): non si ha movimento né del campione né del rivelatore, acquisendo lo spettro di tutta l'immagine contemporaneamente. La risoluzione di questa modalità di acquisizione è la peggiore, ma la velocità di analisi è la più elevata.

Un parametro importante in base all'analisi che si vuole effettuare è la dimensione del pixel [21]. Questa, infatti, contribuisce a definire la risoluzione spaziale dell'immagine, ovvero il più piccolo dettaglio che può essere identificato come un'unità indipendente nell'immagine stessa. Maggiore è questo parametro, maggiore è la capacità di discriminare due punti vicini tra di loro all'interno del campione di interesse. Le dimensioni del pixel dipendono dalle caratteristiche della lente utilizzata nell'analisi e da alcuni parametri di acquisizione. Per quanto riguarda la lente, si ha che maggiore è il suo campo visivo e maggiore sarà la dimensione dei pixel, con una risoluzione spaziale peggiore, e viceversa. I parametri di acquisizione che

possono influenzare la dimensione del pixel sono la frequenza con cui avvengono le acquisizioni (*frame rate*) e la velocità con cui il campione o il rivelatore si muovono nell'acquisizione *whisk broom* e *push broom*.

Le immagini iperspettrali, una volta acquisite, devono andare incontro a una elaborazione spaziale, ovvero un insieme di processi necessari per estrarre l'informazione utile. Un'operazione molto utilizzata è quella del *masking* (mascheramento) che consente di calcolare delle maschere in modo tale da ignorare i pixel che non contengono l'informazione legata al campione. [22]. Una volta che è stata creata la maschera, in alcuni casi è di interesse andare ad individuare specifiche regioni all'interno del campione maggiormente rilevanti per l'analisi, chiamate regioni di interesse (ROI).

Dopo il mascheramento e l'eventuale selezione di ROI, le immagini iperspettrali possono essere studiate secondo diversi approcci in base alle esigenze dell'analisi [23]:

- *Pixel based*: si tratta ogni pixel che costituisce l'immagine o la ROI individuata come se fosse un campione a sé stante.
- *Object based*: si identificano degli oggetti più piccoli all'interno dell'immagine e si analizzano le proprietà globali dei singoli oggetti, che vengono considerati singoli campioni.
- *Image based*: si fa una caratterizzazione dell'immagine nella sua interezza, che viene considerata un campione.

1.2 CHEMIOMETRIA

1.2.1 Generalità

La chemiometria, secondo la definizione della Società Interazionale di Chemiometria, è “una disciplina chimica che usa metodi matematici e statistici per progettare o selezionare procedure o esperimenti ottimali e per fornire la massima informazione chimica, analizzando dati chimici” [24].

Dalla definizione si può dividere la chemiometria in due macroaree. La prima è rappresentata dal disegno sperimentale, che consiste nel progettare le sperimentazioni in modo tale da ottenere la maggiore informazione possibile usando il minor numero di esperimenti.

L'altra macroarea è rappresentata dal *pattern recognition*, che consiste nell'utilizzare modelli matematici e statistici al fine di trovare correlazioni tra i dati, e classificare e identificare le strutture presenti in dati chimici complessi. Ai fini di questa tesi, verranno approfonditi solamente gli aspetti relativi al *pattern recognition*.

La chemiometria trova applicazione in tutti i casi in cui l'informazione da estrarre è presente in dati chimici complessi, multivariati e di grandi dimensioni. Questo la rende una tecnica versatile che può essere applicata in moltissimi campi diversi.

In particolare, la chemiometria è da sempre cresciuta di pari passo con la spettroscopia NIR [5] e questo è dovuto alle proprietà di quest'ultima. Viste le caratteristiche della regione spettrale del NIR, ovvero la presenza di *overtone* e bande di combinazione molto ampie, non è possibile usare una sola lunghezza d'onda per costruire dei modelli. Per ottenere l'informazione necessaria, è invece necessario tipicamente un numero di variabili notevole. Infatti, le metodiche analitiche che attualmente impiegano il NIR sono in grado di registrare spettri con diverse centinaia di lunghezze d'onda in un intervallo di tempo di pochi secondi. Questo determina la presenza di una grandissima quantità di dati che deve essere gestita in modo da trarne la maggiore informazione chimica possibile.

1.2.2 Pretrattamenti

Prima di procedere all'analisi chemiometrica, i dati ottenuti devono essere sottoposti a dei pretrattamenti. Sono delle operazioni matematiche e statistiche che consentono di rimuovere la varianza indesiderata che può essere derivata da questioni strumentali, specie interferenti o da effetti derivanti dalla disposizione del campione durante l'analisi. La scelta dei pretrattamenti influenza fortemente la qualità dei dati che si andranno a utilizzare nelle fasi successive [25].

Per comprendere le tipologie di pretrattamenti bisogna osservare che i dati ottenuti dalle analisi vengono rappresentati sotto forma di matrici. Ciascuna riga che le costituisce rappresenta un campione, mentre ciascuna colonna rappresenta una variabile. I pretrattamenti possono essere applicati sia alle righe che alle colonne.

I pretrattamenti di colonna servono per dare lo stesso peso a tutte le variabili. Questo è di fondamentale importanza perché la chemiometria sfrutta le direzioni di massima varianza,

ma la varianza dipende dalla natura dei dati, e in particolare è influenzata dall'unità di misura e dall'ordine di grandezza di ciascuna variabile. I pretrattamenti di colonna, quindi, hanno l'obiettivo di standardizzare i dati andando a operare dei centraggi e/o delle scalature, in modo tale da rendere le variabili confrontabili tra loro [26].

Tra i pretrattamenti di colonna sono presenti:

- *Mean centering* (centraggio sulla media): corregge una differenza di locazione sistematica tra le variabili. È utile quando la media di colonna è molto diversa tra le variabili [26].
- *Column scaling* (scalatura di colonna): elimina le differenze di dispersione tra le variabili. È utile quando le variabili hanno media simile, mentre la deviazione standard è molto diversa [26].
- *Autoscaling* (autoscalatura): è una combinazione dei precedenti pretrattamenti. Elimina le differenze sia di locazione sia di dispersione. È il pretrattamento di elezione quando le variabili hanno natura diversa [26].

I pretrattamenti di riga, invece, sono particolarmente utili quando sono presenti dei profili continui o pseudo-continui, quindi sono molto utilizzati quando si ha a che fare con degli spettri. Questi pretrattamenti permettono di eliminare, se scelti correttamente, alcuni difetti che i profili spettrali possono presentare, tra cui gli effetti sistematici e il rumore [25].

Negli spettri, compresi quelli NIR, si possono riscontrare la presenza di alcune variazioni sistematiche ricorrenti. Tra queste abbiamo [25]:

- *Baseline shift* (spostamento della linea di base): spostamento della linea di base verso l'alto o verso il basso tra un campione e l'altro.
- *Baseline drift* (deriva della linea di base): cambiamento della pendenza della linea di base.
- Effetto di intensità globale: variazione dell'intensità globale del segnale.

Questi effetti potrebbero essere ridotti tramite correzioni strumentali, ma un trattamento matematico è di norma più rapido ed economico, e di conseguenza è solitamente preferito.

Per correggere gli effetti descritti si possono utilizzare le derivate. In particolar modo, la derivata prima consente di eliminare l'effetto dello spostamento della linea di base [25]. Tuttavia, la derivata prima non è in grado di correggere la deriva. Per ottenere questo risultato è necessario applicare la derivata seconda che è in grado di correggere sia lo spostamento che

la deriva della linea di base [25]. Tuttavia, le derivate hanno lo svantaggio di intensificare il rumore presente negli spettri, dal momento che rispondo alle variazioni di pendenza anche su piccola scala [25].

Per ovviare a questo problema un'alternativa molto utilizzata è la derivata di Savitzky-Golay [25], che oltre ad avere le caratteristiche positive delle derivate applica anche un processo di *smoothing* (smussatura). Si basa sull'utilizzo di una finestra di interpolazione, di ampiezza solitamente dispari, che si muove lungo lo spettro. I punti che, di volta in volta, fanno parte della finestra vengono interpolati con un polinomio dell'ordine desiderato, e la derivata viene calcolata sul nuovo polinomio ottenuto. In questo modo il rumore viene mitigato dall'interpolazione, e il calcolo della derivata non va a intensificarlo ulteriormente.

Un altro pretrattamento molto utilizzato è la trasformata *standard normal variate* (SNV) [25], che viene anche definita l'autoscalatura di riga. Come il corrispondente pretrattamento di colonna, è in grado di correggere sia la locazione che la dispersione. In questo modo elimina lo spostamento della linea di base e corregge l'effetto di intensità globale, ma non ha effetto sulla deriva della linea di base.

Come pretrattamento di riga si possono utilizzare anche i logaritmi, sia in base 10 che quelli naturali. L'uso del logaritmo permette di ridurre l'impatto di valori molto grandi che avrebbero una grande influenza sulla varianza, rendendo le differenze relative minori. Il logaritmo può essere usato solamente nei casi in cui non sono presenti variabili che assumono valori minori o uguali a zero [27].

I pretrattamenti, quindi, hanno un ruolo fondamentale nell'ottenere la maggior informazione possibile dai dati che si sono acquisiti. La loro scelta risulta cruciale per una corretta applicazione delle varie tecniche chemiometriche. Per questo motivo è necessario porre attenzione, oltre che al tipo di trattamento che si utilizza, anche all'ordine in cui questi vengono applicati. In particolare, si devono pretrattare i dati sempre prima con i pretrattamenti di riga e successivamente con i pretrattamenti di colonna [26].

1.2.3 Metodi supervisionati e non supervisionati

All'interno del *pattern recognition* rientrano diversi metodi che possono essere suddivisi a loro volta in due categorie [28]:

- Metodi non supervisionati (*unsupervised*): utilizzati per fare analisi esplorative, in modo da visualizzare il comportamento di dati e di campioni. Tra questi il metodo più importante è l'analisi delle componenti principali (PCA).
- Metodi supervisionati (*supervised*): hanno come obiettivo quello di fare una predizione relativamente a una risposta qualitativa o quantitativa, in modo da trovare una relazione tra le variabili che caratterizzano l'oggetto e una informazione aggiuntiva.

Lo sviluppo di questi metodi necessita di vari passaggi. In primo luogo, soprattutto per i metodi supervisionati, i campioni devono essere divisi in un *training set* (set di allenamento) e in un *test set* (set di verifica). Il training set viene usato per costruire il modello e ottimizzarlo, ovvero nella fase di calibrazione. Il test set viene utilizzato una volta che si ritiene che il modello sia ottimale e ha lo scopo di validarlo, ovvero di verificarne la bontà su un set di dati che non è stato usato nella fase di costruzione. Affinché questa verifica sia ottimale, è bene che il test set sia il più possibile indipendente dal training set. Questo significa le analisi dovrebbero essere fatte in giorni diversi e da operatori diversi su campioni con delle caratteristiche diverse (come origine, periodo, caratteristiche fisiche, etc.). Il test set può essere di due tipologie: interno ed esterno. Il primo è un sottoinsieme dei campioni analizzati per costruire il modello, e avrà delle caratteristiche analoghe al training set. Generalmente si utilizza il 70% dei campioni per la costruzione del modello e il restante 30% per la validazione [26]. Un test set esterno, invece, ha un fattore di variabilità aggiuntivo. Questa fonte di differenza può essere di vario tipo, in base al problema analitico che si considera. Lo scopo di questa validazione è quello di valutare la robustezza del modello che si è costruito su campioni nuovi in cui la variabilità rispetto al training set è massimizzata.

I metodi supervisionati si dividono in due gruppi: quelli che si propongono di predire risposte di tipo qualitativo, detti metodi di classificazione, e quelli che si propongono di predire risposte di tipo quantitativo, detti metodi di regressione.

I metodi di classificazione possono essere di due tipologie. La prima consente di assegnare un campione a una delle classi con cui si è costruito il modello. Il campione verrà assegnato alla classe a cui è più vicino per le sue caratteristiche. In questo tipo di modelli, il campione viene sempre assegnato a una e una sola classe, anche se non appartiene realmente a nessuna di esse [26]. L'altra tipologia è detta modellamento di classe, e in questo caso si modella una classe sola alla volta e lo spazio per questa classe è chiuso. Questo tipo di modelli permettono di definire se un campione appartiene alla classe utilizzata per costruire il modello

o meno [26]. Se sono presenti più classi, un campione può essere assegnato a una, nessuna o più di esse.

Per quanto riguarda la regressione multivariata, essa consiste nel cercare una relazione tra delle variabili predittive e una o più variabili risposta quantitative [26]. Quando si costruiscono dei modelli di regressione multivariata tipicamente si usa una calibrazione inversa. Essa consiste nel misurare la variabile indipendente, in modo da trovare la variabile dipendente. La relazione che lega queste variabili è:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_n X_n + \varepsilon \quad (1.6)$$

dove Y è la risposta che si vuole determinare (variabile dipendente), X sono le variabili indipendenti, β_x rappresenta i coefficienti del modello lineare e ε è l'errore legato alla determinazione di Y .

1.2.3.1 Analisi delle componenti principali

L'analisi delle componenti principali (PCA) rappresenta il metodo esplorativo più utilizzato in chemiometria. Il suo scopo è quello di visualizzare il comportamento dei campioni, in modo da riconoscere raggruppamenti e andamenti particolari, e di comprendere quali sono le variabili che determinano questi fenomeni. È una tecnica che consente di interpretare dataset di grandi dimensioni senza perdere l'informazione contenuta nei dati [26]. Questa tecnica esplorativa consente di analizzare dataset di grandi dimensioni perché è in grado di ridurre la loro grandezza. Compiere questa operazione senza perdere informazione è possibile grazie alla creazione di nuove variabili che prendono il nome di componenti principali.

L'informazione che viene utilizzata dalla PCA è rappresentata dalla variabilità contenuta all'interno dei dati stessi. Di conseguenza, è di fondamentale importanza analizzare come cambiano le variabili. Si analizzano quindi la varianza, che spiega il modo in cui varia una singola variabile tra i vari campioni, e la covarianza, che illustra come variano due variabili l'una rispetto all'altra. In particolare, valori di varianza grandi indicano che un dato parametro cambia molto nei campioni del dataset; valori di covarianza grandi, invece, indicano che le due variabili sono correlate tra di loro.

Varianza e covarianza sono usate per costruire la matrice di varianza-covarianza, una matrice quadrata che contiene la varianza di tutte le variabili lungo la diagonale e la covarianza

delle coppie di variabili nelle altre posizioni. La matrice che si costruisce è simmetrica [26], contiene l'informazione relativa alla variabilità dei dati e rappresenta fondamentalmente il primo calcolo della PCA. Affinché la variabilità derivata da ogni variabile sia comparabile è necessario pretrattare i dati prima di costruire la matrice. Qualora fossero necessari, quindi, si effettuano prima i pretrattamenti di riga, e poi si deve effettuare necessariamente un centraggio di colonna, in modo da rendere confrontabili variabili di natura diversa in termini di unità di misura e ordine di grandezza.[26].

La matrice di varianza e covarianza viene poi utilizzata per andare a costruire delle nuove variabili che vengono dedotte dalla relazione presente tra le variabili misurate. Sono chiamate comunemente componenti principali (PC) e hanno delle caratteristiche fondamentali.

In primo luogo, sono ortogonali tra di loro, e di conseguenza non sono correlate. Applicando la PCA, si parte quindi da variabili numerose e correlate tra di loro e se ne ottengono di nuove, ridotte in numero e non correlate. La covarianza tra le PC è infatti pari a zero. Le componenti principali hanno poi la caratteristica di essere delle combinazioni lineari delle variabili originali, ognuna delle quali contribuisce al calcolo di ciascuna componente, con un contributo diverso dato dal coefficiente che si usa per la combinazione lineare. Questi coefficienti consentono di interpretare l'informazione perché, sapendo la natura della combinazione lineare, si può risalire ai parametri originali. Questo permette di spiegare quali variabili sono responsabili di un certo fenomeno [26].

Un altro aspetto fondamentale è che le componenti principali spiegano una percentuale di varianza decrescente, e quindi la maggior parte della variabilità e di conseguenza dell'informazione sarà contenuta nelle prime componenti. Quindi, le proprietà di interesse possono essere compresse in poche nuove variabili, andando a semplificare notevolmente il sistema di partenza senza perdita dell'informazione [26].

Le componenti principali vengono calcolate tramite degli algoritmi. I due più utilizzati sono NIPALS [26] e SVD [26]. Il primo ricerca la direzione di massima varianza all'interno dei dati e presenta la PCA in forma geometrica; il secondo attua una decomposizione in autovalori e autovettori, rappresentando la PCA in maniera algebrica.

Uno dei vantaggi fondamentali della PCA è che consente di ricavare la relazione presente tra le componenti principali e le variabili originali, come precedentemente accennato. Questo facilita l'interpretabilità dei fenomeni. In particolare, questa relazione è data dal coseno dell'angolo presente tra la componente principale e la variabile originale. Il coseno dell'angolo,

definito *loading*, rappresenta il coefficiente che viene utilizzato per pesare ciascuna variabile nella combinazione lineare che va a costituire la componente principale [26].

I campioni all'interno del nuovo spazio creato dalle variabili latenti occuperanno delle posizioni diverse rispetto a quelle originali. La loro posizione viene definita dalla moltiplicazione matriciale tra il dataset e i loading. In questo modo si ottengono le nuove coordinate dei campioni, dette *scores* [26].

Le informazioni fornite dalla PCA vengono rappresentate graficamente da diversi grafici [26]:

- *Scree plot*: descrive la varianza spiegata dalle componenti principali in funzione del numero della componente. Consente di capire quali sono le componenti principali che contengono l'informazione.
- Grafico degli score: rappresenta i campioni nello spazio definito dalle componenti principali. Permette di riconoscere dei pattern nella distribuzione dei campioni.
- Grafico dei loading: è lo spazio delle variabili. Permette di vedere il contributo di ogni variabile nella determinazione delle PC. È il grafico che mantiene la relazione tra le variabili originali e le componenti principali.

1.2.3.2 Regressione ai minimi quadrati parziali

La regressione ai minimi quadrati parziali (PLS) è una tecnica di regressione lineare basata sulle variabili latenti [26]. Questo metodo si è sviluppato per risolvere i problemi dati da altri metodi di regressione, in particolare OLS (*ordinary least squares*, minimi quadrati ordinari) e MLR (*multiple linear regression*, regressione lineare multipla) [26]. Questi metodi infatti necessitano, per la costruzione del modello, di un numero di campioni almeno pari al numero delle variabili. Nel caso in cui si voglia fare un modello di regressione usando come matrice X dei dati spettrali, rispettare questa condizione risulta essere molto difficile, siccome gli spettri possono contenere anche migliaia di variabili, mentre raramente si hanno a disposizione migliaia di campioni. Inoltre, metodi come MLR (*multiple linear regression*, regressione lineare multipla) [26] non possono essere utilizzati nel caso in cui le variabili siano correlate, per problemi legati al calcolo matriciale.

Il metodo PLS utilizza come algoritmo NIPALS [26]. Se nel caso della PCA questo algoritmo cerca la direzione di massima varianza, nel caso di PLS va a ricercare la direzione di massima covarianza con la variabile risposta da predire (Y) e la matrice dei dati (X). In questo modo, la prima variabile latente che si costruisce spiega la massima informazione correlata con la risposta. Le variabili latenti successive, ortogonali, spiegheranno la covarianza residua.

La costruzione del modello consta, quindi, di due fasi:

1. Proiezione degli oggetti nello spazio delle variabili latenti.
2. Predizione della variabile dipendente dagli scores dei campioni.

Questo porterebbe a una complicazione dell'interpretabilità del modello e del contributo che ogni variabile indipendente ha su di esso perché in questa forma di PLS, definita aperta, la predizione non è effettuata in funzione delle variabili originali, ma in funzione di quelle latenti. Per questo motivo si usa una forma di PLS che viene definita forma chiusa [26]. In questo caso, il modello matematico viene espresso in funzione dei predittori originali rendendo evidente il contributo che ogni variabile originale ha sulla costruzione del modello.

Il contributo delle variabili originarie sulla costruzione del modello può essere rappresentato graficamente. A questo scopo si usano i vettori di regressione. Per ogni variabile originaria si ha un coefficiente che spiega il suo contributo su ogni variabile latente; questo contributo viene inoltre pesato in funzione della covarianza spiegata dalla variabile latente stessa [26]. Un'altra rappresentazione grafica che si usa per identificare il peso delle variabili sul modello è quella dei VIP (*variable importance in the projection*: importanza della variabile nella proiezione). In questo caso si vanno a identificare le variabili che, superando un valore soglia, influenzano in maniera rilevante il modello; se, invece, questo valore non viene superato il parametro viene considerato non rilevante. [29].

La costruzione e validazione di un modello di questo tipo si articola in diverse fasi. In primo luogo, si ha la necessità di dividere i campioni di cui si dispone in due gruppi, il training set e quello di test, come precedentemente descritto. Nella regressione è fondamentale che il training set sia rappresentativo di tutto l'intervallo all'interno del quale si ha intenzione di operare; questo significa che dovrà contenere i campioni che ne rappresentano gli estremi.

Un'altra operazione che viene effettuata per andare a validare il modello è la cross validazione, che viene spesso usata come validazione intermedia. Viene sfruttata per andare ad ottimizzare il modello, sia in termini di pretrattamenti che di scelta del numero di variabili

latenti significative. In questo caso alcuni campioni vengono assegnati, a turno, al training set e altri al set di validazione. Si compiono vari cicli, e per ognuno di essi si costruisce un modello con il training set e lo si applica al set di validazione. Il modello che si ottiene sarà il modello ottenuto su tutti i dati di training, ma l'errore in cross validazione deriverà dalla media degli errori ottenuti dai diversi set di validazione durante la cross validazione [26].

Gli schemi di cross validazione possono essere di vario tipo:

- A blocchi contigui: si dividono i campioni in blocchi di cancellazione. In ogni ciclo, un blocco viene lasciato come set per la validazione, mentre gli altri fanno da training set. Ad ogni iterazione si cambia il blocco di cancellazione che fa parte del set di validazione. Il numero di cicli è pari al numero di blocchi in cui si dividono i campioni [30].
- *Venetian blinds* (a tenda veneziana): in questo caso i blocchi di cancellazione sono costituiti, per ogni ciclo, da campioni non adiacenti tra di loro. In questo modo si riesce a tenere conto della presenza di eventuali effetti strumentali che si possono essere verificati nel proseguire delle analisi [31].
- *Leave one out* (a singolo campione): in ogni ciclo viene lasciato fuori un solo campione che viene usato per la validazione. I modelli che si costruiscono ogni volta sono poco perturbati. La validazione non è ottimale. Questa tecnica si usa solo nel caso in cui il numero di campioni sia estremamente basso [30].
- Random (casuale): i campioni vengono assegnati casualmente al training set e al set di validazione [31].
- *Repeated random* (causale ripetuta): consiste nel ripetere la cross validazione random un numero di volte definito. Il modello che si ottiene sarà derivato dalla media pesata dei singoli modelli [30]. Questo approccio si applica per rendere le performance in cross validazione meno dipendenti possibile dai sottogruppi casuali di dati selezionati per il set di validazione.

Un aspetto fondamentale nella costruzione di questi modelli è rappresentato dalla scelta del numero di variabili latenti significative. Se nella PCA si possono osservare anche componenti principali con percentuale di varianza spiegata molto bassa, nel caso di PLS la scelta del numero delle variabili latenti è cruciale per la costruzione del modello. Tipicamente, si utilizzano nella costruzione del modello tutte le variabili latenti fino a quando si ha un minimo dell'errore in cross validazione. Generalmente, la prima variabile latente che ha la radice

quadrata dell'errore quadratico medio in cross validazione (RMSECV, *root mean square error of cross validation*) maggiore rispetto a quella precedente non deve essere inserita nel modello [26].

La scelta di un numero maggiore di variabili latenti rispetto a quello ottimale porta a ottenere dei modelli che vengono definiti “overfittati”. Infatti, scegliendo più variabili latenti del necessario si inserisce nella costruzione del modello non solo l'informazione di interesse, ma anche il rumore. Questo fa sì che il modello che si ottiene in fase di calibrazione sia all'apparenza molto buono ma, quando si va alla fase di validazione, il modello si rivela poco robusto e la validazione ha un errore molto grande [26]. La spiegazione di questo fenomeno sta proprio nel fatto che utilizzando troppe variabili latenti si va a costruire il modello alla perfezione sui dati che vengono usati per la fase di calibrazione, ma esso, essendo adattato perfettamente su quei dati, non sarà in grado di predire in maniera soddisfacente campioni ignoti, in cui il rumore casuale sarà tendenzialmente diverso.

Una volta che si è costruito il modello, questo deve andare incontro a una fase di validazione esterna, che valuta le abilità predittive del modello su un dataset nuovo, ovvero su un test set, sia esso interno o esterno. Anche in questo caso la performance viene valutata monitorando l'errore, e in particolare si considera la radice quadrata dell'errore quadratico medio in predizione (RMSEP, *root mean square error in prediction*). Valori bassi di questo parametro indicano la bontà del modello.

1.3 PENETRAZIONE DELLA RADIAZIONE NIR

Con le tecniche analitiche di imaging iperspettrale si è arrivati a ottenere informazioni spaziali, lungo due dimensioni, e informazioni spettroscopiche. Questo consente di ottenere delle risposte molto importanti in diversi campi analitici.

Tuttavia, negli ultimi anni sta iniziando a ricevere più attenzione un altro aspetto, che riguarda la capacità della radiazione NIR di penetrare oltre gli strati superficiali di un campione [32]. La possibilità di andare in profondità è collegata all'energia fornita dai fotoni in questa lunghezza d'onda. Il fatto che la regione NIR sia una regione a basso assorbimento, infatti, fa sì che essa possa penetrare all'interno di alcune tipologie di campioni, permettendo, da un punto di vista teorico, di accedere anche a informazioni al di sotto della superficie [32].

La possibilità di combinare questo aspetto del NIR con tecniche HSI permetterebbe di ottenere moltissime informazioni attraverso una sola analisi. Infatti, questo consentirebbe di ricavare da una sola analisi, con un singolo strumento, informazioni spaziali e informazioni chimiche relative sia agli strati superficiali sia a quelli più profondi [32].

Da ciò risulta evidente il motivo per cui, negli ultimi anni, si sta sviluppando un maggiore interesse per approfondire questi aspetti. La possibilità di analizzare campioni stratificati con metodiche analitiche rapide e non distruttive potrebbe essere un vantaggio in molti ambiti. Tuttavia, lo sfruttamento di questa proprietà della radiazione NIR presenta delle criticità legate a una difficoltà nel definire la capacità massima di penetrazione della radiazione [32]. Bisogna inoltre considerare che la possibilità della radiazione di penetrare a una certa profondità non corrisponde necessariamente alla sua capacità di restituire un segnale analitico utile per la determinazione di una certa proprietà. Affinché si possa ottenere informazione relativa agli strati più profondi, oltre a dover penetrare fino a essi, la radiazione diffusa deve anche avere energia residua sufficiente per raggiungere il rilevatore.

1.3.1 Fattori che influenzano la penetrazione

Per riuscire a sfruttare la penetrazione della radiazione NIR è necessario comprendere quali sono i fattori che la influenzano. In questo modo sarà poi possibile cercare di ottimizzare questi parametri per ottenere le informazioni di interesse per l'analisi.

In primo luogo, si deve considerare che la capacità di penetrazione è legata a un basso assorbimento della radiazione. Di conseguenza materiali caratterizzati da gruppi chimici che assorbono molto la radiazione diminuiranno la possibilità di penetrazione della radiazione [32]. Solitamente i materiali biologici sono quelli che determinano una penetrazione minore perché presentano legami che assorbono la radiazione a specifiche lunghezze d'onda. Bisogna considerare che la stessa tipologia di campione, se si trova in condizioni differenti, può determinare una penetrazione diversa. Se, ad esempio, si analizzano campioni analoghi, ma con un contenuto di umidità differente, questi avranno delle penetrazioni differenti. In particolare, in quelli più umidi si avrà penetrazione minore in quanto la maggiore presenza di gruppi O-H determina un maggiore assorbimento della radiazione e, di conseguenza, minore penetrazione [32].

Le caratteristiche chimiche del campione sono quindi importanti per comprendere la profondità di penetrazione, ma non sono l'unico fattore rilevante. Infatti, anche degli aspetti fisici, legati alla geometria del materiale, giocano un ruolo importante in questo fenomeno [32].

Tra i fenomeni fisici che influenzano la profondità di penetrazione abbiamo lo *scattering*. Esso consiste in una diffusione della radiazione in tutte le direzioni. Maggiore è la radiazione che viene deflessa e minore sarà la radiazione che riesce a procedere verso gli strati inferiori del campione. Questo fenomeno è fortemente influenzato dalle caratteristiche fisiche del campione: dimensione, densità, presenza di una superficie liscia o ruvida, organizzazione strutturale interna [32]. L'ultimo aspetto è particolarmente rilevante perché, se le strutture che compongono un campione sono molto più piccole della lunghezza d'onda della radiazione incidente, si avrà lo scattering di Rayleigh [32], in cui la diffusione avviene in maniera analoga in tutte le direzioni, mentre, se esse sono di dimensioni maggiori o uguali alla lunghezza d'onda, lo scattering prevalente sarà quello di Mie [32], in cui la diffusione della luce cambia nelle diverse direzioni in funzione dell'angolo, come è rappresentato nella Figura 1-5. Questo determina anche una perdita energetica della radiazione e quindi diminuisce la possibilità che questa penetri. Questa perdita inoltre è progressiva perché tutte le strutture interne che incontrano la radiazione possono agire da centri di scattering, diminuendo ulteriormente l'energia della radiazione. Gli effetti dovuti allo scattering di Mie sono molto rilevanti quando si analizzano dei materiali fortemente eterogenei che rendono il percorso della radiazione più complesso e di difficile interpretazione [32].

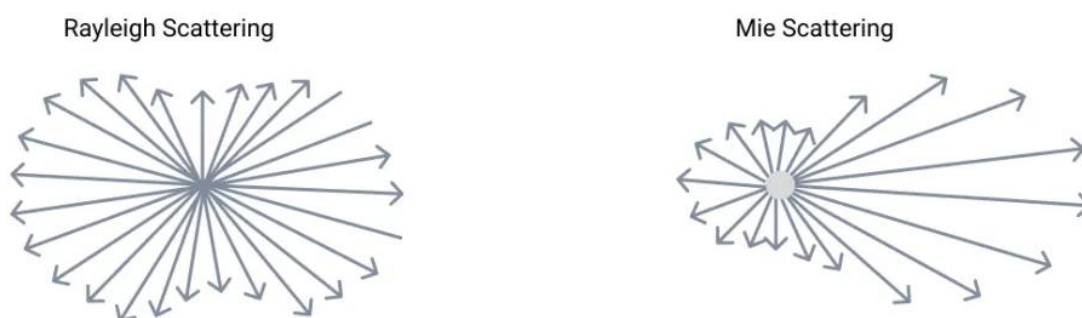


Figura 1-5 Scattering di Rayleigh e Scattering di Mie [33]

Un aspetto strettamente legato ai fenomeni di scattering è la geometria del campione. Strutture con geometrie complesse e superfici ruvide possono comportarsi da centri di scattering

che diffondo la radiazione in direzioni diverse diminuendone la profondità di penetrazione. Questi fenomeni sono accentuati in campioni con densità molto diversa. Questo causa ottiche diverse in punti diversi del campione, che influiscono sul percorso della radiazione all'interno dello stesso. Questi effetti possono essere mitigati impostando i parametri dello strumento in modo da ottimizzare il recupero dell'informazione di interesse. Regolare, ad esempio, l'angolo di incidenza della radiazione o la posizione dei campioni può andare a mitigare gli effetti di scattering [32].

Il fenomeno dello scattering assume caratteristiche diverse in funzione delle dimensioni delle particelle e della lunghezza d'onda della radiazione incidente. In particolare, a lunghezze d'onda maggiori si può avere una riduzione dello scattering di Mie. Questo implica una maggiore possibilità di penetrazione della radiazione [32]. Tuttavia, bisogna anche notare che a lunghezze d'onda maggiori si hanno anche gli assorbimenti di gruppi frequentemente rappresentati nei campioni. Di conseguenza la maggiore penetrazione, dovuta alla diminuzione dello scattering, può essere ridotta notevolmente dai maggiori fenomeni di assorbimento che avvengono in queste zone dello spettro. Una penetrazione maggiore sarà caratteristica di zone spettrali con lunghezza d'onda elevata e basso assorbimento [32]. Anche questo aspetto risulta essere strettamente legato con la struttura chimica e fisica del campione.

1.3.2 Stato dell'arte

L'utilità applicativa della penetrazione della radiazione NIR è testimoniata dai diversi campi tra cui spaziano gli studi a riguardo. L'interesse è registrato dalla chimica alla fisica, dalle applicazioni biochimiche a quelle ingegneristiche [32]. La profondità di penetrazione della radiazione viene intesa come spessore del campione raggiunto dalla radiazione che è in grado di determinare, una volta raggiunto il rilevatore, un segnale analitico. Quello che si ricerca maggiormente è quindi una profondità "utile" piuttosto che andare a esplorare il limite fisico della penetrazione.

Gli studi fatti per rilevare la penetrazione della radiazione confermano come alcune delle proprietà descritte nel paragrafo 1.3.1 possono influenzare la profondità che viene ad essere raggiunta.

In campo biomedico gli studi di Svaasand ed Ellingsen (1983) [34] hanno evidenziato che la penetrazione della radiazione NIR nei campioni di tessuto cerebrale è diversa tra adulti

e neonati. Questa differenza è riconducibile a un diverso livello di mielinizzazione e a un diverso contenuto di acqua, proprietà che, come precedentemente discusso, possono influenzare l'assorbimento e lo scattering della radiazione.

La possibilità di sfruttare la radiazione per conoscere le stratificazioni di un campione è molto rilevante nel campo dei beni culturali. Alcuni studi [35] hanno permesso di ottenere delle informazioni relative agli strati preparatori dei dipinti. Questo obiettivo è stato raggiunto utilizzando la radiazione NIR, cosa che non era riuscita con la spettroscopia MIR. Da un punto di vista spettroscopico ciò è molto rilevante perché conferma che la radiazione NIR è in grado di penetrare e di conseguenza questo influenzerà il segnale spettrale che si ottiene. Inoltre, se si considerano le tecniche di HSI, si può pensare di sviluppare metodiche che consentano di correlare queste informazioni anche con una localizzazione spaziale, aspetto che, fino ad ora, non è stato sviluppato in quanto la maggior parte degli studi sulla penetrazione della radiazione NIR sono stati effettuati con strumenti puntuali.

L'approccio a studi di questo tipo era, nei primi periodi, prevalentemente spettrale. Ci si basava sull'osservare gli spettri e identificare in essi delle differenze [32]. Poi si sono utilizzati dei modelli che sfruttavano delle forme modificate della legge di Lambert-Beer [36]. Attualmente gli studi in questo ambito si basano su una analisi multivariata che consente di elaborare i segnali spettrali [32]. Negli studi effettuati è stata evidenziata una tendenza a utilizzare certe tipologie di pretrattamenti. Tra quelli più utilizzati rientrano la trasformata SNV, la derivata di Savitzky-Golay, MSC (pretrattamento molto utilizzato per rimuovere effetti fisici, tra cui lo scattering) [32].

Attualmente gli studi su questo argomento utilizzano, per analizzare le informazioni relative alla profondità di penetrazione, metodi multivariati sia esplorativi, primo su tutti la PCA [37], che metodi supervisionati [37]. Tra questi sono utilizzati i modelli di regressione e in particolar modo PLS [37]. Tuttavia, si è osservato che il fenomeno della penetrazione presenta un comportamento non lineare [38], di conseguenza si sono riscontrate delle difficoltà nell'utilizzo di metodi di regressione lineare come PLS. Per questa ragione sarebbero necessari anche dei modelli di regressione non lineari, ma al momento in letteratura non sono presenti molti studi a riguardo.

2 SCOPO

Lo scopo di questa tesi è studiare la penetrazione della radiazione NIR, associata all'imaging iperspettrale, attraverso la materia. Tipicamente la spettroscopia NIR, soprattutto nell'ambito dell'imaging, è considerata una tecnica di superficie. Tuttavia, negli ultimi anni stanno iniziando ad essere presentati alcuni studi, in ambito biomedico, farmaceutico e dei beni culturali, che evidenziano la capacità di penetrazione della radiazione NIR.

Tali studi evidenziano l'influenza di parametri chimici, come la composizione del materiale, e fisici, come la geometria delle strutture interne al campione, sulla capacità di penetrazione della radiazione.

Il lavoro qui proposto si pone l'obiettivo di studiare in maniera sistematica il fenomeno. Per raggiungere tale obiettivo si sono utilizzati dei campioni polimerici stratificati a diverso spessore di acido polilattico (PLA) e polietilene tereftalato glicole (PETG), contraddistinti da diverse geometrie di riempimento. I campioni sono caratterizzati da un'altezza totale di 3 o 5 cm, con spessore maggiore rispetto a quella che si ritiene essere la capacità di penetrazione della radiazione NIR. In questo modo si è portato al limite questo concetto, con l'obiettivo di studiare i meccanismi di penetrazione e, idealmente, identificare la profondità limite raggiunta nei materiali considerati.

In particolare in questo lavoro viene presentato nel dettaglio un approccio quantitativo basato sulla regressione ai minimi quadrati parziali (PLS), un metodo di regressione multivariato. PLS, in questo caso studio, non è stato utilizzato a fini predittivi, ma per verificare la profondità raggiunta dalla radiazione. Questo approccio si basa sull'assunto che fino a quando la radiazione è in grado di penetrare attraverso la materia, raggiungendo lo strato di polimero inferiore, si otterranno dei profili spettrali diversi che determineranno la predizione a diversi valori di Y (altezza del polimero posto in alto) nel modello PLS che si è costruito. Quando, invece, si raggiungerà la profondità massima di penetrazione i profili spettrali che si otterranno saranno analoghi e di conseguenza verranno predetti allo stesso valore di Y, rappresentando di fatto una rottura del modello, a indicare il raggiungimento della penetrazione massima.

Lo scopo ultimo è quindi stato quello di definire le profondità raggiunte dalla radiazione e di comprendere l'effetto esercitato da diversi pretrattamenti di riga, in modo da definire il tipo di informazione, chimica o fisica, che essi consentono di ricavare.

L'esito positivo di questo lavoro permetterebbe di definire la radiazione NIR non più superficiale, ma in grado di penetrare in profondità, cosa che ne aprirebbe l'utilizzo nell'analisi di campioni complessi e stratificati. Questo aspetto potrebbe rappresentare un punto di svolta fondamentale in diversi campi, dal farmaceutico al controllo qualità, perché, sviluppato un metodo chemiometrico appropriato, si avrebbe la possibilità di analizzare gli strati profondi di campioni complessi con una tecnica rapida e non distruttiva.

3 MATERIALI E METODI

3.1 CAMPIONI

Per andare a studiare la penetrazione della radiazione NIR si sono utilizzati dei campioni polimerici stratificati che sono stati prodotti tramite stampa 3D. I campioni sono costituiti da due diversi materiali polimerici, PLA e PETG, e sono di due diverse geometrie: dei parallelepipedi a base quadrata, qui chiamati regoli, e dei cilindri. Una caratteristica importante dei campioni è il loro riempimento. Tutti sono caratterizzati dall'avere un riempimento con densità del 100%. Questa caratteristica viene ottenuta tramite il controllo della geometria di riempimento, ovvero la struttura geometrica interna che va a riempire lo spazio alla densità desiderata. A parità di densità totale finale, in questo lavoro si sono utilizzati cinque tipi di *infill* diversi, alcuni caratteristici dei regoli e altri dei cilindri. Si è scelto di utilizzare riempimenti differenti tra regoli e cilindri perché la modalità di deposizione del materiale durante la stampa determina la struttura interna dei provini, e l'uso di strutture interne diverse permette di studiare se queste differenze possano determinare uno scattering della radiazione differente, andando ad influenzare la profondità di penetrazione del NIR.

I campioni sono divisi in due gruppi in base alla loro altezza totale, che può essere o di 3 cm o di 5 cm. In questa tesi gli oggetti ad altezza diversa verranno analizzati separatamente, concentrandosi prima su quelli alti 3 cm e poi su quelli con altezza pari a 5 cm.

Il set di campioni da 3 cm è formato in totale da 20 oggetti, di cui 16 regoli e 4 cilindri. In particolare, per quanto riguarda i regoli, il primo provino è costituito interamente di PETG. I campioni successivi presentano un contenuto crescente di PLA, aggiungendone di volta in volta uno strato pari a 2 mm, e riducendo dello stesso valore l'altezza di PETG, fino ad ottenere un campione contenente solo PLA. La geometria di stampa è uguale per tutti i regoli ed è di tipologia rettilinea. I cilindri, invece, presentano rapporti di altezza dei due polimeri diversi rispetto ai regoli, siccome non presentano un passo costante nel passaggio tra un campione e l'altro in termini di cambiamento dello spessore degli strati, in quanto in alcuni casi il passo è pari a 4 mm e in altri è pari a 6 mm. Le altezze del singolo materiale presentano dei valori dispari, che non sono rappresentati all'interno dei regoli. Inoltre, sono caratterizzati dall'avere due tipologie di riempimento diverso. Le caratteristiche dei cilindri sono schematizzate in Tabella 1.

I campioni da 5 cm sono in totale 34, di cui 26 sono regoli e 8 sono cilindri. I regoli presentano caratteristiche analoghe ai campioni da 3 cm; sono quindi presenti due campioni “puri” e campioni intermedi con un passo pari a 2 mm ottenuti usando una geometria di riempimento rettilinea. Anche in questo caso, i cilindri presentano altezze dei due polimeri diverse rispetto ai regoli e sono caratterizzati dall’avere 4 tipologie di riempimento diverso. Le loro caratteristiche sono riportate in Tabella 2.

Tabella 3-1: Altezze e geometria di riempimento dei cilindri da 3 cm. Ulteriori informazioni circa le notazioni utilizzate nella descrizione dell’altezza si trovano nel paragrafo 3.2.2

Altezza	Geometria di riempimento
7-23 mm PETG-PLA	Octagram Spiral
13-17 mm PETG-PLA	Archimidean Chords
17-13 mm PETG-PLA	Archimidean Chords
23-7 mm PETG-PLA	Octagram Spiral

Tabella 3-2: Altezza e geometria di riempimento dei cilindri da 5 cm. Ulteriori informazioni circa le notazioni utilizzate nella descrizione dell’altezza si trovano nel paragrafo 3.2.2

Altezza	Geometria di riempimento
5-45 mm PETG-PLA	Hillbert Curve
11-39 mm PETG-PLA	Concentric
17-33 mm PETG-PLA	Octagram Spiral
23-27 mm PETG-PLA	Archimidean Chords
27-23 mm PETG-PLA	Archimidean Chords
33-17 mm PETG-PLA	Octagram Spiral
39-11 mm PETG-PLA	Concentric
45-5 mm PETG-PLA	Hillbert Curve

Ulteriori informazioni relative alla struttura dei campioni si possono trovare nella tesi di Bassoli Rebecca.

3.2 ACQUISIZIONE DELLE IMMAGINI

3.2.1 Strumento e specifiche di acquisizione

L'analisi dei campioni è stata effettuata tramite l'utilizzo della camera iperspettrale SWIR3 di Specim (Spectral Imaging Ltd., Oulu; Finlandia). L'acquisizione delle immagini è stata effettuata utilizzando la lente OLES15. La camera è stata montata sul LabScanner 40x20 di Specim (Spectral Imaging Ltd., Oulu; Finlandia). Le analisi sono state effettuate con due set di tre luci alogene (Osram, Styria, Austria) ciascuno. I due gruppi di luci sono stati posizionati, con inclinazione di circa 45°, l'uno di fronte all'altro, in modo da andare ad illuminare la zona in cui si sarebbero disposti i campioni. Lo strumento opera nel range di lunghezze d'onda dai 960 – 2500 nm e utilizza un sensore MCT (mercurio, cadmio e tellurio) in grado di registrare 270 punti.

Il software utilizzato per la gestione delle acquisizioni è il LUMO Scanner V. 2.6 di Specim (Spectral Imaging Ltd., Oulu; Finlandia). Questo permette, tramite tre interfacce (*Setup*, *Adjust*, *Record*), di gestire tutte le operazioni da svolgere per effettuare l'analisi, dalla connessione delle varie componenti strumentali, al setting dei parametri di acquisizione, fino alla misura stessa.

Le immagini sono state acquisite in riflettanza, con i campioni posizionati sul lab scanner. La camera iperspettrale è stata disposta ad una altezza di 22,8 cm rispetto al carrello del lab scanner. Le acquisizioni sono state fatte con un *frame rate*, ovvero il numero di immagini che sono acquisite dalla camera nell'unità di tempo, pari a 70Hz, e un *exposure time*, ovvero il tempo per cui il sensore registra l'intensità della radiazione per ogni lunghezza d'onda, di 6 ms. La velocità del carrello è stata impostata a 28,19 mm/s.

3.2.2 Esecuzione dell'analisi

Per l'esecuzione dell'analisi, dopo l'accensione dello strumento, è necessario aspettare un'ora in modo tale da consentire allo strumento di raggiungere la temperatura di -70°C che consente di effettuare analisi in cui la deriva negli spettri è minimizzata. Trascorso il tempo necessario si apre il software e tramite l'interfaccia "*Setup*" si connettono il Lab Scanner e la camera che si desidera utilizzare.

Passando all'interfaccia "*Adjust*" si ottimizzano i parametri necessari all'analisi. I valori di *frame rate* e di *exposure time* si impostano manualmente in base alle necessità dell'acquisizione. Successivamente si procede a registrare il nero tramite l'opzione "*run dark*"; questa operazione consiste nella chiusura dell'otturatore della camera in modo tale da registrare il rumore elettrico di fondo. Successivamente, accese le luci, si registra il bianco utilizzando come riferimento lo Spectralon, un materiale fluoropolimerico con la più alta riflettanza diffusa conosciuta [6]. Attivando "*Run White*", facendo attenzione che lo Spectralon sia nella posizione impostata, si ha lo spostamento del carrello in modo da fare la lettura del bianco. L'esecuzione del "bianco" e del "nero", a questo punto operativo serve a verificare il buon funzionamento dello strumento, prima di procedere con le altre operazioni. Una scansione di "bianco" e una di "nero" viene poi catturata per ogni immagine raccolta durante la fase di analisi dei campioni.

Successivamente si procede con la messa a fuoco della lente. L'operazione si effettua girando l'anello più esterno della lente, in modo da cambiare millimetricamente la distanza focale e ottimizzare la risoluzione dello strumento. In queste analisi il fuoco è stato impostato in modo tale da ottenere la massima risoluzione sul fondo del carrello. L'ottimizzazione del fuoco viene fatta usando un oggetto con dettagli fini, come un'etichetta di carta con delle piccole scritte. La lente viene regolata manualmente fino a quando non si riesce a creare il massimo contrasto, ovvero fino a quando non si ottengono dei picchi molto netti nel quadrante che rappresenta i pixel. Macroscopicamente, questo significa che facendo scorrere il carrello è possibile visualizzare i dettagli fini dell'oggetto usato per l'ottimizzazione, ovvero, nel nostro caso, è possibile leggere le scritte sull'etichetta.

Dopo aver messo a fuoco la lente si procede a ottimizzare la velocità del carrello, in modo tale da riprodurre fedelmente la forma dell'oggetto. A questo scopo si utilizza un cartoncino bianco raffigurante un cerchio nero e lo si pone sul lab scanner. Utilizzando l'impostazione "*Target Start*" all'interno della voce "*Sensor 1 Position*", si definisce l'inizio del cerchio che si usa come riferimento. Muovendo il carrello si definisce anche la sua fine, andando a impostare la sua posizione alla voce "*Target Stop*". Definita la posizione in cui si trova la circonferenza di riferimento, si va nella sezione "*Scanning Speed Calculation*" e cliccando "*Run*" si ha l'ottimizzazione di questo parametro. Il software, dopo aver registrato l'immagine di prova, e conoscendo la forma che essa deve avere, propone una nuova velocità del carrello che renda l'immagine il più fedele possibile all'originale. Questa nuova velocità può essere impostata selezionando la voce "scanner" che compare all'interfaccia. Al fine di ottimizzare in maniera appropriata questo parametro, in questa sperimentazione si è ripetuta

l'operazione tre volte, per verificare che i valori di velocità che vengono forniti dal software non siano significativamente diversi.

Ottimizzati i parametri dell'analisi, si sono disposti i campioni sul lab scanner. Sono stati posizionati in maniera obliqua l'uno rispetto all'altro e leggermente sfalsati come è schematizzato in Figura 3-1. Ciascun campione è stato affiancato da un'etichetta che indicava l'altezza dello strato di polimero presente nella parte superiore del campione. Inoltre, si è posta attenzione al fatto che i campioni non si facessero ombra a vicenda, in quanto questo determina una minore quantità di luce che raggiunge il campione, influenzando in maniera negativa la possibilità di penetrazione della radiazione. Il numero di campioni acquisiti con ogni immagine è stato variabile in base all'altezza del campione. Per il set alto 3 cm i regoli sono stati acquisiti tre alla volta, fatta eccezione per due immagini che contenevano due soli campioni, mentre i cilindri si sono acquisiti tutti e quattro con una sola immagine. Per il set alto 5 cm si sono invece acquisiti due campioni per volta, sia per i regoli che per i cilindri. Questa necessità è nata dal fatto che i campioni da 5 cm, quando si trovano più distanti dalla lente, restituiscono un'immagine deformata, dato che la lente, in questo caso, non inquadra solo la faccia superiore del campione, ma anche i suoi lati, andando quindi ad alterare la qualità dell'immagine e della successiva elaborazione dei dati.

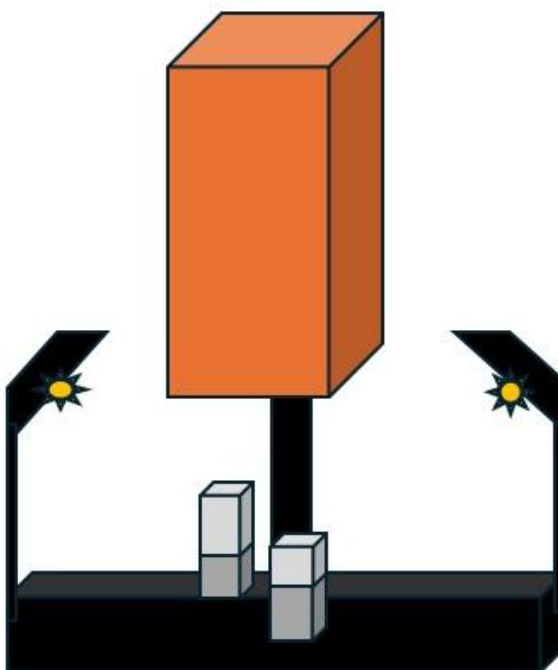


Figura 3-1: Set up delle acquisizioni

Una volta disposti tutti i campioni per un'immagine, si va a impostare un nuovo "Target Start" e "Target Stop" in modo da includere tutti gli oggetti che si desidera registrare all'interno dello spazio di misura dell'immagine.

A questo punto, si procede all'acquisizione. Andando nell'interfaccia "Capture" si ha la possibilità di effettuare un'anteprima per verificare che tutti i campioni siano inclusi nello spazio che si andrà a registrare, verificando così di aver impostato correttamente il target start e il target stop. Successivamente, si procede all'acquisizione tramite il comando "Record". A questo punto, lo strumento procede alla misura del nero, poi del bianco, e infine cattura l'immagine dei campioni veri e propri.

Inoltre, le acquisizioni sono state effettuate ponendo prima un polimero sul fondo del carrello e successivamente l'altro, in modo tale che la radiazione penetrasse in un caso prima attraverso il PETG e nell'altro prima attraverso il PLA. Nelle analisi successive si useranno due notazioni:

1. PETG-PLA: indica che il polimero in alto è PETG e quello in basso PLA.
2. PLA-PETG: indica che il polimero in alto è PLA e quello in basso PETG.

Queste notazioni saranno accompagnate dall'altezza di entrambi i polimeri seguendo lo stesso ordine.

3.3 TRATTAMENTO DEI DATI

Le acquisizioni sono state trattate utilizzando dei codici sviluppati in-house in ambiente MATLAB®, versione R2023b (MathWorks, Natick, Massachusetts, USA). Per effettuare le analisi non supervisionate, ovvero la PCA, e quelle supervisionate, ovvero la regressione PLS, si è utilizzato il software PLS_toolbox 9.5 (Eigenvector Research Inc., Manson, WA, USA).

L'organizzazione e il trattamento chemiometrico dei dati viene illustrato nei seguenti paragrafi.

3.3.1 Dataset

Le analisi sono state effettuate considerando i 4 set di campioni di cui si dispone:

- Campioni da 3 cm PETG-PLA.
- Campioni da 3 cm PLA-PETG.
- Campioni da 5 cm PETG-PLA.
- Campioni da 5 cm PLA-PETG.

Al fine dello svolgimento dell'analisi ciascun set è stato diviso in tre dataset diversi:

- Training set.
- Test set interno.
- Test set esterno.

Per tutte le tipologie di campioni, il training set è composto da tutti i regoli per ciascuna altezza. Il test set interno è invece costituito da una porzione dei regoli; in particolare, per i campioni da 3 cm esso contiene 8 regoli, mentre per quelli da 5 cm ne contiene 10. I campioni che formano il test set interno sono stati selezionati casualmente, ponendo attenzione, solo in un secondo momento, che fossero comunque rappresentativi dell'intero range e che non fossero concentrati in un intervallo ristretto. I campioni utilizzati per il test set interno da 3 cm e da 5 cm sono riportati rispettivamente in Tabella 3 e in Tabella 4, rispettivamente. Si noti che i campioni selezionati per la stessa altezza sono gli stessi, sia per le analisi PLA-PETG che per quelle PETG-PLA. Questo implica che l'altezza del polimero posto in alto sarà differente in base al fatto che le immagini vengano acquisite con PETG sopra o con PLA sopra.

Tabella 3-3: Altezze dei regoli contenuti nel test set interno per i campioni da 3 cm

4-26 mm PETG-PLA	20-10 mm PETG-PLA
12-18 mm PETG-PLA	28-2 mm PETG-PLA
8-22 mm PETG-PLA	26-4 mm PETG-PLA
22-8 mm PETG-PLA	2-28 mm PETG-PLA

Tabella 3-4: Altezze dei regoli contenuti nel test set interno per i campioni da 5 cm

12-38 mm PETG-PLA	26-24 mm PETG-PLA
48-2 mm PETG-PLA	42-8 mm PETG-PLA
36-16 mm PETG-PLA	6-44 mm PETG-PLA
18-32 mm PETG-PLA	38-12 mm PETG-PLA
8-42 mm PETG-PLA	30-2 mm PETG-PLA

Al test set esterno sono stati assegnati tutti i campioni cilindrici dell'altezza di riferimento. I campioni da 3 cm hanno, quindi, un test set esterno costituito da 4 campioni, mentre quelli da 5 cm hanno un test set esterno da 8 campioni. I cilindri si sono utilizzati come mezzo di validazione esterna perché, rispetto ai regoli, presentano una variabilità aggiunta rappresentata dal diverso tipo di riempimento. Questo ha consentito di valutare la validità dei modelli utilizzati anche in campioni in cui lo scattering della radiazione, visto il diverso tipo di geometria di riempimento, potrebbe essere diverso.

3.3.2 Identificazione delle regioni di interesse e analisi object based

Le immagini acquisite tramite HSI necessitano di un trattamento apposito al fine di estrarre da esse le informazioni utili alle analisi da svolgere. Queste operazioni sono rese necessarie dal fatto che le immagini sono state acquisite disponendo i campioni sul carrello e ciascuno di essi era affiancato da un'etichetta. Di conseguenza alcuni dei pixel acquisiti non portano informazione relativa ai campioni, ma rappresentavano gli spettri della carta delle etichette e del carrello. Si sono quindi rese necessarie delle operazioni di mascheramento, che è stato fatto costruendo un modello di PCA sui dati e valutando visivamente quale delle componenti principali portava l'informazione relativa ai campioni, ovvero in quale componente principale i campioni assumessero i valori di scores il più possibile diversi da quelli delle etichette e del carrello. Nel caso presentato, questa condizione si è verificata nella PC2 che ha permesso di differenziare i campioni sia dal lab scanner che dalle etichette di carta, come si può notare dal colore diverso dei regoli rispetto al resto in Figura 3-2. Individuata la PC da utilizzare, si sono costruite le maschere vere e proprie, scegliendo un valore soglia che permettesse di dividere le immagini trattenendo solamente l'informazione relativa ai campioni.

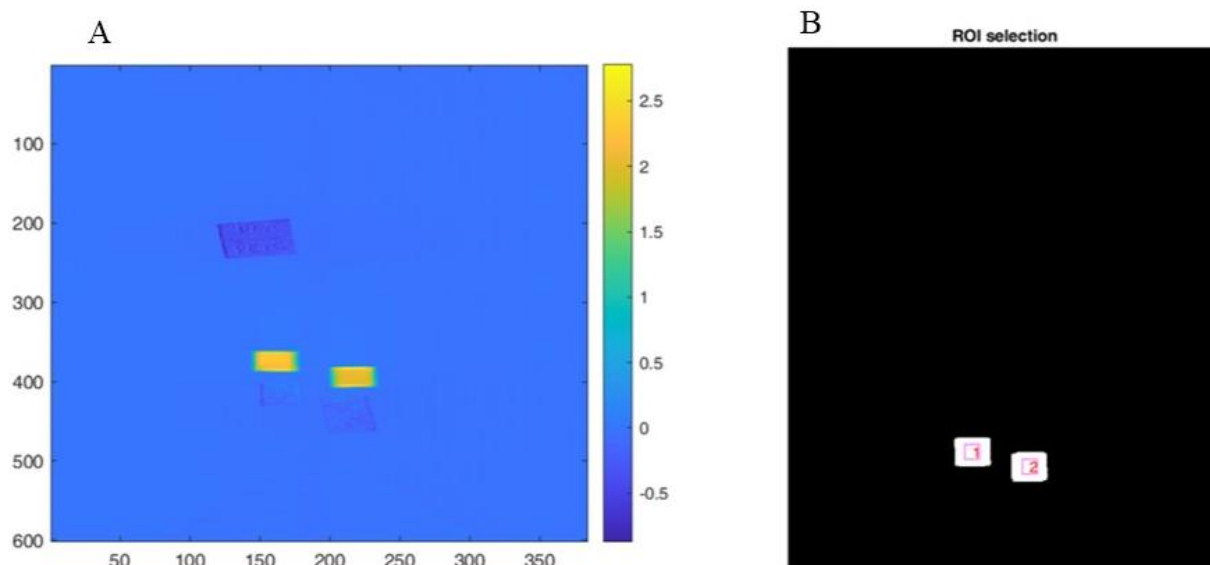


Figura 3-2: A: valori dell'intensità globale lungo PC2. B: individuazione delle ROI all'interno delle maschere

Successivamente si sono andate a definire delle ROI, rappresentate in Figura 3-2. L'identificazione di queste regioni è stata necessaria per eliminare i bordi dei campioni in quanto, visto il modo in cui sono stati stampati, in alcuni punti si aveva una completa saturazione del segnale che andava ad alterare il risultato dell'analisi. Le ROI che si sono estratte sono di forma quadrata, sia per i regoli che per i cilindri.

L'identificazione delle ROI è stata necessaria anche per consentire di passare da un approccio *pixel based* a uno *object based*. Al fine di effettuare questa transizione si è calcolata la mediana di tutti i pixel facenti parte della ROI, ottenendo uno spettro che rappresenta il singolo campione. Si è scelto di utilizzare la mediana e non la media per essere certi di avere uno spettro meno influenzato da pixel con caratteristiche anomale. Questo ci ha garantito una maggiore sicurezza nel caso in cui la ROI selezionata fosse ancora troppo ampia e quindi fossero presenti dei pixel che risentivano ancora dell'effetto dei bordi, che avrebbero influito negativamente sul valor medio.

3.3.3 Pretrattamenti

I dati ottenuti sono stati sottoposti a diversi pretrattamenti, sia di riga che di colonna. In particolare, si sono utilizzati due pretrattamenti di riga e uno di colonna. Il pretrattamento di colonna che si è utilizzato è stato il centraggio sulla media. L'equazione (3.1) descrive l'operazione effettuata da questo pretrattamento:

$$(X_{iv})_{centrato} = X_{iv} - \bar{X}_v \quad (3.1)$$

Dove $(X_{iv})_{centrato}$ è il valore dell'intensità associata alla variabile dopo il pretrattamento, X_{iv} il valore originale della riflettanza associata variabile v per il campione i , e $\bar{X}_v$ è la media della variabile v su tutti i campioni. Di conseguenza, con questo pretrattamento si sottrae a ogni colonna la media della stessa, ottenendo una nuova media della colonna pari a zero. Il centraggio sulla media è molto utilizzato quando due campioni hanno valori medi molto diversi tra di loro, in quanto consente di rendere i dati confrontabili. Il suo effetto è quindi quello di spostare i valori centrandoli sullo zero [26].

Il primo pretrattamento di riga che si è utilizzato è stato la trasformata SNV che è descritta nell'equazione (3.2):

$$(X_{iv})_{SNV} = \frac{X_{iv} - \bar{X}_i}{s_i} \quad (3.2)$$

Con $(X_{iv})_{SNV}$ pari al valore dell'intensità per la variabile v del campione i dopo il pretrattamento, X_{iv} il valore di riflettanza della variabile v per il campione i , $\bar{X}_i$ la media della riga e s_i la deviazione standard della riga.

La trasformata SNV, quindi, opera sottraendo ai valori della riga la loro media e dividendo il valore ottenuto per la loro deviazione standard. Il numeratore della frazione è analogo al centraggio sulla media, anche se lavora su sulla riga anziché sulla colonna, e anche in questo caso svolge la stessa funzione, ovvero quella di rendere la media della riga pari a zero. L'operazione di divisione della riga per la deviazione standard è anch'essa una operazione riconducibile a un altro pretrattamento di colonna, ovvero la scalatura. Questo rende le deviazioni standard pari a 1, correggendo le differenze di dispersione tra le righe. La trasformata SNV è quindi in grado di correggere sia effetti di locazione che di dispersione, e la sua applicazione fornisce dei dati adimensionali. A livello spettrale essa è in grado di correggere lo shift della linea di base e l'effetto di intensità globale degli spettri [25,26].

Il secondo pretrattamento di riga che si è utilizzato è una trasformata logaritmica, e in particolare si è utilizzato il logaritmo in base 10. Questa trasformazione matematica è molto

utilizzata quando si ha a che fare con dei dati che presentano non linearità, per cercare di andare a correggere questa caratteristica.

I dati ottenuti sono stati analizzati usando sempre almeno un pretrattamento di riga e uno di colonna e questi sono sempre stati applicati in questo ordine. In particolare, si sono utilizzate inizialmente due combinazioni:

- SNV – centraggio sulla media
- Log10 – centraggio sulla media

Successivamente, visti gli effetti dei singoli pretrattamenti, si sono utilizzati nella combinazione: Log10 – SNV – centraggio sulla media. Si è scelto di utilizzare prima il logaritmo rispetto alla trasformata SNV in quanto quest'ultima avrebbe prodotto anche dei valori negativi, rendendo complicata la successiva applicazione di log10.

3.3.4 Spettri grezzi e pretrattati

In primo luogo, i dati sono stati osservati andando a rappresentare gli spettri ottenuti. Avendo deciso di approcciare le analisi in maniera *object based* si è ottenuto un numero di spettri pari al numero dei campioni; inoltre, avendo calcolato la mediana della ROI, gli spettri ottenuti sono degli spettri mediani.

Inizialmente si sono analizzati gli spettri grezzi, ovvero quelli ottenuti direttamente dalle acquisizioni senza operare su di essi nessun tipo di pretrattamento. Successivamente, si sono osservati gli spettri pretrattati e in particolare si sono rappresentati gli spettri pretrattati con la trasformata SNV.

3.3.5 Analisi delle componenti principali

Dopo una prima analisi spettrale si sono effettuate delle analisi esplorative, in particolare PCA, per iniziare a valutare la penetrazione della radiazione. Le componenti principali sono state ottenute utilizzando la matrice contenente l'informazione relativa al training set. I dati in questo caso sono stati pretrattati utilizzando la trasformata SNV seguita dal centraggio sulla media. Per ogni training set si sono osservati i grafici dei loading e i grafici

degli score. I dati sono stati riportati in grafici che presentano in ascissa il numero del campione e in ordinata il valore degli score lungo la componente di interesse. Inoltre, per rendere più chiara l'interpretazione dei dati, i campioni sono stati colorati in funzione dell'altezza del polimero posto sopra.

Una volta che il modello è stato costruito e il numero di PC è stato definito, si sono proiettati nello spazio da loro costituito prima il test set interno e poi quello esterno. Ai vari test set sono stati applicati gli stessi pretrattamenti del training set. Le componenti principali analizzate per i test set sono state le stesse che erano state selezionate per il training set. Dal momento che i campioni sono stati proiettati nello spazio creato dalle componenti principali del training set, i loading risultano essere gli stessi per tutti e tre i set di dati.

3.3.6 Regressione ai minimi quadrati parziali

Dopo una prima analisi esplorativa si è proceduto all'utilizzo di un metodo supervisionato e in particolare si è utilizzato PLS. Si sono costruiti i modelli per ogni set di dati utilizzando i rispettivi training set. Per questa operazione si è utilizzato come variabile indipendente X lo spettro dei vari campioni e come variabile dipendente Y l'altezza del polimero posto in alto. I pretrattamenti che si sono utilizzati sugli spettri sono stati:

- SNV – centraggio sulla media
- Log10 – centraggio sulla media
- Log10 – SNV – centraggio sulla media

Sulle Y si è effettuato solamente il centraggio sulla media. Successivamente si è selezionato il metodo di cross validazione, che è stato uno schema a *venetian blinds*. Si è proceduto quindi alla costruzione del modello andando a scegliere il numero di variabili latenti (LV) da utilizzare. Queste sono state scelte valutando i valori di RMSEC e di RMSECV; la scelta di quante LV utilizzare ha permesso di ottenere il modello desiderato. Il RMSEC, infatti, rappresenta la deviazione standard tra il valore stimato dal modello e il valore reale della variabile di interesse. Questo parametro, caratterizzato dall'avere la stessa unità di misura della proprietà che definisce, è calcolato come segue:

$$RMSEC = \sqrt{\frac{\sum_l (y_i - \hat{y}_i)^2}{I - A - 1}} \quad (3.1)$$

Dove y_i rappresenta il valore sperimentale della risposta relativa all'oggetto i , $\hat{y}_i$ è il valore calcolato dal modello di regressione per il campione i , I rappresenta il numero di campioni e A il numero di variabili latenti utilizzate per la costruzione del modello. Per ottenere un buon modello è bene che RMSEC sia basso, a testimoniare la coerenza tra la predizione della risposta e il valore noto a essa associato.

Ottenuti i modelli, se ne sono valutate le caratteristiche andando a considerare i VIP in modo da definire le variabili originali che hanno inciso maggiormente sulla predizione dell'altezza del polimero rivolto verso il rilevatore. Le prestazioni dei modelli sono state valutate considerando i valori di RMSEC e RMSECV e il bias della retta, che corrisponde all'errore sistematico del modello, e che quindi deve essere il più basso possibile. Per analizzare le prestazioni del modello, si sono anche utilizzati i grafici che rappresentano i residui della Y predetta in cross validazione in funzione del campione.

Dal momento che PLS è una tecnica di regressione, e quindi rientra nei metodi supervisionati, i modelli che si sono costruiti sono stati validati. Questa operazione è stata effettuata utilizzando prima i test set interni e poi quelli esterni, a cui sono stati applicati gli stessi pretrattamenti del training set. Per valutare la bontà dei modelli in predizione si è valutato il valore di RMSEP (*root mean square error in prediction*), calcolato in maniera analoga a RMSEC, che rappresenta la deviazione standard tra il valore predetto dal modello e il valore misurato per il test set. Anche in questo caso è bene che il valore di questo parametro sia piccolo e che sia paragonabile a RMSEC, in modo che si abbia la conferma che la prestazione del modello sia analoga sui campioni del training set e su quelli del test set.

4 RISULTATI E DISCUSSIONE

4.1 CAMPIONI DA 3 CM

4.1.1 Spettri grezzi e pretrattati

La prima operazione che si è svolta nell'analisi dei risultati è stata l'osservazione e l'interpretazione degli spettri grezzi. Nella Figura 4-1 possiamo osservare in blu lo spettro puro del PLA e in rosso lo spettro puro del PETG. Una zona di assorbimento importante per i due polimeri è rappresentata dall'intervallo 1100-1250 nm, in cui si possono osservare gli assorbimenti dovuti all'*overtone* dei legami C-H [39]. In questa regione, il PLA presenta una sola banda, dovuta alla presenza dei C-H alifatici, mentre il PETG è caratterizzato da due segnali, dovuti alla presenza di legami C-H sia alifatici che aromatici. Un'altra differenza tra gli spettri dei due polimeri presi in analisi si riscontra intorno ai 1400 nm, in cui cade la banda di *overtone* del legame O-H. In PETG questa è dovuta all'O-H del gruppo glicolico, mentre in PLA la sua presenza è riconducibile al fatto che questo polimero è igroscopico, e quindi rappresenta probabilmente il contributo dell'acqua [40]. Inoltre, si nota come questa banda sia più intensa nel PLA, suggerendo che il contributo dato dall'acqua sia più rilevante rispetto a quello del glicole del PETG. Le altre regioni dello spettro sono caratterizzate dalla sovrapposizione di diverse bande, rendendo di difficile interpretazione l'origine dei segnali. In generale, in questa regione sono presenti informazioni relative alle vibrazioni dei legami C-H, O-H, C=O e C-O [39]. Nonostante la difficoltà dell'attribuzione di queste bande, possiamo notare la presenza di una spalla caratteristica nello spettro del PLA intorno ai 1600 nm e delle differenze tra i due materiali nella zona da 1800 a 2200 nm, che rappresentano una zona di impronta digitale per i due polimeri.

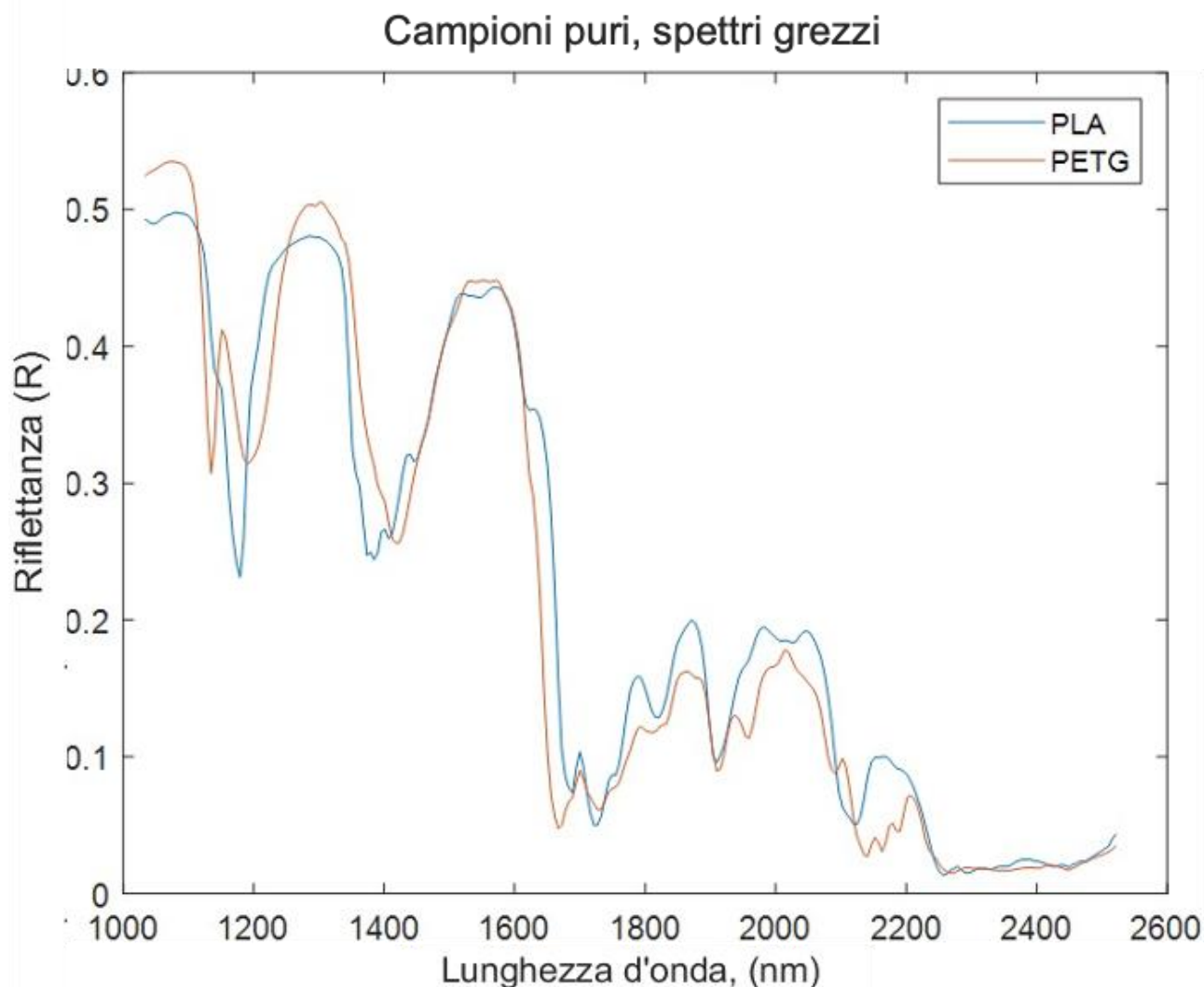


Figura 4-1: Spettri grezzi del PLA e del PETG

Nella Figura 4-2 sono rappresentati gli spettri dei campioni da 3 cm PETG-PLA. In Figura 4-2A vediamo gli spettri grezzi del training set, in cui il segnale in blu rappresenta il campione costituito completamente da PLA, mentre lo spettro in rosso rappresenta quello costituito solamente da PETG. Le colorazioni intermedie che passano dal blu al rosso rappresentano campioni con un contenuto crescente di PETG. Negli spettri grezzi in Figura 4-2A-C-E si può notare un effetto di intensità globale che impedisce di individuare alcuna correlazione tra il segnale e l'altezza dello strato di polimero. Questo effetto è stato corretto applicando il pretrattamento di riga SNV (Figura 4-2B-D-F). Dalla Figura 4-2B, che rappresenta gli spettri pretrattati del training set, si possono individuare delle differenze spettrali rilevanti, non solo tra il campione puro di PLA e i restanti campioni, ma anche per i campioni

seguenti fino a quello avente altezza di PETG pari a 10 mm. Queste differenze risultano particolarmente evidenti nella banda a 1200 nm e, anche se in modo più attenuato, a 1400 nm, dove gli spettri presentano un picco che va ad attenuarsi all'aumentare della quantità di PETG. Un effetto simile si può notare anche in Figura 4-2D, riferita agli spettri pretrattati del test set interno, dove alle stesse lunghezze d'onda si vede una differenza tra i primi due campioni del test set interno e i restanti. Se si considera il test set esterno (Figura 4-2F), questo effetto è meno evidente, probabilmente a causa della maggiore altezza dello strato superiore di PETG che li caratterizza.

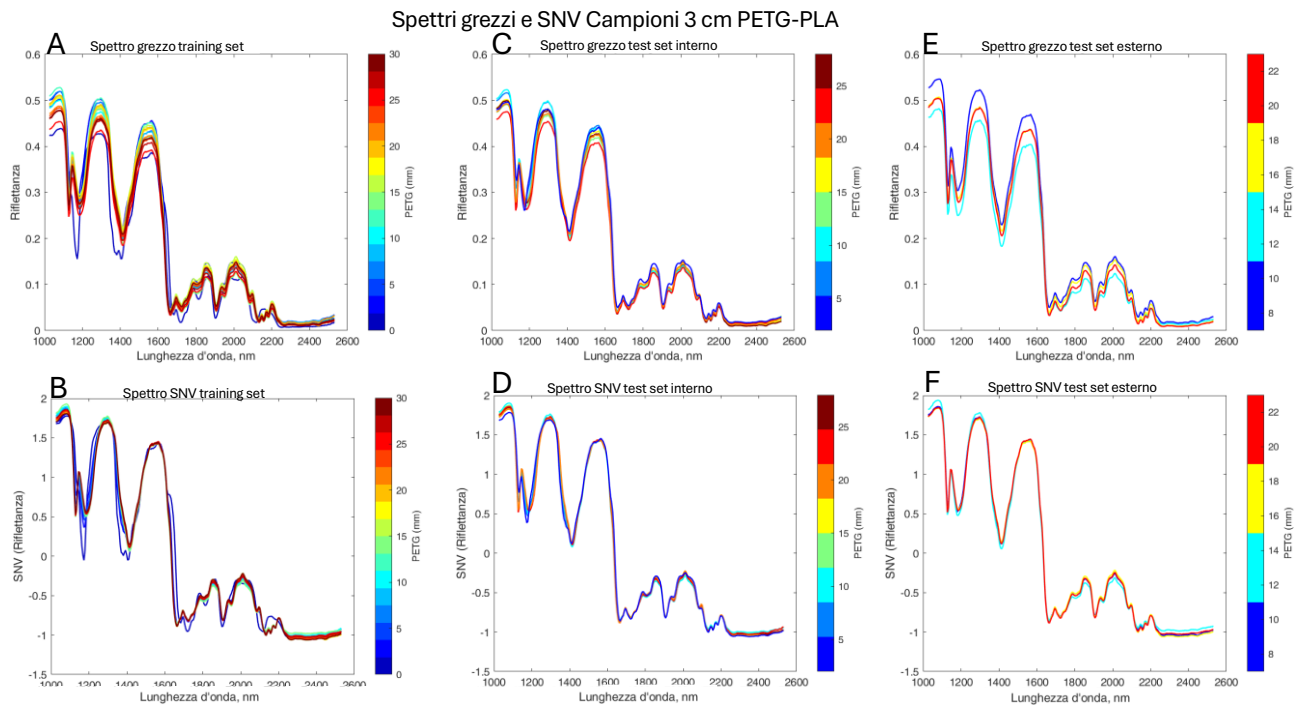


Figura 4-2: Spettri dei campioni con altezza 3 cm in configurazione PETG-PLA. A: Spettro grezzo del training set. B: Spettro SNV del training set. C: Spettro grezzo del test set interno. D: Spettro SNV del test set interno. E: Spettro grezzo del test set esterno. F: Spettro SNV del test set esterno

Analizzando i campioni in configurazione PLA-PETG si possono fare delle osservazioni analoghe. Anche in questo caso osservando Figura 4-3A-C-E, riferite rispettivamente agli spettri grezzi del training set, del test set interno e del test esterno, si nota un effetto di intensità globale. In Figura 4-3A si osserva in questo caso una disposizione ordinata dei campioni fino a un'altezza di 12 mm di PLA in funzione del cambiamento di intensità globale, suggerendo che l'effetto sia in questo caso correlato al rapporto tra i due polimeri costituenti il campione. Tuttavia in Figura 4-3C-E, riferite ai test set, la relazione tra intensità

globale e spessore dello strato di polimero risulta meno ordinata, per cui si è scelto di pretrattare anche in questo caso gli spettri con SNV (Figura 4-3B-D-F). Nella figura riferita al training set (Figura 4-3B) lo spettro costituito interamente da PETG è rappresentato in blu, mentre quello formato da PLA è rappresentato in rosso. Anche in questo caso si evidenziano delle differenze nella regione dei 1200 nm, che sono marcate per i campioni da 0 a 10 mm di PLA, ma si possono notare anche per campioni di altezza maggiore. In questo caso possiamo osservare anche delle differenze rilevanti nella regione dei 1600 nm, in cui la spalla relativa al PLA diventa più evidente all'aumentare dell'altezza del polimero che la definisce. Un ulteriore intervallo di lunghezze d'onda in cui si rilevano differenze è quello 1350-1400 nm, in cui si riescono a distinguere in maniera evidente i campioni fino ai 10 mm di PLA. Per quanto riguarda il test set interno (Figura 4-3D) si notano delle differenze legate ai campioni con contenuto fino a 10 mm di PLA, mentre per il test set esterno (Figura 4-3F) si notano delle differenze nella zona 1350-1400 nm tra il primo campione, colorato di blu, e i restanti cilindri.

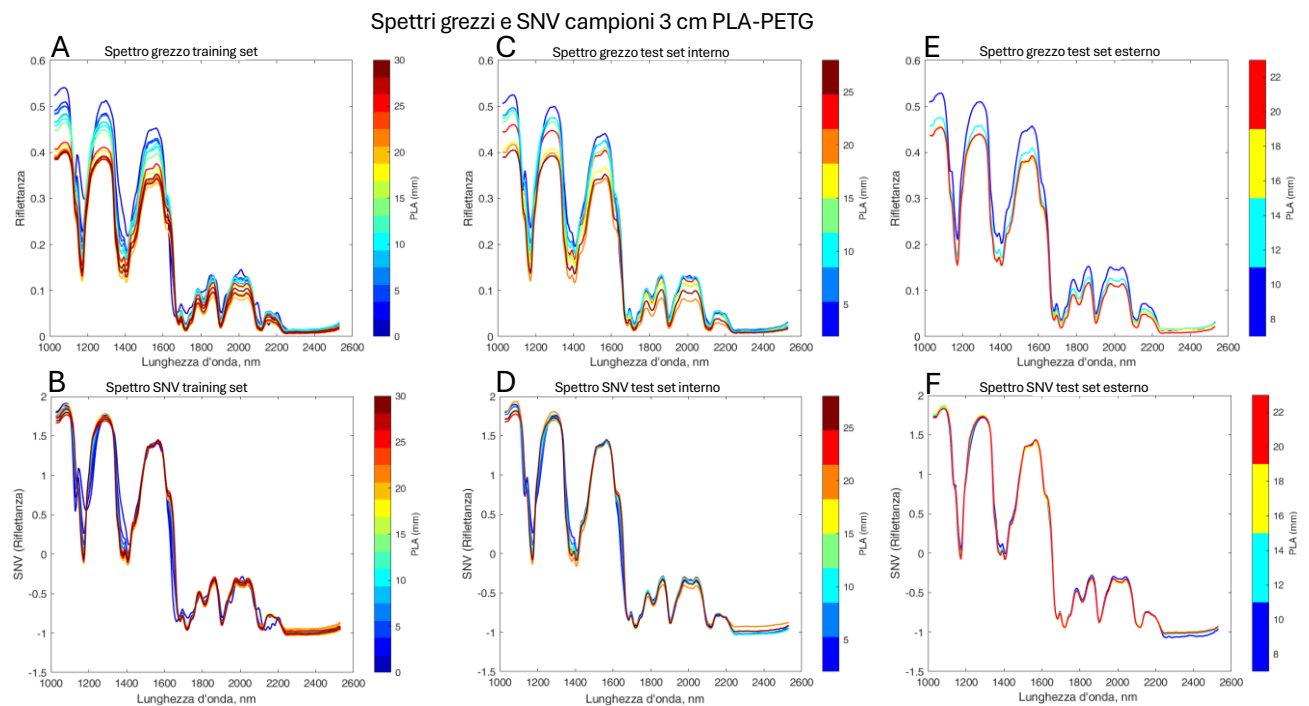


Figura 4-3: Spettri dei campioni con altezza 3 cm in configurazione PLA-PETG. A: Spettro grezzo del training set. B: Spettro SNV del training set. C: Spettro grezzo del test set interno. D: Spettro SNV del test set interno. E: Spettro grezzo del test set esterno. F: Spettro SNV del test set esterno

L'osservazione degli spettri ha portato a supporre che fossero presenti delle differenze nella penetrazione della radiazione attraverso i due materiali. In particolare sembrerebbe che, viste le maggiori differenze spettrali nei campioni PLA-PETG, il NIR penetri meglio attraverso il PLA.

4.1.2 Analisi delle componenti principali

Dopo aver analizzato gli spettri, i dati sono stati elaborati tramite la PCA. Prima del calcolo della PCA gli spettri sono stati pretrattati, e in particolare si è utilizzata la trasformata SNV come pretrattamento di riga per correggere l'effetto di intensità globale presente, come discusso nell'analisi spettrale al capitolo 4.1.1, e il centraggio sulla media come pretrattamento di colonna.

La variabilità di interesse, ovvero l'informazione chimica relativa alla profondità di penetrazione della radiazione NIR, si è trovata per i training set in entrambe le configurazioni in PC1. Di conseguenza, i loading di questa componente sono poi stati usati per proiettare i test set nello stesso spazio. La varianza spiegata dalla PC1 è sempre risultata maggiore dell'80%, confermando che l'informazione oggetto di interesse dello studio, ovvero l'altezza dello strato di polimero, è effettivamente la maggior fonte di variabilità all'interno del dataset.

Osservando i loading in Figura 4-4A possiamo definire le variabili più importanti nello spiegare la varianza tra i campioni. In particolare, un intervallo che influenza la PC1 è quello attorno a 1200 nm, una zona precedentemente riconosciuta come diagnostica nella distinzione spettrale dei due polimeri. La zona intorno ai 1400 nm influenza invece la PC1 negativamente; pertanto, i campioni caratterizzati da una maggiore intensità di questa banda saranno a score negativi di PC1. Intorno a 1600 nm è presente una banda a valori positivi di PC1, a indicare che questa lunghezza d'onda influisce nel collocare i campioni a valori di score positivi. Questa regione spettrale, come è stato sottolineato osservando gli spettri, si riferisce a una spalla caratteristica del PLA. Pertanto, campioni con minore contenuto di PETG, e quindi con un più alto contributo spettrale degli strati di PLA sottostanti, presenteranno valori più alti in questa zona, collocandosi a score positivi su PC1. Questa analisi conferma che le lunghezze d'onda per le quali si erano evidenziate differenze spettrali tra i due polimeri, sono anche responsabili nel definire la componente principale uno.

In Figura 4-4B vengono rappresentati gli scores dei campioni in configurazione PETG-PLA lungo PC1, che spiega il 88,36% di varianza. Si nota come fino al campione numero 6, di altezza pari a 10 mm di PETG, si possa individuare un andamento decrescente degli scores su PC1 in campioni successivi. Dato che i campioni del training set sono disposti in ordine di spessore dello strato superiore di polimero, questo andamento denota che gli scores su PC1 sono correlati a questa variabile di interesse e tale andamento è coerente con quanto si è osservato dai loading. Dal campione numero 7, di altezza del PETG pari a 12 mm, i campioni sono invece caratterizzati da score su PC1 molto simili, il che significa che PCA non rileva differenze tra questi oggetti. Ciò potrebbe implicare che la radiazione oltre questa altezza non è più in grado di penetrare, siccome PCA non individua caratteristiche spettrali legate al polimero che sta sotto. Un altro aspetto da notare è che, nonostante i campioni presentino un passo costante di aggiunta di materiale, questo non si riscontra negli score. Al contrario, si vede come i primi campioni abbiano degli score nettamente diversi, mentre i successivi hanno score che pian piano diventano più vicini tra di loro fino a raggiungere un plateau. Questo suggerisce un andamento non lineare del fenomeno della penetrazione della radiazione.

Analizzando gli scores dei test set, in Figura 4-4C-D, si nota come i campioni rispettivamente del test set interno e del test set esterno si comportano in maniera analoga a quelli del training set lungo PC1. È rilevante sottolineare questo aspetto per il test set esterno che, pur avendo geometria di riempimento diversa, presenta lo stesso comportamento dei campioni a forma di regoli. Analizzando nel dettaglio il test set interno, si riscontra un andamento decrescente lungo PC1 per tre campioni, caratterizzati da un'altezza massima di PETG di 8 mm. Nel test set esterno si evidenzia una disposizione decrescente lungo PC1 dei primi tre regoli. In questo caso si riconoscono quindi delle differenze a livello spettrale fino a 17 mm di altezza di PETG. Questo ci permette di supporre che la geometria di riempimento che caratterizza i campioni influenza la capacità di penetrazione della radiazione, e in particolare che le geometrie usate nei test set promuovano una maggiore penetrazione.

PCA SNV-Centraggio sulla media campioni 3 cm PETG-PLA

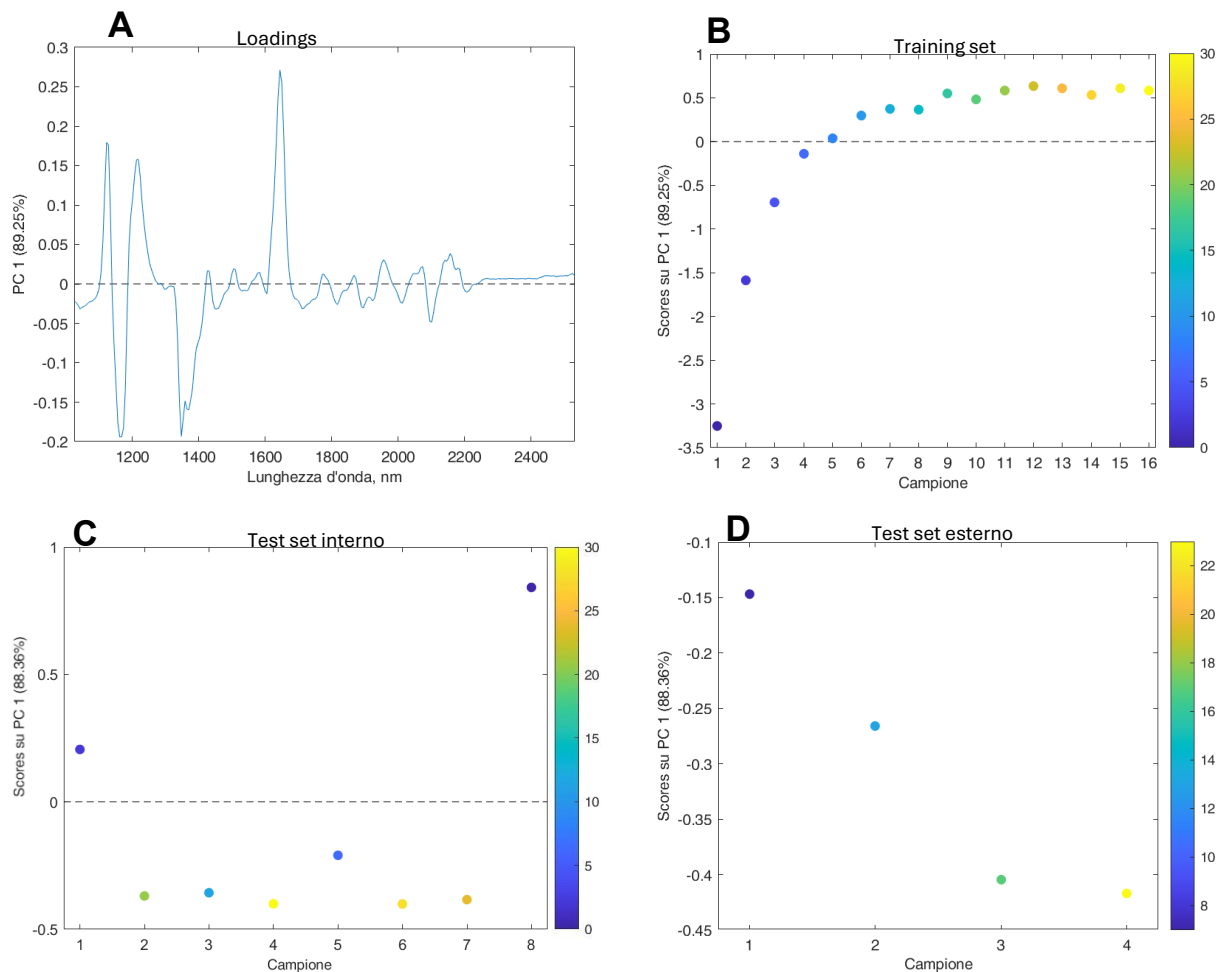


Figura 4-4: PCA dei campioni di altezza 3 cm in configurazione PETG-PLA. A: Loading di PC1. B: Score su PC1 del training set. C: Score su PC1 del test set interno. D: Score su PC1 del test set esterno.

Analizzando i campioni in configurazione PLA-PETG tramite PCA, si notano dei comportamenti simili rispetto a quanto accade in PETG-PLA. La PC1, che spiega 89,25% della varianza, è responsabile, anche in questo caso, di portare l'informazione relativa alla profondità di penetrazione della radiazione NIR. Analizzando la Figura 4-5A notiamo infatti che tra i loading le variabili che hanno maggiore influenza sono le stesse sottolineate in precedenza. In particolare, in questo caso i 1600 nm, che rappresentano la spalla caratteristica dei campioni ricchi del PLA, influiscono positivamente sui valori di PC1 e coerentemente con questo aspetto i campioni con un contenuto maggiore di questo polimero si trovano a valori positivi di PC1.

Un aspetto da sottolineare è come i loading dei campioni PETG-PLA e quelli dei campioni PLA-PETG (Figura 4-4A e 4-5A, rispettivamente) siano molto simili tra di loro, con le stesse variabili che influenzano nello stesso senso la costruzione della PC1. Di conseguenza, per entrambi i set di campioni si trovano a valori positivi di PC1 i campioni ricchi di PLA, e a valori negativi di PC1 quelli costituiti principalmente da PETG (es. Figura 4-4B e Figura 4-5B).

In Figura 4-5B si osservano gli score del training set lungo PC1, e si ritrova l'andamento non lineare notato precedentemente, in cui le differenze tra gli score dei diversi campioni diminuiscono progressivamente all'aumentare dello spessore del polimero che sta sopra. Anche in questo caso si raggiunge un plateau, in cui i campioni oltre circa 16 mm di PLA non presentano più differenze e infatti si trovano a score su PC1 tutti molto simili. Il fatto che, in entrambe le configurazioni in cui si sono analizzati i campioni, si osservi un plateau indica che oltre una certa altezza del polimero superiore non si ha più penetrazione della radiazione. Tuttavia, nel caso dei campioni in configurazione PETG-PLA questo andamento si osserva a 12 mm del PETG, mentre nei regoli PLA-PETG all'altezza di 16 mm del PLA. Ciò significa che la radiazione è in grado di attraversare più strati di PLA, rilevando informazioni spettrali relative al PETG che si trova sotto il PLA. Questo aspetto conferma le supposizioni che si erano fatte analizzando gli spettri, ovvero che ci sia una differenza nella profondità di penetrazione della radiazione tra i due materiali, e che sia maggiore nel PLA.

Per quanto riguarda i test set, si nota che in quello interno (Figura 4-5C) i primi quattro campioni in ordine di altezza (fino a 10 mm di PLA) ripresentano l'andamento crescente lungo PC1 all'aumentare del contenuto di PLA. Il campione successivo, che ha un'altezza di 18 mm di PLA, presenta score su PC1 simili ai campioni successivi e ciò denota che non si riscontrano differenze a livello spettrale tra i regoli di altezza superiore a 10 mm, suggerendo che la radiazione non sia in grado di penetrare a queste profondità. Nel test set esterno (Figura 4-5D) tutti e quattro i campioni, con una altezza massima di PLA pari a 23 mm, sono distribuiti a valori crescenti lungo la PC1, contrariamente a quanto accadeva nei campioni in configurazione PETG-PLA dove gli ultimi due si trovavano a score simili. Questo comportamento sembrerebbe confermare entrambe le ipotesi fatte precedentemente: la maggiore penetrazione della radiazione NIR attraverso il PLA, riconosciuta sia nel training set sia nel test set esterno, e l'effetto della geometria di riempimento sulla penetrazione, dal momento che nel training set l'altezza massima del PLA che si riesce a distinguere è pari a 16 mm, mentre nel test set esterno si hanno informazioni relative agli strati profondi di PETG in un campione che ha uno strato di PLA pari a 23 mm.

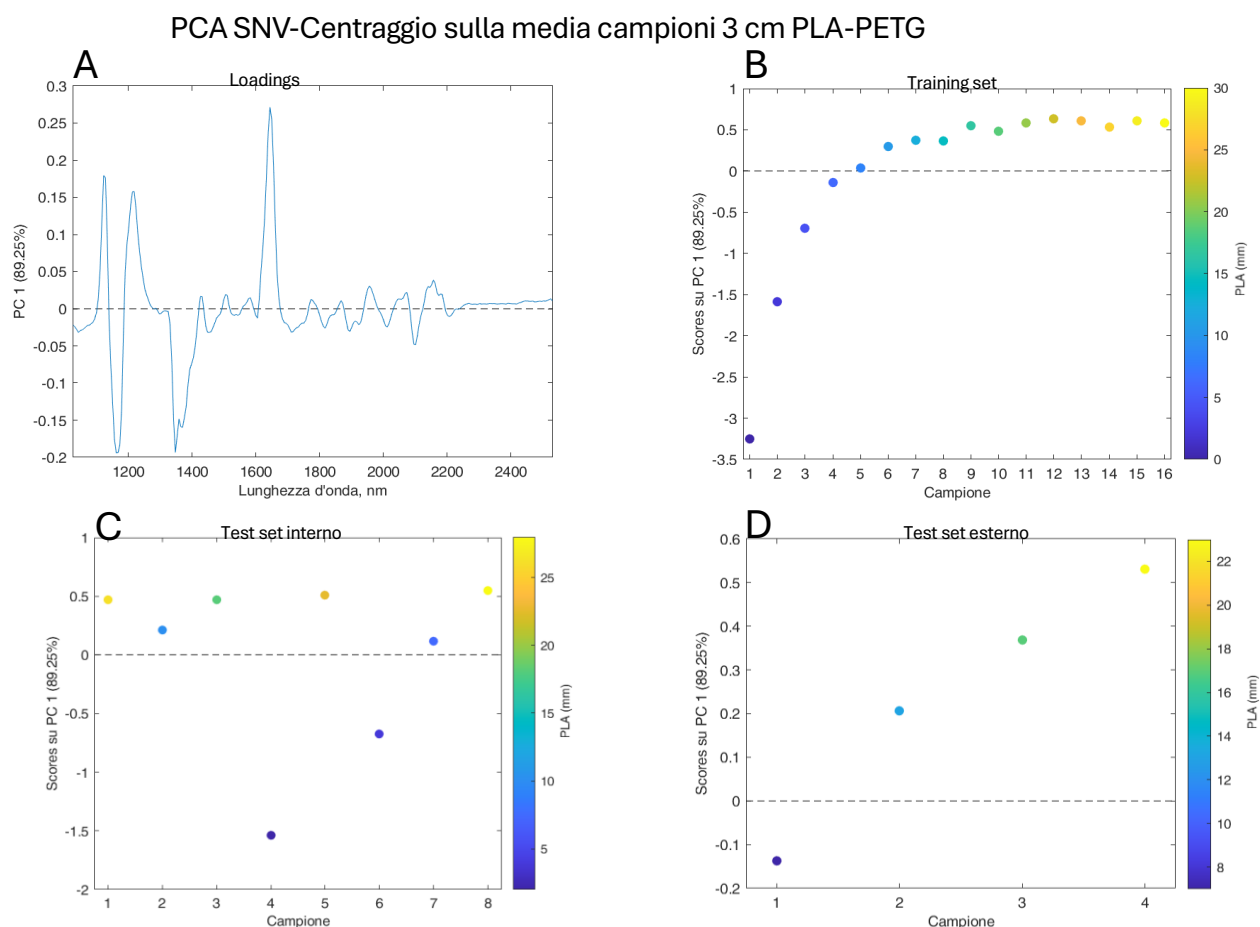


Figura 4-5: PCA dei campioni di altezza 3 cm in configurazione PLA-PETG. A: Loading di PC1. B: Score su PC1 del training set. C: Score su PC1 del test set interno. D: Score su PC1 del test set esterno.

L'esplorazione dei dati tramite la PCA permette di definire in maniera più accurata l'entità della penetrazione della radiazione e conferma le supposizioni relative alla maggiore penetrazione della radiazione attraverso PLA rispetto al PETG. Inoltre, permette di individuare due ulteriori aspetti, ovvero l'andamento non lineare del fenomeno oggetto di studio e l'influenza della geometria di riempimento, e quindi della microstruttura dei campioni, sulla profondità raggiunta dalla radiazione NIR.

4.1.3 Regressione ai minimi quadrati parziali

Dopo l'analisi esplorativa si è proceduto utilizzando dei metodi supervisionati, e in particolare la regressione PLS, con l'obiettivo di andare a definire, tramite il loro utilizzo, la

profondità di penetrazione della radiazione NIR. La scelta di utilizzare dei modelli di regressione non è legata, in questo caso, alla necessità di predire una proprietà, ovvero l'altezza dello strato superiore di polimero, ma all'idea che l'ottenimento di un modello con buone proprietà predittive, fino a certe altezze di polimero, sia sinonimo del fatto che la penetrazione della radiazione sia in grado di raggiungere tali profondità.

Visti i buoni risultati che si sono ottenuti, in PCA, tramite l'utilizzo della trasformata SNV seguita da centraggio sulla media, questa stessa coppia di pretrattamenti è stata utilizzata anche per costruire il modello PLS.

Il modello che si è calibrato sul training set per i campioni PETG-PLA presenta un RMSEC pari a 4,5 mm ed è stato costruito con 3 variabili latenti, selezionate in funzione dei criteri illustrati nel capitolo di materiali e metodi. Questo numero di LV ha infatti consentito di ottenere il miglior rapporto tra RMSEC e RMSECV, tenendo entrambi i valori relativamente bassi, ma senza andare a creare un modello overfittato rischiando di modellare il rumore. Le variabili che hanno influenzato maggiormente la costruzione del modello sono rappresentate in Figura 4-6A, attraverso il grafico dei VIP. L'intervallo di lunghezze d'onda più importante in questo modello è rappresentato dalla zona intorno a 1600 nm, seguito dalle bande intorno a 1200 nm e 1400 nm che hanno un contributo simile tra di loro. Si nota come le lunghezze d'onda più rilevanti sono le stesse che sono state evidenziate sia negli spettri che nei loading delle PC1, confermando ulteriormente che l'informazione che si è modellata in PCA era informazione chimica.

Analizzando il comportamento dei campioni del training set attraverso il grafico dei valori predetti plottati con quelli noti (Figura 4-6B) si possono individuare comportamenti diversi. In particolare, i primi due campioni, caratterizzati dal minor contenuto di PETG, sono predetti a valori negativi di altezza. Questo aspetto è collegato a diversi aspetti a cui si è già accennato nel corso delle analisi. In primo luogo, abbiamo visto una differenza notevole negli spettri dei primi campioni rispetto a quelli successivi. In secondo luogo, abbiamo già evidenziato, osservando le PCA, come la penetrazione della radiazione sembrerebbe essere un fenomeno non lineare, pertanto, utilizzando un metodo di regressione lineare come PLS, ci si aspetta delle difficoltà nel modellare correttamente questo comportamento. L'aspetto della non linearità del fenomeno è anche confermato dalla Figura 4-6C in cui sono rappresentati i residui dei campioni in cross validazione, ovvero la differenza tra il valore predetto e quello misurato per ogni punto della retta di calibrazione. L'analisi dei residui, se si esclude il campione 8 (14

mm di PETG) che ha un comportamento anomalo rispetto agli altri, denota un andamento parabolico, che conferma la non linearità del fenomeno, mostrando che l'altezza di tutti i campioni all'inizio e alla fine della serie è sottostimata, mentre quella dei campioni al centro della serie è sovrastimata. Questo comportamento è un noto segno di criticità nel fitting del modello [41].

Nonostante queste problematiche, dal campione con 4 mm di PETG fino a quello caratterizzato da 12 mm di PETG il modello è in grado di andare a determinare abbastanza bene la composizione dei campioni. Da un'altezza di 14 mm in poi si nota invece una rottura del modello che è rappresentata dal fatto che tutti questi campioni vengono predetti in cross validazione alla stessa altezza. Ciò sta a significare che tra questi campioni non vengono riscontrate delle differenze spettrali, e pertanto suggerisce che oltre l'altezza di 12 mm di PETG la radiazione non penetri ulteriormente.

Il modello è stato poi testato utilizzando il test set interno (Figura 4-6D) e si è evidenziato come la sua performance in predizione sia paragonabile a quella in calibrazione e in cross validazione. Questo aspetto è testimoniato dai valori di RMSEC, RMSECV E RMSEP che sono simili tra di loro. L'andamento dei campioni è analogo a quanto accade nel training set, in quanto si ha una predizione buona, anche se leggermente sovrastimata, per i campioni fino a circa 14 mm di PETG, mentre per i campioni successivi, che vengono sistematicamente sottostimati, si osserva nuovamente la rottura del modello.

Si è poi mandato in predizione il test set esterno (Figura 4-6D) e si è riscontrata una prestazione del modello paragonabile a quanto accade nel training set e nel test set interno, in cui i valori di RMSEP sono simili a quelli di RMSECV. Questo può significare che l'utilizzo di SNV come pretrattamento ha mitigato l'effetto del diverso scattering dovuto ai vari riempimenti. Tuttavia, si nota come i campioni caratterizzati dall'avere lo stesso riempimento vengono entrambi sovrastimati, il primo e il terzo, o sottostimati, il secondo e il quarto, a testimoniare che l'effetto fisico non è stato eliminato completamente.

PLS SNV-Centraggio sulla media
Campioni 3 cm PETG-PLA

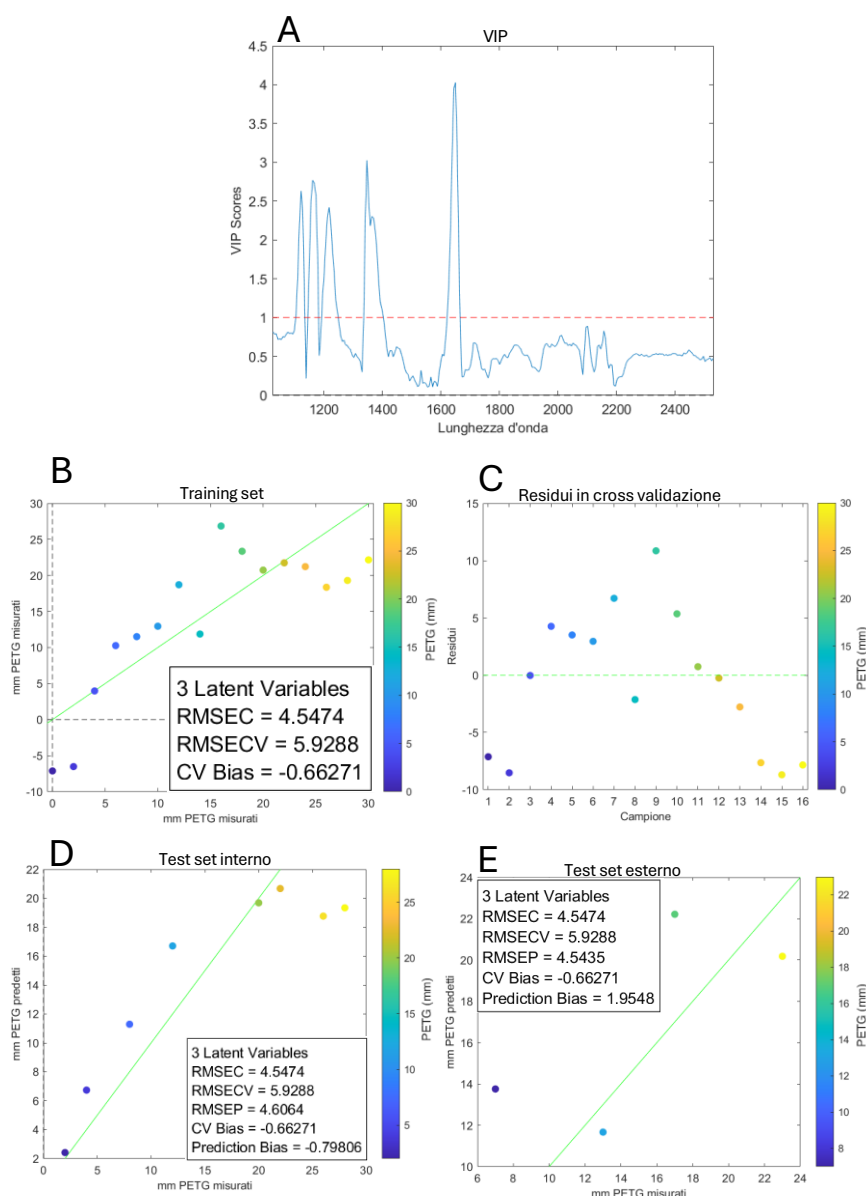


Figura 4-6: Modello PLS, utilizzando come pretrattamento SNV-centraggio sulla media, dei campioni di altezza 3 cm in configurazione PETG-PLA. A: VIP. B: Grafico dell'altezza predetta-misurata del training set. C: Residui in cross validazione del training set. D: Grafico dell'altezza predetta-misurata test set interno. E: Grafico dell'altezza predetta-misurata del test set esterno

Analizzando, con la stessa coppia di pretrattamenti, i campioni in configurazione PLA-PETG si è costruito un nuovo modello PLS caratterizzato da un valore di RMSEC pari a 0,6 mm, utilizzando 6 variabili latenti. Confrontato al precedente modello, riferito alla configurazione PETG-PLA, si ha quindi un aumento del numero delle variabili latenti

significative e una diminuzione drastica dell'errore in calibrazione. Dalla Figura 4-7A, che rappresenta il grafico dei VIP, si nota che l'intervallo di lunghezze d'onda più rilevante nella costruzione del modello è ancora una volta quello attorno a 1600 nm. Altre zone rilevanti, coerentemente a quanto sottolineato analizzando gli spettri, sono quelle intorno a 1200 nm e a 1400 nm. Inoltre, risulta essere significativa anche la zona intorno a 2100-2200 nm, contrariamente a quanto accade nel caso dei campioni in configurazione PETG-PLA. Si nota quindi che i VIP delle due configurazioni di acquisizione dei campioni sono diversi tra loro, diversamente a quanto accade per i loading di PCA. Questo sottolinea che, in PLS, dovendo modellare e non solo esplorare un fenomeno, il modello va a considerare la peculiarità del sistema oggetto di analisi.

Analizzando il comportamento del training set in cross validazione, il cui grafico dei valori predetti contro quelli noti è rappresentato in Figura 4-7B, si osserva che il bias della retta, sia in calibrazione che in cross validazione, è contenuto, il che significa che il modello non è affetto da un errore sistematico. Le prestazioni che si ottengono in questo caso sono migliori rispetto a quelle ottenute per i campioni in configurazione PETG-PLA, con un RMSECV di 2,46 mm contro i 5,93 mm di prima. Oltre alla riduzione dell'errore, risultano meno evidenti anche i problemi di predizione relativi ai primi campioni, non più predetti a valori negativi di Y, pur essendo presenti delle fluttuazioni che, tuttavia, potrebbero essere casuali. Inoltre, il modello mantiene buone prestazioni su tutti i campioni nonostante i primi tre regoli e gli ultimi sei siano caratterizzati da un rumore casuale maggiore che potrebbe essere dovuto all'impiego di un numero maggiore di variabili latenti, che da un lato consente di ottenere un modellamento dei dati migliore, ma dall'altro determina l'introduzione di un po' di rumore. Un aspetto rilevante relativo ai campioni con un'altezza del PLA maggiore di 20 mm è che, pur essendo caratterizzati da una maggiore difficoltà di predizione, presentano comunque un andamento crescente, all'aumentare del PLA, contrariamente a quanto si osserva in Figura 4-6B. I residui in cross validazione (Figura 4-7C) mostrano dal campione 6 al campione 14 un accenno di andamento parabolico, evidenziando la presenza di non linearità che, però, risulta meno marcata rispetto a quella presente nei campioni in configurazione PETG-PLA (Figura 4-6C). Le caratteristiche del modello, relative alle migliori prestazioni predittive e all'andamento dei residui, confermano le supposizioni effettuate precedentemente riguardo alla migliore penetrazione della radiazione attraverso il PLA rispetto a PETG.

Andando a considerare le prestazioni del modello sui set di campioni utilizzati per la validazione, si possono fare alcune osservazioni. Analizzando il test set interno (Figura 4-7D)

si osserva un RMSEP circa doppio rispetto al RMSECV. Questo è dato prevalentemente dai campioni con un'altezza del PLA maggiore di 18 mm, ovvero gli ultimi tre della serie, in quanto il loro contenuto di polimero predetto è molto inferiore rispetto al valore misurato. Al contrario, i primi campioni, fino a circa 12 mm del PLA, sono predetti in maniera soddisfacente dal modello. Nonostante i campioni con un contenuto di PLA maggiore di 18 mm siano sottostimati, si riconosce tra di essi, come nel training set, un andamento crescente all'aumentare del PLA, sottolineando la presenza di informazione utile anche a queste altezze nonostante essa sia ridotta dalla maggiore presenza di rumore.

Se si considera, invece, il test set esterno (Figura 4-7E), le prestazioni del modello sono analoghe se non superiori a quanto accade in cross validazione, in quanto i valori di RMSEP e RMSECV sono paragonabili. In particolare, si nota che tutti e quattro i cilindri sono predetti correttamente, anche se si nota un bias in predizione maggiore rispetto a quanto accade in cross validazione. Un aspetto singolare che caratterizza il test set esterno è rappresentato dal suo valore di RMSEP minore rispetto a quello del test set interno. Ciò è qualcosa di inatteso in quanto il test set interno è caratterizzato dalla stessa tipologia di riempimento dei regoli utilizzati per la costruzione del modello stesso. Le prestazioni migliori sul test set esterno potrebbero essere proprio legate alla differente geometria di riempimento che caratterizza i cilindri e che potrebbe determinare una maggiore penetrazione della radiazione, come si era già evidenziato in PCA (Figura 4-5C), che consentirebbe di bilanciare l'effetto dovuto al diverso riempimento. Un'altra caratteristica particolare del test set esterno è che i campioni tipicamente sovrastimati, ovvero i primi, vengono sottostimati mentre gli ultimi, che sono sottostimati nel training set, vengono sovrastimati. Il fenomeno che potrebbe determinare questo comportamento potrebbe essere di natura fisica in quanto, pur utilizzando la trasformata SNV, si potrebbe risentire di un leggero effetto dovuto al diverso scattering. Nonostante queste differenze, si può dire che il modello si mostra coerente anche su dei set di dati con delle caratteristiche diverse rispetto a quelle usate per la sua costruzione e cross validazione. In particolare, il diverso scattering che è presente tra regoli e cilindri non influisce in maniera significativa sulle prestazioni del modello, e questo è dovuto all'azione di SNV che va a correggere il diverso scattering della radiazione.

PLS SNV-Centraggio sulla media
Campioni 3 cm PLA-PETG

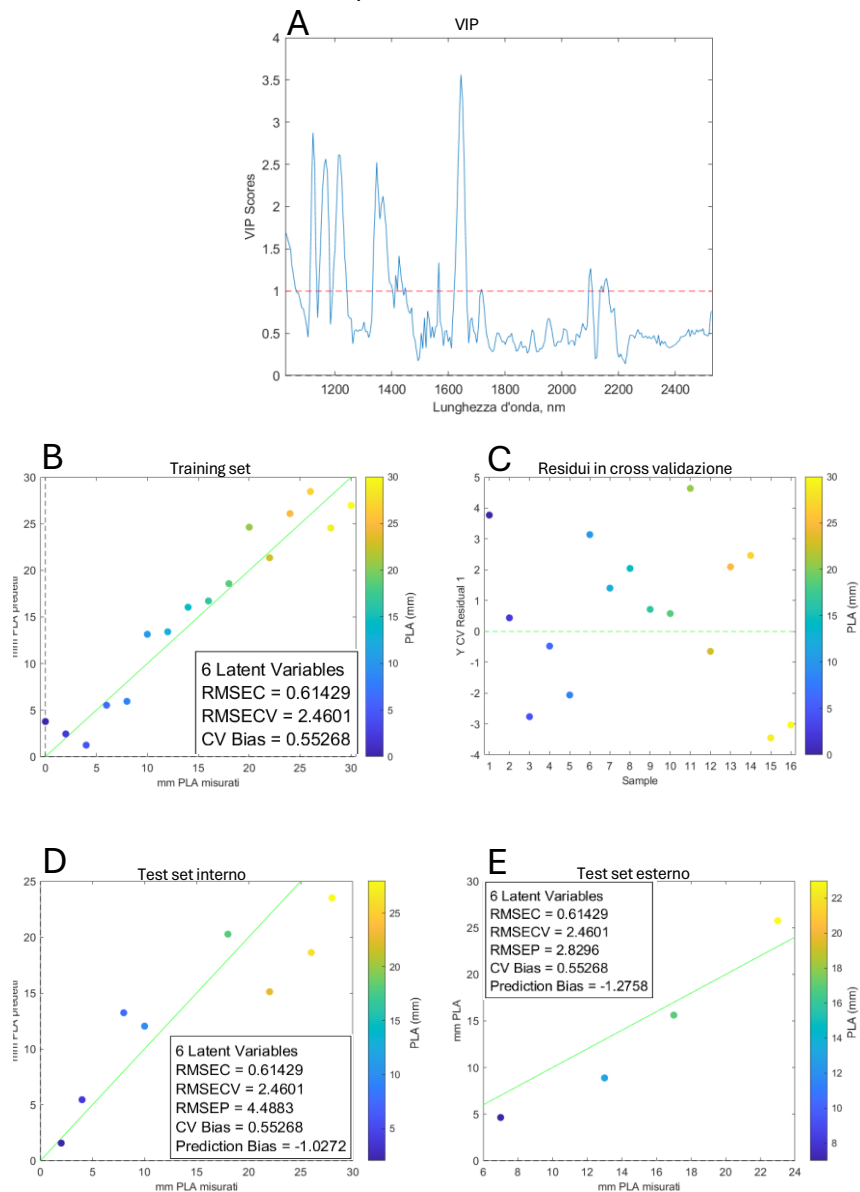


Figura 4-7: Modello PLS, utilizzando come pretrattamento SNV-centraggio sulla media, dei campioni di altezza 3 cm in configurazione PLA-PETG. A: VIP. B: Grafico dell'altezza predetta-misurata del training set. C: Residui in cross validazione del training set. D: Grafico dell'altezza predetta-misurata test set interno. E: Grafico dell'altezza predetta-misurata del test set esterno

L'utilizzo della trasformata SNV seguita dal centraggio sulla media per la creazione di modelli PLS ha permesso di ottenere dei risultati interessanti, soprattutto considerando la configurazione PLA-PETG. Tuttavia, si sono evidenziate anche delle criticità come la difficile predizione dei primi campioni, in particolare nella configurazione PETG-PLA, e la presenza di

un andamento parabolico nei residui in cross validazione. Si ritiene che tali problematiche siano legate al fatto che si stia utilizzando un modello di regressione lineare per descrivere un fenomeno non lineare. Di conseguenza, per cercare di mitigare questo effetto, si è scelto di utilizzare una diversa coppia di pretrattamenti e in particolare si è applicato il $\log_{10}$ come trattamento di riga, mentre come pretrattamento di colonna si è mantenuto il centraggio sulla media.

Nei campioni PETG-PLA trattati con $\log_{10}$ si osservano già delle differenze rispetto a quanto accadeva utilizzando come pretrattamento la trasformata SNV. Innanzitutto, analizzando i VIP (Figura 4-8A) si osserva che le lunghezze d'onda che influiscono sulla creazione del modello, caratterizzato da 3 variabili latenti e con un RMSEC pari a 2,3 nm, sono diverse rispetto a quelle evidenziate nei passaggi di processing precedentemente descritti. In particolare, pur rimanendo costante la rilevanza intorno a 1600 nm, si ha che le zone a 1200 nm e 1400 nm sono meno influenti, pur restando al di sopra del valore di 1 e quindi risultando statisticamente rilevanti. Per contro, si ha che la zona dopo 2100 nm diventa significativa nella creazione del modello. Tale intervallo è in parte legato all'impronta digitale dei campioni, ma oltre i 2200 nm gli spettri presentano per lo più rumore, che viene quindi accentuato tramite l'azione del logaritmo che ne determina il modellamento.

Per quanto riguarda il comportamento del training set (Figura 4-8B) si osserva un bias in cross validazione pari a 1, indicando un peggioramento delle performance rispetto a quanto accadeva con la combinazione di pretrattamenti precedenti, e inoltre si evidenzia una discrepanza maggiore tra RMSEC e RMSECV, a indicare delle difficoltà maggiori nella descrizione del fenomeno da parte del modello. Infatti, osservando i valori di $\hat{Y}$ predetta in cross validazione, si nota che i primi campioni, nei quali è stato dimostrato precedentemente, per mezzo della PCA e dell'analisi degli spettri grezzi, che la radiazione dovrebbe raggiungere il polimero sottostante, vengono predetti ad altezze simili tra di loro. Questo fenomeno avviene anche nei regoli con una quantità di PETG superiore a 15 mm. I due gruppi di campioni presentano però caratteristiche molto diverse. Il comportamento che si osserva è atteso per il secondo gruppo, mentre è inusuale e in contrasto con le osservazioni precedenti per i primi campioni. Ciò rappresenta un peggioramento rispetto ai risultati ottenuti tramite la trasformata SNV seguita dal centraggio sulla media e potrebbe essere determinato dalle nuove lunghezze d'onda rilevanti nella costruzione del modello. Tuttavia, analizzando i residui in cross validazione del training set (Figura 4-8C) non si osserva il netto andamento parabolico che si

riconosceva in Figura 4-6C, sottolineando che il $\log_{10}$ sembrerebbe essere in grado di correggere la non linearità del fenomeno.

Per quanto riguarda il test set interno (Figura 4-8D) si osserva una riduzione di RMSEP rispetto a RMSECV, anche se la prestazione sui campioni e l'andamento dei gruppi iniziali e finali rimangono analoghi a quanto accade nel training set.

Per quanto riguarda il test set esterno (Figura 4-8E), si osserva un bias in predizione pari a 4,4 mm, un valore notevolmente superiore rispetto a quanto accade negli altri set di dati. Inoltre, si osserva un RMSEP maggiore di RMSECV e pari al triplo di RMSEC. Questo indica una maggiore difficoltà del modello nel predire i campioni del test set esterno, cosa probabilmente dovuta al diverso riempimento di questi campioni, che causa uno scattering diverso. L'effetto legato al diverso scattering, che non è corretto da nessun pretrattamento in questa configurazione, impedisce quindi al modello di descrivere il fenomeno.

PLS Log10-Centraggio sulla media
Campioni 3 cm PETG-PLA

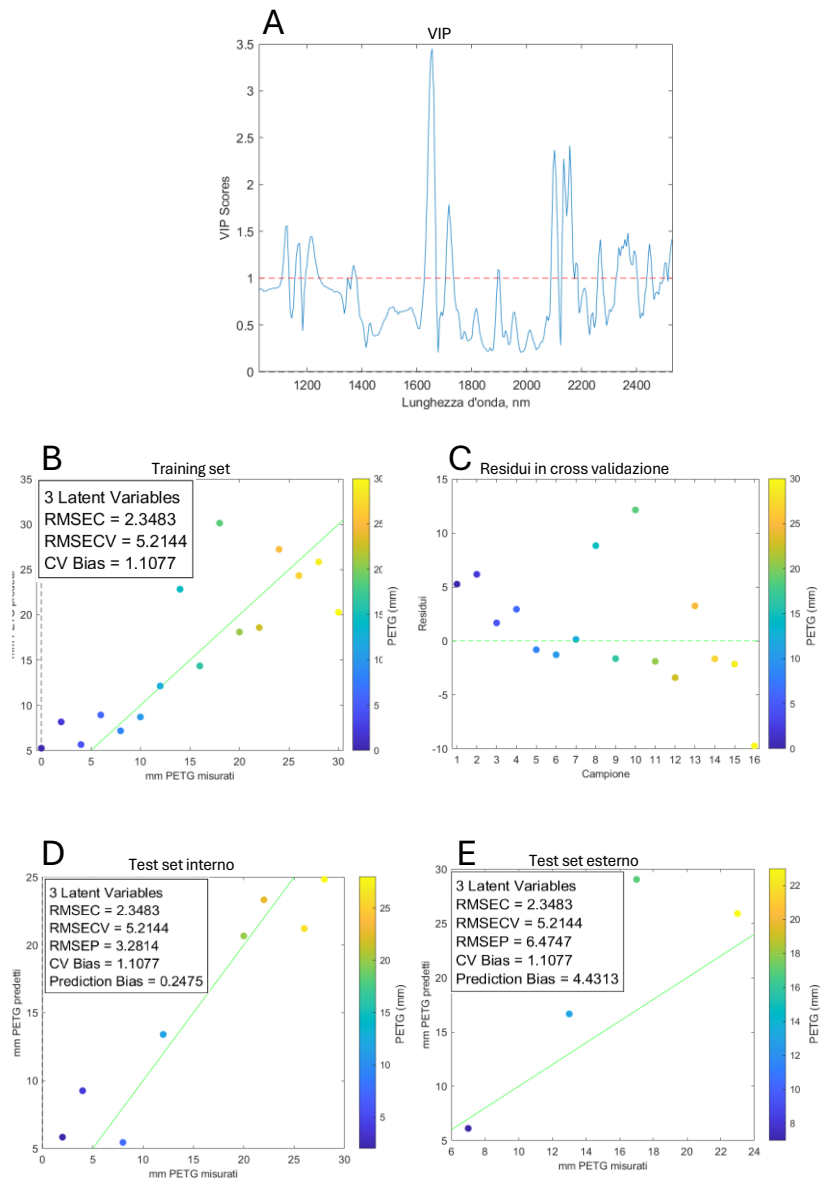


Figura 4-8: Modello PLS, utilizzando come pretrattamento Log10-centraggio sulla media, dei campioni di altezza 3 cm in configurazione PETG-PLA. A: VIP. B: Grafico dell'altezza predetta-misurata del training set. C: Residui in cross validazione del training set. D: Grafico dell'altezza predetta-misurata test set interno. E: Grafico dell'altezza predetta-misurata del test set esterno

Nel modello costruito per i campioni in configurazione PLA-PETG, caratterizzato dall'utilizzo di 5 variabili latenti e da un RMSEC pari a 2,47 mm, si ha che le variabili più rilevanti (Figura 4-9A) sono diverse rispetto a quanto osservato nel caso dei campioni PETG-PLA. Le lunghezze d'onda più importanti nella costruzione del modello sono quelle comprese

tra 1650-1800 nm e quelle intorno a 2100 nm. Bande rilevanti sono anche rappresentate nella zona intorno a 1200 nm, 1400 nm, 1900 nm e 2200-2400 nm. In questo caso si vede quindi come le bande a lunghezze d'onda maggiori, che sono di difficile attribuzione a livello spettrale e caratterizzano la regione dell'impronta digitale, diventino particolarmente rilevanti. La loro importanza crescente, rispetto al modello che utilizza la trasformata SNV, iniziava a notarsi già per i campioni PETG-PLA, diventando evidente in questo caso.

Il modello che è stato costruito è caratterizzato dall'avere un bias, sia in calibrazione che in cross validazione, più contenuto rispetto ai campioni PETG-PLA pretrattati in maniera analoga. Un aspetto che rimane costante è invece rappresentato da RMSECV che ha un valore quasi doppio rispetto a RMSEC. Per quanto riguarda la predizione in cross validazione del training set (Figura 4-9B) si possono notare tre gruppi di campioni che vengono predetti ad altezze analoghe. Il primo gruppo è rappresentato dai campioni con un basso contenuto di PLA (fino a 4 mm) che vengono descritti tutti e tre come privi di PLA. Il secondo gruppo è rappresentato dai campioni intermedi, dai 6 mm ai 14 mm di PLA e anche in questo caso i campioni sono predetti tutti allo stesso valore di Y attorno a 10-12 mm, mentre nel modello PETG-PLA questa era la zona meglio descritta. Infine, l'ultimo gruppo contiene i campioni caratterizzati dall'avere un contenuto di PLA pari a 16 mm o maggiore, e anche in questo caso la predizione avviene all'incirca alla stessa Y per ognuno di essi, il che è un comportamento che ci si aspetta perché è dove si supponeva che la radiazione non penetrasse ulteriormente. Il comportamento del modello sugli altri campioni, tuttavia, risulta essere non in linea con quanto osservato con i precedenti trattamenti.

La performance del modello sui set di campioni usati per la validazione è, per entrambi i test set, fondamentalmente analoga quando si vanno a considerare i valori di RMSEP rispetto a quelli di RMSECV. Tuttavia, si deve considerare che nel caso del test set interno (Figura 4-9D) si ha un bias in predizione minore, seppure abbastanza elevato, rispetto a quello presente nel test set esterno (Figura 4-9E). Nel test set interno individuiamo lo stesso comportamento del training set con il riproporsi in maniera pressoché equivalente dei tre gruppi individuati precedentemente. Nel caso del test set esterno tutti i campioni sono sottostimati rispetto al loro valore reale, come sottolineato dal bias in predizione che ha valore negativo. In generale il modello non fornisce delle prestazioni soddisfacenti se si considerano i valori di RMSECV e di RMSEP e se lo si paragona con quanto si riesce a osservare, sullo stesso set di dati, utilizzando la trasformata SNV al posto del logaritmo.

PLS Log10-Centraggio sulla media
Campioni 3 cm PLA-PETG

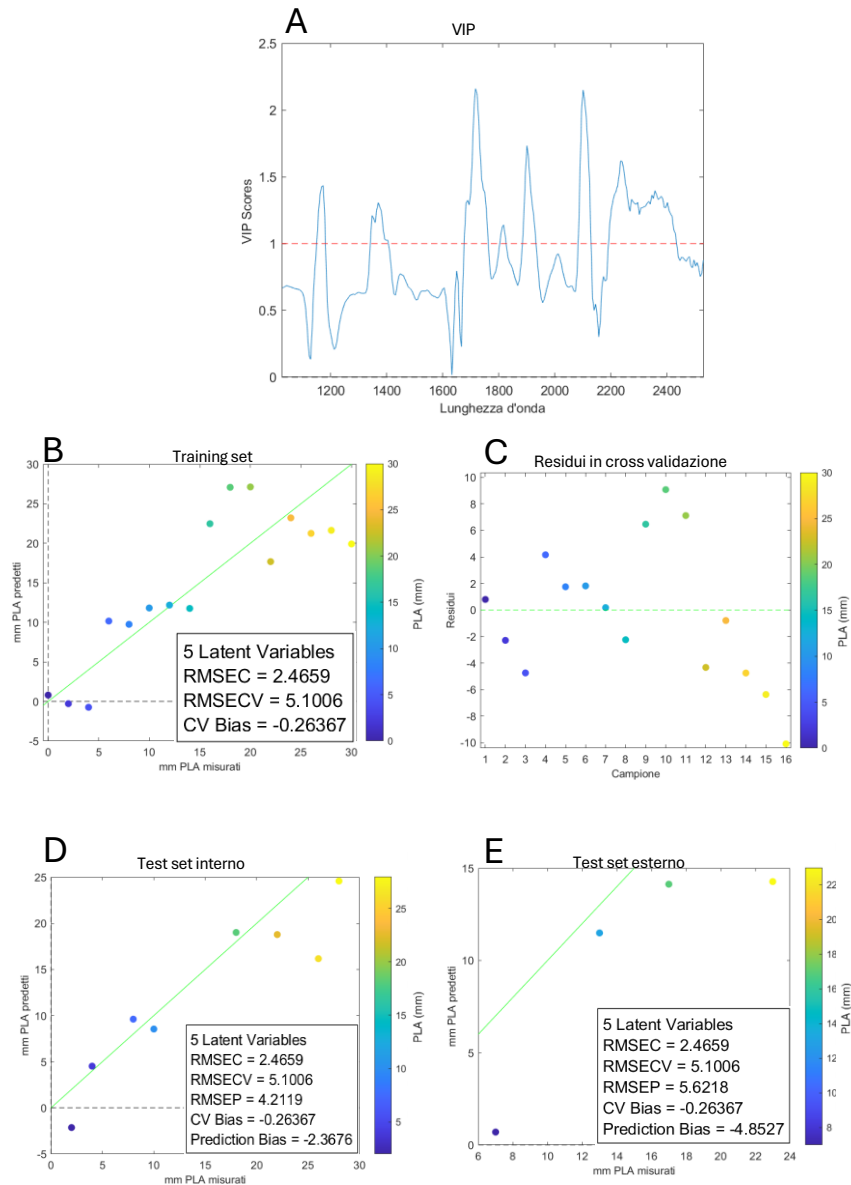


Figura 4-9: Modello PLS, utilizzando come pretrattamento Log10-centraggio sulla media, dei campioni di altezza 3 cm in configurazione PLA-PETG. A: VIP. B: Grafico dell'altezza predetta-misurata del training set. C: Residui in cross validazione del training set. D: Grafico dell'altezza predetta-misurata test set interno. E: Grafico dell'altezza predetta-misurata del test set esterno

Alla luce di quanto illustrato fino a ora si possono andare a definire gli effetti principali delle due coppie di pretrattamenti utilizzati, e in particolare del pretrattamento di riga. La trasformata SNV consente di correggere gli effetti legati a un aspetto fisico dei campioni e in particolare rende il diverso scattering dei cilindri rispetto ai regoli un fattore non influente sulle

abilità predittive del modello, tanto che la predizione sul test set esterno presenta delle performance analoghe a quella sul test set interno. La trasformata SNV, tuttavia, non è in grado di correggere l'andamento non lineare del fenomeno che si osserva e questo ne definisce un limite, perché modellare un fenomeno non lineare con un modello lineare è complesso. Al contrario, il $\log_{10}$ è in grado di mitigare la natura non lineare del fenomeno, ma non corregge l'effetto dovuto al diverso scattering della radiazione, rendendo la predizione del test set esterno più complessa. Visto il diverso effetto dei due pretrattamenti, che risolvono due problemi differenti ed entrambi rilevanti, si è pensato di utilizzarli in combinazione, nella successione $\log_{10}$ -SNV-centraggio sulla media, per verificare se insieme fossero in grado di correggere entrambe le problematiche.

Nel caso dei campioni in configurazione PETG-PLA si ha che l'applicazione di entrambi i pretrattamenti di riga porta ad ottenere un modello con 3 variabili latenti, con un RMSECV pari a 13,4 mm, valore pari a circa la metà dell'altezza dei campioni in esame. Un errore quasi al 50% è indice di una prestazione del modello non accettabile. Quando si va ad analizzare il grafico dei VIP (Figura 4-10A), si individuano tra le lunghezze d'onda rilevanti le stesse identificate fino ad ora, con le più importanti rappresentate sempre dall'intorno dei 1600 nm. Come si è accennato, l'errore effettuato dal modello in cross validazione è elevato e, come si può notare dalla Figura 4-10B dei valori predetti contro quelli noti per il training set, questo è dovuto principalmente al campione costituito interamente da PLA, che viene predetto a valori fortemente negativi. Per quanto riguarda i restanti campioni si riconosce comunque una difficoltà nella descrizione delle loro caratteristiche da parte del modello che è stato costruito. Inoltre, osservando i residui in Figura 4-10C si può individuare comunque un andamento leggermente non lineare, a sottolineare che il pretrattamento non ha avuto l'effetto desiderato.

Nella validazione del modello con il test set interno, rappresentata in Figura 4-10D, si osserva che RMSEP assume un valore paragonabile al corrispondente valore in calibrazione. Le migliori prestazioni del modello sul test set interno sono riconducibili all'assenza in esso del campione puro. Questo comportamento porta a supporre che la rimozione del primo campione del training set consentirebbe di ottenere delle performance soddisfacenti anche in cross validazione. I campioni del test set interno vengono infatti predetti in maniera ottimale, nonostante si possa riconoscere un incremento del rumore nei campioni ad altezza del PETG maggiore.

Quando si considera il test set esterno (Figura 4-10E) il valore di RMSEP aumenta nuovamente fino a un valore pari a 7,7 mm e si ha un bias in predizione molto elevato, di due ordini di grandezza maggiore rispetto al test set interno. Questo si riflette nella predizione dei campioni, che presenta delle difficoltà per tutti i campioni che sono stati analizzati. Tale risultato è inaspettato, dal momento che singolarmente i due pretrattamenti mostravano effetti positivi e di conseguenza ci si aspettava che la loro combinazione consentisse di unire gli aspetti positivi di entrambi. In realtà sembra che l'utilizzo del logaritmo vada ad intensificare gli effetti dello scattering nei dati, legati ai diversi riempimenti, e che SNV non riesca a ribilanciare in modo efficace questo effetto.

PLS Log10-SNV-Centraggio sulla media
Campioni 3 cm PETG-PLA

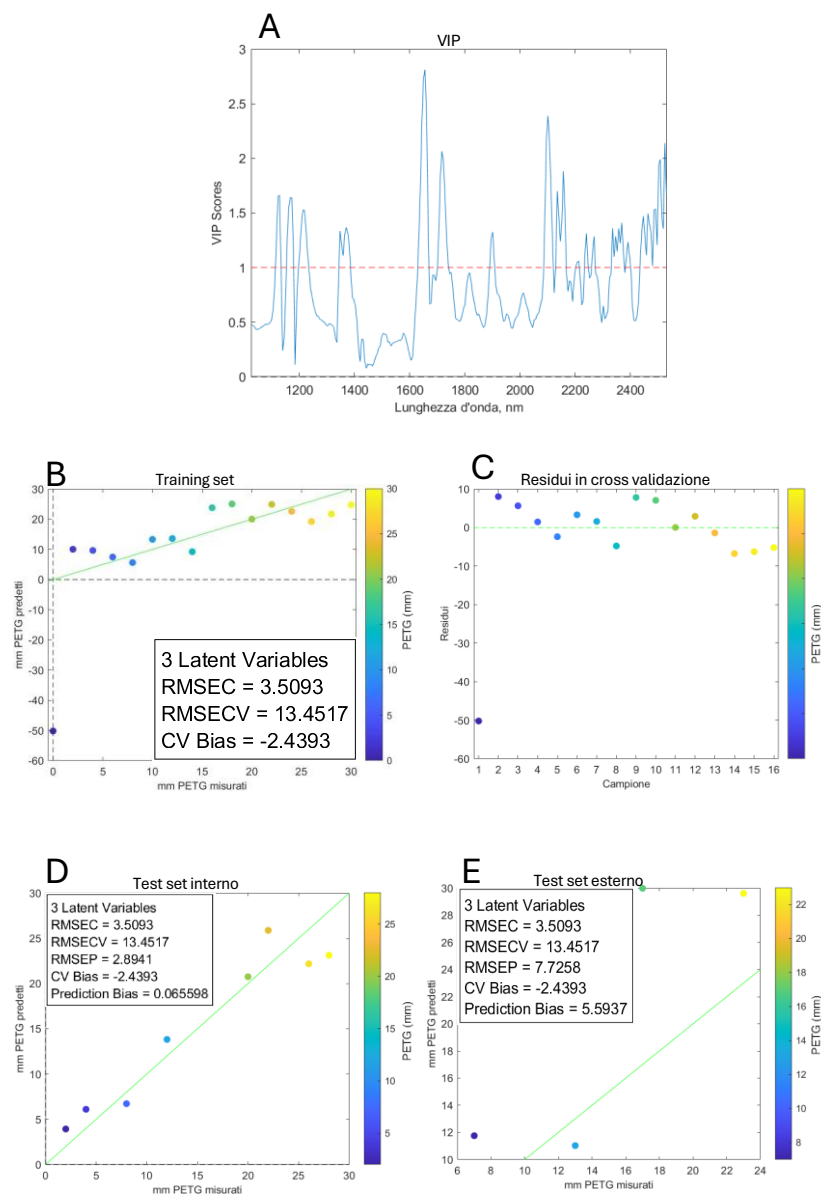


Figura 4-10: Modello PLS utilizzando come pretrattamento Log10-SNV-centraggio sulla media dei campioni di altezza 3 cm in configurazione PETG-PLA. A: VIP. B: Grafico dell'altezza predetta-misurata del training set. C: Residui in cross validazione del training set. D: Grafico dell'altezza predetta-misurata test set interno. E: Grafico dell'altezza predetta-misurata del test set esterno.

Il comportamento dei campioni in configurazione PLA-PETG sottoposti alla stessa associazione di pretrattamenti è diverso rispetto a quanto accade per i campioni PETG-PLA, a conferma del fatto che il materiale influenza il fenomeno della penetrazione.

Le lunghezze d'onda che influenzano la costruzione del modello, ottenuto usando 4 variabili latenti, rappresentano una delle poche analogie con la configurazione PETG-PLA. Si nota come il grafico dei VIP (Figura 4-11A) nei due casi sia pressoché identico. Ricoprono grande importanza le lunghezze d'onda intorno a 1600 nm e 2200 nm. Risultano significative anche quelle oltre 2400 nm, che probabilmente sono legate principalmente al rumore. Appaiono, invece, meno significative le zone intorno ai 1200 nm e ai 1400 nm, che dall'analisi spettrale sappiamo che portano informazione chimica di interesse. Nella predizione in cross validazione del training set (Figura 4-11B) abbiamo un valore di RMSECV pari a 5,4 mm, che risulta essere molto più contenuto rispetto a quanto osservato per la configurazione PETG-PLA con gli stessi pretrattamenti. Pur avendo ridotto notevolmente l'errore, rimane la difficoltà di predizione dei primi due campioni, 0 mm di PLA e 2 mm di PLA. Inoltre, i due risultano anche essere incoerenti tra loro, con 2 mm di PLA molto sottostimato. Ciò rappresenta un comportamento anomalo rispetto a quanto è accaduto utilizzando le due coppie di pretrattamenti precedenti sugli stessi campioni. I restanti campioni sono generalmente descritti dal modello in maniera abbastanza coerente con i loro valori reali. Pur non essendo la performance migliore ottenuta in questa trattazione, un aspetto interessante da notare è che i residui della Y in cross validazione (Figura 4-11C) hanno quasi del tutto perso l'andamento non lineare, che si ravvisa solo nel secondo e negli ultimi due campioni, il che testimonia che, in questo caso e sotto questo punto di vista, la coppia di pretrattamenti svolge la funzione desiderata.

Le abilità del modello in predizione sul test set interno (Figura 4-11D) sono analoghe a quanto accade in cross validazione, come testimoniato dai valori di RMSECV e RMSEP che sono simili tra di loro. La predizione diventa, invece, più complicata per quanto riguarda il test set esterno (Figura 4-11E). In questo caso il valore di RMSEP è pari a 14,3 mm, indicando che il modello non riesce a predire i campioni. Un altro fattore che sottolinea la difficoltà in predizione del modello su questo set di campioni è rappresentato dal bias in predizione che ha

un valore pari a -12,37 mm. Queste considerazioni sono confermate dal fatto che tre campioni su quattro sono predetti a valori negativi di Y. La motivazione di questo comportamento è legata all'utilizzo del logaritmo che ha intensificato gli effetti dello scattering che non sono poi corretti in maniera ottimale dalla trasformata SNV, come già suggerito nella configurazione PETG-PLA. Tale esacerbazione fa sì che i campioni del test set esterno, che presentano riempimento e quindi scattering diverso rispetto al training set, non riescano a essere predetti in modo accettabile dal modello.

PLS SLog10-NV-Centraggio sulla media
Campioni 3 cm PLA-PETG

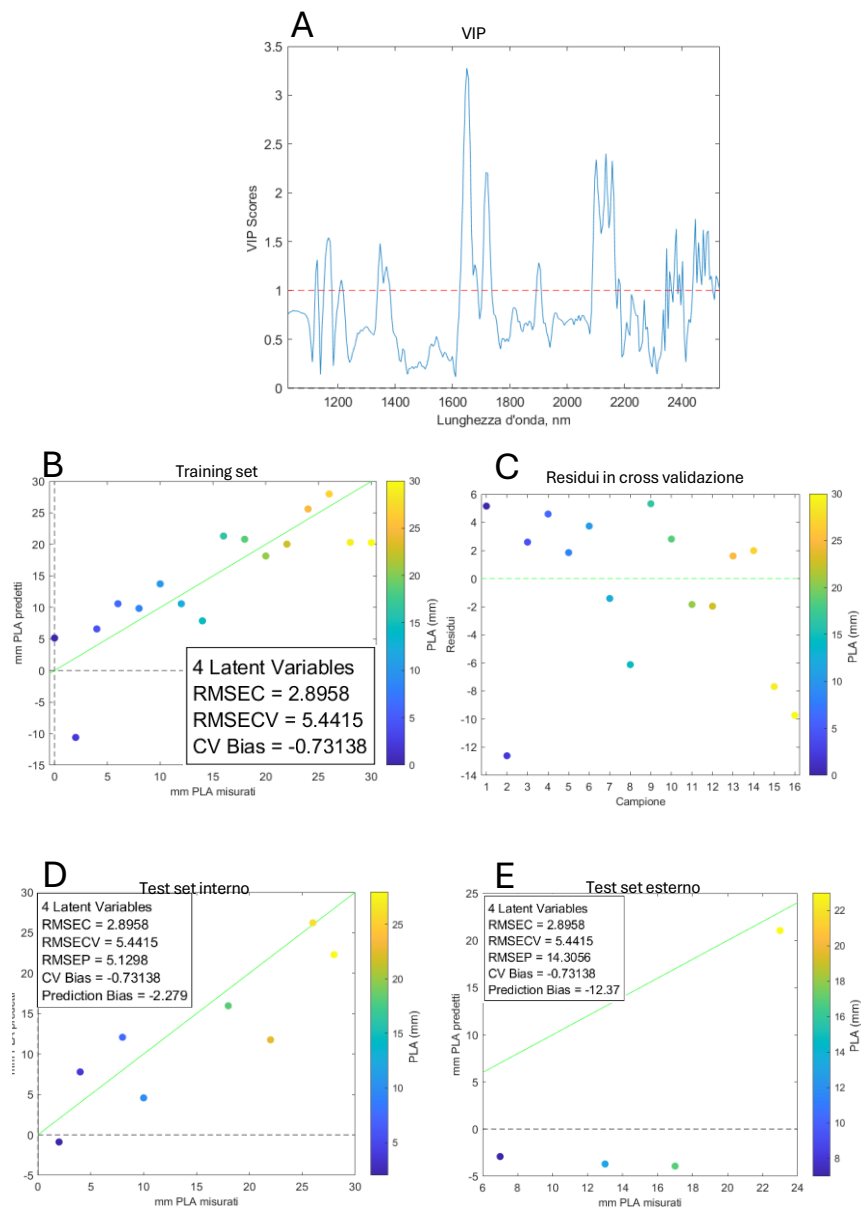


Figura 4-11: Modello PLS utilizzando come pretrattamento Log10-SNV-centraggio sulla media dei campioni di altezza 3 cm in configurazione PLA-PETG. A: VIP. B: Grafico dell'altezza predetta-misurata del training set. C: Residui in cross validazione del training set. D: Grafico dell'altezza predetta-misurata test set interno. E: Grafico dell'altezza predetta-misurata del test set esterno.

Dai risultati ottenuti sui campioni da 3 cm tramite l'applicazione dei pretrattamenti Log₁₀-SNV-centraggio sulla media si possono trarre alcune considerazioni. Innanzitutto, si nota come la combinazione dei due pretrattamenti, che singolarmente si sono rivelati utili per risolvere due problemi distinti, non abbia avuto l'effetto migliorativo desiderato, anzi si sono ottenuti anche dei risultati peggiori rispetto ai singoli pretrattamenti. Successivamente si può notare come l'ordine in cui i pretrattamenti vengono applicati influisce sul modello che si ottiene. Nel nostro caso la scelta di applicare prima il logaritmo e poi SNV è legata al fatto che SNV restituisce anche dei valori negativi che sono incompatibili con l'applicazione del logaritmo. Applicando i pretrattamenti in questo ordine, però, si osserva come il problema della non linearità del fenomeno è attenuato, mentre lo scattering della radiazione non è stato correttamente risolto, come si deduce dal fatto che la predizione sul test set esterno si è rivelata spesso insoddisfacente, come dimostrato dagli elevati valori di RMSEP.

Confrontando tra loro i vari modelli si possono definire gli aspetti positivi e negativi di ciascuna combinazione di pretrattamenti utilizzati. In generale la coppia trasformata SNV-centraggio sulla media riesce a mitigare gli effetti fisici presenti nei campioni, ma non è in grado di limitare la non linearità del fenomeno. Al contrario la combinazione logaritmo in base 10 e centraggio sulla media gestisce buona parte della non linearità ma fallisce nel limitare i fenomeni di natura fisica. La terza combinazione di pretrattamenti utilizzata, ovvero Log₁₀-SNV-centraggio sulla media, non presenta particolari vantaggi rispetto ai singoli pretrattamenti. Infatti, l'effetto di entrambi i pretrattamenti risulta essere mitigato correggendo solo parzialmente la non linearità e i fenomeni di natura fisica.

In Tabella 4-1 vengono riassunti i valori di RMSE e bias per tutti i set per tutti i modelli calcolati, in modo da riassumere i risultati per i campioni da 3 cm. Analizzando i valori di RMSE sui vari set di dati si possono riassumere le caratteristiche dei diversi pretrattamenti. In particolare si nota come l'utilizzo del log₁₀ fornisca dei valori di RMSEP sul test set interno più bassi rispetto a quelli che si ottengono utilizzando la trasformata SNV. Questo aspetto conferma la capacità del logaritmo di gestire la non linearità che caratterizza il fenomeno e consente di migliorare le prestazioni dei modelli rappresentati dallo stesso tipo di scattering. Al contrario,

come testimoniato dai valori di RMSEP maggiori rispetto a quelli ottenuti utilizzando SNV, la performance sul test set esterno peggiora. Tale comportamento è dovuto al fatto che il logaritmo non consente di gestire il diverso scattering generato dalle varie geometrie di riempimento dei campioni. D'altra parte la trasformata SNV, rispetto al logaritmo, fornisce delle performance leggermente peggiori sul training set e sul test interno perché non mitiga la non linearità del fenomeno, ma consente di ottenere modelli meno overfittati che permettono di descrivere anche campioni diversi. Le osservazioni fatte sui singoli pretrattamenti avevano portato alla supposizione che la loro combinazione avrebbe potuto portare dei benefici nella descrizione del fenomeno. Tuttavia si riscontra, fatta eccezione per il test set interno in configurazione PETG-PLA, un peggioramento delle performance dei modelli ottenuti, come testimoniato dai valori di RMSE più elevati per tutti i set di dati.

L'analisi dei valori di RMSE permette anche di individuare il miglior pretrattamento per ciascuna configurazione. Nel caso dei campioni PETG-PLA la coppia di pretrattamenti SNV-centraggio sulla media fornisce dei valori di RMSEC, RMSECV e RMSEP, su entrambi i test set, paragonabili tra di loro, suggerendo una buona prestazione solida su tutti i set di dati. Per quanto riguarda le altre combinazioni di pretrattamenti si hanno dei valori di RMSEC minori, rispetto a quello di SNV-centraggio sulla media, ma i valori di RMSECV e RMSEP sono significativamente maggiori rispetto all'errore in calibrazione. Pertanto, la coppia di pretrattamenti SNV-centraggio sulla media fornisce una migliore descrizione dell'andamento dei campioni, nonostante non corregga la non linearità del fenomeno. La coppia di pretrattamenti SNV-centraggio sulla media fornisce una buona descrizione del fenomeno anche per i campioni PLA-PETG. In particolare, nonostante il rapporto tra RMSECV e RMSEC sia più elevato di quanto auspicabile, fornisce una buona descrizione del fenomeno anche nei test set, specialmente in quello esterno. La predizione dei cilindri è l'area in cui le restanti combinazioni di pretrattamenti riscontrano più difficoltà, fino a raggiungere un errore di 14,3 mm con l'utilizzo del $\log_{10}$ -SNV-centraggio sulla media.

Tabella 4-1: Riassunto dei valori di RMSE e di bias per ciascuna configurazione e ciascun pretrattamento per i campione di altezza 3 cm.

Configurazione	Pretrattamento	Training set		Cross validazione		Test set interno		Test set esterno	
		RMSE (mm)	Bias (mm)	RMSE (mm)	Bias (mm)	RMSE (mm)	Bias (mm)	RMSE (mm)	Bias (mm)
PETG/PLA	SNV +MC	4,5	0	5,9	-0,6	4,6	-0,8	4,5	1,9

PETG/PLA	log10 +MC	2,3	0	5,2	1,1	3,3	0,2	6,5	4,4
PETG/PLA	log10+ SNV +MC	3,5	0	13,4	-2,4	2,8	-2,4	7,7	5,6
PLA/PETG	SNV +MC	0,6	0	2,5	0,6	4,5	-1,0	2,8	-1,3
PLA/PETG	log10 + MC	2,5	0	5,1	-0,3	4,2	-2,4	5,6	-4,8
PLA/PETG	log10+ SNV + MC	2,9	0	5,4	-0,7	5,1	-2,3	14,3	-12,4

4.2 CAMPIONI 5 CM

Successivamente si sono effettuate le stesse analisi chemiometriche sui campioni da 5 cm, per verificare ulteriormente la profondità massima di penetrazione della radiazione e per valutare se lo spessore del campione sia un fattore che influenza il fenomeno. Nel condurre queste osservazioni si sono utilizzati gli stessi pretrattamenti già descritti per i campioni da 3 cm, in virtù delle osservazioni scaturite da questi.

4.2.1 Spettri grezzi e pretrattati

Nell'analisi degli spettri si riscontrano dei fenomeni analoghi rispetto a quanto riscontrato nei campioni da 3 cm, e in particolare per quanto riguarda all'interpretazione spettrale delle bande si rimanda al paragrafo 4.1.1.

Osservando gli spettri grezzi dei campioni in configurazione PETG-PLA si è evidenziato anche in questo caso un effetto di intensità globale (Figura 4-12A-C-E) generalmente scorrelato allo spessore del polimero sovrastante, che è stato corretto applicando come pretrattamento la trasformata SNV (Figura 4-12B-D-F). Dagli spettri del training set (Figura 4-12B) si nota, in particolar modo nelle zone intorno ai 1200 nm, 1400 nm e 1600 nm, che i primi campioni presentano un andamento più simile a quello del PLA rispetto che a quello del PETG. Questo conferma che si rilevano informazioni chimiche relative agli strati di PLA che si trovano al di sotto del PETG. Di conseguenza, la radiazione penetra attraverso tutto lo strato sovrastante del PETG fino ad altezze di circa 10 mm. Al contrario, sia per i campioni del test set interno (Figura 4-12D) sia per quelli del test set esterno (Figura 4-12E), non è stato

possibile individuare differenze spettrali correlate all'altezza dello strato polimerico superiore nei campioni. Infatti, gli spettri relativi al test set interno sono risultati completamente sovrapposti. I campioni del test set esterno hanno invece mostrato delle leggere differenze in termini di intensità globale, anche dopo pretrattamento con la trasformata SNV, ma queste differenze non sono correlate con l'informazione di interesse e derivano probabilmente dall'uso di geometrie di riempimento diverse.

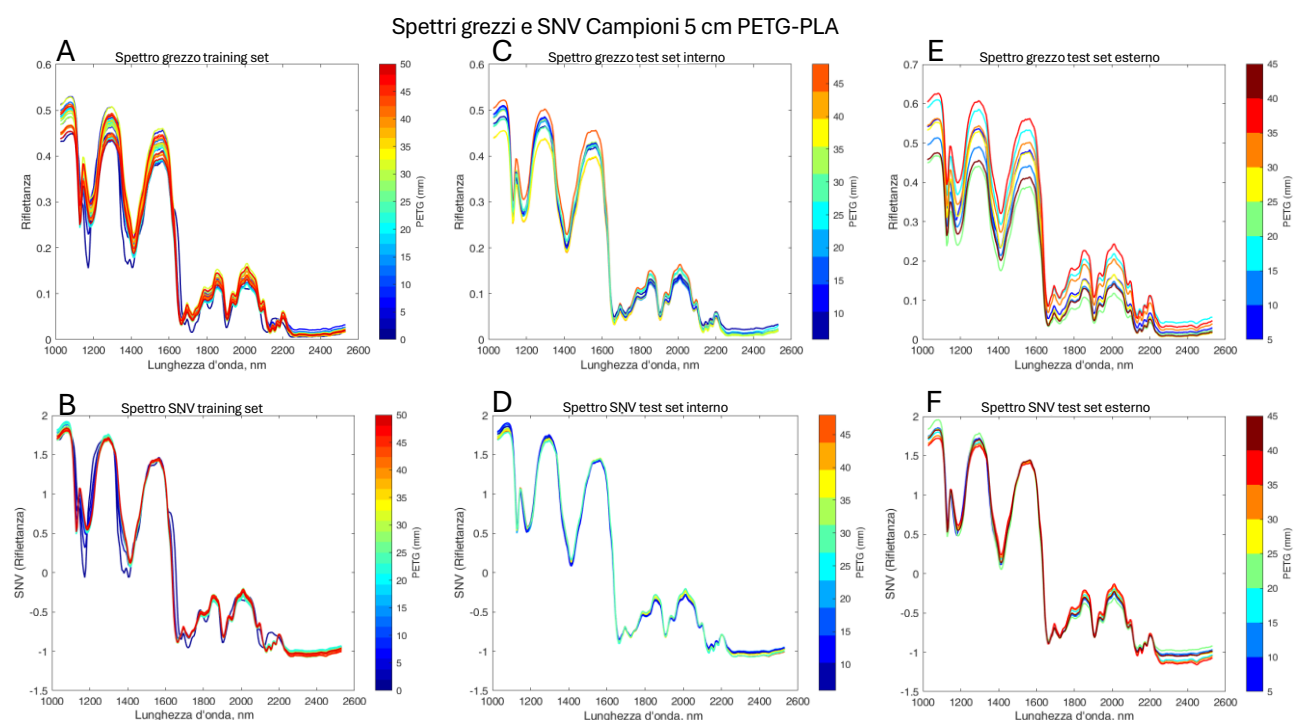


Figura 4-12: Spettri dei campioni con altezza 5 cm in configurazione PETG-PLA. A: Spettro grezzo del training set. B: Spettro SNV del training set. C: Spettro grezzo del test set interno. D: Spettro SNV del test set interno. E: Spettro grezzo del test set esterno. F: Spettro SNV del test set esterno.

Nei campioni in configurazione PLA-PETG, gli spettri grezzi del training set e del test interno mostrano un effetto di intensità globale parzialmente correlato all'informazione chimica di interesse, come già precedentemente riscontrato nei campioni da 3 cm. Questo effetto non si riscontra nel test set esterno, in cui l'intensità globale sembra essere correlata ad effetti fisici. Per questo motivo e perché nel training set e nel test set interno il trend non è completamente ordinato, si è scelto di pretrattare ugualmente i dati con SNV. In questo modo si riesce a porre l'attenzione solamente sugli effetti chimici e non su quelli fisici. Le differenze spettrali che caratterizzano il training set (Figura 4-13B) sono particolarmente evidenti negli spettri colorati di blu (altezza di PLA fino a circa 10 mm). Le lunghezze d'onda più rilevanti si trovano intorno

a 1200 nm, in cui si individua la presenza di una banda tipica di PETG fino a valori di y di circa 1, e attorno a 1600 nm, in cui per gli stessi campioni è assente la spalla della banda che caratterizza i campioni ricchi di PLA. Inoltre, in questo caso, si notano delle differenze anche nella zona attorno a 2000 nm, dove si riconoscono alcuni spettri caratterizzati dall'aver un solo massimo centrale, rispetto ai due massimi che si osservano per i campioni con un elevato contenuto di PLA. A differenza di quanto accade per i campioni in configurazione PETG-PLA, in questo caso si possono intravedere delle differenze spettrali anche per i campioni che costituiscono sia il test set interno (Figura 4-13D) sia il test set esterno (Figura 4-13F). In entrambi i casi si può individuare un campione che presenta il picco del PETG a 1200 nm circa e l'assenza della spalla del PLA a 1600 nm. Questo aspetto confermerebbe ulteriormente l'ipotesi per cui la radiazione penetri meglio attraverso il PLA rispetto al PETG.

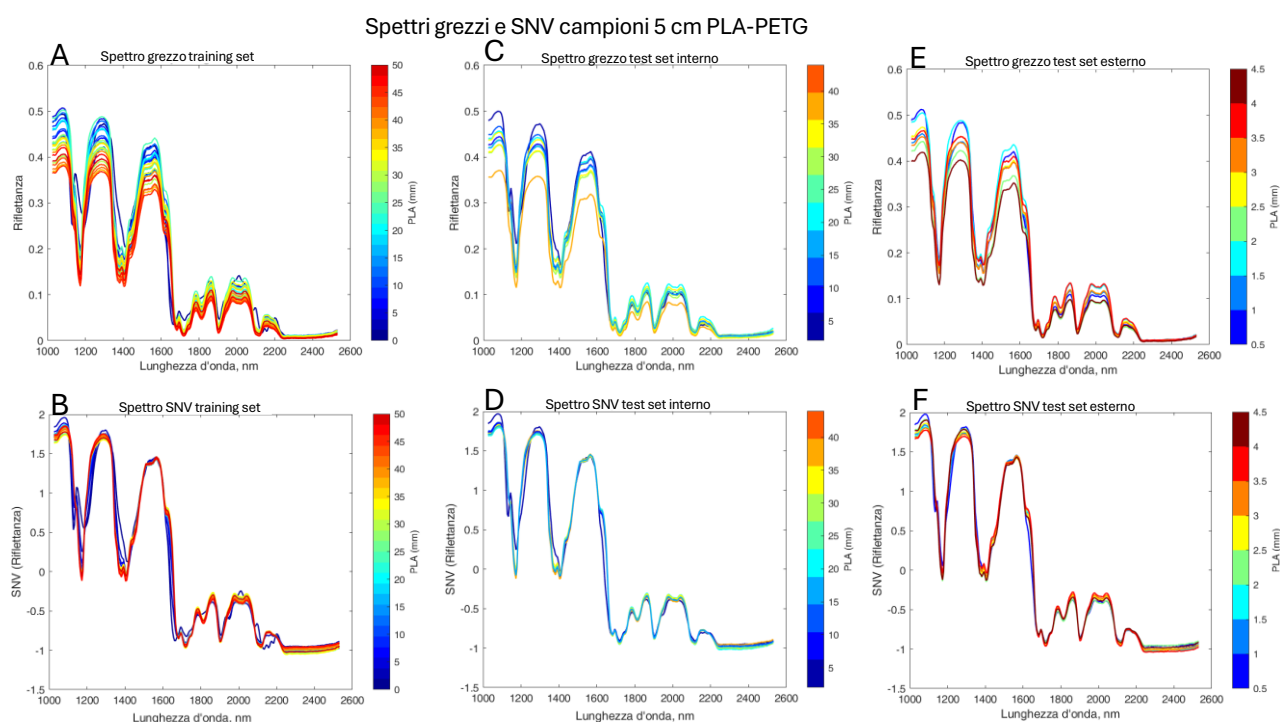


Figura 4-13: : Spettri dei campioni con altezza 5 cm in configurazione PLA-PETG. A: Spettro grezzo del training set. B: Spettro SNV del training set. C: Spettro grezzo del test set interno. D: Spettro SNV del test set interno. E: Spettro grezzo del test set esterno. F: Spettro SNV del test set esterno.

Osservando gli spettri dei campioni da 5 cm si osserva come entrambi i materiali sembrino essere caratterizzati da una profondità di penetrazione significativa, dal momento che in entrambe le configurazioni si riconoscono delle differenze fino a 10 mm di polimero superiore. Inoltre, troverebbe conferma l'ipotesi che la radiazione NIR penetri meglio attraverso

il PLA. Questa, infatti, è sostenuta dal fatto che all'interno del test set interno ed esterno della configurazione PLA-PETG si individuino delle differenze spettrali, aspetto che non viene riscontrato nella configurazione PETG-PLA.

4.2.2 Analisi delle componenti principali

Per verificare e ampliare le osservazioni e le ipotesi fatte osservando gli spettri si è effettuata, su entrambi i set di dati una analisi delle componenti principali. Come sui campioni da 3 cm, anche in questo caso si è utilizzata la trasformata SNV come pretrattamento di riga e il centraggio sulla media come pretrattamento di colonna.

Nei campioni in configurazione PETG-PLA, i cui risultati si trovano in Figura 4-14, si ha che l'informazione relativa alla profondità di penetrazione della radiazione si trova in PC1, che spiega l'82,02% della varianza, confermando che l'informazione che si sta ricercando all'interno dei campioni è la fonte più rilevante di variabilità. L'influenza delle variabili nella costruzione di tale componente principale è rappresentata dai loading presentati in Figura 4-14A. Le lunghezze d'onda intorno a 1200 nm sono caratterizzate dalla presenza di due bande positive e di una negativa. In particolare la banda negativa si trova in corrispondenza del picco caratteristico del PETG a queste lunghezze d'onda. Di conseguenza, i regoli con un contenuto maggiore di PETG si dovrebbero trovare a valori negativi di PC1. Al contrario, la zona intorno a 1600 nm determina valori positivi di PC1 e questa regione corrisponde alla spalla che si osserva nello spettro del PLA. Di conseguenza, i campioni che presentano questo andamento si troveranno a valori positivi di PC1 e sono, coerentemente con quanto osservato a livello spettrale, i primi campioni (quelli con maggiore quantità di PLA). L'andamento dei loading è pressoché identico a quello descritto per i campioni da 3 cm (Figura 4-4A), a confermare che il fenomeno che viene descritto è analogo per entrambi i set di dati.

Le osservazioni tratte dai loading trovano riscontro nel grafico degli score del training set, come si vede in Figura 4-14B. I campioni caratterizzati da un basso contenuto di PETG si trovano a valori positivi di PC1, mentre al crescere del contenuto di PETG si ha un progressivo spostamento verso valori negativi di PC1. Questo accade fino a ottenere un andamento asintotico, intorno al campione numero 7, con uno spessore dello strato di PETG pari a 12 mm. Oltre questo regolo si ha che tutti i punti si trovano allo stesso valore di PC1, suggerendo il raggiungimento di un limite nella penetrazione della radiazione. Inoltre, si individua anche in

questo caso un andamento non lineare del fenomeno, dal momento che la quantità di polimero aumenta con passo costante, ma questo non avviene per gli score su PC1 che variano in modo non lineare.

Nel proiettare i campioni che costituiscono il test set interno (Figura 4-14C) e il test set esterno (Figura 4-14D) sulle componenti principali del training set si osserva un comportamento dei dati analogo a quello descritto per la Figura 4-14B. Nel test set interno, in particolare, si osserva un andamento decrescente per gli score dei primi tre campioni, con altezza massima di PETG pari a 12 mm, mentre dal quarto in poi tutti i campioni iniziano a essere caratterizzati da score simili. Anche per il test set esterno si ha un andamento simile agli altri set di dati, in cui i primi due campioni si trovano a score molto diversi lungo PC1, mentre i successivi non presentano differenze lungo questa componente.

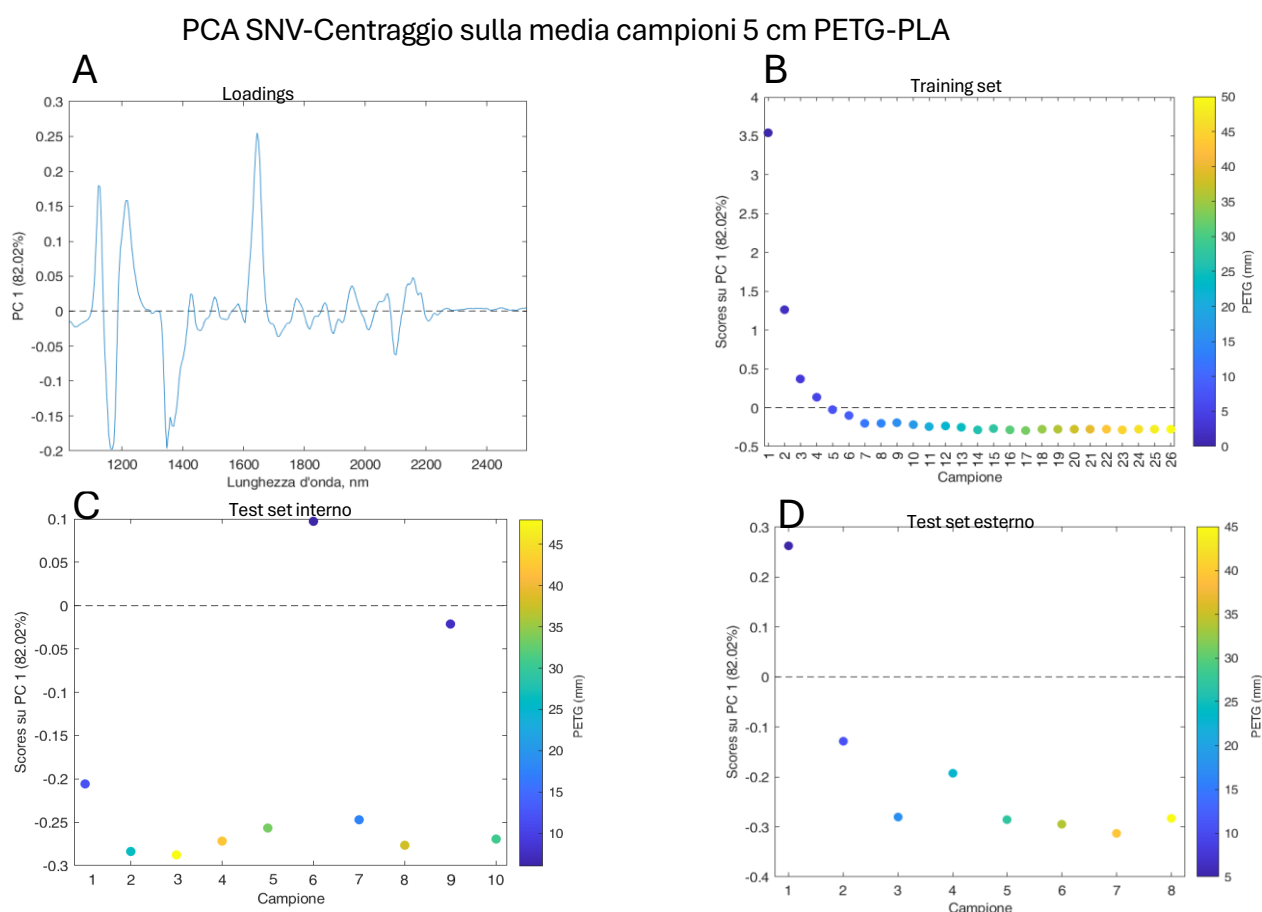


Figura 4-14: PCA dei campioni di altezza 5 cm in configurazione PETG-PLA. A: Loading di PC1. B: Score su PC1 del training set. C: Score su PC1 del test set interno. D: Score su PC1 del test set esterno.

Per quanto riguarda i campioni in configurazione PLA-PETG, l'informazione di interesse si trova di nuovo sulla PC1 che spiega l'83.81% della varianza presente nei dati. Analogamente a quanto accadeva nel caso dei campioni da 3 cm, anche in questo caso i loading dei campioni acquisiti nelle due configurazioni sono pressoché uguali, come si può notare confrontando Figura 4-14A e Figura 4.15A. Si ha quindi che in entrambi i casi le PC1 sono influenzate in maniera analoga dalle stesse lunghezze d'onda, e questo si traduce in un andamento speculare dei campioni lungo la PC1. Infatti, nella configurazione PLA-PETG i campioni costituiti maggiormente da PLA, ovvero gli ultimi della serie, si presentano sempre a valori positivi di score su PC1, mentre quelli caratterizzati da un contenuto maggiore di PETG si trovano a valori negativi di score. Osservando il training set (Figura 4-15B), si nota che i campioni iniziano a stabilizzarsi su valori costanti di score su PC1 intorno al campione numero 9, che corrisponde a uno strato di PLA pari a 16 mm. Nel test set interno (Figura 4-15C) il campione caratterizzato da 2 mm di PLA si trova a score minori rispetto ai restanti. Un aspetto inusuale, che si evidenzia per la prima volta e che è in contrasto con le osservazioni fatte, è la presenza di un campione caratterizzato da un elevato contenuto di PLA a valori negativi di PC1. Questo comportamento inusuale è probabilmente spiegato dalla presenza di rumore spettrale inaspettato, come ad esempio un'ombra nell'acquisizione. Quando si analizza il test set esterno (Figura 4-15D), si ritrova l'andamento atteso in cui i primi tre campioni, con altezza di PLA massima pari a 17 mm, sono disposti a valori crescenti lungo la PC1.

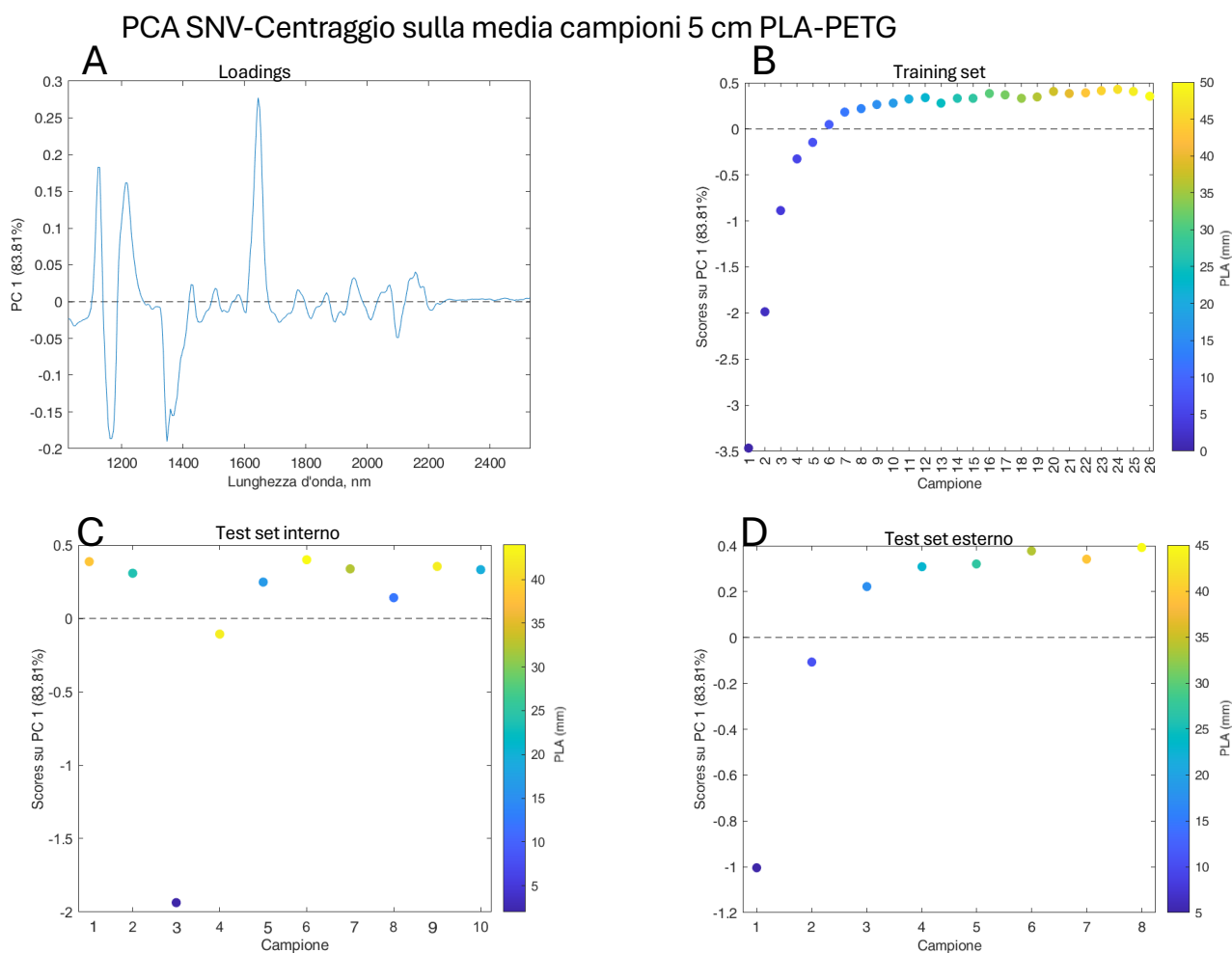


Figura 4-15: PCA dei campioni di altezza 5 cm in configurazione PLA-PETG. A: Loading di PC1. B: Score su PC1 del training set. C: Scores su PC1 del test set interno. D: Score su PC1 del test set esterno.

L'analisi delle PCA dei campioni da 5 cm ha mostrato dei risultati simili a quanto accade nei campioni alti 3 cm. Si riconosce anche in questo caso una maggiore profondità di penetrazione attraverso il PLA. Risulta evidente anche la natura non lineare del fenomeno come dimostrato dagli score molto diversi che caratterizzano i primi tre campioni.

4.2.3 Regressione ai minimi quadrati parziali

Dopo aver analizzato il fenomeno tramite un metodo chemiometrico non supervisionato, si è proceduto andando a utilizzare un metodo supervisionato per dare rilevanza ancora maggiore all'informazione di nostro interesse. L'obiettivo, analogamente a quanto fatto per i

campioni da 3 cm, è quello di andare a valutare la profondità di penetrazione della radiazione con l'idea che un modello con buone performances indichi una penetrazione efficace nel range di altezza considerato.

Visti i buoni risultati ottenuti in PCA con l'applicazione della trasformata SNV seguita dal centraggio sulla media, questi due pretrattamenti sono stati usati anche in questo caso come primo passo per l'analisi di regressione PLS.

Nei campioni in configurazione PETG-PLA il numero di variabili latenti che consente di ottenere il modello con il miglior rapporto tra RMSECV e RMSEC è pari a tre. Le variabili che si sono rilevate più importanti in questa costruzione sono evidenziate nel grafico dei VIP (Figura 4-16A) e sono, anche in questo caso, in primo luogo le zone intorno a 1600 nm seguite da quelle a 1200 e 1400 nm.

Andando ad analizzare i grafici dei valori predetti contro quelli noti, si osserva che il modello presenta delle difficoltà nella descrizione dei primi due regoli del training set (Figura 4-16B), che vengono predetti a valori di Y negativi. I restanti campioni seguono in linea di massima l'andamento atteso. Il motivo di questo risultato potrebbe essere legato alle grandi differenze spettrali dei primi due campioni, derivanti dalla non linearità del fenomeno. Tale osservazione sarebbe anche sostenuta dalla PCA, in cui questi campioni (Figura 4-14B) sono caratterizzati da score molto diversi rispetto ai restanti regoli. Pertanto, la difficoltà della predizione dei primi due campioni deriverebbe dal tentativo di modellare un fenomeno non lineare con un metodo di regressione lineare. Per quanto riguarda i campioni caratterizzati da un contenuto di PETG maggiore o uguale a 4 mm si ha una predizione in cross validazione relativamente buona, e questo sembra accadere fino a circa 20-25 mm. Oltre questo valore si osserva la rottura del modello, che si manifesta con la predizione dei campioni restanti alla stessa altezza, cosa che suggerisce una fine della penetrazione efficace della radiazione in questo polimero.

L'analisi dei residui (Figura 4-16C) permette di individuare un andamento simil-parabolico, con i primi e gli ultimi campioni a residui negativi e i campioni centrali quasi tutti a residui positivi o solo leggermente negativi. Questo comportamento va a sostenere l'ipotesi della non linearità del fenomeno di penetrazione della radiazione NIR.

Nella validazione con il test set interno (figura 4-16D), in cui il valore di RMSEP è paragonabile a quello di RMSEC, si osserva che i campioni oltre i 25 mm del PETG vengono predetti a valori simili di Y, in accordo con il training set. Il comportamento del test set interno

sembrerebbe confermare quanto si è osservato nel training set, ovvero una buona predizione fino a 20 mm del PETG. Nel test set esterno si ha invece un peggioramento della prestazione del modello, come è indicato dall'aumentare del valore di RMSEP e dalla presenza di un bias positivo in predizione. Nonostante questo, ad eccezione di due campioni che vengono fortemente sottostimati, si può riconoscere la presenza di un andamento crescente su quasi tutto il range di calibrazione.

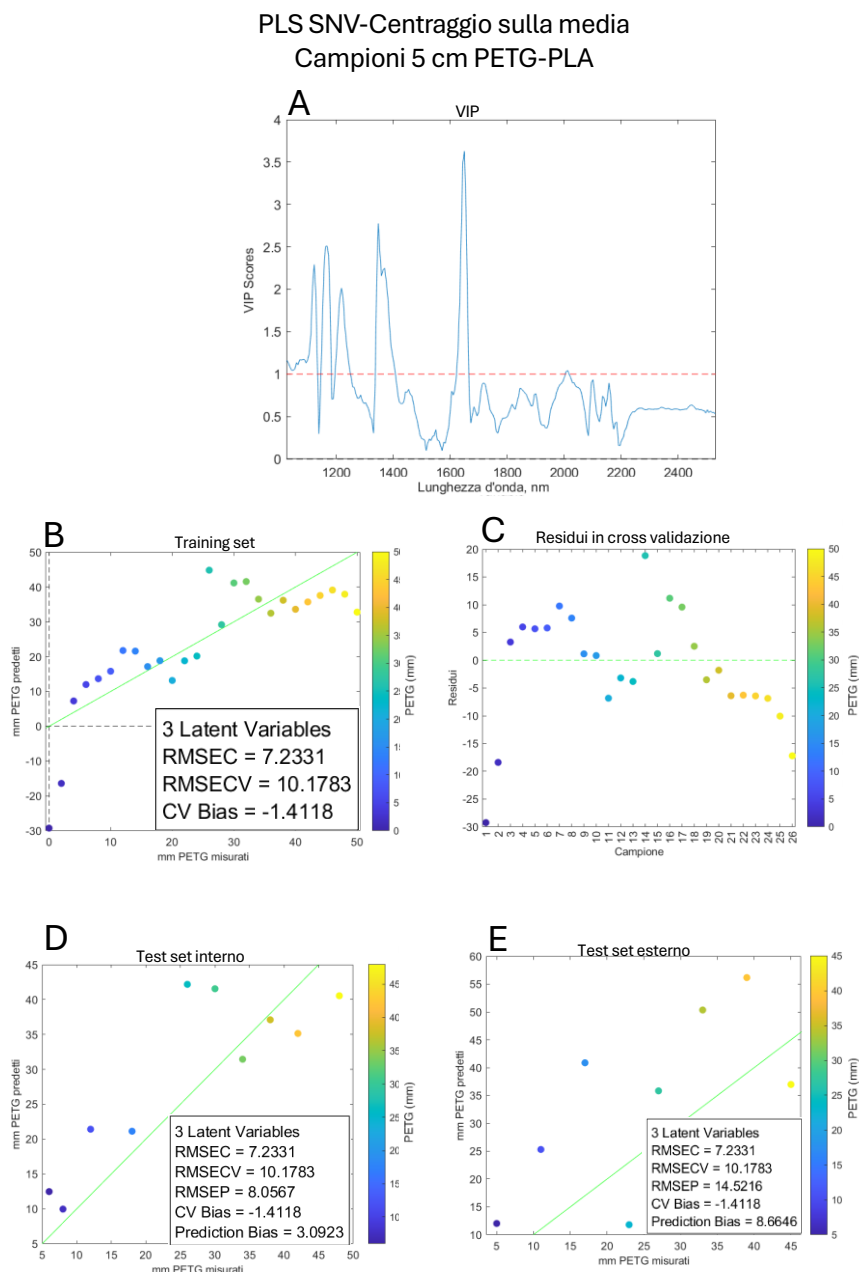


Figura 4-16: Modello PLS, utilizzando come pretrattamento SNV-centraggio sulla media, dei campioni di altezza 5 cm in configurazione PETG-PLA. A: VIP. B: Grafico dell'altezza predetta-misurata del training set. C: Residui in cross validazione del training set. D: Grafico dell'altezza predetta-misurata test set interno. E: Grafico dell'altezza predetta-misurata del test set esterno.

Per i campioni analizzati in configurazione PLA-PETG, il modello è stato costruito utilizzando 5 variabili latenti, valore che consente di avere RMSEC e RMSECV paragonabili e relativamente contenuti per lo scopo di interesse. Nello sviluppo del modello, tra gli intervalli spettrali che si sono rilevati più importanti, come rappresentato nel grafico dei VIP (Figura 4-17A), oltre quelli evidenziati fino ad ora si ha anche la presenza delle lunghezze d'onda comprese nell'intervallo 2100-2200 nm. A queste lunghezze d'onda negli spettri del training set (Figura 4-13B) si possono individuare delle differenze nello spettro del campione puro del PETG e dei campioni a basso contenuto di PLA rispetto a quelli con una quantità maggiore di PETG, suggerendo che anche questa zona spettrale faccia riferimento a informazione chimica relativa ai due polimeri.

Le prestazioni del modello, su questo set di dati, sono migliori rispetto a quelle ottenute per i campioni in configurazione PETG-PLA trattati con gli stessi pretrattamenti, sia nella fase di calibrazione che in quella di validazione. Nel training set (Figura 4-17B) si riscontra sempre una certa difficoltà nella predizione del primo campione, anche se l'effetto è più contenuto rispetto ai campioni in configurazione PETG-PLA, come si nota dai valori minori di RMSEC e RMSECV. Le prestazioni non ottimali su questo campione possono essere spiegate dal fatto che a livello spettrale è nettamente diverso, essendo l'unico a non contenere PLA, e, data la non linearità dei dati, il suo modellamento risulta complicato. L'aspetto anomalo da sottolineare è che questo campione viene predetto a valori di PLA pari a 25 mm, cosa che porterebbe a supporre che siano presenti delle similitudini a livello chimico tra il campione puro di PETG e i regoli ricchi di PLA. Ciò tuttavia non è supportato a livello spettrale, in quanto le differenze tra i due profili sono molto marcate. Maggiori analisi saranno necessarie in futuro per comprendere al meglio questo evento.

Passando ai campioni successivi al primo, essi vengono predetti correttamente dal modello, nonostante la presenza di rumore, fino ad una altezza di circa 36 mm di PLA, valore oltre il quale si presenta una rottura del modello dimostrato dall'assenza di un andamento crescente. Nel modello sembrano inoltre riconoscersi due gruppi di campioni: il primo costituito dai campioni con altezza da 0 a 25 mm di PLA e il secondo dai campioni con un contenuto di PLA superiore a 25 mm, segnando una perfetta divisione a metà del set di dati. Questo comportamento può essere legato a un fenomeno fisico derivato dalla modalità di stampa dei regoli. Analizzando i residui in cross validazione (Figura 4-17C) si riconosce un andamento

non lineare. Infatti, escludendo il primo campione, si può individuare la presenza di due andamenti parabolici ciascuno dei quali descrive uno dei due gruppi individuati.

Nei campioni in configurazione PLA-PETG il test set interno (Figura 4-17D) è descritto coerentemente con quanto accade nel training set, come si può osservare dai valori paragonabili di RMSECV e RMSEP. Inoltre, si ha una predizione buona anche del campione con 2 mm di PLA, che nella fase di costruzione del modello è predetto con maggiore difficoltà. Anche nel test set interno si riconosce una zona di rottura del modello, che caratterizza i campioni oltre i 30 mm di PLA, a conferma di quanto si è osservato nel training set.

Nel test set esterno (Figura 4-17E) si ha un RMSEP più simile a RMSECV, se confrontato con quanto accade nei campioni in configurazione PETG-PLA, pur rimanendo elevato il bias in predizione. In generale, si ha una buona predizione di tutti i campioni del test set esterno, eccetto quelli con il contenuto rispettivamente minore e maggiore di PLA, il che è comprensibile considerando le osservazioni fatte per i regoli.

Pertanto, considerando il test set esterno, si potrebbe affermare che la radiazione sia in grado di penetrare fino circa a 40 mm di profondità nel PLA. Questa è una profondità molto elevata, che non si è osservata precedentemente né in questo studio né in letteratura. Riferendoci al presente studio, analizzando i campioni alti 3 cm solitamente si individuava un deterioramento del modello attorno a profondità massime di circa 20 mm, cosa che aveva portato a supporre che questo fosse il limite di penetrazione in questo materiale. Il raggiungimento di buone predizioni a valori così alti di profondità per i campioni da 3 cm potrebbe essere dovuto a una disponibilità maggiore di campioni nel range di altezze che sono caratterizzate da bassa penetrazione, ovvero con un elevato contenuto di PLA. Questo aumento del campionamento potrebbe consentire al modello di avere dati a sufficienza per rilevare anche piccole variazioni spettrali non visibili a occhio nudo, e di conseguenza permetterebbe di modellare meglio il comportamento in questa zona rispetto a quanto accade per i regoli da 3 cm, in cui il campionamento a profondità elevate è più limitato. Inoltre, il fatto che questo comportamento si sia osservato nei campioni in configurazione PLA-PETG potrebbe essere un ulteriore elemento a conferma dell'ipotesi per cui la radiazione penetra meglio attraverso il PLA.

PLS SNV-Centraggio sulla media
Campioni 5 cm PLA-PETG

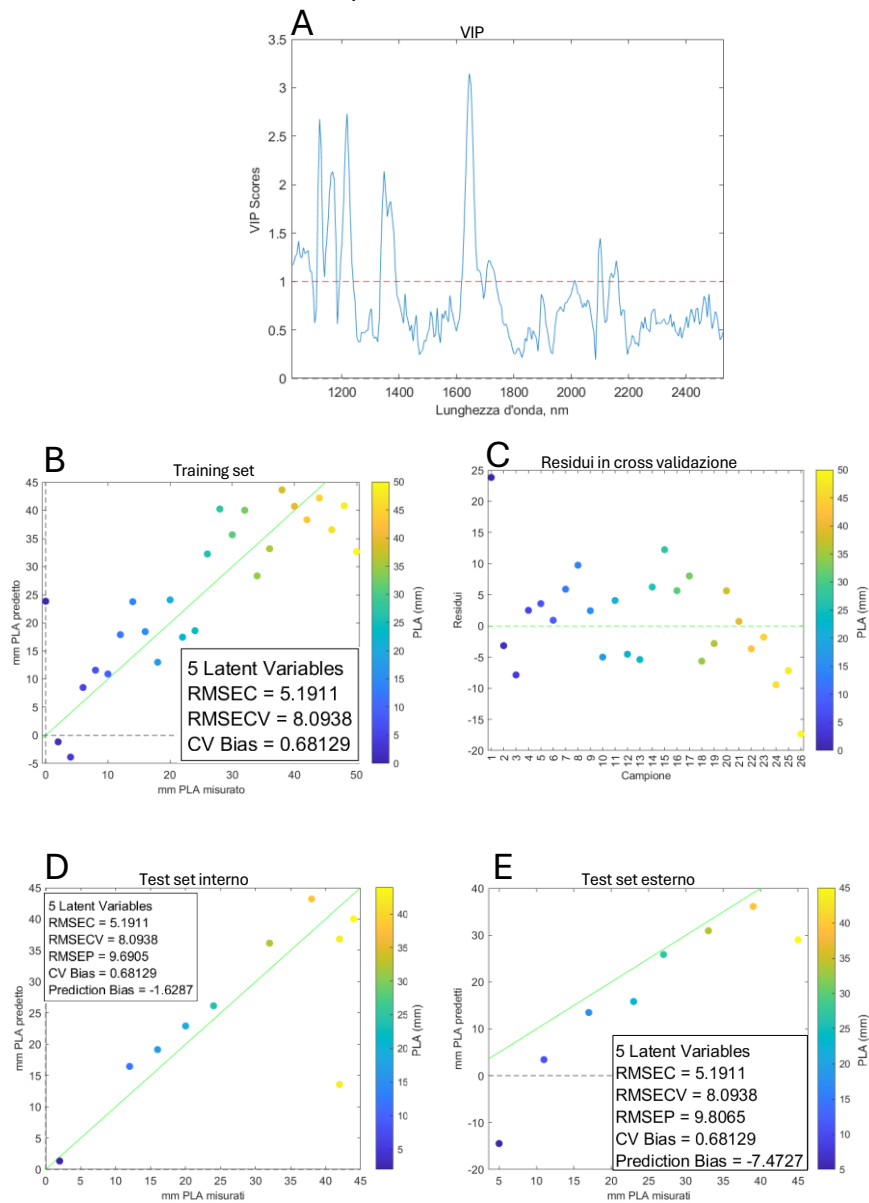


Figura 4-17: Modello PLS, utilizzando come pretrattamento SNV-centraggio sulla media dei campioni, di altezza 5 cm in configurazione PLA-PETG. A: VIP. B: Grafico dell'altezza predetta-misurata del training set. C: Residui in cross validazione del training set. D: Grafico dell'altezza predetta-misurata test set interno. E: Grafico dell'altezza predetta-misurata del test set esterno.

Dal momento che il comportamento non lineare si è osservato in entrambe le configurazioni anche per i campioni da 5 cm, si è scelto di provare anche in questo caso ad applicare il logaritmo in base decimale per correggere questo effetto.

Utilizzando il $\log_{10}$ e il centraggio sulla media sui campioni analizzati in configurazione PETG-PLA si è quindi costruito un nuovo modello di regressione con tre variabili latenti. Nell'analizzare i VIP (Figura 4-18A) si osserva che le lunghezze d'onda nella regione attorno a 1200 nm e 1400 nm, che tipicamente sono molto rilevanti nel modello, perdono di importanza, tanto da superare di poco il limite di significatività. Si mantiene costante invece l'importanza della zona a 1600 nm, e acquisisce particolare peso l'intervallo di lunghezze d'onde che va da 2100 nm fino a 2500 nm. Se intorno a 2200 nm c'era ancora la possibilità di riscontrare delle differenze spettrali tra i campioni (Figura 4-12A), a 2400 nm si osserva prevalentemente rumore negli spettri. L'importanza di queste variabili nella costruzione del modello suggerisce un maggiore modellamento del rumore con questo pretrattamento, cosa che potrebbe poi rivelarsi importante nella successiva prestazione del modello, in particolar modo durante la fase di validazione.

Nell'analizzare la predizione del training set in cross validazione (Figura 4-18B) si riscontra in generale una performance abbastanza buona del modello, nonostante i valori siano abbastanza rumorosi. Fatta eccezione per il campione puro di PLA, predetto a valori negativi di Y, si ha che i regoli fino a quasi 3 cm vengono descritti bene, mentre oltre a questa altezza non vengono più individuate differenze tra i campioni, che conseguentemente vengono predetti alla stessa Y. Osservando i residui (Figura 4-18C) non si osserva un andamento particolare, tranne marginalmente per gli ultimi campioni, a testimoniare che l'utilizzo del logaritmo è stato utile nel limitare la non linearità insita nei dati.

Procedendo alla validazione con il test set interno (Figura 4-18D) si ottengono delle prestazioni del modello conformi con la calibrazione, in particolare si hanno valori molto simili di RMSECV e RMSEP. Ciò porterebbe a considerare il modello come ottimale, ma la validazione con il test set esterno (Figura 4-18E) smentisce questo fatto. Infatti, si arriva ad ottenere un bias pari a -13 mm e RMSEP di 28 mm. Questi valori, soprattutto l'errore in predizione, indicano che il modello non è in grado di riconoscere nessuna proprietà dei campioni e questo si osserva in maniera evidente dai valori di Y che vengono predetti per diversi cilindri. Questo evento è spiegato dalla mancanza di un pretrattamento che consenta di correggere il diverso scattering che i cilindri presentano per via del diverso riempimento che li caratterizza. Non correggendo questo effetto, lo scattering della radiazione viene modellato durante la fase di calibrazione rendendo possibile la predizione del test set interno, mentre quella del test set esterno risulta compromessa.

PLS Log10-Centraggio sulla media
Campioni 5 cm PETG-PLA

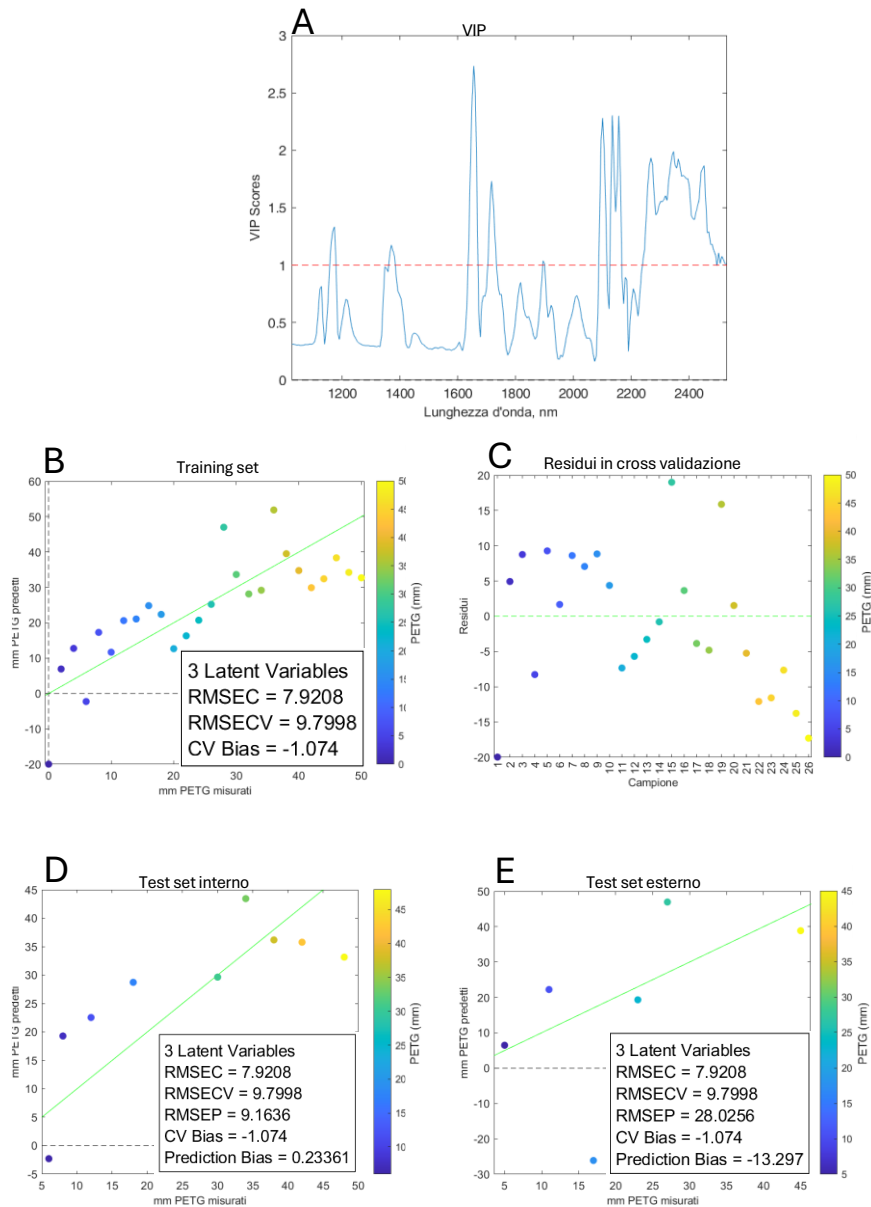


Figura 4-18: Modello PLS utilizzando come pretrattamento Log_{10} -centraggio sulla media dei campioni di altezza 5 cm in configurazione PETG-PLA. A: VIP. B: Grafico dell'altezza predetta-misurata del training set. C: Residui in cross validazione del training set. D: Grafico dell'altezza predetta-misurata test set interno. E: Grafico dell'altezza predetta-misurata del test set esterno.

Nel caso dei campioni acquisiti in configurazione PLA-PETG, utilizzando come pretrattamento il logaritmo in base 10 e il centraggio sulla media, si riscontrano delle similitudini con quanto accade nel caso precedente. In particolare, le analogie sono presenti a livello delle variabili che influenzano la costruzione del modello (Figura 4-19A), in cui oltre

alle variabili indicate precedentemente si osserva una rilevanza ancora maggiore delle lunghezze d'onda intorno ai 2400 nm che acquisisce circa la stessa importanza dei 1600 nm, pur non essendo una zona diagnostica a livello spettrale. Aumenta quindi anche in questo caso la modellazione del rumore.

Il modello che si è costruito è caratterizzato da 4 variabili latenti e presenta dei valori di RMSEC e RMSECV paragonabili. La predizione dei campioni del training set in cross validazione (Figura 4.19B), tuttavia, non risulta essere ottimale. Si riscontrano infatti delle caratteristiche simili ai casi precedenti, ovvero la difficoltà nella predizione dei campioni a basso contenuto del polimero che sta sopra e la presenza di molto rumore. Nonostante queste caratteristiche si può individuare, escludendo il primo campione, un andamento crescente lungo tutto il range di interesse, nonostante nella seconda parte del range molti campioni successivi siano predetti a valori costanti di Y. Sebbene nel modello si possa individuare l'andamento crescente, si riconosce infatti anche la presenza elevata di rumore che determina una prestazione del modello peggiore rispetto a quella ottenuta sullo stesso set di dati utilizzando la trasformata SNV e il centraggio sulla media. Nei residui della Y in cross validazione (Figura 4-18C) inoltre si può riconoscere, soprattutto nella seconda metà dei campioni, un andamento non lineare che sottolinea che in questo caso il fenomeno che si voleva cercare di correggere con il pretrattamento è ancora almeno in parte presente.

Durante la fase di validazione del modello con il test set interno (Figura 4-19D) le prestazioni che si ottengono sono analoghe a quanto accade durante la calibrazione, predicendo in maniera soddisfacente i campioni eccetto gli ultimi 3. Il valore di RMSEP è leggermente più alto rispetto al corrispondente errore in cross validazione e questo può essere attribuito prevalentemente al campione con 42 mm di PLA che viene predetto a valori di Y abbastanza anomali rispetto alla sua composizione (Figura 4-19D). L'aspetto più interessante da notare si ritrova nella fase di validazione del modello con il test set esterno. Infatti, dalle prestazioni del modello dei campioni PETG-PLA con gli stessi pretrattamenti, ci si aspetterebbe che i campioni del test set esterno vengano predetti in modo casuale, vista l'influenza dello scattering della radiazione che non viene corretta da nessun pretrattamento. Invece, in questo caso il comportamento del modello si dimostra migliore rispetto a quanto accade nel test set interno. Il valore di RMSEP del test set esterno è paragonabile con RMSECV e RMSEC, il bias in predizione è minore rispetto allo stesso parametro del test set interno e in generale si osserva una buona predizione dei cilindri lungo quasi tutte le altezze di PLA. Sembrerebbe che anche in questo caso, come accadeva per i campioni nella stessa configurazione pretrattati con SNV,

si possa ritenere accettabile la predizione dei cilindri fino ad un'altezza di 35 mm di PLA, evidenziando nuovamente una profondità che non si era riscontrata precedentemente.

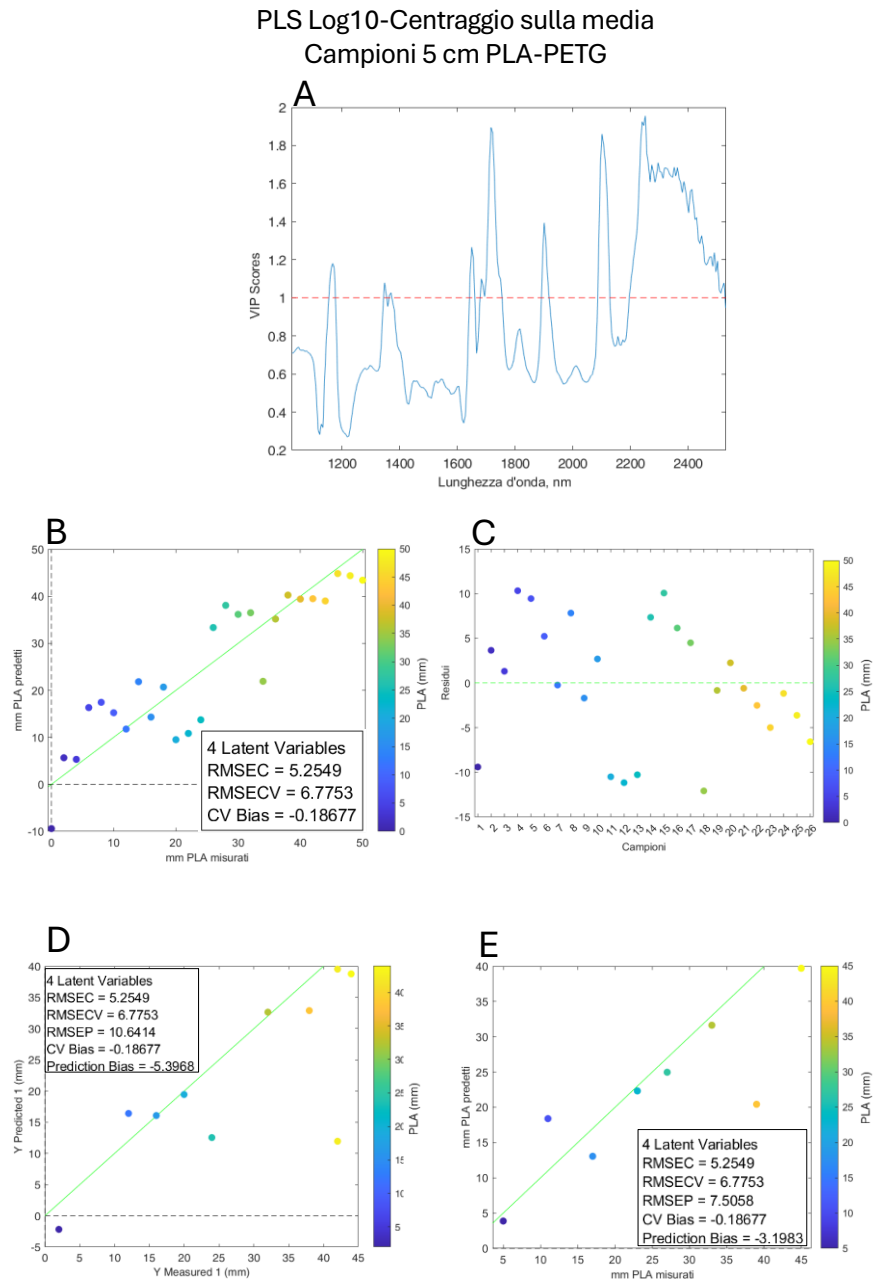


Figura 4-19: Modello PLS utilizzando come pretrattamento Log₁₀-centraggio sulla media dei campioni di altezza 5 cm in configurazione PLA-PETG. A: VIP. B: Grafico dell'altezza predetta-misurata del training set. C: Residui in cross validazione del training set. D: Grafico dell'altezza predetta-misurata test set interno. E: Grafico dell'altezza predetta-misurata del test set esterno.

Nel tentativo di ottimizzare i risultati che si sono ottenuti con i diversi pretrattamenti, anche per i campioni da 5 cm si è deciso di provare ad applicare, come pretrattamenti di riga,

sia il logaritmo in base 10 che SNV, mantenendo sempre come pretrattamento di colonna il centraggio sulla media.

Nei campioni in configurazione PETG-PLA si sono scelte 3 variabili latenti per il modello in questa configurazione. Le lunghezze d'onda più rilevanti nella sua costruzione si trovano oltre i 1600 nm, che, pur essendo presenti delle bande caratteristiche dei due polimeri, non è sicuramente la zona spettrale in cui le differenze sono più evidenti. Inoltre, l'applicazione di questi pretrattamenti determina una riduzione della rilevanza dell'area intorno a 1200 nm e a 1400 nm, che sono invece zone particolarmente caratteristiche. Una zona che invece sembra acquisire molta rilevanza è quella intorno ai 2100 nm, che rappresenta l'impronta digitale dello spettro. Rispetto ai VIP dei corrispondenti campioni pretrattati con il solo logaritmo in base 10 si ha una leggera riduzione della rilevanza della zona oltre i 2200 nm. Si dovrebbe quindi avere una riduzione del modellamento del rumore, probabilmente grazie all'effetto esercitato da SNV.

Nella calibrazione del modello (Figura 4-20B) si osservano dei fenomeni simili rispetto ai casi precedenti. In particolare, si ripresenta la predizione del campione puro a valori negativi, sempre per via delle evidenti differenze spettrali rispetto agli altri regoli. Risulta invece attenuata la difficoltà di predizione dei campioni successivi, che sono, in generale, descritti bene dal modello fino a circa 10 mm di PETG, valore oltre il quale si osserva una prima rottura del modello. Infatti, i campioni da 12 mm di PETG fino a 24 mm di PETG vengono predetti allo stesso valore di Y, a testimoniare che non si riconoscono più differenze tra questi campioni. Oltre i 24 mm si riconosce un altro gruppo di regoli, predetti tutti alla stessa altezza anche se separati rispetto a quelli precedenti. La presenza di questi due gruppi diversi è probabilmente legata a un fenomeno di natura fisica dovuto alla modalità con cui sono stati stampati i campioni. Infatti, il polimero presente in minore quantità è sempre stato stampato per primo, rimanendo quindi aderente alla superficie di stampa. Questo fa sì che l'estremità del regolo con il polimero presente in minore quantità sia liscia, mentre quella che presenta il polimero maggioritario è ruvida. Questo fenomeno determina quindi un diverso scattering della radiazione tra la prima e la seconda metà della serie di campioni, che può essere responsabile della generazione dei due diversi gruppi che si osservano in questo caso. Infatti, la generazione dei due diversi gruppi si presenta proprio alla metà dei campioni, dove è stata effettuata l'inversione di stampa. Questo aspetto sembra suggerire che i fenomeni fisici, con questa combinazione di pretrattamenti, non vengono corretti completamente. L'effetto fisico che determina la divisione nei due gruppi si riscontra anche nei residui. Infatti, osservando i residui

di Y in cross validazione (Figura 4-20C) si nota come questi abbiano, in ognuno dei due gruppi, prima tutti valori positivi che tendono poi a diminuire fino a diventare negativi.

Gli effetti legati ai pretrattamenti descritti fino a ora si possono osservare anche nei risultati che si ottengono con la validazione. Infatti, il test set interno (Figura 4-20D) viene predetto con delle prestazioni simili a quanto accade durante la fase di calibrazione del modello, come dimostrato dal valore di RMSEP paragonabile rispetto al valore di RMSECV. Anche il valore del bias in predizione risulta contenuto. In questo caso, si potrebbe anche ipotizzare che il modello sia in grado di predire fino a 20 mm del PETG, a differenza del training set, dove a questa altezza si riscontrava già il primo gruppo predetto a valori costanti. Tuttavia, si deve considerare che nel training il primo gruppo di rottura viene predetto proprio a un'altezza di circa 20 mm. Di conseguenza questa considerazione deve essere fatta con cautela.

Per il test set esterno (Figura 4-20E) vale invece un discorso diverso, che si inizia a individuare già da un valore di RMSEP maggiore. Nel caso dei cilindri si riscontra una maggiore difficoltà nella predizione, a testimonianza del fatto che i fattori fisici che influenzano il fenomeno della penetrazione della radiazione NIR non sono stati corretti completamente. Tuttavia, bisogna sottolineare che l'effetto non è così rilevante come accade in Figura 4-18E e questo è dovuto al fatto che comunque si ha l'effetto dovuto all'applicazione di SNV.

PLS Log10-SNV-Centraggio sulla media
Campioni 5 cm PETG-PLA

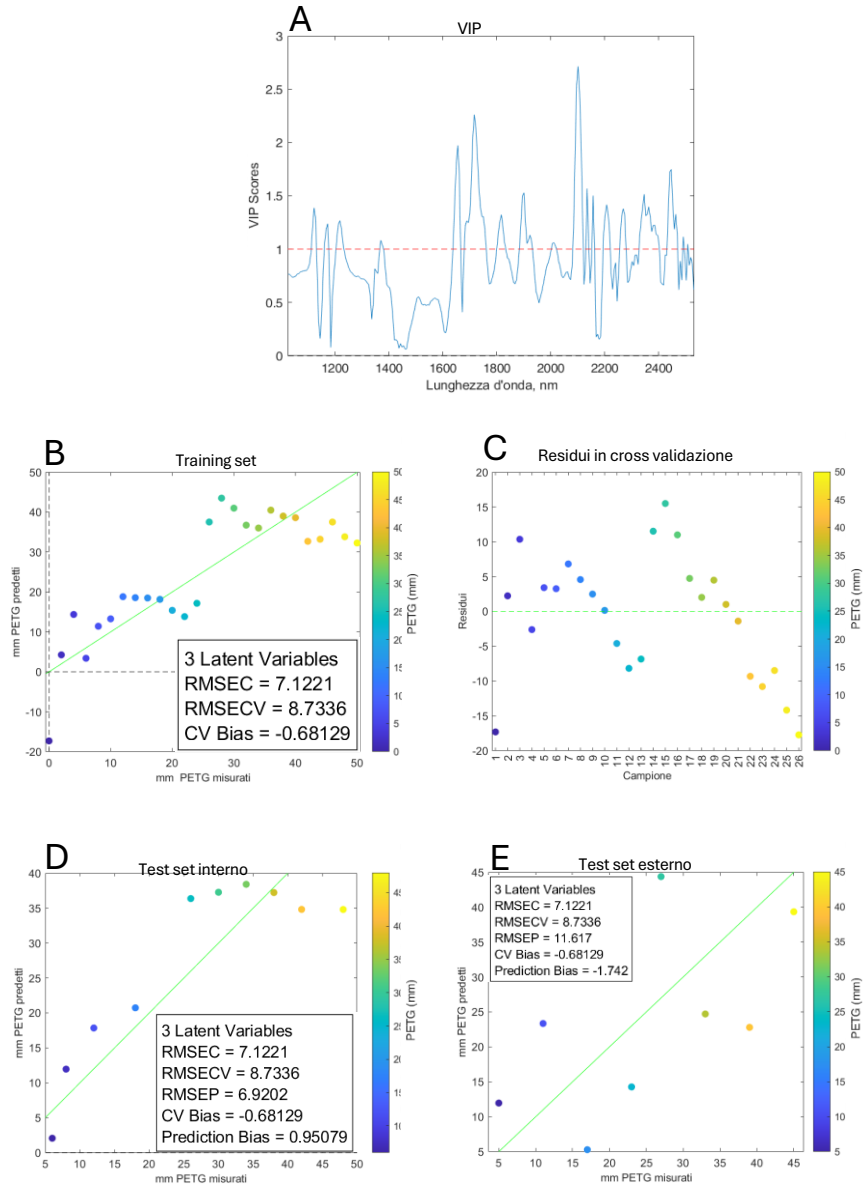


Figura 4-20: Modello PLS, utilizzando come pretrattamento Log₁₀-SNV-centraggio sulla media, dei campioni di altezza 5 cm in configurazione PETG-PLA. A: VIP. B: Grafico dell'altezza predetta-misurata del training set. C: Residui in cross validazione del training set. D: Grafico dell'altezza predetta-misurata test set interno. E: Grafico dell'altezza predetta-misurata del test set esterno.

Nel caso dei campioni acquisiti in configurazione PLA-PETG si riscontra un comportamento simile a quanto accade nei campioni in configurazione PETG-PLA. Infatti, anche in questo caso si ha che le zone più rilevanti nella costruzione del modello (Figura 4-21A) si trovano dopo i 1600 nm. Un'altra similitudine è che le variabili intorno ai 1200 nm e ai

1400 nm, importanti per la costruzione del modello con i pretrattamenti SNV-centraggio sulla media, non ricoprono più un ruolo centrale. La similitudine dei VIP tra i modelli costruiti per le due configurazioni si riflette anche in analogie con la prestazione che si ottiene nella fase di calibrazione (Figura 4-21B). Una delle costanti è rappresentata dalla difficoltà nella descrizione dei primi campioni. Nella configurazione PLA-PETG questa difficoltà è maggiore in quanto caratterizza i primi tre campioni (altezza di PLA pari a 4 mm). Un'altra caratteristica comune tra le due configurazioni è la presenza di due gruppi di campioni predetti ad altezze costanti. La differenza tra i due modelli ottenuti sta nel fatto che, nel caso di quello costruito per i dati PETG-PLA, si aveva una buona predizione dei regoli fino a circa 10 mm di PETG. In questo caso, invece, si individuano prevalentemente i due gruppi in cui sembrerebbe rompersi il modello, anche se si riconosce comunque un leggero andamento crescente. Questo trend è leggermente più evidente nel secondo gruppo, caratterizzato da un'altezza di PLA superiore a 25 mm. La presenza dell'andamento crescente nei due gruppi si riscontra anche a livello dei residui (figura 4-21C) in quanto è meno evidente, soprattutto nel secondo gruppo, la diminuzione costante del residuo.

Nella fase di validazione si riconosce per il test set interno (Figura 4-21D) una prestazione paragonabile al training set, testimoniata dai valori analoghi di RMSEP e RMSECV. La predizione risulta difficoltosa nel caso del campione da 2 mm di PLA, come accade nel training set, e per gli ultimi tre campioni (oltre i 40 mm di PLA). I restanti campioni sono descritti correttamente dal modello. Nel test set esterno (Figura 4-21E) si ha un aumento di RMSEP, seppur non in maniera critica come quando si utilizza solamente il logaritmo. Questo errore maggiore si traduce in difficoltà di predizione di tutti i campioni, in cui risulta evidente l'effetto del bias in predizione, pari a -9 mm, che determina una sottostima di gran parte dei cilindri. Nonostante i cilindri siano sottostimati, si riconosce all'interno di essi un andamento crescente. Questi fattori sottolineano l'influenza delle variabili fisiche sulla descrizione del fenomeno, ma suggerirebbero anche una penetrazione efficace della radiazione in tutto l'intervallo di interesse, fenomeno già notato sui campioni in questa configurazione con entrambe le combinazioni di pretrattamenti precedenti.

PLS SLog10-NV-Centraggio sulla media
Campioni 5 cm PLA-PETG

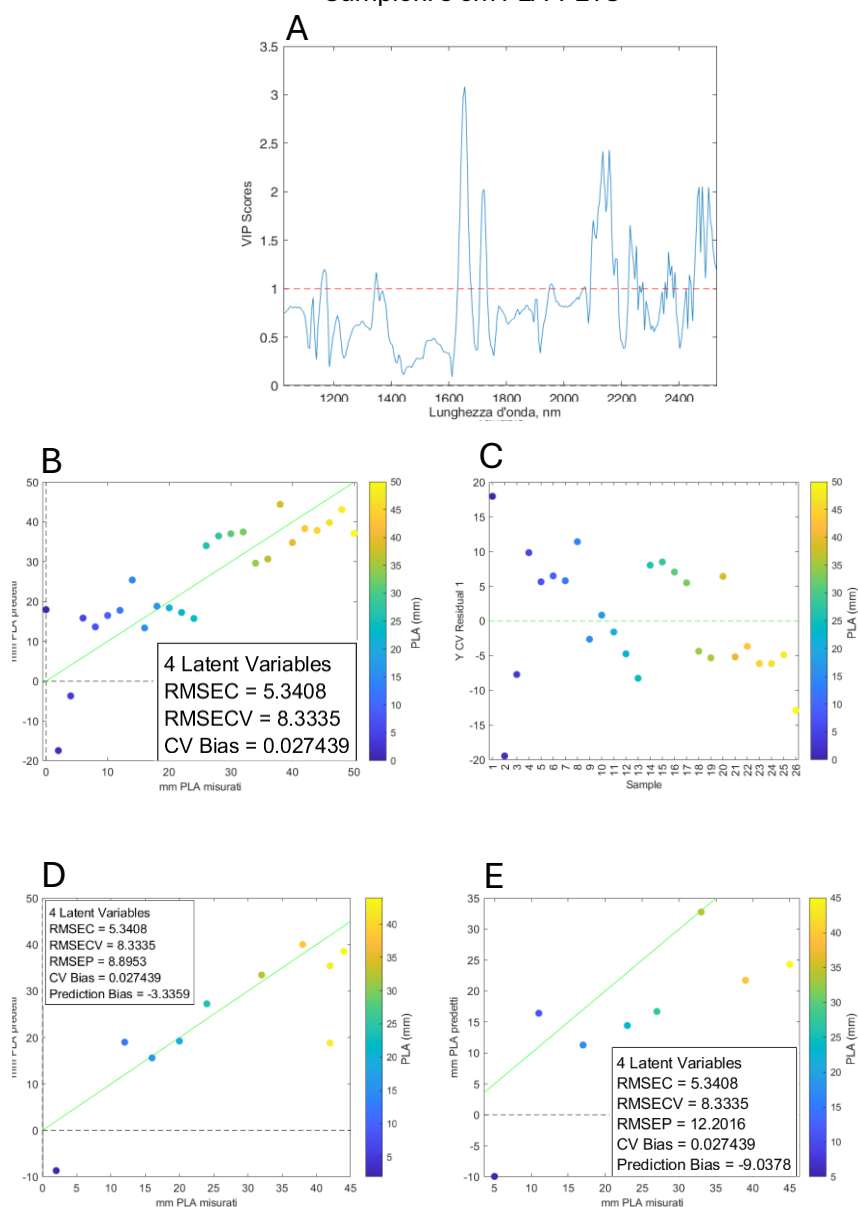


Figura 4-21: Modello PLS utilizzando come pretrattamento Log₁₀-SNV-centraggio sulla media dei campioni di altezza 5 cm in configurazione PLA-PETG. A: VIP. B: Grafico dell'altezza predetta-misurata del training set. C: Residui in cross validazione del training set. D: Grafico dell'altezza predetta-misurata test set interno. E: Grafico dell'altezza predetta-misurata del test set esterno.

In Tabella 4-2 vengono riassunti i valori di RMSE e di bias dei modelli costruiti per i campioni da 5 cm in modo da riassumerne le caratteristiche. Nel caso dei campioni da 5 cm l'effetto principale di ciascuna combinazione di pretrattamenti è analogo a quello descritto nel paragrafo 4.1.3. Se si considera quello che accade su ciascuna delle due configurazioni si possono fare diverse considerazioni. Analizzando i campioni in configurazione PETG-PLA si

nota come gli errori in calibrazione dei tre modelli siano simili tra di loro. Queste prestazioni si riscontrano similmente anche in cross validazione e nel test set interno. Quando si considerano i test set esterni si ha un aumento dell'errore. In questo caso è presente una notevole differenza che caratterizza i dati pretrattati con Log_{10} e il centraggio sulla media in quanto il valore di RMSEP diventa pari a 28 mm. Questo dato testimonia la necessità di un pretrattamento che sia in grado di correggere i fenomeni fisici legati allo scattering, compito che non viene svolto dal logaritmo. Gli altri due pretrattamenti forniscono prestazioni simili, anche se si può sottolineare che nel caso dei campioni pretrattati con Log_{10} -SNV-centraggio sulla media si hanno dei valori di RMSE minori. Questo porterebbe a presupporre una migliore descrizione del fenomeno da parte di questa combinazione di pretrattamenti. Tuttavia, in questo studio l'interesse è posto prevalentemente sullo studio dell'andamento dei campioni che si osserva nei grafici dell'altezza predetta-misurata, portando a ritenere il modello migliore quello costruito con SNV-centraggio sulla media.

Nella configurazione PLA-PETG tutti e tre i modelli sono, generalmente, caratterizzati da buone prestazioni in validazione, sia sul test set esterno che su quello interno. In particolare, si nota che tutti i valori di RMSE sono minori rispetto ai corrispondenti valori della configurazione PETG-PLA suggerendo una migliore penetrazione della radiazione attraverso il PLA. Anche nei casi in cui le performance del modello sembrano essere leggermente peggiori perché caratterizzate da valori di RMSE più alti o da bias elevati, come nel caso del test set esterno pretrattato con Log_{10} -SNV-centraggio sulla media, si riconosce comunque un andamento crescente dei campioni che potrebbe essere una conferma della migliore penetrazione della radiazione NIR attraverso il PLA rispetto al PETG. In questa configurazione il modello Log_{10} -centraggio sulla media è caratterizzato da buoni valori di RMSE, specialmente per quanto riguarda il test set esterno. Questo aspetto va in contrasto con quanto osservato sia nei campioni da 3 cm che per i campioni da 5 cm in configurazione PETG-PLA. La differenza rispetto a quanto accade nei campioni da 3 cm è dovuta al maggior numero di regoli, nel set da 5 cm, con altezze di polimero superiore in cui la radiazione penetra poco. Un maggior numero di campioni consente di rilevare delle differenze minori. L'utilizzo di un pretrattamento che mitiga la non linearità permette di modellare in maniera soddisfacente queste piccole differenze. Le migliori prestazioni del logaritmo sui campioni da 5 cm PLA-PETG, rispetto alla configurazione opposta, sono dovute alla migliore penetrazione della radiazione attraverso il PLA che consente di individuare maggiori differenze, rendendo più efficace l'utilizzo del logaritmo.

Tabella 4-2: Riassunto dei valori di RMSE e di bias per ciascuna configurazione e ciascun pretrattamento per i campione di altezza 5 cm

Configurazione	Pretrattamento	Training set		Cross validazione		Test set interno		Test set esterno	
		RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
PETG/PLA	SNV +MC	7,2	0	10,1	-1,4	8,1	3,0	14,5	8,6
PETG/PLA	log10 +MC	7,9	0	9,8	-1,1	9,1	0,2	28,0	-13,2
PETG/PLA	log10+SNV +MC	7,1	0	8,7	-0,7	6,9	0,9	11,6	-1,7
PLA/PETG	SNV +MC	5,2	0	8,1	0,7	9,7	-1,6	9,8	-7,5
PLA/PETG	log10 + MC	5,5	0	6,8	-0,2	10,6	-5,4	7,5	-3,2
PLA/PETG	log10+SNV + MC	5,3	0	8,3	0,03	8,9	-3,3	12,2	-9,0

5 CONCLUSIONI

Questo lavoro ha permesso di individuare e di confermare la rilevanza di fattori chimici e fisici nell'influenzare il fenomeno oggetto di studio. Tra i fattori chimici si è evidenziata la diversa penetrazione in funzione del materiale polimerico incontrato dalla radiazione, come dimostrato dalla maggior penetrazione attraverso il PLA quando confrontato con il PETG. Tra i fenomeni fisici si è sottolineata, in particolare, la rilevanza di fattori connessi alla microstruttura interna della materia da penetrare, in questo caso studio determinata dal diverso riempimento dei regoli e dei cilindri che, causando uno scattering diverso della radiazione, determina proprietà diverse nella sua penetrazione. Queste proprietà, se non sono corrette in maniera appropriata per mezzo di pretrattamenti chemiometrici, influenzano la possibilità di modellare il fenomeno oggetto di studio. La rilevanza dei fenomeni fisici si è osservata anche nel momento in cui si è usata la combinazione di pretrattamenti Log_{10} -SNV-centraggio sulla media sui campioni alti 5 cm, dove si sono individuati due gruppi di campioni predetti alla stessa altezza, probabilmente a causa della natura della superficie, liscia o ruvida, con cui la radiazione entra in contatto per prima.

Attraverso l'analisi delle componenti principali è stata evidenziata una natura non lineare del fenomeno di penetrazione della radiazione NIR. Preso atto di questa natura, si è comunque esplorata la possibilità di andare a studiare la profondità della penetrazione avvalendosi di un modello di regressione lineare quale è PLS, con l'idea che l'uso di un algoritmo che indirizzi l'analisi dei dati verso l'informazione di interesse potrebbe contribuire a linearizzare il fenomeno. A questo scopo ci si è concentrati sull'effetto che diversi pretrattamenti comportano e l'utilità che essi hanno nel poter descrivere il fenomeno. Si è osservato come la trasformata SNV sia in grado di correggere l'effetto fisico del diverso scattering subito dalla radiazione, dimostrandosi un pretrattamento ottimale quando l'informazione di interesse è di natura chimica. L'utilizzo del Log_{10} ha invece permesso di mitigare la non linearità del fenomeno, ma non si è dimostrato efficace nella correzione dei fenomeni fisici. Il suo utilizzo, quindi, consente di ricavare informazione chimica solamente quando i fenomeni fisici che caratterizzano i diversi campioni sono analoghi, altrimenti si ha il prevalere dell'informazione fisica che permette di individuare la presenza di diverse strutture interne al campione. L'utilizzo sequenziale di Log_{10} e della trasformata SNV, utilizzato con l'obiettivo di combinare gli effetti dei singoli pretrattamenti, ha permesso di mitigare

parzialmente sia i fenomeni fisici sia la non linearità presente, senza però correggere completamente nessuno dei due. Nonostante ciò, anche da questo aspetto, si sono tratti degli spunti rilevanti ossia il fatto che, quando usati in combinazione, si osserva prevalentemente l'effetto del pretrattamento applicato per primo, mentre quello utilizzato per secondo risulta essere meno rilevante. Questo si traduce nel fatto che il test set esterno dei campioni pretrattati con Log₁₀-SNV-centraggio sulla media presenti delle difficoltà maggiori nella sua predizione legate ai fenomeni fisici.

Queste osservazioni indicano che, per sviluppare dei modelli PLS che descrivano accuratamente la penetrazione della radiazione della radiazione NIR, sono necessari maggiori studi per trovare i pretrattamenti migliori che consentono di correggere tutte le problematiche che si incontrano. Nel condurre questi studi si dovrà considerare che in funzione della componente di interesse, sia essa quella chimica o quella fisica, sarà necessario utilizzare pretrattamenti differenti in modo da esaltare un effetto piuttosto che un altro. Un aspetto che deve essere ulteriormente approfondito, indipendentemente dalla componente di interesse, è il modo con cui limitare la non linearità del fenomeno in modo tale da poter utilizzare un modello di regressione lineare.

Tuttavia, nonostante tutte le migliorie che possono essere apportate in questo ambito di studio, i risultati permettono di affermare con certezza che la radiazione NIR non è una radiazione superficiale. Grazie sia alle informazioni ottenute da PCA sia a quelle ricavate da PLS, si riconosce che la radiazione attraverso PETG penetra per almeno 12 mm, mentre la radiazione NIR è in grado di penetrare maggiormente attraverso PLA raggiungendo profondità fino a 16 mm nei campioni da 3 cm e, sembrerebbe, fino a quasi 40 mm nei campioni da 5 cm. Questo aspetto suggerisce che la penetrazione della radiazione NIR è influenzata, oltre che dal materiale e dai fenomeni fisici, dalla quantità di dati disponibili nella regione profonda. Tale osservazione permette di definire l'importanza delle tecniche di imaging nella descrizione del fenomeno dal momento che questa metodica analitica è in grado di ottenere spettri per ogni pixel dell'immagine acquisita.

Per concludere, questo lavoro di tesi ha permesso di definire che la radiazione NIR e l'imaging spettroscopico in questa regione non sono tecniche di superficie. Alla luce di questa osservazione sarà necessario effettuare ulteriori studi in questo ambito per sfruttare a pieno le potenziali ricadute in campo analitico. Tra queste si potrebbero avere delle applicazioni nel settore agro-alimentare, in cui questo tipo di approccio potrebbe

permettere di analizzare gli alimenti direttamente dall'interno del packaging, minimizzando l'impatto del controllo qualità nella linea produttiva e quindi favorendo una maggior quantità di controlli. In ambito farmaceutico, analogamente, si aprirebbe la possibilità di analizzare alcune tipologie forme farmaceutiche, come le compresse, direttamente nei blister, e di valutare la composizione e l'omogeneità delle formulazioni anche in campioni massicci.

Bibliografia

- [1] E.W. Ciurczak, Handbook of Near- Infrared Analysis, n.d.
<https://www.routledge.com/Practical->.
- [2] J.W. Ellis, J. Bath, Modifications in the near infra-red absorption spectra of protein and of light and heavy water molecules when water is bound to gelatin, J. Chem. Phys. 6 (1938). <https://doi.org/10.1063/1.1750157>.
- [3] A.A. Michelson, XXXVIII. On the application of interference-methods to spectroscopic measurements. —I , The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 31 (1891).
<https://doi.org/10.1080/14786449108620117>.
- [4] D.L. Wetzel, Sleeper Among Spectroscopic Techniques, 1983. [https://doi.org/0003-2700/83/A351-1165\\$01.50/0](https://doi.org/0003-2700/83/A351-1165$01.50/0).
- [5] D.A. Burns, E.W. Ciurczak, Handbook of Near-Infrared Analysis 3rd edition, (2008).
- [6] C. Pasquini, Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications, 2003. <https://doi.org/10.1590/S0103-50532003000200006>.
- [7] T. Davies, The history of near infrared spectroscopic analysis: Past, present and future - From sleeping technique to the morning star of spectroscopy, Analusis 26 (1998) 17–19. <https://doi.org/10.1051/analusis:199826040017>.
- [8] Y. Ozaki, Near-infrared spectroscopy--Its versatility in analytical chemistry, Analytical Sciences 28 (2012) 545–562. <https://doi.org/10.2116/analsci.28.545>.
- [9] J.E. D, Mode of Vibrations, Encyclopedia of Vibrations, (n.d.).
- [10] Mark Somoza March, Oscillatore armonico e anarmonico, (2024).
- [11] Jan Skvaril, Konstantinos G. Kyprianidis, Erik Dahlquist, Modes of Mesurments in NIRS, Applied Spectroscopy Review (2017).
<https://doi.org/10.1080/05704928.2017.1289471>.
- [12] V.H. da Silva, J.J. da Silva, C.F. Pereira, Portable near-infrared instruments: Application for quality control of polymorphs in pharmaceutical raw materials and calibration transfer, J. Pharm. Biomed. Anal. 134 (2017).
<https://doi.org/10.1016/j.jpba.2016.11.036>.
- [13] A. Peinado, J. Hammond, A. Scott, Development, validation and transfer of a Near Infrared method to determine in-line the end point of a fluidised drying process for commercial production batches of an approved oral solid dose pharmaceutical product, J. Pharm. Biomed. Anal. 54 (2011). <https://doi.org/10.1016/j.jpba.2010.07.036>.
- [14] P.R. Wahl, G. Fruhmann, S. Sacher, G. Straka, S. Sowinski, J.G. Khinast, PAT for tableting: Inline monitoring of API and excipients via NIR spectroscopy, European

- Journal of Pharmaceutics and Biopharmaceutics 87 (2014).
<https://doi.org/10.1016/j.ejpb.2014.03.021>.
- [15] M. De Géa Neves, R.J. Poppi, M.C. Breitzkreitz, Authentication of plant-based protein powders and classification of adulterants as whey, soy protein, and wheat using FT-NIR in tandem with OC-PLS and PLS-DA models, *Food Control* 132 (2022).
<https://doi.org/10.1016/j.foodcont.2021.108489>.
 - [16] A. Malvandi, H. Feng, M. Kamruzzaman, Application of NIR spectroscopy and multivariate analysis for Non-destructive evaluation of apple moisture content during ultrasonic drying, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 269 (2022).
<https://doi.org/10.1016/j.saa.2021.120733>.
 - [17] R. Vitorino, A.S. Barros, S. Guedes, D.C. Caixeta, R. Sabino-Silva, Diagnostic and monitoring applications using near infrared (NIR) spectroscopy in cancer and other diseases, *Photodiagnosis Photodyn. Ther.* 42 (2023).
<https://doi.org/10.1016/j.pdpdt.2023.103633>.
 - [18] G. ElMasry, D.W. Sun, Principles of Hyperspectral Imaging Technology, in: *Hyperspectral Imaging for Food Quality Analysis and Control*, Elsevier Inc., 2010: pp. 3–43. <https://doi.org/10.1016/B978-0-12-374753-2.10001-2>.
 - [19] Freij Hannes, *Hyperspectral Image Registration and Construction From Irregularly Sampled Data*, Linköping University, 2021.
 - [20] M. Manley, Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials, *Chem. Soc. Rev.* 43 (2014) 8200–8214.
<https://doi.org/10.1039/c4cs00062e>.
 - [21] A. Bhargava, A. Sachdeva, K. Sharma, M.H. Alsharif, P. Uthansakul, M. Uthansakul, Hyperspectral imaging and its applications: A review, *Heliyon* 10 (2024).
<https://doi.org/10.1016/j.heliyon.2024.e33208>.
 - [22] E. Arbash, A. De Lima Ribeiro, S. Thiele, N. Gnann, B. Rasti, M. Fuchs, P. Ghamisi, R. Gloaguen, Masking Hyperspectral Imaging Data with Pretrained Models, in: *Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing*, 2023.
<https://doi.org/10.1109/WHISPERS61460.2023.10431074>.
 - [23] Z.D. Uca Avcı, M. Karaman, E. Ozelkan, I. Papila, A comparison of pixel-based and object-based classification methods, a case study: Istanbul, Turkey, in: *34th International Symposium on Remote Sensing of Environment - The GEOSS Era: Towards Operational Environmental Monitoring*, 2011.
 - [24] S.C.I. Gruppo di Chemiometria, Cos'è la chemiometria,
<https://Gruppochemiometria.It/Index.Php/La-Chemiometria> (n.d.).
<https://gruppochemiometria.it/index.php/la-chemiometria> (accessed April 4, 2026).

- [25] P. Oliveri, C. Malegori, R. Simonetti, M. Casale, The impact of signal pre-processing on the final interpretation of analytical outcomes – A tutorial, *Anal. Chim. Acta* 1058 (2019) 9–17. <https://doi.org/10.1016/j.aca.2018.10.055>.
- [26] Forina Michele, *ELEMENTI di CHEMIOMETRIA*, n.d.
- [27] Eigenvector, *Advanced Preprocessing: Simple Mathematical Operations*, (2015). https://www.eigenvectordocs.com/index.php?title=Advanced_Preprocessing:_Simple_Mathematical_Operations#:~:text=The%20absolute%20value%20method%20is,=%2010%2D5%20by%20default. (accessed April 8, 2026).
- [28] E.J. Abraham, J.J. Kellogg, Chemometric-Guided Approaches for Profiling and Authenticating Botanical Materials, *Front. Nutr.* 8 (2021). <https://doi.org/10.3389/fnut.2021.780228>.
- [29] A. Oussama, F. Elabadi, S. Platikanov, F. Kzaiber, R. Tauler, Detection of olive oil adulteration using FT-IR spectroscopy and PLS with variable importance of projection (VIP) scores, *JAOCS, Journal of the American Oil Chemists' Society* 89 (2012). <https://doi.org/10.1007/s11746-012-2091-1>.
- [30] E. Lopez, J. Etxebarria-Elezgarai, J.M. Amigo, A. Seifert, The importance of choosing a proper validation strategy in predictive models. A tutorial with real examples, *Anal. Chim. Acta* 1275 (2023). <https://doi.org/10.1016/j.aca.2023.341532>.
- [31] A. Rácz, D. Bajusz, K. Héberger, Modelling methods and cross-validation variants in QSAR: a multi-level analysis, *SAR QSAR Environ. Res.* 29 (2018) 661–674. <https://doi.org/10.1080/1062936X.2018.1505778>.
- [32] Z. Li, C. Scagliarini, A. Mazzoleni, S. Gariglio, E. Catelli, C. Malegori, S. Prati, E. Alladio, G. Sciutto, P. Oliveri, Exploiting the penetration depth of XRF and NIR radiation: from 2D to 3D spectral imaging, *TrAC - Trends in Analytical Chemistry* 197 (2026). <https://doi.org/10.1016/j.trac.2026.118693>.
- [33] D. Haberthür, M. Frenz, B. Photonik, *Use of Lasers in Medicine: Tissue Soldering and Precise Cutting*, n.d.
- [34] L.O. Svaasand, R. Ellingsen, *OPTICAL PROPERTIES OF HUMAN BRAIN*, *Photochem. Photobiol.* 38 (1983). <https://doi.org/10.1111/j.1751-1097.1983.tb02674.x>.
- [35] M. Longoni, B. Genova, A. Marzanni, D. Melfi, C. Beccaria, S. Bruni, FT-NIR Spectroscopy for the Non-Invasive Study of Binders and Multi-Layered Structures in Ancient Paintings: Artworks of the Lombard Renaissance as Case Studies, *Sensors* 22 (2022). <https://doi.org/10.3390/s22052052>.
- [36] F. Faris, M. Thorniley, Y. Wickramasinghe, R. Houston, P. Rolfe, N. Livera, A. Spencer, Non-invasive in vivo near-infrared optical measurement of the penetration depth in the neonatal head, *Clinical Physics and Physiological Measurement* 12 (1991). <https://doi.org/10.1088/0143-0815/12/4/005>.

- [37] I.O. Afara, R. Shaikh, E. Nippolainen, W. Querido, J. Torniainen, J.K. Sarin, S. Kandel, N. Pleshko, J. Töyräs, Characterization of connective tissues using near-infrared spectroscopy and imaging, *Nat. Protoc.* 16 (2021). <https://doi.org/10.1038/s41596-020-00468-z>.
- [38] M. Vega-Castellote, M.T. Sánchez, J.P. Wold, N.K. Afseth, D. Pérez-Marín, Near infrared light penetration in watermelon related to internal quality evaluation, *Postharvest Biol. Technol.* 204 (2023). <https://doi.org/10.1016/j.postharvbio.2023.112477>.
- [39] J. Workman Jr., L. Weyer, *Practical Guide and Spectral Atlas for Interpretive Near-Infrared Spectroscopy*, 2012. <https://doi.org/10.1201/b11894>.
- [40] S. Muroga, Y. Hikima, M. Ohshima, Near-Infrared Spectroscopic Evaluation of the Water Content of Molded Polylactide under the Effect of Crystallization, *Appl. Spectrosc.* 71 (2017). <https://doi.org/10.1177/0003702816681011>.
- [41] Z. Zhang, Residuals and regression diagnostics: Focusing on logistic regression, *Ann. Transl. Med.* 4 (2016). <https://doi.org/10.21037/atm.2016.03.36>.