

**POLITECNICO DI TORINO**

Collegio di Ingegneria Chimica e dei Materiali

**Corso di Laurea Magistrale  
in Ingegneria Chimica e dei Processi Sostenibili**

Tesi di Laurea Magistrale

**Sviluppo di un modello multivariato  
per l'analisi del sudore basato su misure  
Raman**



**Relatori**

Prof. Leonardo IANNUCCI  
Prof.ssa Sabrina GRASSINI

**Candidato**

Francesco GIRAUDO

Marzo 2024



## Abstract

La misura del livello di componenti metabolici nel sangue è un punto di interesse di prim'ordine nel campo medico, clinico e terapeutico. La ricerca di nuovi metodi di analisi e modalità di campionamento ha portato ad innumerevoli soluzioni, esplorando non solo vie dirette per il raggiungimento dello scopo, ma anche metodi di misura indiretti tramite l'analisi della concentrazione degli analiti nel sudore, rendendo così il rilevamento non invasivo, più agevole e soprattutto continuo. Gli analiti di interesse per questo studio sono glucosio, lattato ed urea componenti principali del sudore eccrino e metaboliti di notevole interesse clinico e fisiologico. In particolare, l'utilizzo di metodi spettrometrici, come la spettroscopia Raman, si è rivelato l'approccio più promettente per lo sviluppo di nuovi strumenti di campionamento ed analisi che possano rappresentare una valida alternativa alle misure effettuate in seguito a prelievi di notevoli quantità di sangue da parte di personale medico. Il principale limite di questa metodologia è fornito dall'efficienza delle analisi degli spettri ottenuti in quanto viene tipicamente effettuata individuando un picco specifico per l'analita di interesse, restringendo quindi il campo di dati presi in considerazione nel procedimento. Tale metodo risulta in misure troppo sensibili a interferenze date da molecole con caratteristiche simili alla ricercata. Nell'analisi mediante spettroscopia Raman, questi tre componenti si comportano da interferenti reciproci ed il metodo proposto mira anche a diminuire questo effetto sfruttando tecniche chemiometriche di analisi multivariata. Lo scopo di questo lavoro di tesi è la produzione di un modello analitico e matematico per il calcolo della concentrazione di glucosio, lattato e urea in soluzioni simulanti il sudore umano, mediante l'analisi dello spettro Raman del campione. Il procedimento comprende la scrittura del codice in Python, la fase di istruzione del modello, costruendo delle rette di taratura tramite soluzioni a concentrazione degli analiti nota e crescente, ed infine una fase di testing per la validazione dell'efficacia del modello per soluzioni simulanti il sudore e contenenti gli interferenti delle misure. I risultati sono stati positivi, con ottimi campi di linearità ( $R^2 > 0.99$ ) ed una buona stabilità del modello nell'analizzare tutti e tre i componenti di interesse.



# Indice

<b>Elenco delle figure</b>	V
<b>1. Introduzione</b>	1
1.1. Obiettivi e metodologia	1
<b>2. Stato dell'arte</b>	3
2.1. Il glucosio	3
2.1.1. La problematica del diabete	3
2.1.2. Misure convenzionali	4
2.1.3. Misure indirette non invasive	6
2.2. Il lattato	8
2.2.1. Metodi di misura del lattato	9
2.3. L'urea	9
2.3.1. Metodi di misura dell'urea	10
2.4. La spettroscopia Raman	10
2.5. Metodi di analisi chemiometrici	14
2.5.1. Principal Component Analysis	15
2.5.2. Principal Component Regression e Principal Least Square Regression	19
<b>3. Materiali e metodi</b>	21
3.1. Materiali utilizzati e strumentazione	21
3.2. Il metodo	22
3.3. Il modello di regressione	23
<b>4. Risultati</b>	25
4.1. Misure Raman in acqua distillata	27
4.1.1. Glucosio	27
4.1.2. Lattato	33
4.1.3. Urea	37
4.2. Misure Raman in sudore artificiale	42
4.2.1. Analisi del glucosio con interferenti	43
4.2.2. Analisi del lattato con interferenti	48
4.2.3. Analisi dell'urea con interferenti	52
4.3. Misura degli interferenti al variare dell'analita principale	57
4.3.1. Analisi del glucosio in soluzioni di lattato a concentrazione variabile	57
4.3.2. Analisi del glucosio in soluzioni di urea a concentrazione variabile	58
4.3.3. Analisi del lattato in soluzioni di glucosio a concentrazione variabile	60
4.3.4. Analisi dell'urea in soluzioni di glucosio a concentrazione variabile	62
4.3.5. Analisi del lattato in soluzioni di urea a concentrazione variabile	64
4.3.6. Analisi dell'urea in soluzioni di lattato a concentrazione variabile	65
<b>5. Conclusioni e sviluppi futuri</b>	69
<b>Bibliografia</b>	71
<b>Ringraziamenti</b>	75
	III



## Elenco delle figure

2.1:	Andamento livello glicemico di due pazienti affetti da diabete mellito (curve 1 e 2) e di un individuo sano (curva 3). Nel grafico sono evidenziate le zone di ipoglicemia (rossa), normoglicemia (verde) e iperglicemia (gialla) . . . . .	3
2.2:	Riflettometro Ames . . . . .	5
2.3:	Schema riassuntivo delle metodologie di misura del glucosio [17] . . . . .	7
2.4:	Sensore pieghevole per la misura del lattato. La sudorazione può essere indotta tramite ionoforesi, ovvero una breve scarica elettrica a bassa intensità [21] . . . . .	9
2.5:	Schema di funzionamento di uno spettrometro Raman . . . . .	11
2.6:	Rappresentazione dell'effetto di scattering Raman. Sono visibili i tre meccanismi di Rayleigh, Stokes e anti-Stokes (a) e gli scostamenti in termini di frequenza della radiazione diffusa (b) . . . . .	13
2.7:	Effetto della fluorescenza su un generico spettro Raman. Lo spettro verde è affetto da una forte fluorescenza e i picchi presenti nella metà sinistra dello spettro risultano totalmente nascosti. Lo spettro rosso è ottenuto con una sorgente laser a 785 nm, mentre quello verde a 532 nm . . . . .	13
2.8:	Effetto dei passaggi di <i>Pre-Processing</i> su uno spettro Raman, con identificazione della linea di fondo gialla (a), dopo la sua rimozione (b), dopo lo <i>smoothing</i> con filtro di Savitzky-Golay (c) e dopo normalizzazione (d) . . . . .	16
2.9:	Esempio grafico biplot per la determinazione dei <i>cluster</i> fra le prove eseguite su un set di lattato in acqua a concentrazione crescente . . . . .	17
2.10:	Esempio di grafico per lo studio degli <i>outliers</i> per campioni di glucosio in acqua con concentrazione crescente . . . . .	19
3.1:	Componenti principali dello strumento utilizzato nelle misure. Spettrofotometro BTC675N (a), laser BRM-785 (b), sonda probe a trasmissione in fibra ottica (c) e porta-cuvette Raman BCR100A (d) . . . . .	22
4.1:	Spettri Raman di soluzioni acquose di glucosio (a), lattato (b) e urea (c) a concentrazione crescente. La scala cromatica va dal viola, corrispondente al campione 0 mM, fino al rosso scuro, rappresentante i campioni 200 mM . . . . .	25
4.2:	Componenti principali del modello (a) e loro varianza espressa (b) nell'analisi di campioni di glucosio in acqua a concentrazione variabile . . . . .	27
4.3:	Spettri Raman del glucosio in soluzioni acquose all'aumentare della concentrazione. La scala cromatica va dal viola, corrispondente al campione 0 mM, fino al rosso scuro, rappresentante i campioni 200 mM . . . . .	28
4.4:	Determinazione degli <i>outliers</i> per il set di glucosio a concentrazione variabile in acqua deionizzata . . . . .	29
4.5:	Diagrammi biplot del set di glucosio a concentrazione variabile in acqua deionizzata . . . . .	29

4.6:	Rappresentazione tridimensionale dei diagrammi di <i>clustering</i> del glucosio in acqua .....	30
4.7:	PCR effettuata su soluzioni di glucosio in acqua a concentrazione crescente ...	31
4.8:	Ingrandimento della PCR sul set di glucosio in acqua ed errore assoluto .....	31
4.9:	PLSR effettuata sul set di glucosio in acqua a concentrazione crescente .....	32
4.10:	Ingrandimento PLSR sul set di glucosio in acqua nella zona 0-10 mM .....	32
4.11:	Componenti principali del modello (a) e loro varianza espressa (b) nell'analisi di campioni di lattato in acqua a concentrazione variabile .....	33
4.12:	Spettri Raman del lattato in soluzioni acquose all'aumentare della concentrazione. La scala cromatica va dal viola, corrispondente al campione 0 mM, fino al rosso scuro, rappresentante i campioni 200 mM .....	33
4.13:	Determinazione degli <i>outliers</i> per il set di lattato a concentrazione variabile in acqua deionizzata .....	34
4.14:	Diagrammi biplot del set di lattato a concentrazione variabile in acqua deionizzata .....	34
4.15:	Rappresentazione tridimensionale dei diagrammi di <i>clustering</i> del lattato in acqua .....	35
4.16:	PCR effettuata su soluzioni di lattato in acqua a concentrazione crescente ...	35
4.17:	Ingrandimento della PCR sul set di lattato in acqua ed errore assoluto .....	36
4.18:	PLSR effettuata sul set di lattato in acqua a concentrazione crescente .....	36
4.19:	Ingrandimento PLSR sul set di lattato in acqua nella zona 0-10 mM .....	37
4.20:	Componenti principali del modello (a) e loro varianza espressa (b) nell'analisi di campioni di urea in acqua a concentrazione variabile .....	37
4.21:	Spettri Raman dell'urea in soluzioni acquose all'aumentare della concentrazione. La scala cromatica va dal viola, corrispondente al campione 0 mM, fino al rosso scuro, rappresentante i campioni 200 mM .....	38
4.22:	Determinazione degli <i>outliers</i> per il set di urea a concentrazione variabile in acqua deionizzata .....	39
4.23:	Diagrammi biplot del set di urea a concentrazione variabile in acqua deionizzata .....	39
4.24:	Rappresentazione tridimensionale dei diagrammi di <i>clustering</i> dell'urea in acqua .....	40
4.25:	PCR effettuata su soluzioni di urea in acqua a concentrazione crescente .....	40
4.26:	Ingrandimento della PCR sul set di urea in acqua ed errore assoluto .....	41
4.27:	PLSR effettuata sul set di urea in acqua a concentrazione crescente .....	41
4.28:	Ingrandimento PLSR sul set di urea in acqua nella zona 0-10 mM .....	42
4.29:	Spettri Raman del glucosio in soluzioni con interferenti. La scala cromatica va dal viola, corrispondente al campione 0 mM, fino al rosso scuro, rappresentante i campioni 200 mM .....	43

4.30: Determinazione degli <i>outliers</i> per il set di glucosio a concentrazione variabile con interferenti . . . . .	44
4.31: Diagrammi biplot del set di glucosio a concentrazione variabile in presenza di interferenti . . . . .	45
4.32: Rappresentazione tridimensionale dei diagrammi di <i>clustering</i> del glucosio in presenza di interferenti . . . . .	45
4.33: PCR effettuata su soluzioni di glucosio con interferenti a concentrazione crescente . . . . .	46
4.34: Ingrandimento della PCR sul set di glucosio con interferenti ed errore assoluto . . . . .	46
4.35: PLSR effettuata sul set di glucosio con interferenti e a concentrazione crescente . . . . .	47
4.36: Spettri Raman del lattato in soluzioni con interferenti. La scala cromatica va dal viola, corrispondente al campione 0 mM, fino al rosso scuro, rappresentante i campioni 200 mM . . . . .	48
4.37: Determinazione degli <i>outliers</i> per il set di lattato a concentrazione variabile con interferenti . . . . .	49
4.38: Diagrammi biplot del set di lattato a concentrazione variabile in presenza di interferenti . . . . .	49
4.39: Rappresentazione tridimensionale dei diagrammi di <i>clustering</i> del lattato in presenza di interferenti . . . . .	50
4.40: PCR effettuata su soluzioni di lattato con interferenti a concentrazione crescente . . . . .	50
4.41: Ingrandimento della PCR sul set di lattato con interferenti ed errore assoluto	51
4.42: PLSR effettuata sul set di lattato con interferenti e a concentrazione crescente	52
4.43: Spettri Raman dell'urea in soluzioni con interferenti. La scala cromatica va dal viola, corrispondente al campione 0 mM, fino al rosso scuro, rappresentante i campioni 200 mM . . . . .	53
4.44: Determinazione degli <i>outliers</i> per il set di urea a concentrazione variabile con interferenti . . . . .	53
4.45: Diagrammi biplot del set di urea a concentrazione variabile in presenza di interferenti . . . . .	54
4.46: Rappresentazione tridimensionale dei diagrammi di <i>clustering</i> dell'urea in presenza di interferenti . . . . .	54
4.47: PCR effettuata su soluzioni di urea con interferenti a concentrazione crescente	55
4.48: Ingrandimento della PCR sul set di urea con interferenti ed errore assoluto . . .	55
4.49: PLSR effettuata sul set di urea con interferenti e a concentrazione crescente . .	56
4.50: Determinazione del glucosio in soluzioni a concentrazione crescente di lattato e concentrazione fissa di urea. La linea rossa indica il valore reale di concentrazione del glucosio . . . . .	57

<b>4.51:</b> Ingrandimento della determinazione del glucosio in soluzioni a concentrazione crescente di lattato e concentrazione fissa di urea. La linea rossa indica il valore reale di concentrazione del glucosio . . . . .	58
<b>4.52:</b> Determinazione del glucosio in soluzioni a concentrazione crescente di urea e concentrazione fissa di lattato. La linea rossa indica il valore reale di concentrazione del glucosio . . . . .	59
<b>4.53:</b> Ingrandimento della determinazione del glucosio in soluzioni a concentrazione crescente di urea e concentrazione fissa di lattato. La linea rossa indica il valore reale di concentrazione del glucosio . . . . .	59
<b>4.54:</b> Determinazione del lattato in soluzioni a concentrazione crescente di glucosio e concentrazione fissa di urea. La linea rossa indica il valore reale di concentrazione del lattato . . . . .	60
<b>4.55:</b> Ingrandimento della determinazione del lattato in soluzioni a concentrazione crescente di glucosio e concentrazione fissa di urea. La linea rossa indica il valore reale di concentrazione del lattato . . . . .	61
<b>4.56:</b> Determinazione dell'urea in soluzioni a concentrazione crescente di glucosio e concentrazione fissa di lattato. La linea rossa indica il valore reale di concentrazione dell'urea . . . . .	62
<b>4.57:</b> Ingrandimento della determinazione dell'urea in soluzioni a concentrazione crescente di glucosio e concentrazione fissa di lattato. La linea rossa indica il valore reale di concentrazione dell'urea . . . . .	63
<b>4.58:</b> Determinazione del lattato in soluzioni a concentrazione crescente di urea e concentrazione fissa di glucosio. La linea rossa indica il valore reale di concentrazione del lattato . . . . .	64
<b>4.59:</b> Ingrandimento della determinazione del lattato in soluzioni a concentrazione crescente di urea e concentrazione fissa di glucosio. La linea rossa indica il valore reale di concentrazione del lattato . . . . .	65
<b>4.60:</b> Determinazione dell'urea in soluzioni a concentrazione crescente di lattato e concentrazione fissa di glucosio. La linea rossa indica il valore reale di concentrazione dell'urea . . . . .	66
<b>4.61:</b> Ingrandimento della determinazione dell'urea in soluzioni a concentrazione crescente di lattato e concentrazione fissa di glucosio. La linea rossa indica il valore reale di concentrazione dell'urea . . . . .	66

# 1. Introduzione

La determinazione degli analiti nel sangue a livello clinico e medico è indubbiamente uno dei più praticati esami di routine e su richiesta specifica. Tale misura ha estrema importanza sia per pazienti non affetti da alcuna patologia, sia per pazienti con malattie metaboliche o acute che possano modificare la concentrazione nel flusso sanguigno.

## *1.1. Obiettivi e metodologia*

L'obiettivo di questo lavoro è proporre un nuovo approccio per affrontare il problema, mediante l'utilizzo di metodi di analisi chemiometrici a partire da misure Raman ottenute da metodi indiretti, ovvero non analizzando direttamente campioni di sangue, ma tramite l'analisi del sudore umano, il quale fornisce informazioni sullo stato fisiologico del soggetto strettamente legate alla composizione del flusso sanguigno. Lo scopo ultimo di questo procedimento è quello di valutare la possibilità di migliorare le già note tecniche di analisi su spettri Raman, e di creare un modello in grado di processare grandi quantità di dati in modo da valutare la concentrazione di glucosio, lattato e urea.

La suddivisione dei capitoli è così organizzata:

- Capitolo 2: presentazione dello stato dell'arte sulla misura della concentrazione metabolica del glucosio, del lattato e dell'urea. Vengono presentati dispositivi già in uso e studi di progettazione e ricerca per nuove metodologie, analizzandone i miglioramenti apportati alle tecniche più comuni;
- Capitolo 3: viene presentato il nuovo modello elaborato in questo lavoro, discutendone la struttura ed il funzionamento, oltre che la strumentazione utilizzata per l'ottenimento dei dati di calibrazione e di test;
- Capitolo 4: vengono presentati i risultati della fase di addestramento e di testing del modello;
- Capitolo 5: si presentano le considerazioni finali e i possibili sviluppi futuri.



## 2. Stato dell'arte

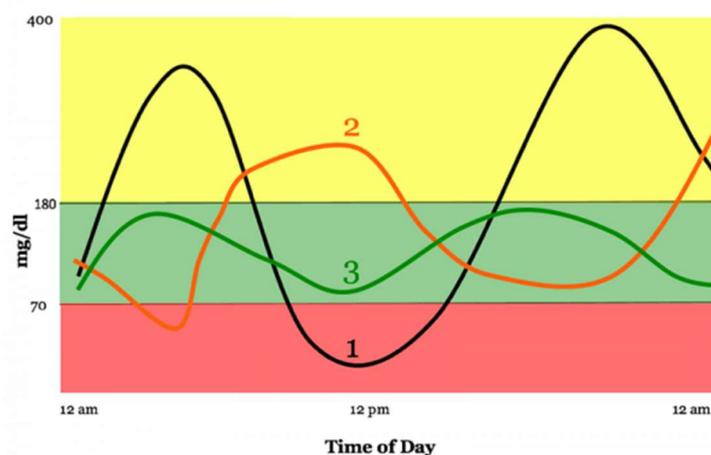
Viene dapprima presentata una panoramica dei tre principali analiti trattati in questo lavoro, ovvero glucosio, lattato e urea, seguita poi da un'introduzione alle tecniche strumentali ed analitiche di cui ci si è avvalsi nella realizzazione del modello, evidenziandone vantaggi attesi e le motivazioni che hanno condotto ad ipotizzare che potesse portare ad un miglioramento rispetto ai metodi presenti.

### 2.1. Il glucosio

La molecola del glucosio è una delle più conosciute e studiate in vari campi, a partire dalla chimica di base fino ad applicazioni industriali. Dal punto di vista metabolico, il glucosio rappresenta la principale fonte di energia per le cellule dell'organismo. La regolazione del livello di glucosio avviene mediante l'ormone insulina [2], molecola prodotta dalle cellule  $\beta$  a livello del pancreas e che agisce come attivatore di questi processi metabolici.

#### 2.1.1. La problematica del diabete

Il più comune esempio di problematica relativa al glucosio è il diabete mellito. Con questo termine si indica l'insieme di più varianti della patologia, comunemente nota come diabete, che porta ad avere una concentrazione di glucosio nel sangue molto superiore alla norma, condizione nota come iperglicemia. In maniera diametralmente opposta, si può anche avere la condizione di ipoglicemia, dovuta ad una carenza di glucosio, che può condurre a sudorazione eccessiva, tremore e sensazione di freddo, fino a problemi cardiocircolatori e nervosi. Il problema principale è legato al fatto che un paziente affetto da diabete è soggetto a variazioni non regolari del livello di glucosio in entrambe le direzioni [1], anche in un arco di tempo molto ristretto (Figura 2.1).



**Figura 2.1:** Andamento livello glicemico di due pazienti affetti da diabete mellito (curve 1 e 2) e di un individuo sano (curva 3). Nel grafico sono evidenziate le zone di ipoglicemia (rossa), normoglicemia (verde) e iperglicemia (gialla).

La regolazione del livello di glucosio avviene mediante l'ormone insulina [2], molecola prodotta dalle cellule  $\beta$  a livello del pancreas e che agisce come attivatore di questi processi metabolici. Un soggetto affetto da diabete mellito può presentare problemi di sintesi e secrezione di insulina, oppure può sviluppare una resistenza all'azione di questo ormone, che porta dunque all'inefficacia della sua funzione metabolica e, dunque, all'accumulo eccessivo di zuccheri nel sangue. Conseguenza di questa condizione sono l'iperosmolarità iperglicemica, ovvero l'aumento della pressione osmotica del sangue dovuta alla carenza di idratazione, e la chetoacidosi, che comporta l'accumulo di chetoni nell'organismo ed il conseguente abbassamento del pH fisiologico [3].

Alla fine del XX secolo il diabete mellito era già riconosciuto come una delle principali patologie non infettive a livello mondiale, soprattutto in paesi economicamente sviluppati, fino ad essere annoverato come la quinta maggior causa di morte negli Stati Uniti per numero di decessi nel 2000 [4]. Dall'inizio del nuovo millennio, con l'ulteriore sviluppo economico, l'aumento statistico di fattori come età media, obesità, stile di vita sedentario ed errate abitudini alimentari hanno portato ad un ulteriore aggravamento di una condizione già critica, per arrivare al 2021 in cui l'American Diabetes Association ha rilevato che l'11,6% della popolazione totale statunitense è affetta da diabete mellito, di cui più di 350000 sono giovani sotto i vent'anni di età [5]. In paesi con condizioni di vita meno favorevoli, la più comune forma presente risulta essere il diabete di tipo 2, tipico di pazienti più anziani e spesso affetti anche da altre patologie metaboliche. Questi dati preoccupanti hanno portato ad un sempre maggior sforzo nella ricerca per soluzioni mediche, ma anche e soprattutto di metodologie di monitoraggio della concentrazione del glucosio nel sangue dei pazienti affetti o dichiarati a rischio [6].

### 2.1.2. Misure convenzionali

Le prime misure venivano effettuate a livello clinico, mediante prelievi di notevoli quantità di sangue per effettuare poi delle analisi di ricerca in laboratorio, con metodi anche piuttosto vari vista la disponibilità di operazioni svolgibili in questi ambienti, spaziando da semplici misure colorimetriche fino alla cromatografia liquida ad alta prestazione (HPLC, High Performance Liquid Chromatography). Queste procedure si svolgono ancora oggi, ma solo per analisi su richiesta specifica ed isolata, come in preparazione ad un intervento chirurgico o in caso si sospetti la presenza di sostanze estranee al metabolismo tipico umano. Il problema principale delle analisi presso un ospedale o uno studio medico è la scomodità, in quanto vanno svolte presso una struttura apposita, vi sono procedure definite da seguire e non è possibile ripetere la misura con la frequenza desiderata per un paziente soggetto a condizioni specifiche come il diabete. Si rende dunque necessaria la presenza di metodi di monitoraggio affidati al paziente, la cui introduzione sul mercato, se pur con dispositivi piuttosto semplici, risale all'inizio degli anni '70 con il Riflettometro Ames (Figura 2.2). Questo strumento è molto somigliante ai dispositivi attuali, in quanto funzionava con una striscia reattiva che a contatto con il sangue cambiava colore in base alla concentrazione di glucosio. La variazione nella riflettanza dovuta al cambio di colore veniva poi analizzata dallo strumento che elaborava il valore finale mostrato su un display. Sebbene questo dispositivo fosse molto costoso, voluminoso e pesante, ha

comunque rappresentato un notevole punto di svolta nel monitoraggio autonomo della glicemia.



**Figura 2.2:** Riflettometro Ames.

Negli anni a seguire gli sviluppi si concentrarono maggiormente sulla miniaturizzazione dello strumento, oltre che sulla diminuzione dei quantitativi di sangue richiesti per le analisi. Solo con l'inizio del ventunesimo secolo e i notevoli avanzamenti tecnologici, si cominciò a pensare a nuovi metodi di misura, passando da metodi colorimetrici come quello del Riflettometro di Ames a metodi elettrochimici. I più comuni fra questi sono metodi enzimatici, come quelli che coinvolgono gli enzimi esochinasi per catalizzare una reazione di degradazione del glucosio con la produzione di elettroni, che vengono poi rilevati da un amperometro apposito e la concentrazione finale ricavata sulla base dell'intensità di corrente sviluppata. Al giorno d'oggi il metodo globalmente più diffuso per la misura del livello glicemico è ancora l'analisi diretta del sangue mediante metodi elettrochimici enzimatici, ma gli strumenti sono stati ottimizzati sia dal punto di vista della comodità di trasporto, che dal punto di vista dei software utilizzati e quindi delle prestazioni. Per questi strumenti permangono due problemi di grande importanza. In primis, le misure richiedono l'ottenimento di almeno una goccia di sangue capillare, ottenuta di solito con una lancetta che crea un piccolo foro sul dito del paziente. Questo rende la misura ancora invasiva e parzialmente dolorosa, soprattutto considerando che gran parte dei soggetti è anziana e quindi si possono anche avere problemi di infezione o addirittura di ottenimento del campione a causa di problemi circolatori in zone periferiche dell'organismo. Inoltre, la misura non è continua, ma viene effettuata a orari precisi, a discrezione del piano medico. Negli intervalli di tempo fra le misure si potrebbe avere una variazione importante e grave e la mancanza di una misura continua potrebbe essere causa di complicazioni. Infine, per questi strumenti è necessario avere una riserva di strisce reattive che, considerando fino a sei misure giornaliere per pazienti affetti da diabete mellito, possono portare a spese importanti nel tempo.

Negli ultimi anni la ricerca si è concentrata sull'ottenimento di misure meno invasive e soprattutto continue [7, 8]. La più recente soluzione, presente ancora in minor parte rispetto ai classici glucometri, è quella rappresentata dai sistemi di monitoraggio continuo (CGM) ovvero l'utilizzo di un piccolo dispositivo impiantato

sotto la pelle, con funzionamento analogo a quelli elettrochimici-enzimatici già discussi. In tal modo non si ha la necessità di estrarre un campione di sangue, visto che la misura è fatta direttamente nell'organismo. Proprio per questo, si può rendere il procedimento continuo. Permane il problema della necessità di un intervento chirurgico per l'installazione del dispositivo e per eventuali problematiche e sostituzioni. Inoltre, a seconda della condizione del paziente e della procedura, vi è ancora un rischio di infezione non nullo. Tale metodologia è nota come sistema ad invasività minima ed è la base di partenza per metodi ancora in fase di sviluppo noti come metodi non invasivi.

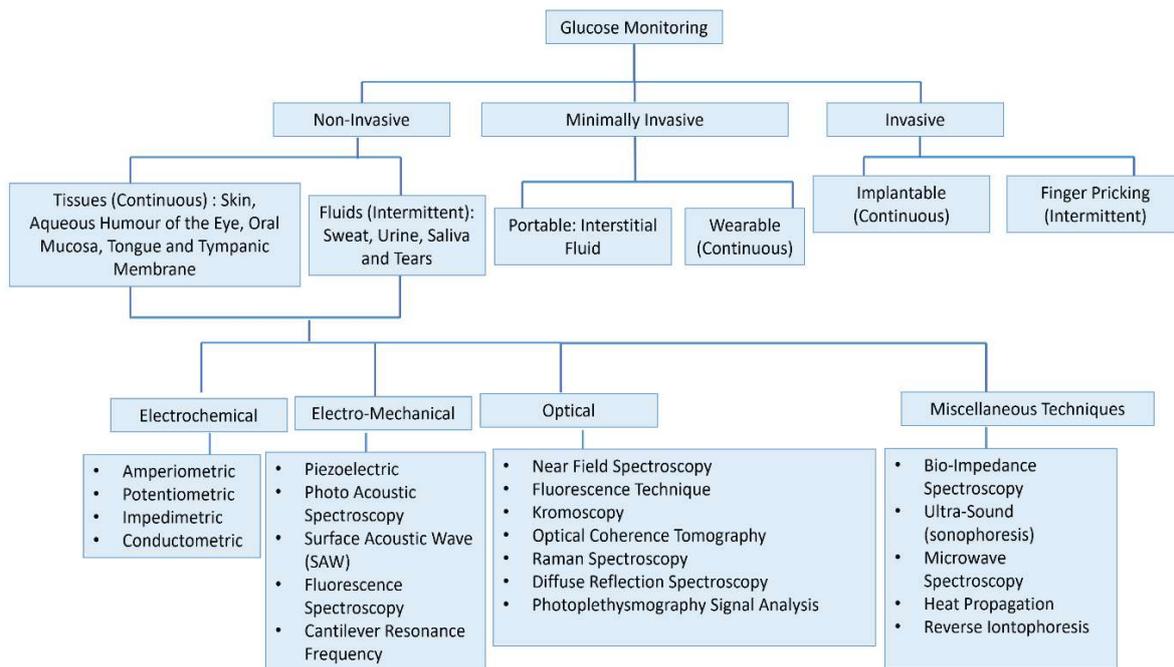
### 2.1.3. Misure indirette non invasive

Le misure non invasive rappresentano l'ultima frontiera di ricerca in questo campo. Le metodologie più analizzate fino ad oggi sono quelle che sfruttano metodi spettroscopici come la spettrometria infrarossa, nel vicino infrarosso o la spettroscopia Raman. Le misure spettroscopiche sono una vasta famiglia di metodi che sfruttano una radiazione inviata sul campione in analisi per la produzione di uno spettro caratteristico mediante vari metodi. Tale spettro funge da impronta digitale per l'analita considerato e sarà composto da picchi ad intensità variabile a seconda della concentrazione della sostanza stessa. Vi sono comunque alternative come metodi a bioimpedenza, che misurano la variazione della resistenza elettrica del tessuto preso in esame al variare della concentrazione di glucosio, oppure tecniche elettro-meccaniche, ad esempio metodi fotoacustici e metodi di fluorescenza. Questi ultimi sono però campi secondari e meno interessanti rispetto alle misure spettrometriche.

Un campo di applicazione estremamente promettente per questa categoria di misure è quello delle misure indirette. Per misura indiretta si intende la misura della concentrazione del glucosio fisiologico non effettuata direttamente sul sangue, ma su un fluido corporeo secondario che agisca da vettore dell'analita desiderato. In tal senso, sono state molte le proposte avanzate, passando dalla misura in campioni di saliva, a campioni lacrimali o urine. Tutti questi presentano però notevoli variazioni che possono anche non essere date dalla patologia per cui si esegue la misura. Ad esempio, urina e saliva sono strettamente correlate all'alimentazione, anche abitudini alimentari sane possono dare grandi fluttuazioni nei valori di concentrazione. Inoltre, l'analisi delle urine può anche essere influenzata da eventuali patologie al sistema urinario, assolutamente frequenti soprattutto in pazienti affetti da diabete mellito di tipo 2. Per quanto riguarda le secrezioni lacrimali, la misura sarebbe difficile sia per il campionamento molto scomodo, che per la discontinuità nella produzione di tale medium.

Tutto questo ha portato alla proposta della misura tramite analisi del sudore. La correlazione tra composizione del sudore e del sangue è stata ampiamente studiata [9, 10], sottolineando uno stretto legame fra i due. La variazione dei parametri all'interno del sudore è un chiaro indice dell'andamento dei processi metabolici interni e dunque riflette le oscillazioni di molti fattori, come appunto la concentrazione del glucosio, nel sangue. Le misure indirette sul sudore garantiscono il livello minimo di invasività, essendo praticate all'esterno dell'epidermide del paziente, ed i dispositivi finora proposti e realizzati hanno avuto riscontri molto soddisfacenti. Nella maggior parte dei casi si parla di apparecchiature di dimensioni molto contenute, adesive,

adattabili alla pelle e assolutamente prive di effetti dolorosi o fastidiosi. Sono stati messi a punto strumenti di vario genere, a partire da orologi simili ai cardiofrequenzimetri commerciali, fino a capsule circolari di pochi centimetri di diametro e spesse fino a mezzo centimetro, tipicamente applicate sulla parte posteriore del braccio. Molto interessanti sono anche alcuni strumenti paragonabili a cerotti adesivi e flessibili, facilmente applicabili e comprendenti sia il sistema di acquisizione del campione che il dispositivo di analisi. I risultati vengono trasmessi ad una apposita applicazione e possono essere visualizzati direttamente su dispositivi personali come smartphones e computer, per essere monitorati direttamente anche dal paziente. In letteratura è possibile trovare numerosi brevetti e proposte per strumenti di questo tipo, denominati sensori indossabili [11, 12, 13]. Non è fondamentale la scelta della posizione dello strumento, in quanto l'unica richiesta è che vada a trovarsi in una zona in cui si abbia una sufficiente produzione di sudore eccrino, ovvero quello designato alla regolazione della temperatura corporea [15]. Sono stati condotti studi sull'influenza della condizione di produzione del sudore, ovvero se il campione sia stato prodotto mediante intensa attività fisica [16], oppure tramite riscaldamento, ad esempio durante un ciclo di sauna. Per quanto riguarda il glucosio, va considerato che il sudore prodotto mediante attività fisica presenterà concentrazioni minori in quanto durante tale attività questo viene consumato maggiormente per produrre energia. Durante un ciclo di sauna invece non si ha questo effetto ed il livello glicemico sarà in generale più alto. In Figura 2.3 è riportato uno schema riassuntivo delle tipologie di sensori e metodi discussi.



**Figura 2.3:** Schema riassuntivo delle metodologie di misura del glucosio [17].

Tali misure introducono però delle variabili aggiuntive. Va infatti considerato che, anche se la correlazione tra la concentrazione del glucosio nel sangue e nel sudore è ormai nota, le due quantità non sono uguali, ma il livello presente nel sudore è

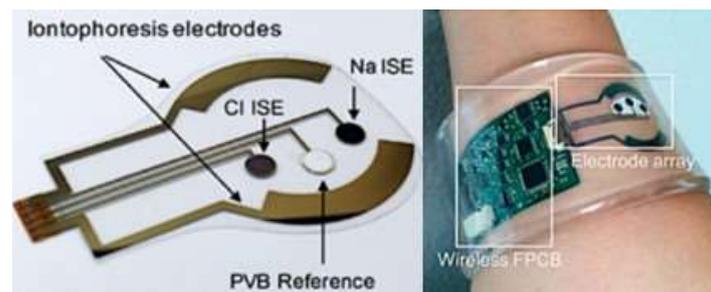
sensibilmente inferiore rispetto alla glicemia sanguigna [18]. Si parla di concentrazioni medie di 2 millimoli di glucosio in un litro di sudore per pazienti con livelli glicemici normali. I metodi utilizzati dovranno avere dunque una elevata sensibilità a basse concentrazioni. Il principale limite dei prototipi di oggi, e motivo per cui non siano ancora ampiamente diffusi, sono infatti le basse performance che forniscono rispetto ai metodi invasivi o poco invasivi come conseguenza dei bassi range che devono essere raggiunti. La ricerca sta infatti puntando all'aumento dell'efficienza di tali apparecchiature, in termini sia di precisione che di accuratezza. Inoltre, altri componenti metabolici mostrano un comportamento opposto e possono anche essere presenti delle sostanze che nel sangue erano presenti solamente in tracce o assenti. Ad esempio, urea e lattato di sodio [19] sono due molecole presenti in quantità maggiori nel sudore piuttosto che nel sangue, soprattutto se si considerano attività fisiche intense, che modificano temporaneamente il metabolismo in maniera da produrne maggiori quantità e, per contro, consumare più glucosio. Si consideri che le concentrazioni fisiologiche medie nel sudore per un individuo sano, a riposo ed a temperatura ambientale si aggirano intorno alle mM di lattato e 20 mM per l'urea. Questo porta ad avere più componenti nella miscela e più possibili interferenze, ovviamente a seconda della metodologia considerata. Uno degli obiettivi principali di questo studio è stato infatti proporre un metodo che non sia compromesso da queste interferenze. Un altro aspetto ancora carente di questi strumenti è la necessità di una ricalibrazione periodica dopo determinati intervalli di tempo, oltre che al fatto di dover sostituire interamente lo strumento dopo un certo periodo di attività, operazione comunque facile considerata la non invasività dello stesso. Entrambi questi parametri temporali sono molto variabili a seconda del brevetto considerato e anche a seconda del metodo di misura sfruttato, elettrochimico o spettrometrico [20]. I primi richiedono frequenze di manutenzione maggiori, viste le componenti enzimatiche presenti.

## *2.2. Il lattato*

Il lattato, presente in forma anionica legata a diversi cationi metallici o in forma di acido lattico all'interno del metabolismo umano, ha un'importanza metabolica piuttosto differente rispetto al glucosio. Infatti, non è una fonte primaria di energia, ma viene impiegato per quest'uso solo in condizioni critiche, producendo piruvato, specialmente nei muscoli e nel fegato. Una via alternativa è la gluconeogenesi, processo che avviene a livello del fegato nel ciclo Cori e con cui si consuma lattato per produrre glucosio, più utile per la produzione di energia. In generale, il lattato è un prodotto di scarto del metabolismo anaerobico, legato dunque a condizioni di scarsa ossigenazione dei tessuti, possibile conseguenza di attività fisica intensa o patologie specifiche. Un'ulteriore funzione del lattato è quella di regolatore metabolico, visto che ha influenza sulla vasodilatazione e, dunque, sul trasporto da parte del sangue dell'ossigeno ai tessuti dell'organismo. Le principali complicanze cliniche riguardanti questo metabolita sono dovute al suo accumulo. Malattie mitocondriali possono indurre la cellula ad affidarsi al metabolismo anaerobico ed al conseguente aumento eccessivo della concentrazione di lattato. Nei casi più gravi, l'accumulo potrebbe sfociare nell'acidosi lattica, con conseguente abbassamento del pH sanguigno.

### 2.2.1. Metodi di misura del lattato

Come per il glucosio, esistono metodi di misura diretti, ovvero effettuati in seguito al prelievo di sangue da parte di personale medico e la seguente analisi in laboratorio mediante spettrometria o cromatografia liquida ad alta prestazione. Da oltre un decennio sono però diventati comuni sensori indossabili, molto richiesti nell'ambito sportivo, in quanto coniugano la praticità della misura continua e gestibile dall'utilizzatore ed un discreto livello di comodità e discrezione che permette di effettuare attività fisica con un impatto nullo. Sono stati progettati sensori elettrochimici-enzimatici, che sfruttano enzimi come lattato ossidasi. Il contatto dell'enzima con il lattato presente nel sudore produce un flusso di elettroni proporzionale alla concentrazione del metabolita, la quale viene letta dal sensore ed elaborata. Un'interessante variante già piuttosto diffusa è quella di dispositivi noti come *'Wearable Smart'*, simili ad orologi con lo strumento di acquisizione posizionato nella parte posteriore o nel cinturino e a contatto con l'epidermide del polso, zona in cui si ha una buona produzione di sudore durante attività fisica. L'avvento della tecnologia flessibile ha reso possibile anche la produzione di strumenti ultra-sottili specifici per attività che richiedano il minimo livello di invasività possibile (Figura 2.4).



**Figura 2.4:** Sensore pieghevole per la misura del lattato. La sudorazione può essere indotta tramite ionoforesi, ovvero una breve scarica elettrica a bassa intensità [21].

### 2.3 L'urea

L'urea è un metabolita sensibilmente diverso rispetto a glucosio e lattosio. La sua funzione principale è, infatti, legata al cosiddetto ciclo dell'urea, ovvero quella serie di processi con cui l'azoto derivante dal catabolismo di amminoacidi e proteine viene utilizzato per sintetizzare l'urea, che viene espulsa attraverso l'urina. Tale cammino metabolico è anche noto come detossificazione dell'ammoniaca, visto che serve ad evitare accumuli di questa molecola tossica e altamente dannosa. Si deve considerare che l'urea è molto meno nociva e tossica dell'ammoniaca, ma in grandi concentrazioni può portare a problematiche simili. Difetti nel ciclo dell'urea, come malattie epatiche, potrebbero portare ad un suo eccessivo accumulo, che causerebbe gravi danni neurologici vista la nocività specifica nei confronti del sistema nervoso centrale. Anche in caso di insufficienza renale si avrebbero conseguenze di questa portata, visto che l'organismo non riuscirebbe a disfarsi dell'accumulo formatosi. Un secondo contributo fondamentale dell'urea è quello di regolazione dell'equilibrio idrico dell'organismo

veicolando il movimento dell'acqua attraverso vari tessuti. La variazione dei livelli di urea nel sudore può dunque essere un indice di problemi di idratazione o alimentazione, possibili sintomi di patologie più gravi. Al contrario del lattato, esercizio fisico di modesta intensità porta alla diminuzione della concentrazione dell'urea nel sudore. Questo avviene in quanto l'organismo tenderà ad espellere più liquidi, principalmente acqua, diluendo dunque la quantità di urea espulsa dall'epidermide. Con l'aumentare dell'intensità dell'attività fisica, però, si avrà l'aumento del catabolismo proteico in unione con un certo livello di disidratazione, con conseguente produzione di maggiori quantitativi di urea che verrà espulsa via sudore.

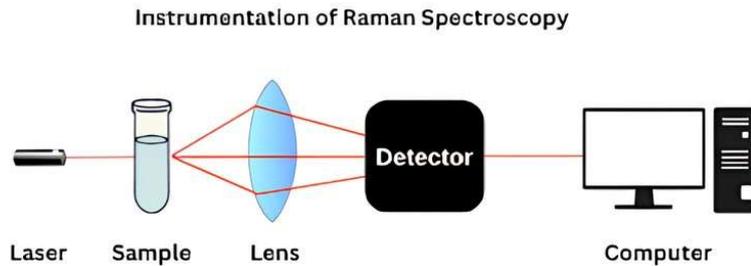
### 2.3.1. Metodi di misura dell'urea

Per quanto riguarda le analisi in laboratorio, molto comuni sono le analisi colorimetriche, che sfruttano reagenti che cambiano colore in presenza di urea, permettendo l'analisi mediante successiva spettroscopia. Anche per l'urea esistono metodi enzimatici, basati sull'azione dell'enzima ureasi, che catalizza la degradazione dell'urea ad ammoniaca e anidride carbonica. Per la determinazione è sufficiente la misura della variazione del pH. In alternativa, si possono avere sistemi basati su anticorpi specifici, con funzionamento simile. Esistono, inoltre, sensori non invasivi e indiretti di tipo elettrochimico, che misurano la variazione della conduttività del sudore al variare del livello di urea. Tali metodi sono incorporabili in strumenti indossabili per il monitoraggio continuo.

## 2.4. *La spettroscopia Raman*

Come descritto in precedenza, una delle possibilità più interessanti per lo sviluppo di strumenti non invasivi per il monitoraggio continuo della glicemia tramite misure indirette è il campo delle misure ottiche. Fra queste, la spettroscopia Raman è la tecnica più promettente [22], in quanto coniuga una relativa semplicità nella strumentazione, denotata dagli innumerevoli studi e brevetti correlati, con una buona precisione nella misura. Oggigiorno, infatti, i metodi basati su questa tecnologia non possiedono ancora prestazioni soddisfacenti per sopravanzare i metodi elettrochimici poco invasivi o invasivi, ma se si riuscisse a migliorare questo aspetto il cambio sarebbe netto, visti tutti i vantaggi in termini di comodità già discussi.

Il metodo prende il nome dal noto fisico indiano Chandrasekhara Venkata Raman, Premio Nobel per la fisica nel 1930, proprio grazie agli studi sulla diffusione della luce che portarono allo sviluppo di questa tecnica. La spettroscopia Raman è una tecnica analitica non distruttiva che sfrutta il fenomeno dello scattering, anche noto come Effetto Raman. In figura 2.5 è illustrato lo schema sintetico dell'apparecchiatura.



**Figura 2.5:** Schema di funzionamento di uno spettrometro Raman.

Una radiazione luminosa monocromatica, emessa da un apposito laser, e denominata radiazione incidente, viene inviata contro il campione, che può essere liquido o solido per studiarne le vibrazioni molecolari. Il campo elettromagnetico prodotto dai fotoni della radiazione causa una polarizzazione della nube elettronica delle molecole del campione, le quali vengono dunque eccitate fino ad uno stadio energetico superiore. Tale effetto è stato descritto come la formazione di un complesso fotone-molecola virtuale, denominato semplicemente stadio virtuale o fittizio ed è equivalente ad un dipolo elettrico momentaneo governato dalla legge

$$\boldsymbol{\mu} = \boldsymbol{\alpha} \cdot \boldsymbol{E} \quad (2.1)$$

Dove  $\boldsymbol{\mu}$  è il momento di dipolo indotto sulla molecola come effetto della polarizzazione,  $\boldsymbol{E}$  è il campo elettrico prodotto dalla radiazione e  $\boldsymbol{\alpha}$  è definito come il tensore di polarizzabilità. La molecola si trova dunque in uno stadio molto instabile e riemette l'energia del fotone in tempi estremamente brevi come luce diffusa, che viene convogliata, tramite un sistema di specchi più o meno complesso a seconda dello strumento, e inviata allo spettrometro per l'ottenimento dello spettro finale. È anche presente un sistema di lenti ottiche per l'amplificazione del segnale e l'aumento di risoluzione dei risultati.

Dal punto di vista macromolecolare si osserva quindi che la radiazione incidente crea una perturbazione nella struttura elettronica del materiale in esame, il che risulta in vibrazioni molecolari specifiche a seconda della struttura chimica del campione stesso. L'energia luminosa viene poi restituita come luce diffusa ed il processo avviene secondo tre meccanismi differenti. Per la maggior parte, si ha l'equivalente di un urto elastico, con il fotone che non scambia energia con la molecola e la luce viene diffusa con la stessa lunghezza d'onda della radiazione iniziale, e dunque con la stessa energia. Tale meccanismo viene detto diffusione di Rayleigh. Il secondo meccanismo vede invece la molecola assorbire parte dell'energia della radiazione incidente, che verrà riemessa come luce diffusa con una lunghezza d'onda maggiore, e dunque energia minore. Si parla quindi di effetto di scostamento di Raman mediante diffusione di Stokes, assimilabile ad un urto anelastico tra la molecola e il fotone. In maniera opposta, si può avere un terzo caso in cui l'impatto tra la radiazione incidente e la molecola risulta nell'aumento di energia della luce, che verrà riemessa con una

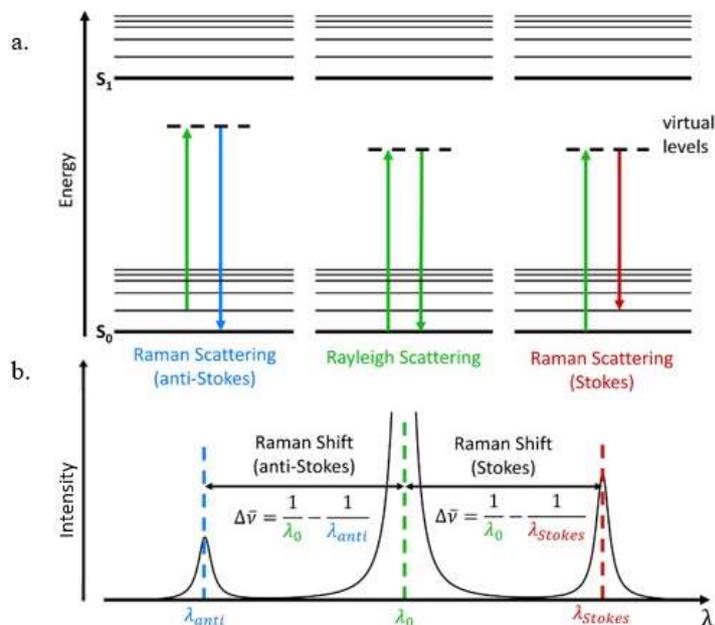
lunghezza d'onda minore. Questo effetto è denominato scostamento Raman mediante diffusione anti-Stokes. Nel totale si avranno dunque tre radiazioni riemesse, con tre differenti lunghezze d'onda. Riprendendo l'equazione (2.1) ed espandendo il tensore di polarizzabilità in serie di Taylor si arriva alla seguente formulazione

$$\mu = E_0\alpha_0\sin 2\pi\nu t + \frac{1}{2}\beta E_0[\cos 2\pi(\nu - \nu_{vib})t - \cos 2\pi(\nu + \nu_{vib})t] \quad (2.2)$$

dove si può notare facilmente la dipendenza dalla frequenza della luce diffusa nei tre meccanismi. In particolare,  $\nu$  è la frequenza della radiazione incidente iniziale,  $\nu_{vib}$  rappresenta la frequenza della vibrazione molecolare indotta. Quest'ultima compare nel termine  $\nu - \nu_{vib}$ , rappresentante l'effetto Stokes, in quanto la frequenza finale è minore di quella incidente e dunque vi è stato un trasferimento di energia dalla molecola alla radiazione. Vi è poi l'analogo termine  $\nu + \nu_{vib}$  relativo alla diffusione anti-Stokes, con passaggio energetico in verso opposto ed una frequenza risultante maggiore. Va sottolineato che fra i meccanismi di Stokes e anti-Stokes vi è una propensione statistica verso il primo. Questo avviene perché allo stato rilassato, la maggior parte delle molecole del campione si troveranno al livello energetico di base secondo la distribuzione di Boltzmann, e dunque non potranno passare ad un livello energetico ancora più basso alla fine del processo. I termini  $E_0$ ,  $\alpha_0$  e  $\beta$  sono invece fattori dipendenti dalle caratteristiche del sistema, come la tipologia di luce utilizzata, l'intensità e l'angolo di impatto.

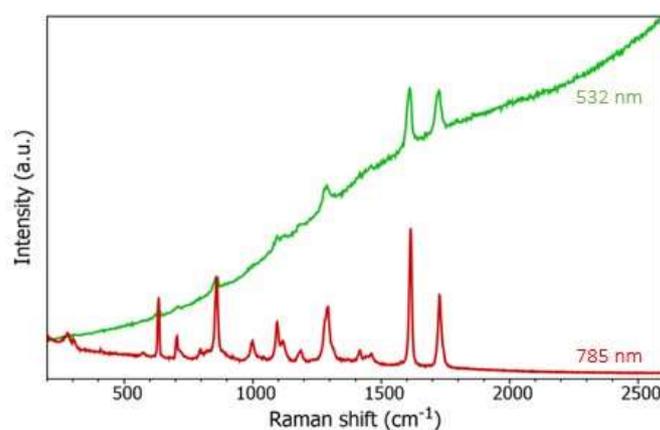
La variazione della frequenza fra la radiazione incidente e quella riflessa, data dalla somma degli effetti di diffusione di Rayleigh, Stokes e anti-Stokes, globalmente nota come Raman shift, è proporzionale alla quantità di energia scambiata, la quale dipende a sua volta dalla tipologia di vibrazione molecolare che la causa. Determinate configurazioni molecolari produrranno corrispondenti scostamenti Raman e questo principio fornisce la base per questa metodologia: analizzare l'entità dello scattering per risalire alla tipologia di struttura molecolare. Inoltre, a seconda delle intensità delle frazioni riflesse, si può risalire anche alla concentrazione del campione, visto che più legami di un determinato genere produrranno vibrazioni molecolari dello stesso tipo, ma con intensità di riflessione della luce maggiori.

Va sottolineato che fra i meccanismi di Stokes e anti-Stokes vi è una propensione statistica verso il primo. Questo avviene perché allo stato rilassato che si ha a temperatura ambiente, la maggior parte delle molecole del campione si troveranno al livello energetico base secondo la distribuzione di Boltzmann, e dunque non potranno risultare ad un livello energetico ancora più basso alla fine del processo (Figura 2.6).



**Figura 2.6:** Rappresentazione dell'effetto di scattering Raman. Sono visibili i tre meccanismi di Rayleigh, Stokes e anti-Stokes (a) e gli scostamenti in termini di frequenza della radiazione diffusa (b).

Una delle principali difficoltà nell'utilizzo di questa tecnica è dato dalla fluorescenza, ovvero il fenomeno per cui si ha il decadimento di un elettrone dallo stato fondamentale del livello elettronico eccitato, dopo l'azione della radiazione incidente, ad uno dei livelli vibrazionali dello stato elettronico fondamentale. Questo fa sì che una grande parte di energia venga emessa ad una lunghezza d'onda maggiore andando quindi a sommarsi all'effetto della diffusione di Stokes, causando una copertura dei picchi di questo effetto che sono molto meno intensi per natura (Figura 2.7).



**Figura 2.7:** Effetto della fluorescenza su un generico spettro Raman. Lo spettro verde è affetto da una forte fluorescenza e i picchi presenti nella metà sinistra dello spettro risultano totalmente nascosti. Lo spettro rosso è ottenuto con una sorgente laser a 785 nm, mentre quello verde a 532 nm.

Il risultato è che si avrà una linea di base dello spettro molto più alta e dunque l'intensità dei picchi sarà molto minore, in quanto la parte bassa viene coperta dal più forte effetto di fluorescenza. In casi estremi alcuni picchi possono essere addirittura nascosti dalla curva di fluorescenza, rendendo quindi totalmente impossibile la misura quantitativa ed andando ad invalidare anche la misura qualitativa, visto che determinate vibrazioni molecolari in realtà presenti non risulteranno visibili. Per minimizzare questo effetto, la contromisura più efficace è utilizzare una sorgente luminosa con una lunghezza d'onda maggiore, ovvero meno energetica, spostandosi quindi verso il campo delle radiazioni rosse o nel vicino infrarosso, motivo per cui le sorgenti predilette sono laser e diodi laser.

Esistono due varianti della spettroscopia Raman estremamente interessanti e promettenti nel campo della misura della concentrazione in campioni di derivazione biologica. Il problema principale di queste misure si rivela essere la bassa concentrazione dell'analita ricercato nella maggior parte dei fluidi fisiologici. Per cercare una soluzione a questo dilemma, va considerato che il principio di base della spettroscopia Raman, cioè lo scattering elastico ed anelastico, è un fenomeno di superficie, che dipende dunque anche dalla sezione d'urto della radiazione. Queste due varianti agiscono in tal senso.

- *Surface Enhanced Raman Scattering (SERS)*: questa tecnica utilizza un substrato metallico specifico, posto a distanza nanometrica dal campione, per aumentare la densità elettronica. Questo porta all'aumento dell'intensità della radiazione diffusa e quindi ad una sensibilità maggiore anche a basse concentrazioni. Per questo motivo gran parte degli studi sulla concentrazione del glucosio nel sangue e nel sudore sono effettuati con questa metodologia [23, 24].
- *Microspettroscopia Raman*: questo metodo sfrutta l'accoppiamento della spettroscopia Raman tradizionale con un microscopio ottico, il quale cattura la luce diffusa dal campione e la invia allo spettrometro. In questo modo si amplifica la risoluzione sulla superficie del campione, riuscendo a lavorare con superfici anche inferiori al micrometro. In campo biologico e medico rappresenta un enorme vantaggio, visto che non si necessita più di volumi di liquido di qualche millilitro, ma una semplice goccia risulta sufficiente.

## 2.5. Metodi di analisi chemiometrici

Le analisi Raman producono una notevole quantità di dati, visto che per ogni campione viene prodotto uno spettro completo e dunque si analizza il comportamento del campione a vari numeri d'onda riferiti alla radiazione. Perciò, ogni set di dati acquisito sarà costituito da una matrice in cui ciascuna riga rappresenta uno spettro e ogni colonna si riferisce ad uno specifico numero d'onda. È possibile utilizzare metodi chemiometrici per determinare quale parte dei dati acquisiti è realmente necessaria per rappresentare l'informazione nello spettro.

### 2.5.1. Principal Component Analysis

La *Principal Component Analysis* (PCA) è una tecnica di analisi dei dati che studia la correlazione fra le variabili del sistema multivariato preso in esame per determinare quali possano essere rimosse poiché contenenti informazioni ridondanti o, in generale, di poco peso nella creazione del modello finale. L'algoritmo consiste nel trasporre la matrice del problema iniziale, contenente le misure acquisite, in un nuovo spazio vettoriale costruito sulla base dei *Principal Component*, ovvero delle nuove variabili che sostituiscono quelle di partenza e che contengono la maggior parte della varianza dei dati iniziali [25]. Questo vuol dire che tali componenti, anche denominati *Loadings*, rappresentano un nuovo modello di dimensioni minori rispetto a quello di partenza, ma contenente un'approssimazione sufficientemente buona dell'informazione contenuta nel set base, tipicamente almeno il 90% del totale. Il problema si può rappresentare come un calcolo matriciale

$$\mathbf{X} = \mathbf{T} \cdot \mathbf{P}^T + \mathbf{E} \quad (2.3)$$

dove  $\mathbf{X}$  è la matrice del data set originale di dimensioni  $m \times n$ ,  $\mathbf{T}$  è la matrice composta dagli autovalori del metodo e di dimensione  $m \times k$ ,  $\mathbf{P}$  è il vettore dei *Principal Components*, di dimensione  $n \times k$ , che rappresentano gli autovettori del nuovo spazio vettoriale, ed  $\mathbf{E}$  è la matrice degli errori, che racchiude le informazioni non espresse dal modello PCA, ovvero l'informazione non rappresentata dai *Loadings* prescelti, ovviamente di dimensione  $m \times n$ . Nel nostro caso avremo  $m$  pari al numero di campioni e  $n$  pari al numero di variabili del modello, ovvero i numeri d'onda. Diagonalizzando la matrice di covarianza relativa al set di dati originali  $\mathbf{X}$ , si ottiene una nuova matrice, i cui autovalori ed autovettori saranno rispettivamente i componenti della matrice  $\mathbf{T}$  e del vettore  $\mathbf{P}$ . Durante un'analisi mediante PCA, l'operatore può anche variare il numero di componenti principali necessario. Aumentando tale parametro, si otterrà una varianza totale maggiore, e quindi un modello più accurato, a discapito però della pesantezza, in quanto il modello avrà una variabile in più. Fondamentale è dunque trovare un numero di componenti principali che rappresenti un compromesso tra accuratezza e riduzione di dimensioni del set base.

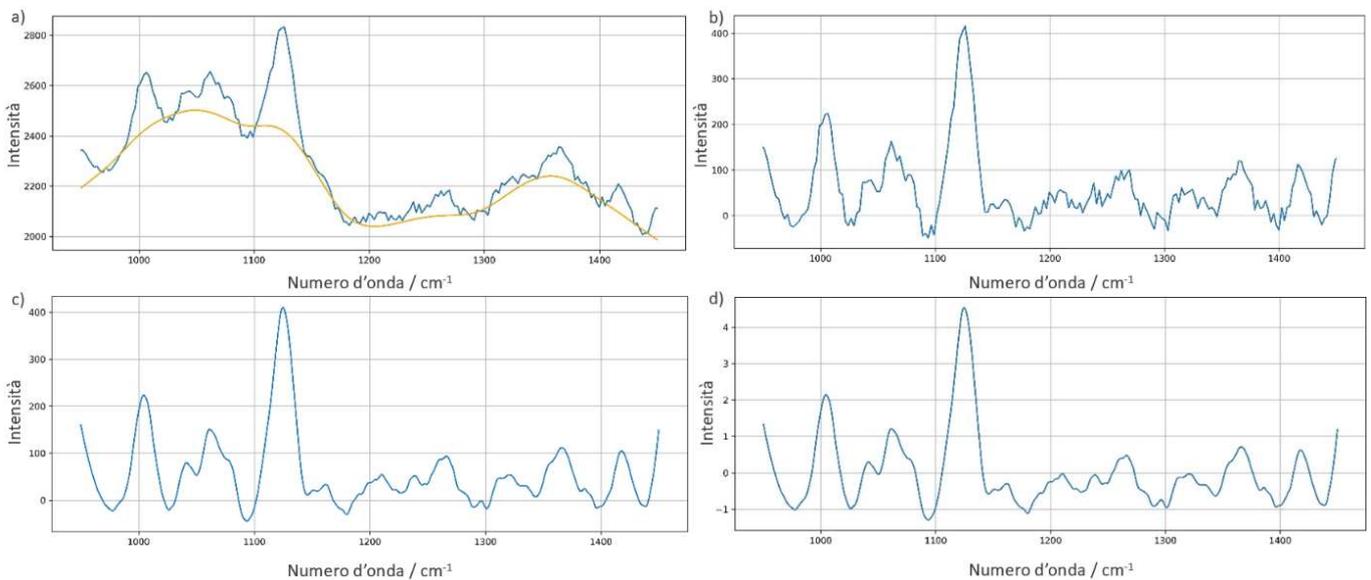
Questo metodo richiede una fase di manipolazione preliminare dei dati, poiché risulta essere molto sensibile al rumore di fondo delle misure e all'intensità del segnale, che nella spettroscopia Raman varia anche a seconda della scelta della radiazione incidente. La procedura, nota come *Pre-Processing*, consiste in quattro passaggi consecutivi (Figura 2.8):

- Selezione dell'intervallo: a partire dal set base, si possono rimuovere degli intervalli considerati di poco interesse. Quando si analizzano campioni chimici relativamente semplici, questi possono presentare picchi Raman localizzati in zone definite dello spettro. Rimuovendo le parti dello spettro in cui non si hanno picchi interessanti si riduce la possibilità di errori di misura o influenze di interferenti indesiderati.
- Rimozione linea di fondo: effetti secondari della spettroscopia Raman, principalmente legati alla fluorescenza, portano ad avere un segnale di base non nullo e lo spettro risulta avere una caratteristica forma a campana. Questo

influisce sulla determinazione del punto di inizio di un picco e, dunque, rende più difficile la misura dell'intensità reale. Esistono algoritmi specifici per effettuare questa operazione.

- **Smoothing:** questo processo consiste nel diminuire, o teoricamente addirittura annullare, il rumore di fondo della misura, aumentando quindi il più possibile il rapporto rumore-segnale. In tal modo si ottiene uno spettro meno rumoroso, aumentando dunque la risoluzione della curva e rendendo più agevole ed accurata la valutazione dell'intensità di un picco. Anche per questa tecnica esistono molti algoritmi differenti, denominati filtri. Il più comune è il filtro di Savitzky-Golay, che costruisce una polinomiale a tratti ai minimi quadrati locali che approssimi lo spettro grezzo. Per maggiore precisione, lo spettro viene scomposto in sezioni di uguale ampiezza e su ognuna di queste sezioni viene costruita una polinomiale a sé. Come input di questo metodo occorre fornire il grado desiderato della polinomiale e l'ampiezza dell'intervallo, in termini di numero di dati, su cui costruire l'approssimazione. Si deve prestare attenzione in questo passaggio, visto che uno smoothing troppo aggressivo potrebbe causare l'abbassamento parziale anche dei picchi d'interesse e, dunque, una notevole perdita di informazione.
- **Normalizzazione:** utile per rendere tutti gli spettri centrati rispetto al valor medio e riportati ad unità di varianza, ovvero al posto dell'intensità si visualizzerà la varianza rispetto al valor medio. La formulazione matematica di questo passaggio è rappresentata dall'equazione

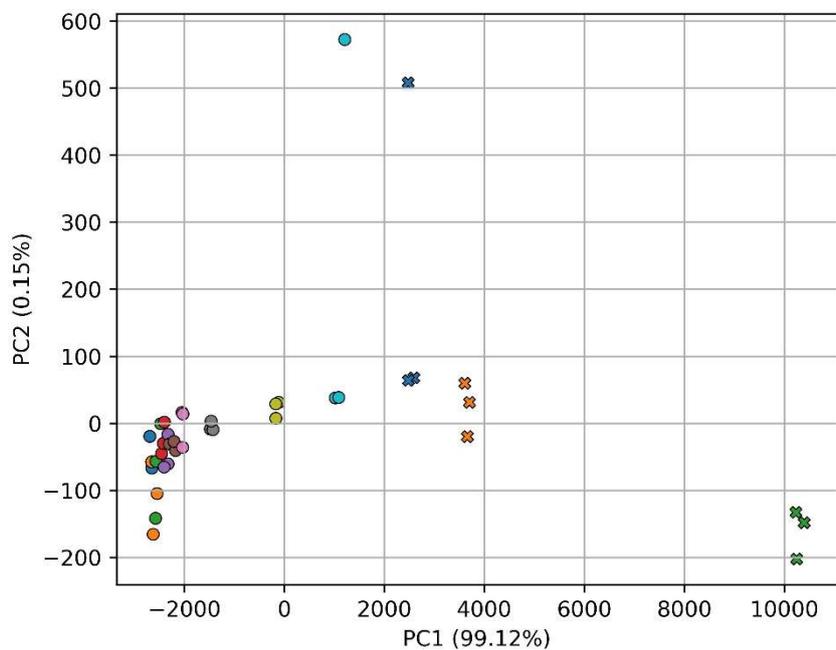
$$y_N = \frac{y - \bar{y}}{\sigma} \quad (2.4)$$



**Figura 2.8:** Effetto dei passaggi di *Pre-Processing* su uno spettro Raman, con identificazione della linea di fondo gialla (a), dopo la sua rimozione (b), dopo lo *smoothing* con filtro di Savitzky-Golay (c) e dopo normalizzazione (d).

Eseguendo tali passaggi si ottengono spettri di più facile lettura e, soprattutto, si elimina una gran parte di fattori che modificano lo spettro senza apportare informazioni utili per il tipo di misura eseguita.

Una volta eseguito il *Pre-Processing* ed ottenuta la nuova rappresentazione semplificata della matrice di dati iniziale, quindi ottenuti anche gli autovettori e gli autovalori corrispondenti, si possono effettuare analisi statistiche per determinare la natura delle misure di partenza, distinguendo cioè spettri statisticamente simili tra loro, oppure al contrario si può notare se più misure effettuate su uno stesso campione, e che dovrebbero essere dunque statisticamente simili, presentano errori di grandi proporzioni. Uno strumento molto utile per questo scopo sono i diagrammi di *clustering* o *biplots*. Tali diagrammi riportano in ordinata e ascisse gli autovalori riferiti a due diversi autovettori del modello. Ogni singolo campione analizzato è rappresentato su questi diagrammi da un punto e diversi punti vicini (che formano quindi un *cluster*) saranno da considerare statisticamente simili. Si supponga di avere eseguito cinque misure spettroscopiche Raman su un singolo campione a concentrazione nota. Su questi diagrammi ci si aspetta di trovare i cinque punti in prossimità l'uno dell'altro. Se quattro fossero vicini e uno solo molto distante, questo sarebbe indice di un errore nella misura e dunque si potrebbe procedere con l'esclusione del campione dalla costruzione del modello. In figura 2.9 si nota per esempio una buona distribuzione delle prove (più punti di pari forma e colore rappresentano ripetizioni diverse della stessa concentrazione). Solo due ripetizioni si distaccano dalle simili, ma considerando il basso valore di peso del componente principale 2, riportato in ascissa, questo distacco non rappresenta una differenza rilevante nella prova.



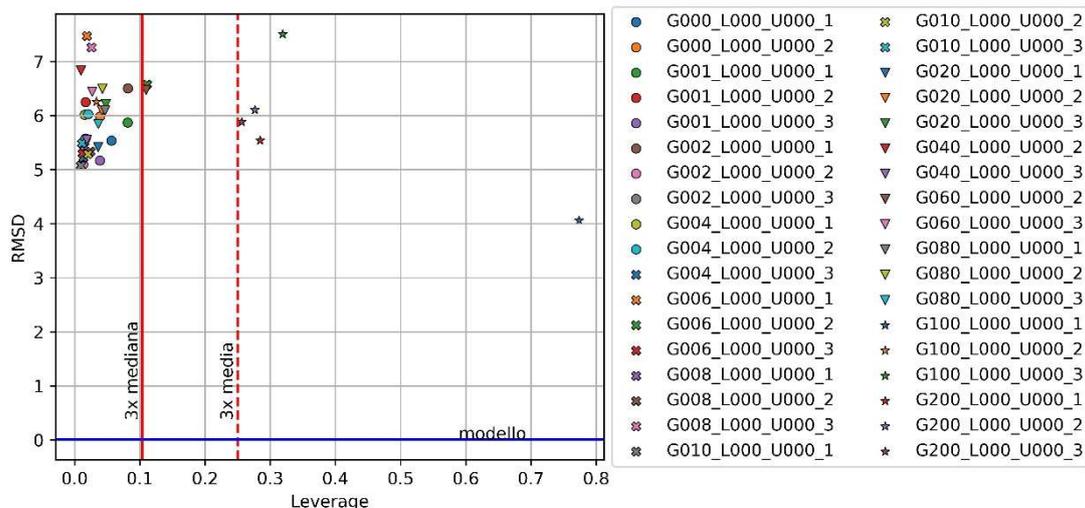
**Figura 2.9:** Esempio grafico biplot per la determinazione dei *cluster* fra le prove eseguite su un set di lattato in acqua a concentrazione crescente.

In maniera analoga, si può costruire un secondo diagramma, noto come diagramma dei valori anomali o degli *outliers*. Tale diagramma riporta in ascisse il cosiddetto “Effetto Leva” o *Leverage*, rappresentato da H nell’equazione (2.5), mentre in ordinata è riportata la radice dello scarto quadratico medio RMSD, ottenibile come illustrato nella (2.6)

$$H = T(T^T T)^{-1} T^T \quad (2.5)$$

$$RMSD = \sqrt{\frac{\sum_{i=1}^J (X - TP^T)^2}{J}} \quad (2.6)$$

dove J rappresenta il numero di valori di numero d’onda rimasti dopo la selezione dell’intervallo nel *Pre-Processing*. Il valore del *leverage* è rappresentativo dell’influenza che il campione preso in considerazione ha sulla costruzione del modello finale. Uso comune è imporre una soglia di accettabilità di questo valore, tipicamente posta a tre volte il valore medio di *leverage* fra tutti i campioni, oppure addirittura a tre volte la mediana. Campioni che eccedano tale soglia sono denominati *outlier* e saranno da considerare troppo influenti rispetto agli altri nella costruzione del modello. Questa condizione è indesiderata in quanto si otterrebbe un modello sbilanciato, ovvero molto preciso nell’intorno dell’*outlier* preponderante e molto impreciso per campioni molto diversi dall’*outlier*. Il valore di RMSD deve essere il più basso possibile, in quanto rappresenta quanto un campione si distacca dalla previsione del modello. In figura 2.10 si può notare una distribuzione omogenea di RMSD su valori piuttosto bassi, quindi non problematici, mentre il *leverage* evidenzia qualche problema nella prima ripetizione del campione di glucosio 100 mM, il cui simbolo compare molto più a destra degli altri valori. Questo porta a pensare di eliminare la misura dal set, per poi ripetere il calcolo e verificare che non vi siano più prove così preponderanti. I campioni oltre la linea di 3x media sono anch’essi possibili *outliers*, ma tenendo conto delle alte concentrazioni e che si trovano in vicinanza della soglia e non spostati come la misura precedentemente discussa, possono essere mantenuti nel modello.



**Figura 2.10:** Esempio di grafico per lo studio degli *outliers* per campioni di glucosio in acqua con concentrazione crescente.

In seguito a questi passaggi, si possono ricostruire gli spettri a partire dal nuovo sistema di dati, utilizzando l'equazione (2.7). Se necessario, si potrebbe sovrapporre uno dei grafici del set di dati iniziale con l'equivalente ritrasformato in maniera da verificare la bontà del modello.

$$\mathbf{X}' = \sum_i t_i \mathbf{p}_i^T \quad (2.7)$$

dove  $\mathbf{X}'$  è il grafico ricostruito,  $t_i$  è l' $i$ -esimo autovalore corrispondente all' $i$ -esimo autovettore  $\mathbf{p}_i$ .

### 2.5.2. Principal Component Regression e Partial Least-Square Regression

Anche nota come PCR, la *Principal Component Regression* è una tecnica di regressione analitica [26]. Il modello di base è quello della classica regressione lineare, ma la differenza è nel fatto che la PCR non viene effettuata sul set di dati base, bensì sulla nuova rappresentazione ottenuta dopo aver effettuato la PCA sul modello di partenza. La nuova predizione avrà la seguente formulazione

$$\mathbf{Y} = \beta_0 + \sum_i \beta_i \mathbf{PC}_i + \mathbf{E} \quad (2.8)$$

Dove  $\mathbf{Y}$  è il valore della variabile di risposta, i termini  $\beta_i$  sono i coefficienti determinati dal metodo e i termini  $\mathbf{PC}_i$  sono i *loadings* dei componenti principali derivati dalla PCA. In particolare, il vantaggio principale è che non si effettua una regressione

lineare sul set di dati intero, che come sempre può essere di notevoli dimensioni, ma su uno di dimensioni ridotte, così da diminuire il tempo e il requisito computazionale dei calcoli. Inoltre, avendo eseguito la PCA in precedenza, possiamo preventivamente evitare di inserire in una classica regressione delle misure errate o rumorose, vista la fase di *Pre-Processing* e di individuazione degli *outliers*. Questo processo si inserisce nella PCA prima della trasformazione inversa descritta in precedenza come ultimo step della *Principal Component Analysis*. La tecnica risulta molto efficiente nel caso in cui si abbia una elevata multicollinearità, ovvero quando molte delle variabili del set di partenza siano correlate tra loro. Questo caso avrebbe grande influenza sulla classica regressione lineare, rendendone inaffidabili le valutazioni.

Una valida alternativa è rappresentata dalla *Partial Least Square Regression* (PLSR). La differenza di base fra questi due modelli è che la PCR agisce in maniera da massimizzare la varianza tra le variabili predittive, visto che si basa sulla PCA, mentre la PLSR è volta a massimizzare la covarianza tra le variabili predittive  $X$  e le variabili di risposta del processo  $Y$ . Questo obiettivo viene raggiunto tramite la costruzione di componenti latenti, ovvero combinazioni lineari delle variabili predittive  $X$  realizzate in maniera da massimizzare la correlazione tra i campi  $X$  e  $Y$ . Questo fa sì che la capacità predittiva di questa tecnica sia superiore alla PCR soprattutto in casi in cui la correlazione tra le variabili predittive e di risposta sia molto complessa. Per modelli in cui tale correlazione non sia limitante, è invece consigliabile l'utilizzo della PCR, vista la sua maggior versatilità nel diminuire la dimensione del problema considerato. L'utilizzo di PLSR per modelli troppo semplici potrebbe condurre a situazioni di overfitting, rendendo quindi il modello troppo suscettibile a variazioni sperimentali, come anche solo l'effetto del rumore di fondo delle misure.

### 3. Materiali e metodi

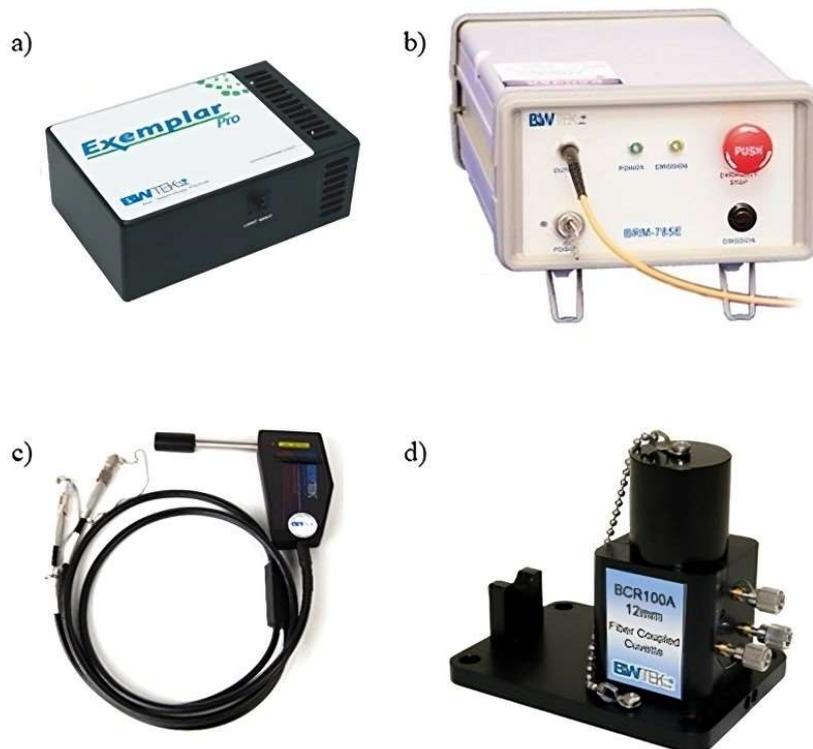
In questo capitolo vengono presentati i materiali, lo strumento e i metodi di analisi utilizzati per la costruzione del modello finale di calcolo della concentrazione di glucosio, lattato e urea in soluzione.

#### 3.1. Materiali e strumentazione

Per la produzione delle soluzioni utilizzate nelle analisi si è utilizzata acqua deionizzata mediante il purificatore e deionizzatore Milli-Q di Millipore Corporation (Massachusetts, USA). Il glucosio e il lattato di sodio sono stati acquistati da Merck KGaA Sigma-Aldrich, Missouri (USA), ad una purezza superiore al 99,5% e 99,0% rispettivamente. L'urea è stata ottenuta da Honeywell Fluka™, con purezza superiore al 99%.

Le misure sono state effettuate con uno spettrofotometro BTC675N (Figura 3.1a), coniugato con un laser BRM-785 (Figura 3.1b) ed utilizzando una cuvetta standard 1-Q in quarzo di base quadrata con lato esterno di 12,5 mm, cammino ottico interno di 10 mm e volume nominale di 3,5 ml proveniente dalla Starna Scientific Ltd, Essex (UK). Lo strumento è stato utilizzato nella sua configurazione con sonda probe a trasmissione in fibra ottica (Figura 3.1c). Per le misure è stata scelta la sorgente laser rossa a 785 nm. Per rendere la misura più riproducibile ed isolare il campione da qualsiasi radiazione esterna si è adoperato una porta-cuvette Raman BCR100A (Figura 3.1d) con sezione dell'alloggio di 12 mm estendibile. Tale apparecchiatura consiste in un contenitore di pareti nere impenetrabili alla luce, con un coperchio dello stesso materiale per richiudere l'apertura d'ingresso della cuvette dopo l'inserimento. È anche presente un sistema di specchi, regolabile tramite viti poste sul retro del pezzo, per l'intensificazione della radiazione in un punto specifico della cuvette: regolato a dovere, garantisce che il punto di focus della radiazione incidente sia posto nel centro della sezione della cuvette, evitando le pareti del contenitore stesso. Sul lato è presente l'ingresso per la sonda probe dello spettrofotometro, con un quarzo di bloccaggio per evitare che la sonda entri in contatto con la cuvette potenzialmente danneggiando entrambi. Lo spettrofotometro, il laser ed il porta-cuvette sono stati entrambi acquistati da B&W Tek, New Jersey (USA). Le cuvette sono state invece ottenute da White Bear Photonics LLC, Minnesota (USA).

Per quanto riguarda le misure, sono state effettuate delle prove per ottimizzare i parametri di acquisizione, scegliendo infine misure di 90 secondi con quattro ripetizioni e potenza del laser di 105 mW (30% del valore massimo).



**Figura 3.1:** Componenti principali dello strumento utilizzato nelle misure. Spettrofotometro BTC675N (a), laser BRM-785 (b), sonda probe a trasmissione in fibra ottica (c) e porta-cuvette Raman BCR100A (d).

### 3.2. Il metodo

Ogni misura consiste nella prima acquisizione dello spettro del campione, seguita da una fase di acquisizione del *Dark*, ovvero un rilevamento effettuato in assenza della sorgente allo scopo di creare lo spettro di fondo o rumore della misura, che viene automaticamente sottratto alla prima acquisizione per dare lo spettro finale.

Per la costruzione delle rette di taratura sono stati realizzati tre distinti set di soluzioni. Per ogni analita si è infatti realizzata una scala di concentrazioni crescenti a partire da 0 mM (campione bianco) fino a 200 mM, diluendo sempre con acqua distillata. La distribuzione dei campioni in questo range è stata eseguita in modo da essere più densa nel campo dei valori tra 0 mM e 10 mM, così da avere una maggior risoluzione nel campo delle soluzioni utili a livello fisiologico. Da 20 mM a 100 mM si è scelto uno step di 20 mM fra due misure consecutive, mentre soluzione è stata preparata fra i campioni a 100 mM e 200 mM. La scelta di un'ultima misura così distante dalle altre è stata eseguita per verificare in maniera semplice se il metodo avesse presentato una perdita di linearità per valori molto alti. L'analisi di ogni campione è stata effettuata tre volte con nuove aliquote della stessa soluzione, allo scopo di studiare la ripetibilità della misura. La stessa concentrazione non è stata creata da nuovo per le ripetizioni per annullare gli errori di preparazione delle soluzioni.

Per l'analisi dell'effetto degli interferenti si è adottato un procedimento simile, ma le soluzioni preparate contenevano uno dei tre analiti, quello di riferimento per quel determinato set, a concentrazione variabile, con la stessa distribuzione del primo passaggio, più una quantità costante degli altri due analiti, pari alla quantità fisiologica già discussa. I set creati sono riassunti in tabella 3.1

**Tabella 3.1:** Set di soluzioni preparati per l'analisi con interferenti. Le lettere G, L e U indicano le colonne dei dati di concentrazione rispettivamente di glucosio, lattato e urea in millimoli al litro per ciascun set di dati.

SET GLUCOSIO VARIABILE			SET LATTATO VARIABILE			SET UREA VARIABILE		
G (mM)	L (mM)	U (mM)	G (mM)	L (mM)	U (mM)	G (mM)	L (mM)	U (mM)
0	40	20	1	0	20	1	40	0
1	40	20	1	1	20	1	40	1
2	40	20	1	2	20	1	40	2
4	40	20	1	4	20	1	40	4
6	40	20	1	6	20	1	40	6
8	40	20	1	8	20	1	40	8
10	40	20	1	10	20	1	40	10
20	40	20	1	20	20	1	40	20
40	40	20	1	40	20	1	40	40
60	40	20	1	60	20	1	40	60
80	40	20	1	80	20	1	40	80
100	40	20	1	100	20	1	40	100
200	40	20	1	200	20	1	40	200

Per l'ultima analisi di validazione del modello non sono state effettuate nuove misure, ma si sono analizzati gli spettri ottenuti nell'analisi con interferenti. Si è effettuata l'analisi di ogni set andando a ricercare la concentrazione degli interferenti, e non del componente variabile come nella fase precedente. A titolo di esempio, si riporta il ragionamento applicato al set del glucosio: per questa validazione si è imposto al modello di calcolare la concentrazione di urea e lattato, così da poterle poi confrontare con i valori noti dalla creazione delle soluzioni e riportati in tabella 3.1.

### 3.3. Il modello di regressione

Gli input da fornire al modello matematico sono molti. Dapprima, si inseriscono in una cartella gli spettri di riferimento, ovvero quelli che verranno utilizzati per costruire il modello multivariato. Per ognuna delle analisi effettuate, questi spettri sono sempre quelli ottenuti dalle misure sui set dei singoli analiti in acqua. Così operando, il modello impara a riconoscere quali picchi deve ricercare e, paragonando l'intensità del campione incognito con questo set, fornisce la predizione finale. In una cartella separata vanno poi inseriti gli spettri da analizzare, ovvero quelli che vengono scomposti secondo gli autovettori calcolati precedentemente. Vanno inoltre inseriti i parametri per i passaggi di *Pre-Processing*, come l'intervallo di numeri d'onda di

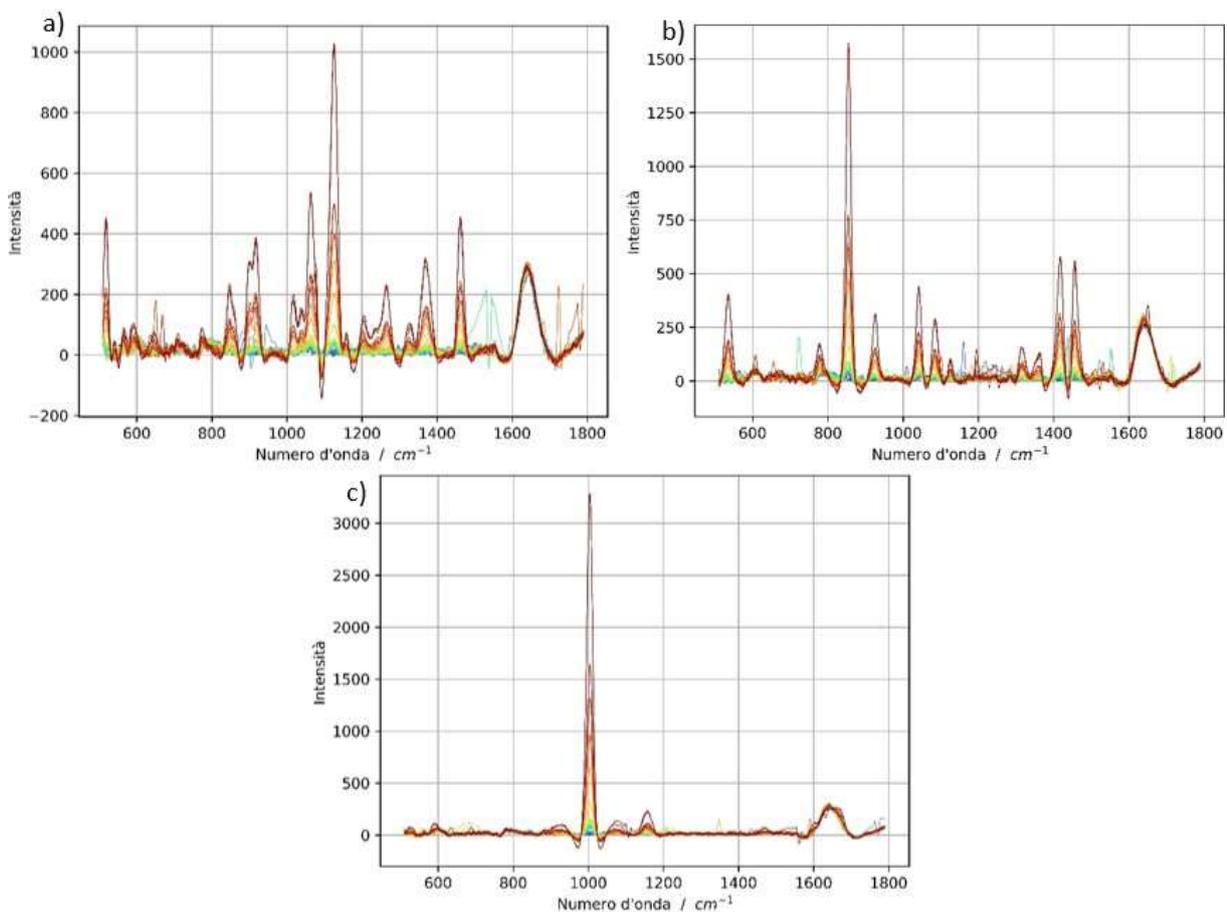
interesse, l'ordine della polinomiale e l'ampiezza della finestra desiderati per lo smoothing e, infine, il numero di *Principal Components* da utilizzare per la costruzione del modello.

Nelle analisi, si è scelto di calcolare gli autovettori lavorando sulla derivata prima degli spettri e non sugli spettri originali. Questo accorgimento è piuttosto frequente nell'analisi multivariata ed ha lo scopo di aumentare la precisione del metodo generale, andando a diminuire ulteriormente l'effetto della linea di base e di *spikes*, ovvero picchi molto stretti ed improvvisi che possono essere causati da vari fattori come dipoli momentanei in soluzioni di elettroliti deboli oppure da errori temporanei nell'elettronica del sistema di misura.

Per la sottrazione della linea di fondo è stato adoperato l'algoritmo di smoothing simmetrico ai minimi quadrati [27] importabile su Python dalla libreria libera *Scipy*. Lo smoothing integrale è invece stato effettuato con il metodo di Savitzky-Golay [28] utilizzando una polinomiale di secondo ordine su intervalli di 15 valori di numero d'onda dello spettro. Anche questo algoritmo è ottenibile dalla libreria *Scipy* tramite la funzione *savgol\_filter*. Le funzioni di base per effettuare le analisi chemiometriche sono invece ricavate dalla libreria *Scikit-learn*, nella classe di comandi denominata `sklearn.decomposition.PCA`.

## 4. Risultati

Come già scritto precedentemente, gli spettri Raman di glucosio, lattato ed urea possono presentare interferenze reciproche. Infatti, la vicinanza in termini di numeri d'onda fra alcuni dei picchi principali caratterizzanti ciascuna delle tre specie può far sì che, analizzando una soluzione in cui siano presenti contemporaneamente tutti e tre gli analiti, si abbia una certa sovrapposizione degli spettri. Questo fenomeno rende più complicato determinare il punto base esatto da cui iniziare a misurare l'area del picco, il che a sua volta causa un errore nel calcolo della concentrazione finale. Per analizzare quali possibili sovrapposizioni sono da tenere in conto, si sono esaminati dapprima gli spettri ottenuti per i tre analiti disciolti singolarmente in acqua deionizzata, con concentrazione variabile da 0 a 200 mM (Figura 4.1).



**Figura 4.1:** Spettri Raman di soluzioni acquose di glucosio (a), lattato (b) e urea (c) a concentrazione crescente. La scala cromatica va dal viola, corrispondente al campione 0 mM, fino al rosso scuro, rappresentante i campioni 200 mM.

In figura si possono notare i principali picchi dei tre analiti. Per quanto riguarda il glucosio, sono presenti diversi picchi di possibile interesse, nonostante il principale sia

quello a  $1120\text{ cm}^{-1}$ . Questo picco, in correlazione col complesso presente tra  $1000$  e  $1200\text{ cm}^{-1}$  è tipico delle vibrazioni da *stretching* del legame singolo C-O, presente sia in gruppi idrossilici che in forma di etere nell'anello glucosidico. In particolare, il legame C-O-H è il responsabile del picco a  $1120\text{ cm}^{-1}$  circa, il quale risulta più intenso degli altri proprio perché tale gruppo è più presente nella molecola rispetto al gruppo etere, il quale causa i picchi tra  $1000$  e  $1100\text{ cm}^{-1}$  [29]. Il complesso presente tra gli  $800$  e i  $1000\text{ cm}^{-1}$  è associato alla configurazione anomeric del glucosio, ovvero dall'orientamento del gruppo -OH presente sul carbonio adiacente al legame etere dell'anello acetalico, che può essere in posizione  $\alpha$  oppure  $\beta$ . Inoltre, può anche essere ricondotto a vibrazioni fuori dal piano della molecola dei legami C-H. Intorno ai  $1450\text{ cm}^{-1}$  si ha invece il picco relativo al movimento a forbice o *scissoring* dovuto ai gruppi -CH<sub>2</sub>-, i quali seguono il complesso a circa  $1350\text{ cm}^{-1}$  relativo alla vibrazione da piegamento o *bending* di -CH- e -CH<sub>2</sub>-. Per lo spettro del lattato è immediatamente identificabile il picco più intenso a  $850\text{ cm}^{-1}$  circa, dovuto alla torsione dei gruppi -CH<sub>3</sub>, seguito a  $920\text{ cm}^{-1}$  dal picco legato allo stiramento del doppio legame C=C. La serie presente a  $1020$ ,  $1060$  e  $1120\text{ cm}^{-1}$  rappresenta invece l'impronta lasciata dallo stiramento dei legami C-O, presenti in forma di gruppi alcolici C-O-H. La distribuzione in tre picchi distinti è data dalla distinzione dei gruppi alcolici semplici e quelli della parte carbossilica della molecola. A  $1400\text{ cm}^{-1}$  circa è invece presente il picco relativo alla deformazione del gruppo COO-. Nel grafico dell'urea, molto più semplice rispetto ai primi due, si nota il picco a  $1000\text{ cm}^{-1}$  dovuto alle vibrazioni per stiramento del legame -C-N. Il picco a  $1150\text{ cm}^{-1}$  circa è invece correlato a vibrazioni di piegamento del gruppo amminico -NH<sub>2</sub>. Per tutti e tre gli spettri si nota un picco a  $1600\text{ cm}^{-1}$  a intensità non variabile con la concentrazione, dovuto alla matrice di acqua utilizzata per preparare le soluzioni, in particolare alla deformazione per piegamento dell'angolo formato nella struttura H-O-H.

I metodi di analisi univariata utilizzano solo il picco principale di ciascun analita, ovvero quello più intenso, in quanto presenta maggiore sensibilità alla variazione di concentrazione e solitamente permette una buona predizione anche per soluzioni piuttosto diluite. Nonostante ciò, l'alta sensibilità si può rivelare un'arma a doppio taglio, in quanto, in caso di interferenza da parte di un altro composto chimico in miscela, si avrebbe una misura troppo dipendente anche da questa seconda molecola. Se si considera una miscela glucosio-urea, il picco a  $1000\text{ cm}^{-1}$  dell'urea andrebbe a sommarsi con il complesso presente nella stessa zona del grafico del glucosio, portando ad un aumento della sua intensità e dunque ad una predizione finale maggiore del valore corretto. Allo stesso modo il picco a circa  $850\text{ cm}^{-1}$  del lattato si trova in corrispondenza di un analogo complesso di picchi nel glucosio. Ovviamente, per l'analisi del glucosio il picco più interessante è quello a  $1120\text{ cm}^{-1}$ , ma anche in questa zona vi sono picchi secondari di urea e lattato. Infatti, il picco a circa  $1150\text{ cm}^{-1}$  del lattato e il complesso a  $1160\text{ cm}^{-1}$  dell'urea possono interferire col principale picco a  $1120\text{ cm}^{-1}$  del glucosio. Tali interferenze sono state ampiamente studiate [30], sottolineando che una variazione significativa nella predizione della concentrazione del glucosio in soluzioni con tutti e tre gli analiti viene rilevata per concentrazioni di lattato superiori a  $70\text{ mM}$  e concentrazioni di urea superiori a  $60\text{ mM}$ . In entrambi i casi, i valori limite per la zona di interferenza sono superiori ai livelli medi di concentrazione di questi analiti nel sudore umano. Nonostante ciò, si deve considerare che in campioni reali, la concentrazione del glucosio è notevolmente inferiore rispetto a quelle di lattato e urea, il che significa che nello spettro otterremo un picco molto

basso, sovrapposto a picchi di intensità non indifferente dovuti a lattato e urea. Questo farebbe sì che il contributo del glucosio nell'intensità del picco risultante sia trascurabile rispetto agli interferenti, rendendo il metodo meno efficace. Sfruttando l'analisi multivariata, questo effetto viene diminuito considerando una più ampia zona del grafico e tenendo quindi anche conto dei picchi secondari che non presentano interferenze così forti come quelle dei picchi principali. In seguito a queste osservazioni, sono stati scelti degli intervalli di numeri d'onda specifici per ciascun analita, in maniera da minimizzare le interferenze di picchi minori in zone periferiche dello spettro ed escludere picchi relativi ad altre molecole presenti (in questo caso solo l'acqua delle soluzioni, ma in campioni reali si avrebbero molti altri componenti metabolici). In particolare, nell'analisi del glucosio si è scelto l'intervallo fra 950 e 1450  $\text{cm}^{-1}$ , per il lattato si è optato per l'intervallo tra 750 e 1150  $\text{cm}^{-1}$  mentre per l'urea l'intervallo 875-1500  $\text{cm}^{-1}$ . In generale, il limite a 1500  $\text{cm}^{-1}$  è stato imposto per evitare di comprendere il picco dell'acqua a 1600  $\text{cm}^{-1}$  circa, eliminando anche eventuali *spike* in questa zona.

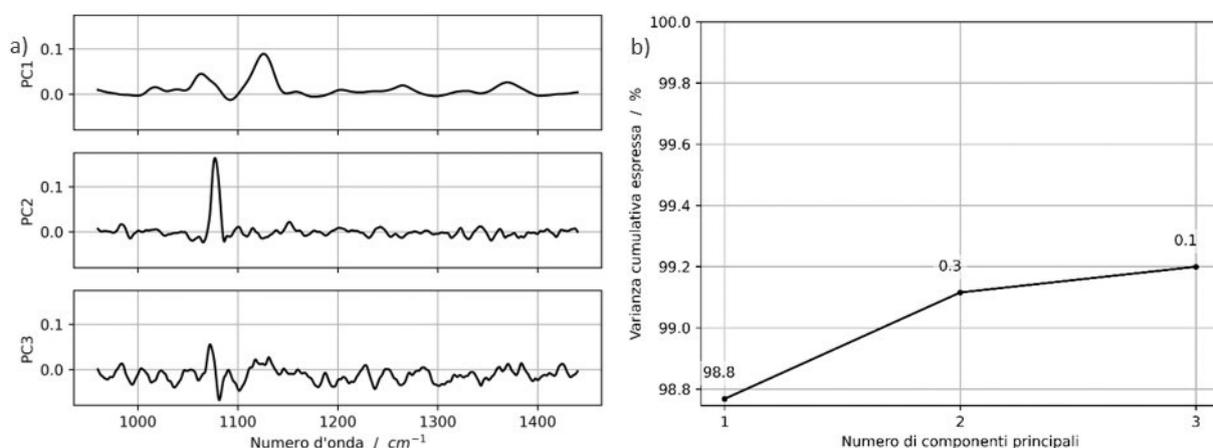
Di seguito si presentano i risultati ottenuti dalla costruzione e applicazione del modello, discutendo il loro significato ed interpretando gli aspetti inattesi. Sono state preventivamente eliminate le prove con evidenti errori, lasciando solo quelle in cui la inclusione ha richiesto ragionamenti ulteriori.

#### 4.1. Misure Raman in acqua distillata

Le prime analisi, svolte su set di soluzioni dei tre analiti presi singolarmente in acqua deionizzata sono servite alla creazione di un valido set di riferimento per le misure successive. Gli spettri acquisiti per i tre set distinti sono stati forniti in input alla PCA, designandoli sia come modello di riferimento che come spettri da analizzare. Tale procedura è necessaria essendo il primo set costruito. Gli autovettori costruiti vengono in seguito forniti alla PCR per creare la regressione lineare del modello.

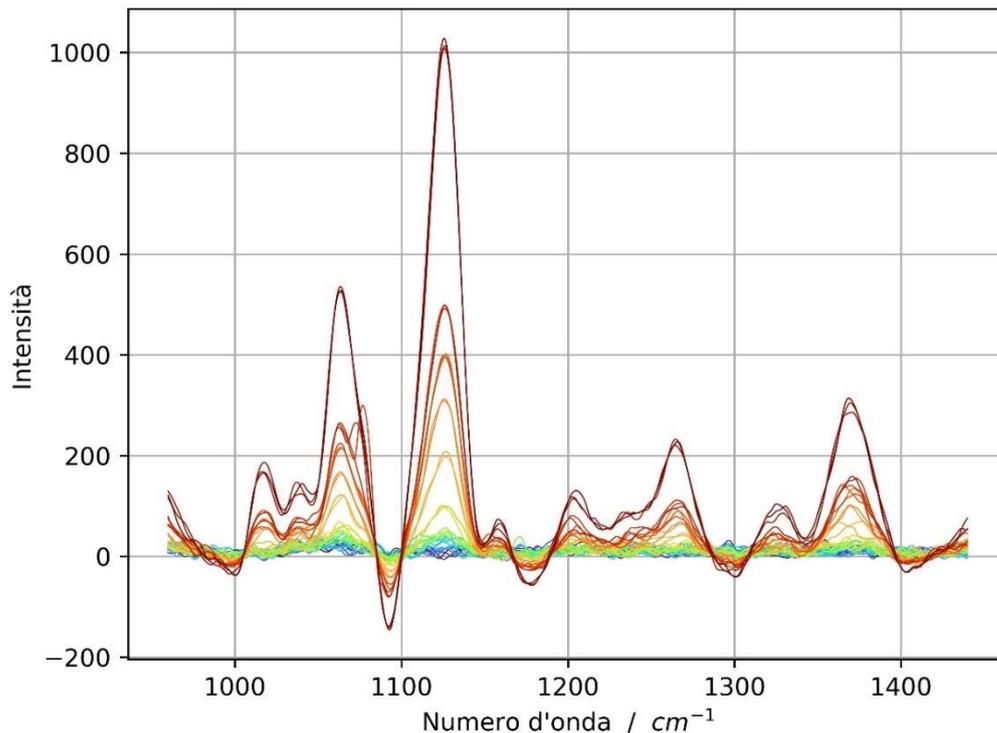
##### 4.1.1. Glucosio

Dapprima sono stati visualizzati la forma degli autovettori utilizzati, la loro influenza sul modello (Figura 4.2) e i grafici ottenuti dalle misure (Figura 4.3).



**Figura 4.2:** Componenti principali del modello (a) e loro varianza espressa (b) nell'analisi di campioni di glucosio in acqua a concentrazione variabile.

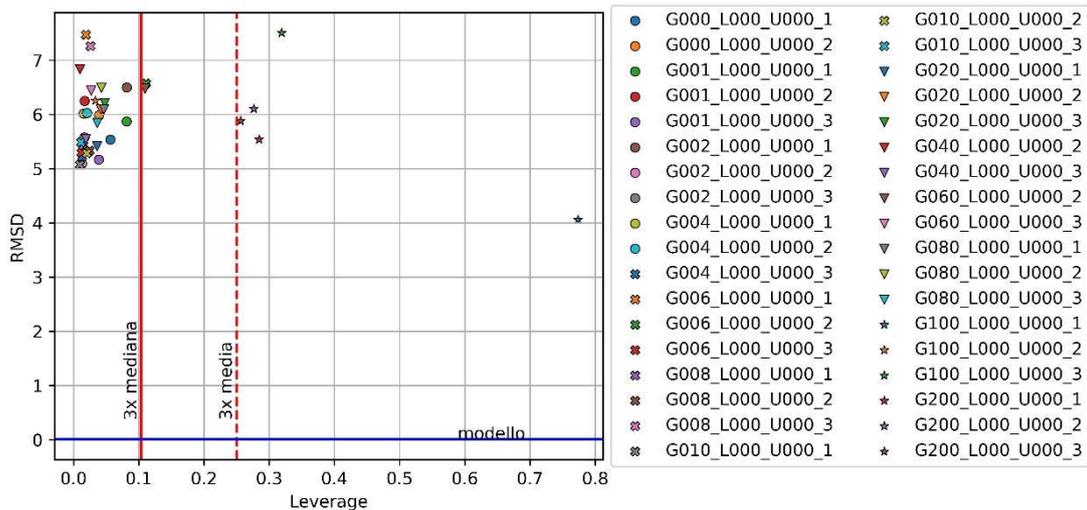
Sono stati selezionati tre componenti principali, visto che già il terzo ha un apporto alla varianza totale dello 0,1%, rispetto al 98,8% del primo. Poiché il grafico è riferito al solo glucosio in acqua, lo spettro risulta essere piuttosto semplice, per cui anche due soli componenti principali risultano sufficienti ad ottenere una buona stima della varianza del set base. In figura 4.2a si nota infatti che il terzo componente principale rappresenta principalmente rumore di fondo della misura, motivo del basso peso che avrà nel modello.



**Figura 4.3:** Spettri Raman del glucosio in soluzioni acquose all'aumentare della concentrazione. La scala cromatica va dal viola, corrispondente al campione 0 mM, fino al rosso scuro, rappresentante i campioni 200 mM.

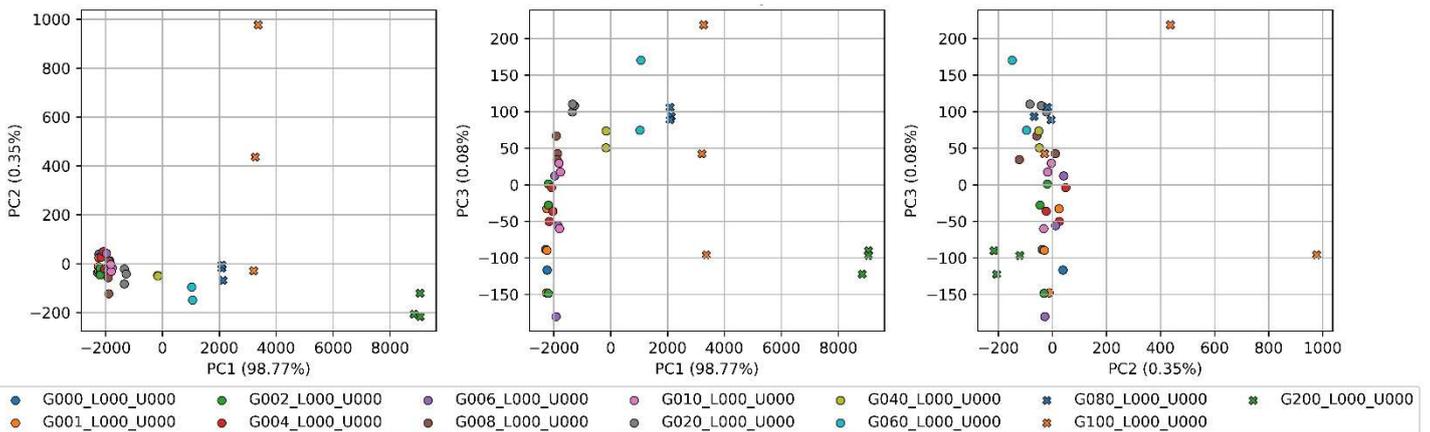
Il campo di numeri d'onda è stato limitato al campo da 950 a 1450  $\text{cm}^{-1}$  sia per evitare il picco dell'acqua a 1650  $\text{cm}^{-1}$  sia per eliminare picchi secondari molto poco sensibili alla variazione di concentrazione. La visualizzazione dei grafici è utile per identificare velocemente misure anomale. Ad esempio, uno degli spettri presenta un picco anomalo a circa 1075  $\text{cm}^{-1}$ , probabilmente dato da uno *spike* sovrapposto al grafico principale, ma questo non porta un errore evidente sul risultato del modello, in quanto provando ad eliminarlo i parametri di linearità rimanevano immutati.

Per quanto riguarda l'analisi degli *outliers*, e dunque la significatività di ciascuna misura, vengono riportati il diagramma *leverage*/deviazione quadratica media (Figura 4.4) e i biplot ottenuti dall'accoppiamento di due dei tre componenti principali, ottenendo dunque tre combinazioni possibili (Figura 4.5).



**Figura 4.4:** Determinazione degli *outliers* per il set di glucosio a concentrazione variabile in acqua deionizzata.

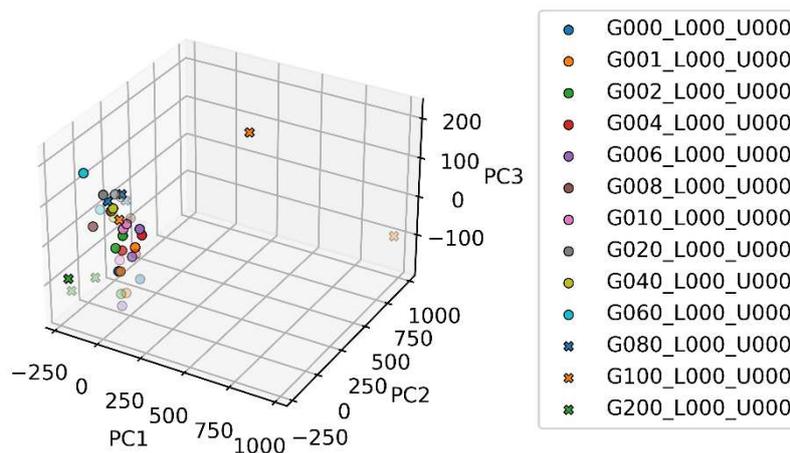
Fra tutte le misure effettuate, sono state eliminatele quattro ripetizioni, in particolare per i campioni a concentrazione 0, 4, 40 e 60 mM, in quanto presentavano valori di *leverage* molto maggiore dei campioni rimasti. Nel set finale di misure, si evidenzia solo la prima ripetizione del campione di concentrazione 100 mM, la quale ha un valore di *leverage* più alto rispetto alle altre misure, probabilmente dato dall'anomalia vista dai grafici di figura 4.3. Per quanto riguarda i campioni immediatamente a destra della linea soglia tratteggiata di tre volte il valore medio, non sono da scartare. Questo avviene poiché sono campioni concentrati, molto differenti dai campioni più diluiti. Si consideri che il campione 200 mM è isolato rispetto alle altre prove, per cui è normale che le sue ripetizioni vengano rilevate come significativamente diverse rispetto alle altre.



**Figura 4.5:** Diagrammi biplot del set di glucosio a concentrazione variabile in acqua deionizzata.

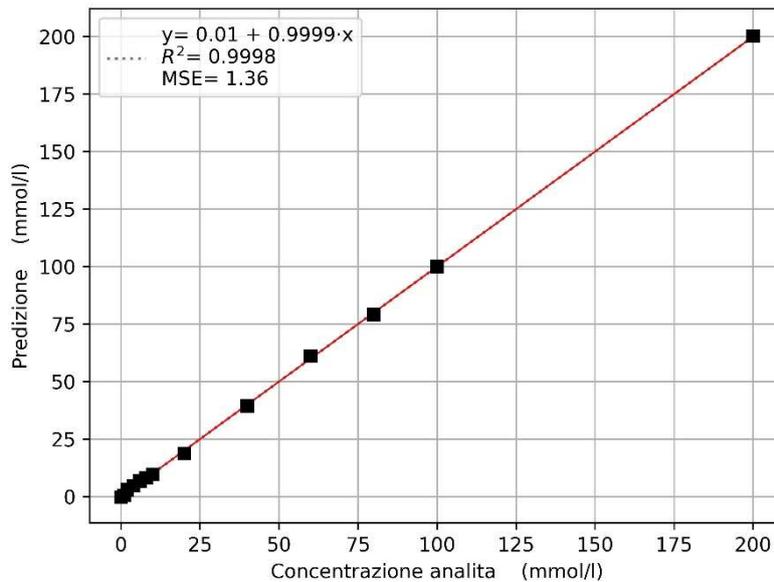
Dai diagrammi biplot si nota che la corrispondenza fra prove simili è piuttosto buona, visto che nel primo grafico i simboli dello stesso colore, e quindi rappresentanti

più ripetizioni della stessa concentrazione, sono molto vicini fra loro. L'unico campione che devia da questo comportamento è quello 100 mM, che presenta infatti tre ripetizioni piuttosto distanziate l'una dall'altra. Possiamo dunque dire che vi sia una deviazione alta se comparata rispetto alle altre soluzioni, ma l'errore assoluto che deriva da questo non è così influente. Infatti, come riportato sull'asse delle ordinate, la componente PC2 su cui è evidente questo distacco influisce sul modello solo per lo 0,35% del peso totale. Questo si nota in quanto le tre croci arancioni sono praticamente sulla stessa verticale, quindi poco distaccati rispetto a PC1, mentre si trovano a tre altezze diverse, leggibili sull'asse di PC2. Nonostante ciò, prendendo come riferimento PC1, si può dire che le soluzioni con concentrazione diversa appaiono nel corretto ordine crescente, e sono quindi statisticamente diverse l'una dall'altra. Il confronto tra PC1 e PC3 porta a risultati analoghi. Il diagramma tra PC2 e PC3 presenta invece una distribuzione meno lineare, con cluster non evidenti fra campioni simili. Va però considerato che questo è costruito su due componenti che rappresentano lo 0,43% della varianza totale, e dunque non sono significativi per il modello. I tre diagrammi biplot possono anche essere riassunti in un unico grafico tridimensionale che comprende tutti e tre i componenti principali (Figura 4.6).



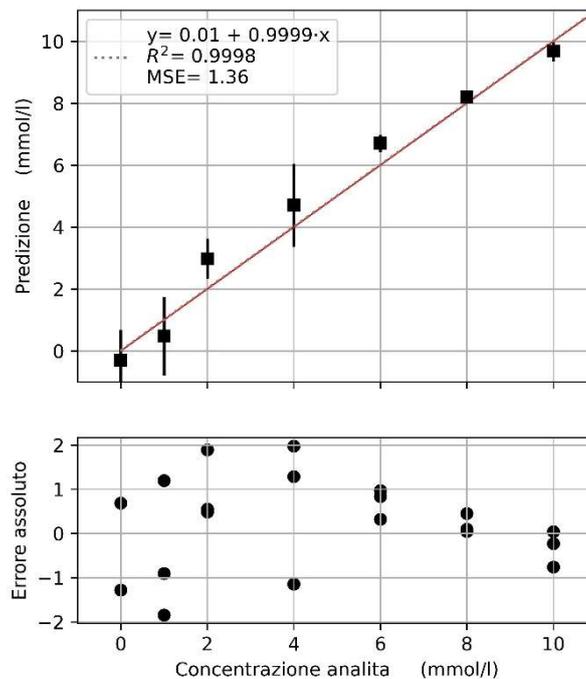
**Figura 4.6:** Rappresentazione tridimensionale dei diagrammi di *clustering* del glucosio in acqua.

Per la costruzione del modello di regressione sono quindi stati utilizzati solo i primi due autovettori. La regressione è stata ottenuta fra le medie delle predizioni di ciascun campione, ovvero si ottiene dapprima la concentrazione predetta per ciascuna misura, si calcola la media fra i campioni a pari concentrazione e poi si esegue la regressione fra gli elementi rimasti, ovvero uno spettro per ogni concentrazione (Figura 4.7).



**Figura 4.7:** PCR effettuata su soluzioni di glucosio in acqua a concentrazione crescente.

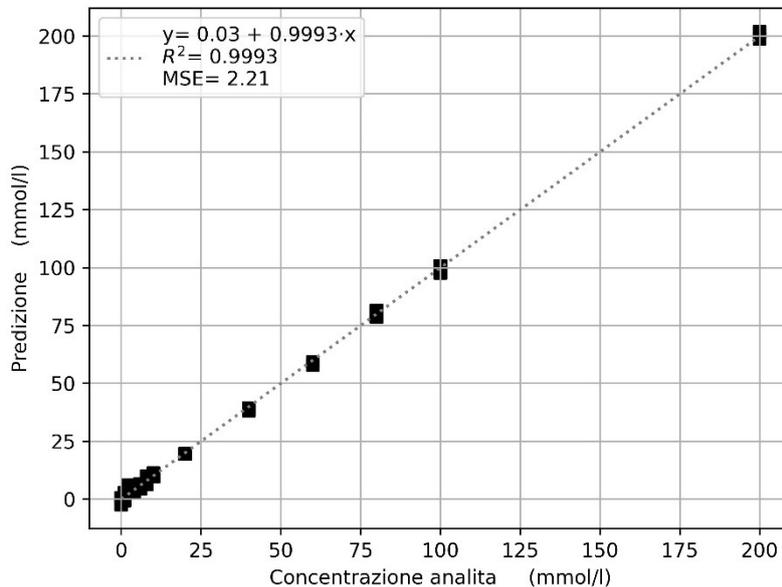
È stata raggiunta una buona linearità della curva di predizione come mostrato dal valore di  $R^2$ , molto vicino a 1. Infatti, anche per alti valori di concentrazione la predizione rimane lineare, senza raggiungere valori limite di stazionarietà. Il valore dell'intercetta, praticamente nullo, indica che il modello non necessita di fattori correttivi nel calcolo della predizione finale. Inoltre, il valore della pendenza della retta, tendente all'unità, indica una buona sensibilità nel riconoscimento della concentrazione, indice del fatto che soluzioni con concentrazioni anche lievemente differenti tra loro verrebbero ben distinte dal modello. Ingrandendo la zona delle soluzioni più diluite, si può analizzare meglio il campo delle soluzioni più utili a livello fisiologico (Figura 4.8).



**Figura 4.8:** Ingrandimento della PCR sul set di glucosio in acqua ed errore assoluto.

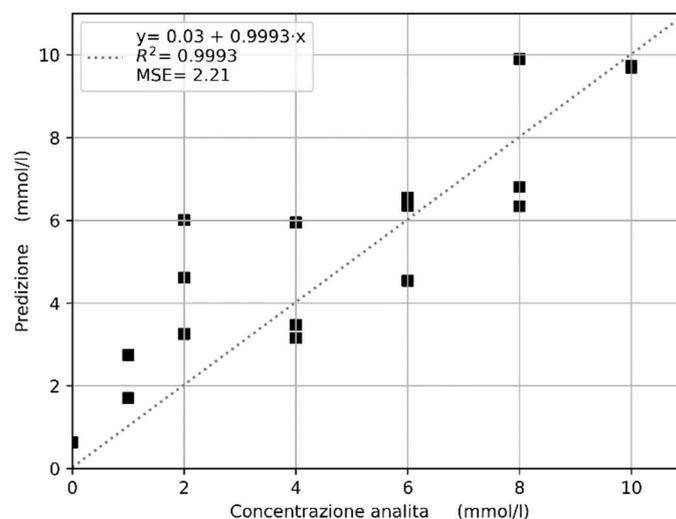
Anche in questo intervallo il modello è piuttosto preciso e si visualizza un errore massimo di 2 mM su singole prove, molto meno considerato sulla media delle ripetizioni simili.

Per quanto riguarda la PLSR sono stati ottenuti buoni risultati (Figura 4.9).



**Figura 4.9:** PLSR effettuata sul set di glucosio in acqua a concentrazione crescente.

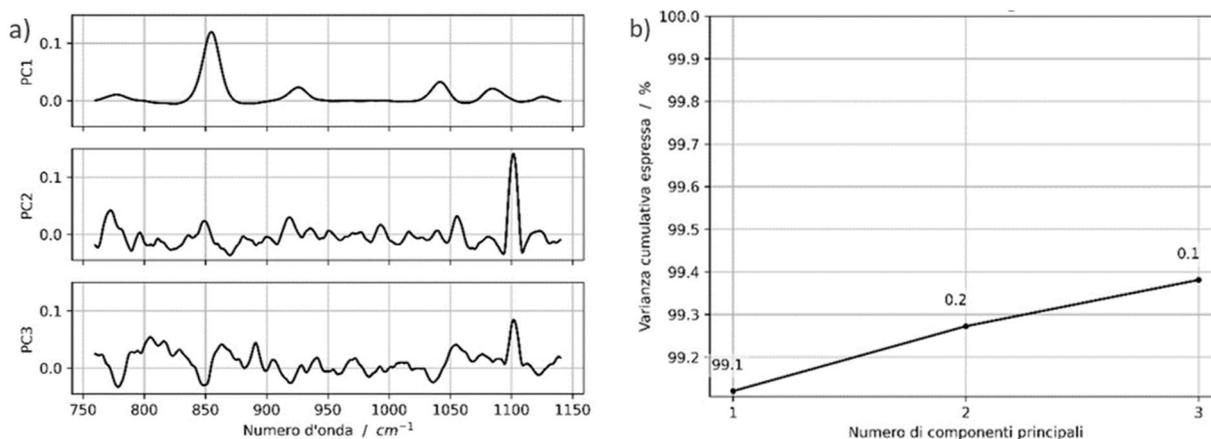
Anche per la PLSR sono stati ottenuti buoni valori di linearità, sia per  $R^2$  che per pendenza ed intercetta della retta. I valori sono leggermente inferiori rispetto alla PCR, ma comunque tendono ai limiti ottimali teorici con scarti minimi. Ingrandendo la zona di concentrazioni inferiori a 10 mM (Figura 4.10), si notano discostamenti maggiori rispetto alla PCR, con errori sui campioni più diluiti anche fino a 4 mM.



**Figura 4.10:** Ingrandimento PLSR sul set di glucosio in acqua nella zona 0-10 mM.

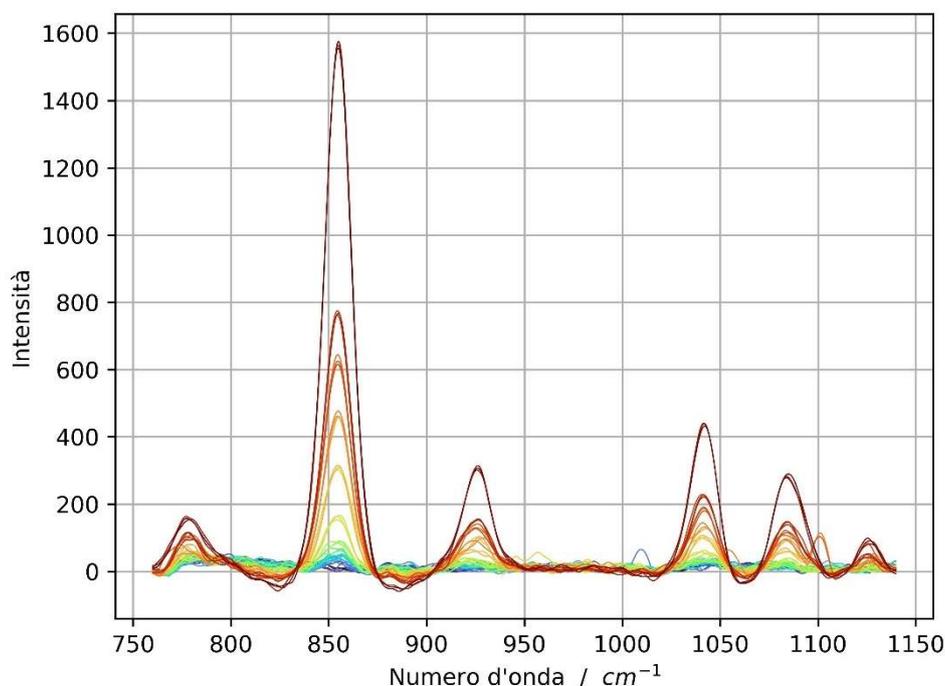
#### 4.1.2. Lattato

Anche nella costruzione della curva di calibrazione del lattato sono stati dapprima visualizzati i componenti principali e la loro varianza espressa (Figura 4.11).



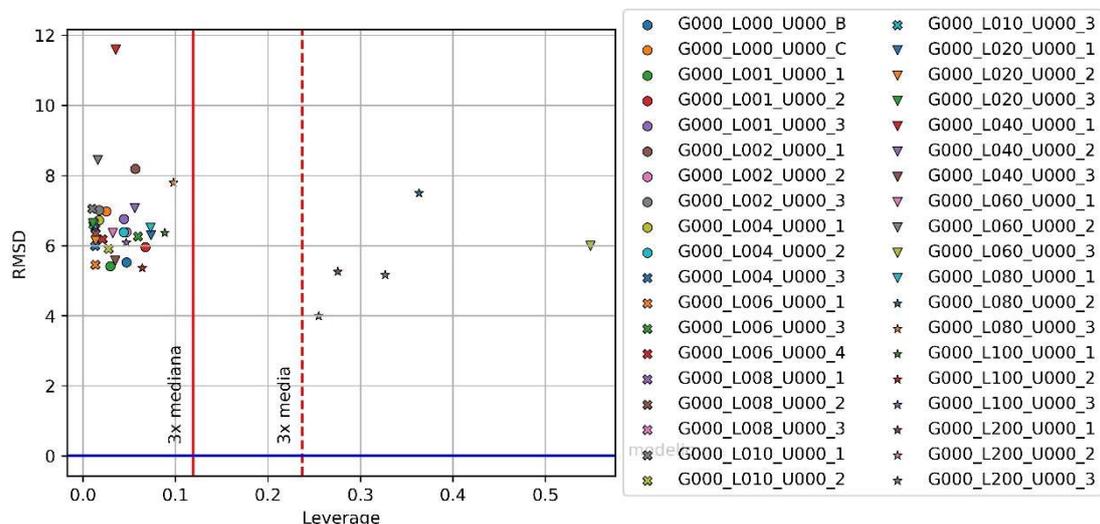
**Figura 4.11:** Componenti principali del modello (a) e loro varianza espressa (b) nell'analisi di campioni di lattato in acqua a concentrazione variabile.

Si nota che il primo componente rappresenta una percentuale di varianza superiore anche al caso del glucosio. Dall'analisi della forma dei componenti principali si nota però che il componente PC2 porta l'informazione principale della varianza del picco a 1100  $cm^{-1}$  e va dunque considerato. Lo spettro rimane comunque piuttosto semplice da descrivere con poche componenti. Vengono tracciati i grafici delle prove (Figura 4.12).



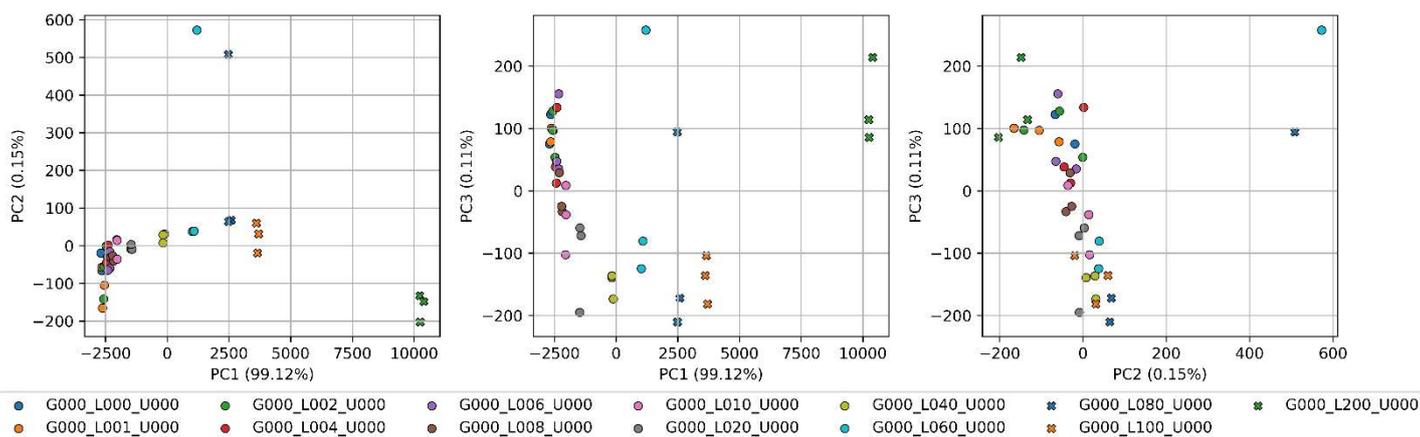
**Figura 4.12:** Spettri Raman del lattato in soluzioni acquose all'aumentare della concentrazione. La scala cromatica va dal viola, corrispondente al campione 0 mM, fino al rosso scuro, rappresentante i campioni 200 mM.

Dai diagrammi si sono notati un picco anomalo a circa  $1010\text{ cm}^{-1}$  e due a  $1100\text{ cm}^{-1}$  circa, leggermente sovrapposti al picco noto a  $1080\text{ cm}^{-1}$ . In particolare, i secondi sono anche visibili nel grafico degli *outliers* (Figura 4.13).

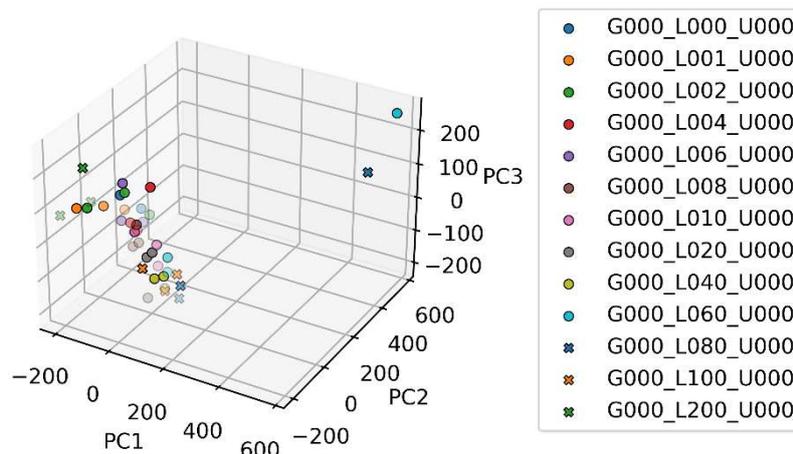


**Figura 4.13:** Determinazione degli *outliers* per il set di lattato a concentrazione variabile in acqua deionizzata.

Si nota infatti che la terza ripetizione della prova a concentrazione 60 mM rappresentata da un triangolo verde chiaro risulta avere il valore di *leverage* più alto del set. Inoltre, anche la seconda ripetizione della prova a concentrazione 80 mM ha un valore di *leverage* piuttosto alto, sebbene non quanto la prima. Ulteriore conferma è fornita dai diagrammi di *clustering biplot* (Figura 4.14) e tridimensionale (Figura 4.15).



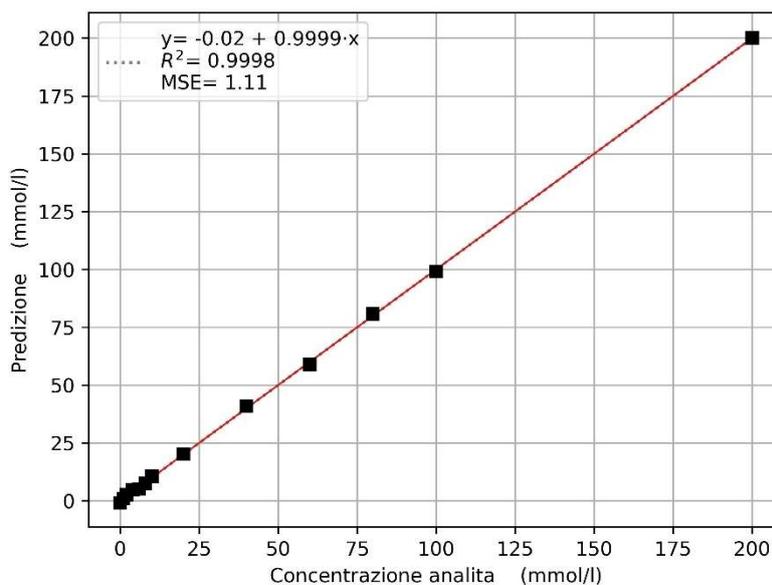
**Figura 4.14:** Diagrammi biplot del set di lattato a concentrazione variabile in acqua deionizzata.



**Figura 4.15:** Rappresentazione tridimensionale dei diagrammi di *clustering* del lattato in acqua.

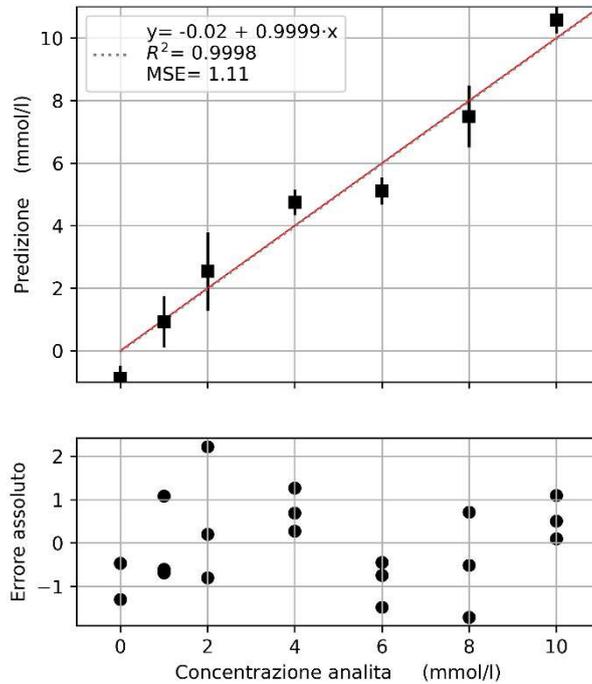
Nel grafico di confronto tra PC1 e PC2 si nota infatti un buon *clustering* generale delle prove simili, molto ravvicinate tra di loro ed ordinate coerentemente alla scala di concentrazioni. Fanno eccezione solo due ripetizioni, a 60 mM e 80 mM, le quali si distaccano dalle simili rispetto a PC2. Questo risultato è ancora più evidente nel diagramma tridimensionale. Nonostante tutto ciò, è stato deciso di mantenere le due ripetizioni, in quanto la discriminazione eseguita da PC2 ha peso pari allo 0,15% della varianza totale, quindi molto poco. Inoltre, anche provando a rimuovere tali prove, i risultati sulla PCR e PLSR subiscono variazioni dell'ordine della quarta cifra decimale, peggiorando addirittura il modello, visto che vi è un riferimento in meno.

Si riporta il risultato della PCR eseguita con due sole componenti principali come base (Figura 4.16).



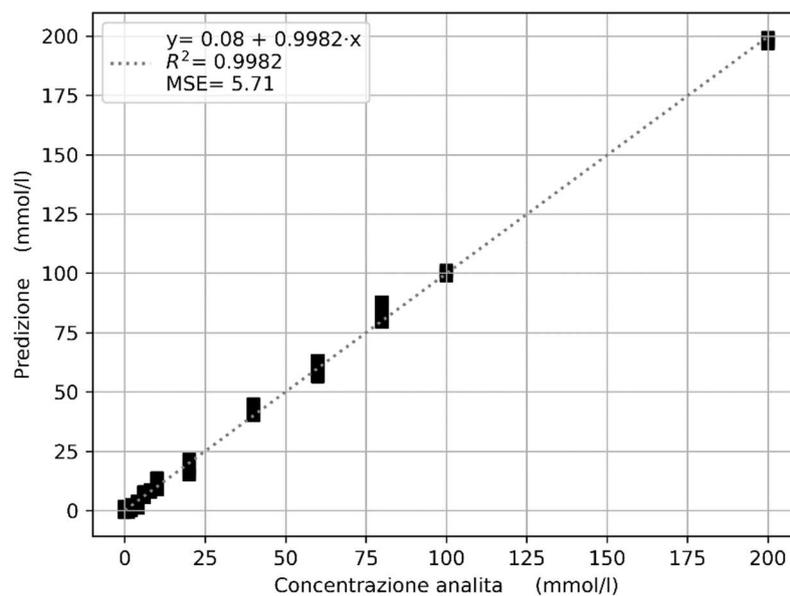
**Figura 4.16:** PCR effettuata su soluzioni di lattato in acqua a concentrazione crescente.

Anche in questo caso sono stati ottenuti buoni riscontri di linearità, simili al caso del glucosio e con uno scarto medio quadratico addirittura inferiore. Ingrandendo la zona delle concentrazioni inferiori a 10 mM, si evidenziano errori assoluti massimi di 2 mM solo su determinate ripetizioni, e non sulla media della misura (Figura 4.17).



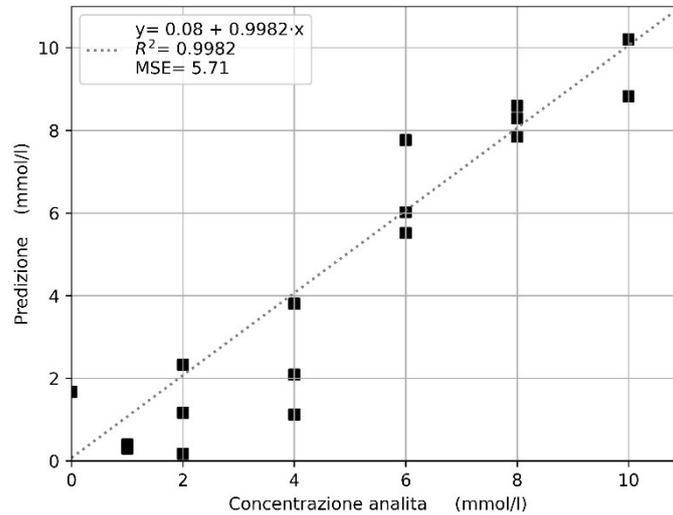
**Figura 4.17:** Ingrandimento della PCR sul set di lattato in acqua ed errore assoluto.

Analogamente si riportano i risultati della regressione eseguita con PLSR (Figura 4.18).



**Figura 4.18:** PLSR effettuata sul set di lattato in acqua a concentrazione crescente.

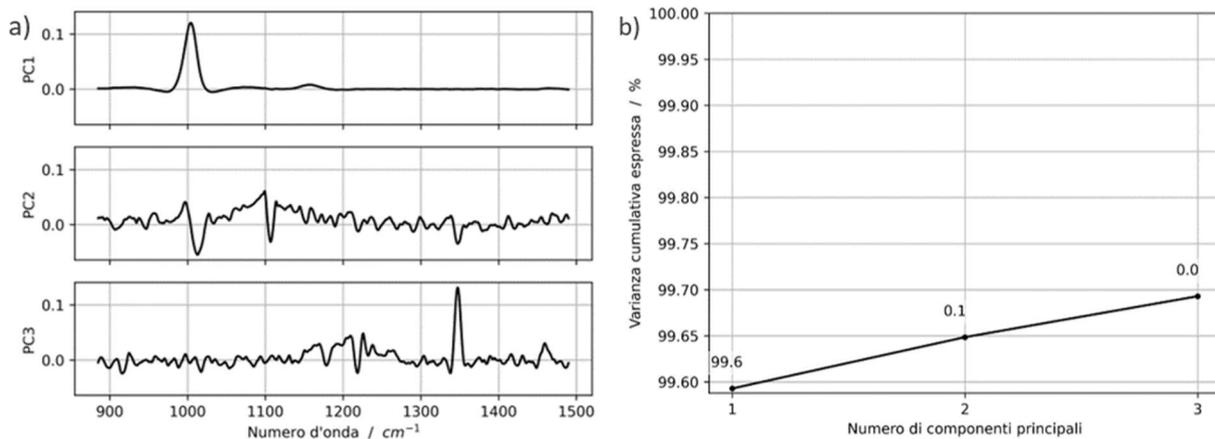
Per la PLSR si ottengono ancora buoni valori di linearità, anche se con uno scarto quadratico medio maggiore rispetto al caso del glucosio, più del doppio rispetto a quel caso. Questo effetto è visibile anche nell'ingrandimento della zona tra 0 e 10 mM (Figura 4.19) in cui si nota una perdita di linearità più marcata soprattutto per concentrazioni minori di 4 mM.



**Figura 4.19:** Ingrandimento PLSR sul set di lattato in acqua nella zona 0-10 mM.

#### 4.1.3. Urea

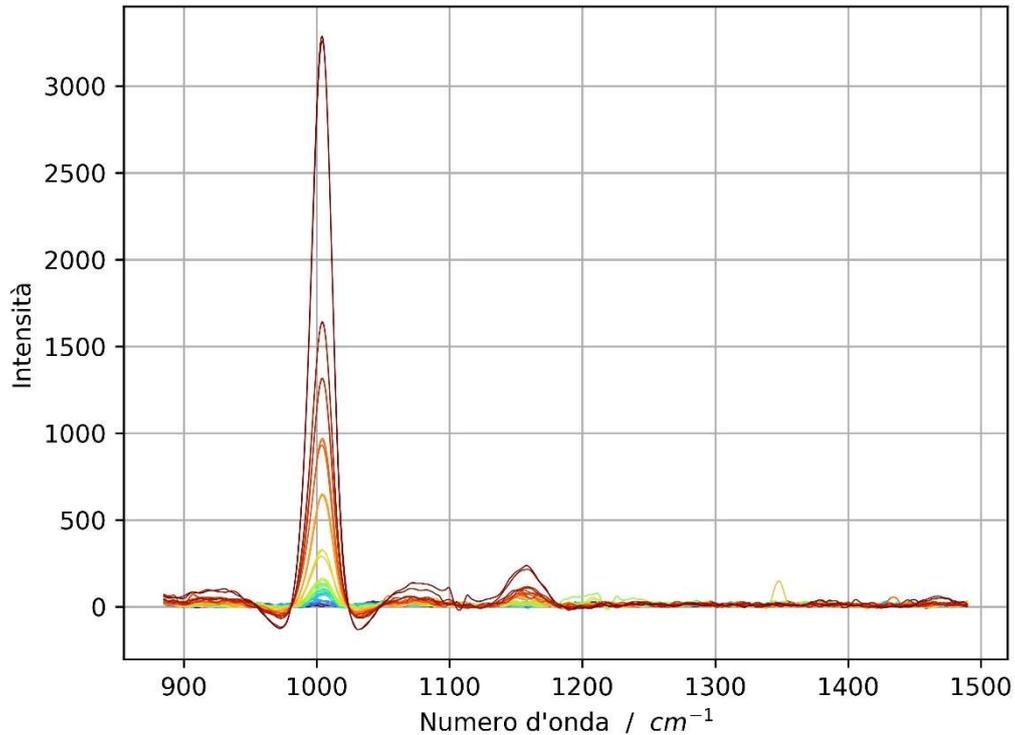
Come per i due casi precedenti, si visualizzano per primi gli autovettori ottenuti dalla PCA eseguita sul set e la quantità di varianza totale che ciascuno di essi esprime (Figura 4.20).



**Figura 4.20:** Componenti principali del modello (a) e loro varianza espressa (b) nell'analisi di campioni di urea in acqua a concentrazione variabile.

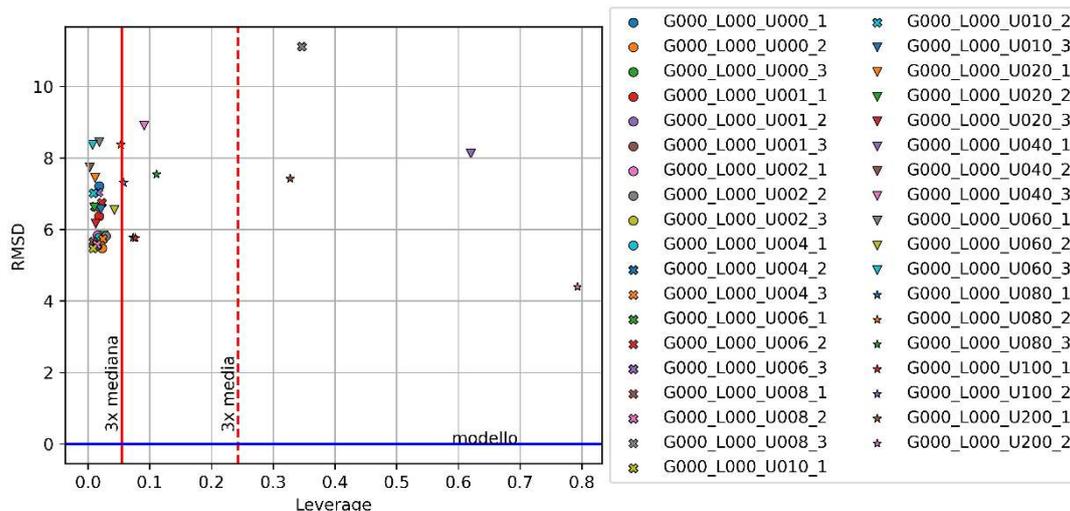
Anche per quest'analisi il 99,6% della varianza totale degli spettri è esprimibile con un solo autovettore, sebbene il secondo componente identifichi il secondo picco tipico dell'urea a circa 1200 cm<sup>-1</sup>. Questo secondo picco è però molto meno sensibile alla

variazione di concentrazione fra i vari campioni e dunque possiede una varianza trascurabile rispetto al picco principale. Si è scelto dunque di utilizzare solo i primi due componenti principali per la regressione. Gli spettri ottenuti per la costruzione del modello evidenziano un solo picco anomalo a circa  $1350\text{ cm}^{-1}$  (Figura 4.21).



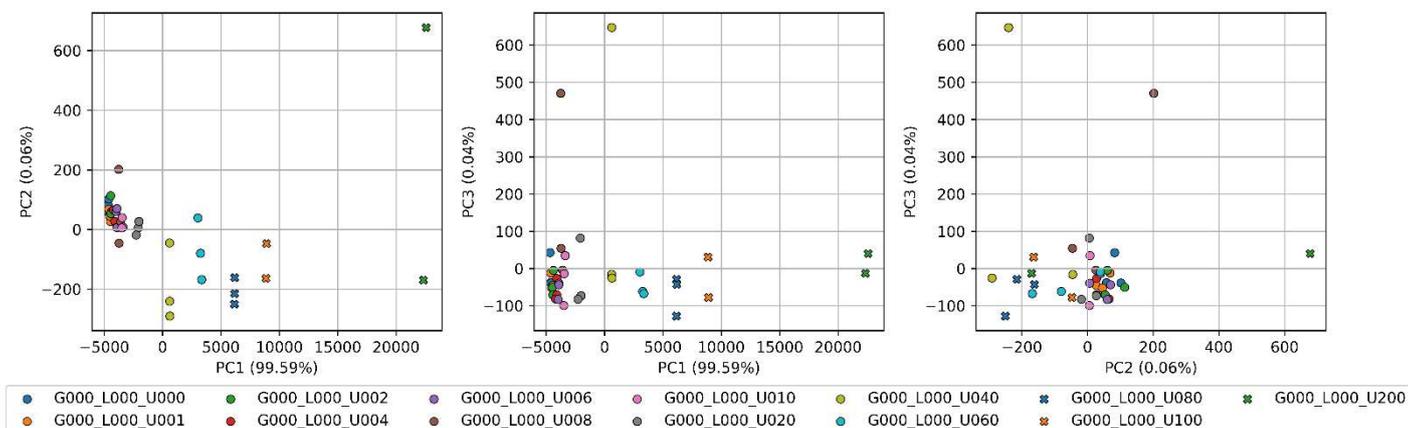
**Figura 4.21:** Spettri Raman dell'urea in soluzioni acquose all'aumentare della concentrazione. La scala cromatica va dal viola, corrispondente al campione 0 mM, fino al rosso scuro, rappresentante i campioni 200 mM.

Provando a limitare la PCA a circa  $1250\text{ cm}^{-1}$ , eliminando così il picco anomalo, non porta miglioramenti significativi, visto che non è sufficientemente intenso per interferire con la misura. Inoltre, tale operazione ridurrebbe la robustezza del modello, in quanto lasciare questo range di numeri d'onda fa sì che questo tenga conto che l'urea non ha segnali degni di nota nella zona tra  $1250$  e  $1500\text{ cm}^{-1}$ . Lo studio degli *outliers* (Figura 4.22) e dei diagrammi di *clustering* bidimensionali (Figura 4.23) e tridimensionale (Figura 4.24) ha confermato la scelta fatta, evidenziando solo due ripetizioni devianti dal comportamento medio, in particolare la prima ripetizione del campione 40 mM e la terza ripetizione del campione 8 mM.

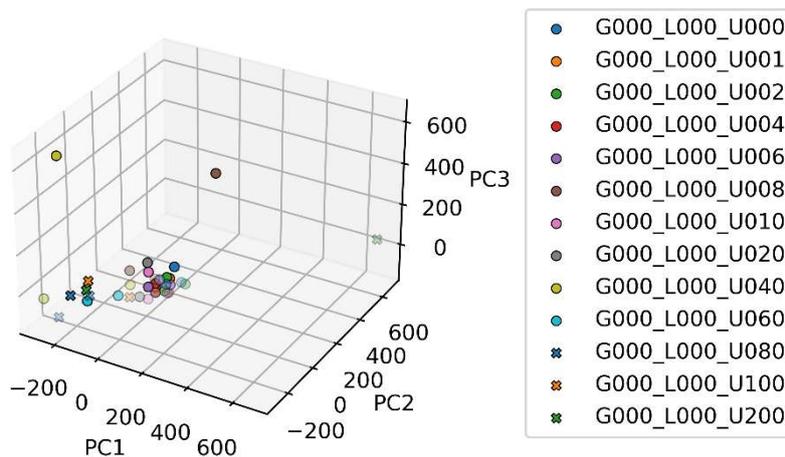


**Figura 4.22:** Determinazione degli *outliers* per il set di urea a concentrazione variabile in acqua deionizzata.

In questo caso la maggior parte dei campioni presenta valori di *leverage* minori di 0,1. Questo fa sì che media e mediana siano molto basse, e di conseguenza anche le soglie da essi dipendenti. Per questo, campioni con valori di *leverage* generalmente basso si trovano ad essere facilmente fuori dai campi di accettabilità, come la terza ripetizione della soluzione 8 mM. Questo è un ulteriore ragionamento che ha portato a mantenere le due prove che erano state identificate come possibili *outliers*.



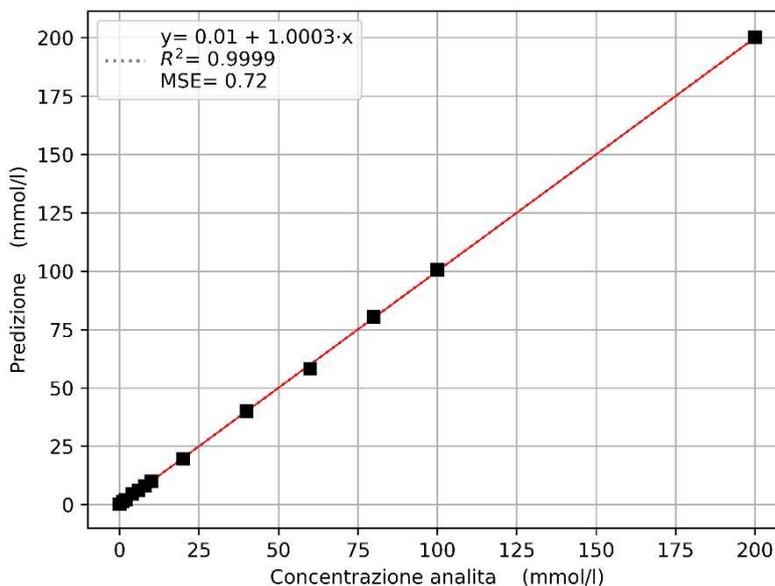
**Figura 4.23:** Diagrammi biplot del set di urea a concentrazione variabile in acqua deionizzata.



**Figura 4.24:** Rappresentazione tridimensionale dei diagrammi di *clustering* dell'urea in acqua.

Dai grafici di *clustering* si ha un'ulteriore conferma per il trattamento di questi campioni, in quanto guardando i valori rispetto alla PC1 si nota che i campioni potenzialmente anomali non si distaccano dai loro simili. Vengono discriminati solo dal grafico PC2/PC3, che però rappresenta solo lo 0,1% della varianza totale (sommando le due componenti). Il campione è stato dunque mantenuto nel set.

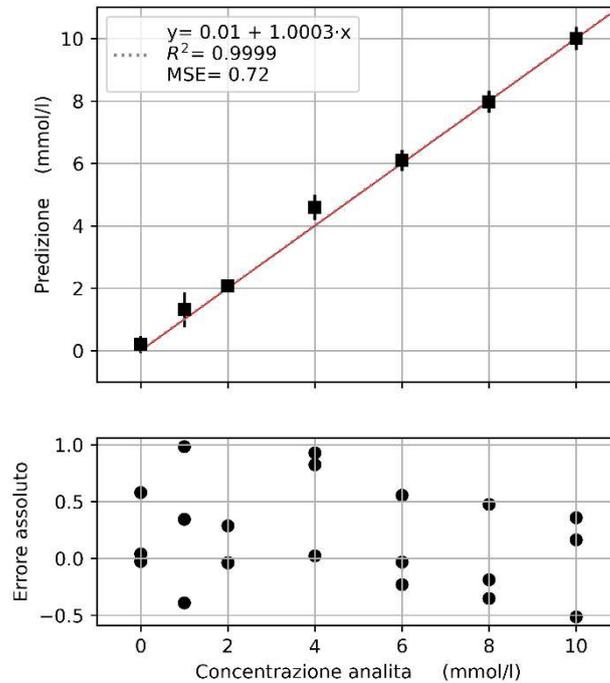
Per la regressione lineare sono dunque state scelte due componenti principali, come nei casi precedenti (Figura 4.25).



**Figura 4.25:** PCR effettuata su soluzioni di urea in acqua a concentrazione crescente.

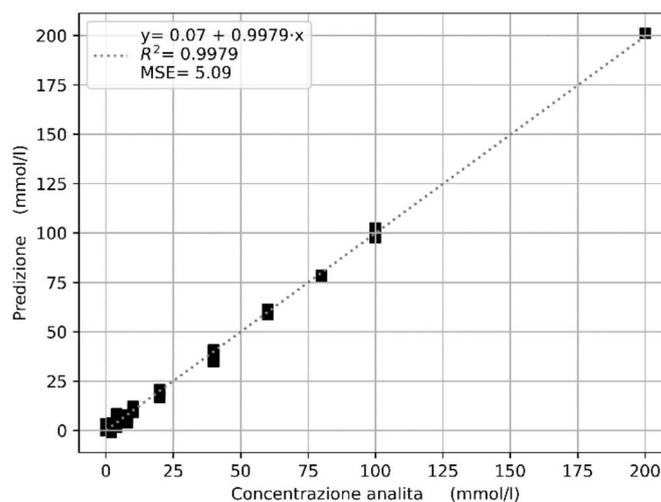
È stata ottenuta una buona linearità, migliore dei due casi precedenti, vista la semplicità degli spettri formati praticamente da un solo picco principale. Anche lo

scarto quadratico medio è sensibilmente inferiore a quello di glucosio o lattato, segno di buona precisione della calibrazione. Ingrandendo la zona sotto 10 mM si notano errori assoluti dimezzati rispetto agli altri due analiti, arrivando al massimo a 1 mM di differenza sulla singola misura (Figura 4.26).



**Figura 4.26:** Ingrandimento della PCR sul set di urea in acqua ed errore assoluto.

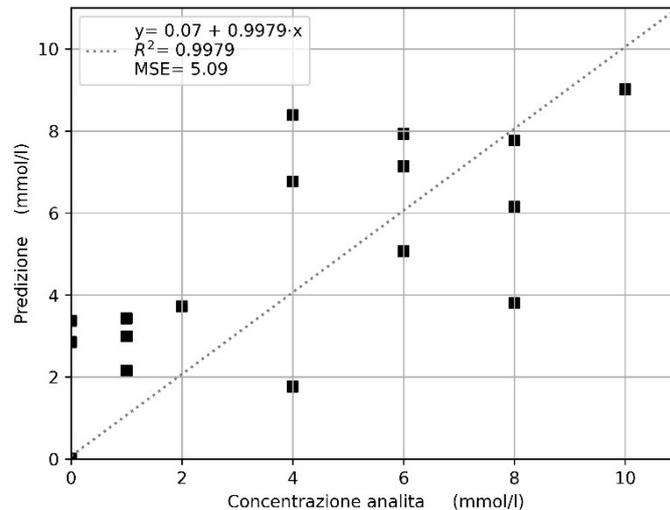
Con la regressione mediante PLSR (Figura 4.27) è stata ottenuta una buona linearità, con uno scarto medio quadratico non molto elevato e buoni valori di intercetta e pendenza della retta.



**Figura 4.27:** PLSR effettuata sul set di urea in acqua a concentrazione crescente.

Anche in questo caso, i risultati della PLSR sono lievemente peggiori rispetto alla PCR. In più, la differenza tra PCR e PLSR è più marcata rispetto ai casi precedenti, probabilmente dovuta alla maggior semplicità degli spettri dell'urea rispetto a glucosio e lattato.

Piuttosto differente è la zona di concentrazioni minori di 10 mM (Figura 4.28), che, se presa separatamente dal resto del set, non presenta un'ottima linearità, paragonata ai casi precedenti. Si ha perdita di linearità per soluzioni più diluite di 8 mM, campo in cui comunque si è sotto i valori fisiologici di concentrazione per questo analita.



**Figura 4.28:** Ingrandimento PLSR sul set di urea in acqua nella zona 0-10 mM.

#### 4.2. Misure Raman in sudore artificiale

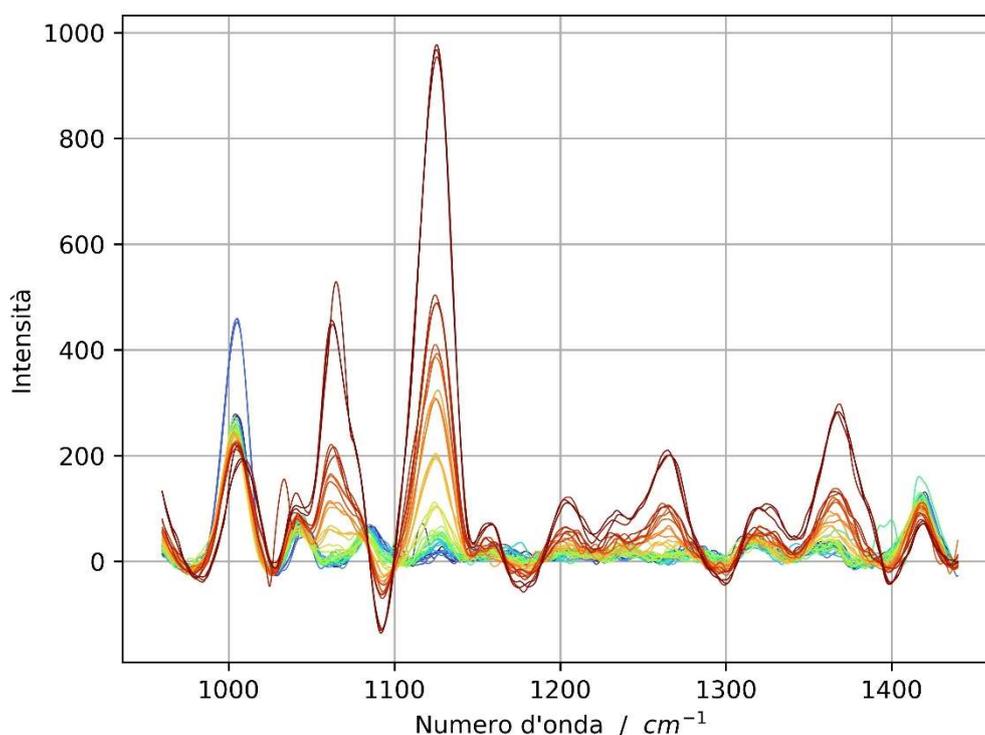
Per queste misure sono stati utilizzati gli spettri degli analiti in acqua, analizzati nella sezione precedente, come riferimento per confrontare gli spettri dei set dei tre analiti in soluzioni contenenti gli altri due in quantità fissa e nota come interferenti, secondo le concentrazioni riportate in tabella 3.1. Gli spettri saranno dunque piuttosto simili a quelli dei singoli analiti, ma saranno presenti anche i picchi caratteristici dei due interferenti ad intensità fissa. Inoltre, per queste analisi non sono più visualizzati gli autovettori e la loro varianza espressa in quanto questi vengono costruiti sul set di riferimento, che è esattamente il set di soluzioni dell'analita in esame in acqua deionizzata. Ciò significa che i diagrammi sono gli stessi delle analisi presentati nella sezione precedente a parità di analita.

Il resto del procedimento di analisi è il medesimo, analizzando dapprima la validità delle misure tramite PCA, producendo anche gli autovettori necessari in seguito, per poi eseguire la regressione dei dati sia con PCR che con PLSR.

#### 4.2.1. Analisi del glucosio con interferenti

In quest'analisi, il glucosio è presente in quantità variabile, esattamente con la stessa distribuzione vista nelle prove precedenti, mentre il lattato e l'urea sono fissi a 40 mM e 20 mM rispettivamente per ciascun campione.

Dapprima sono stati visualizzati i grafici acquisiti, i quali hanno rivelato alcune anomalie (Figura 4.29)

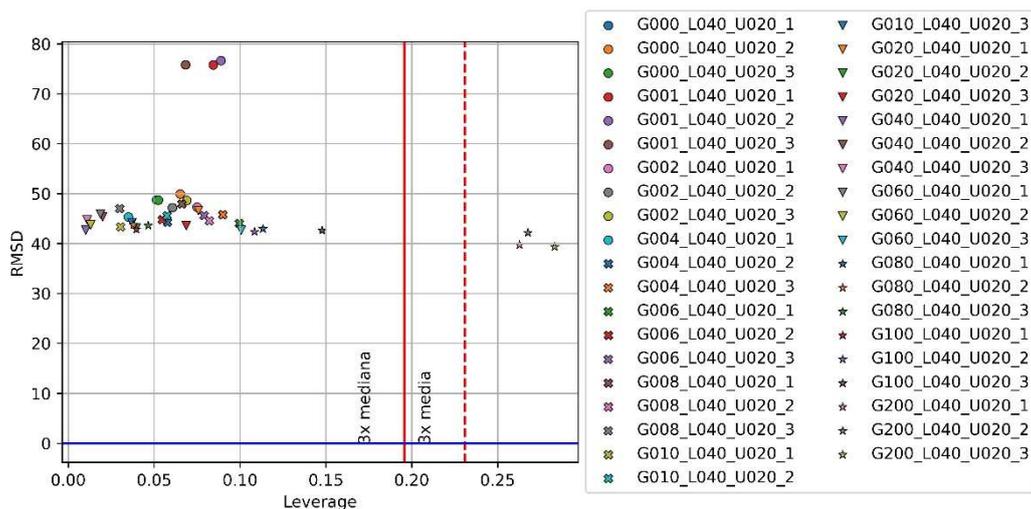


**Figura 4.29:** Spettri Raman del glucosio in soluzioni con interferenti. La scala cromatica va dal viola, corrispondente al campione 0 mM, fino al rosso scuro, rappresentante i campioni 200 mM.

Innanzitutto, sono ben visibili i picchi a  $1010\text{ cm}^{-1}$  e  $1425\text{ cm}^{-1}$  discussi in Figura 4.1 e relativi rispettivamente all'urea ed al lattato i quali, presenti come interferenze, hanno concentrazione costante in tutti i campioni e, quindi, intensità costante. Si è anche rilevato che il picco dell'urea ha intensità molto maggiore rispetto a quello del lattato, nonostante la concentrazione del primo analita sia la metà rispetto al secondo. Questo è dovuto sia al fatto che il picco del lattato non è il principale dello spettro, ma anche al fatto che il metodo sembra avere una maggior sensibilità rispetto all'urea, motivo per cui, a parità di concentrazione, i suoi picchi risultano più intensi. Tale comportamento è visualizzabile paragonando anche gli spettri di un campione 200 mM di urea e uno 200 mM di lattato dalle figure 4.21 e 4.12, per cui il picco del lattato arriva ad un'intensità di 1600 mentre l'urea supera le 3000 unità. Inoltre, analizzando con attenzione la banda tra  $1030\text{ cm}^{-1}$  e  $1100\text{ cm}^{-1}$  si nota che il tipico complesso del glucosio presente in questa zona è visibile solo per concentrazioni maggiori di 20 mM, in quanto i picchi a  $1040\text{ cm}^{-1}$  e  $1080\text{ cm}^{-1}$  circa, relativi al lattato, risultano prevalenti

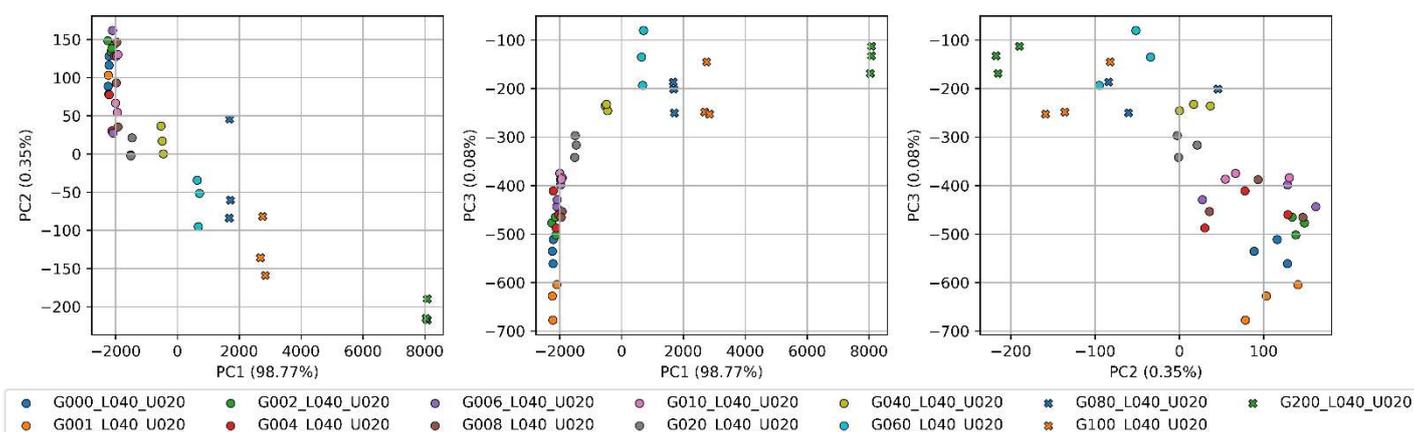
rispetto al glucosio. Il picco a  $1010\text{ cm}^{-1}$  dell'urea influisce solo parzialmente sulla base del picco a  $1050\text{ cm}^{-1}$  del glucosio, creando quindi un'interferenza minima. Lo stesso non si può dire per il picco a  $1080\text{ cm}^{-1}$  del lattato, che risulta invece totalmente sovrapposto al picco a  $1070\text{ cm}^{-1}$  del glucosio. Nella parte finale del picco principale dello spettro, quello a  $1150\text{ cm}^{-1}$  circa, si nota inoltre un picco di bassa intensità dovuto nuovamente al lattato. Questo non ha molta interferenza, in quanto si ha una sovrapposizione solo parziale alla base e, in più, il picco del lattato presenta area trascurabile rispetto a quello del glucosio.

Si nota, infine, che uno dei campioni a bassa concentrazione, comprendente tutte e tre le sue ripetizioni rappresentate da spettri blu, presenta un picco di altezza quasi doppia rispetto agli altri a circa  $1000\text{ cm}^{-1}$ . Questo è la conseguenza di un errore nella preparazione della soluzione, in cui è stata dosata una quantità di urea doppia rispetto al dovuto. Nonostante ciò, la misura è stata mantenuta in quanto non va a influire sul significato statistico della prova, che è quello di valutare le prestazioni del modello in presenza di interferenti. Se lo stesso picco fosse stato in concomitanza di uno del glucosio, allora la misura sarebbe stata fortemente influenzata ed il campione avrebbe dovuto essere sostituito. Tali ragionamenti sono avvalorati anche dall'analisi degli *outliers* (Figura 4.30).

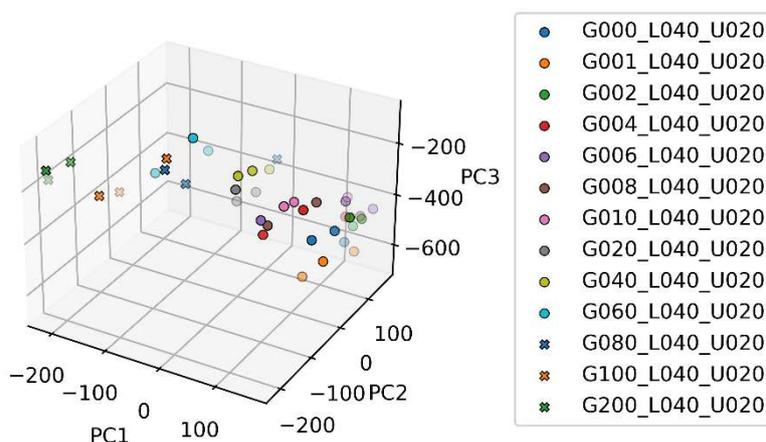


**Figura 4.30:** Determinazione degli *outliers* per il set di glucosio a concentrazione variabile con interferenti.

Non sono presenti forti discostamenti per quanto riguarda il *leverage*, salvo, come previsto, i tre campioni a concentrazione massima. Dal punto di vista del RMSD si hanno invece tre campioni con un valore più alto della media degli altri, pur rimanendo nello stesso ordine di misura. Questi campioni sono stati analizzati anche tramite i diagrammi di *clustering* bidimensionali (Figura 4.31) e tridimensionali (Figura 4.32).

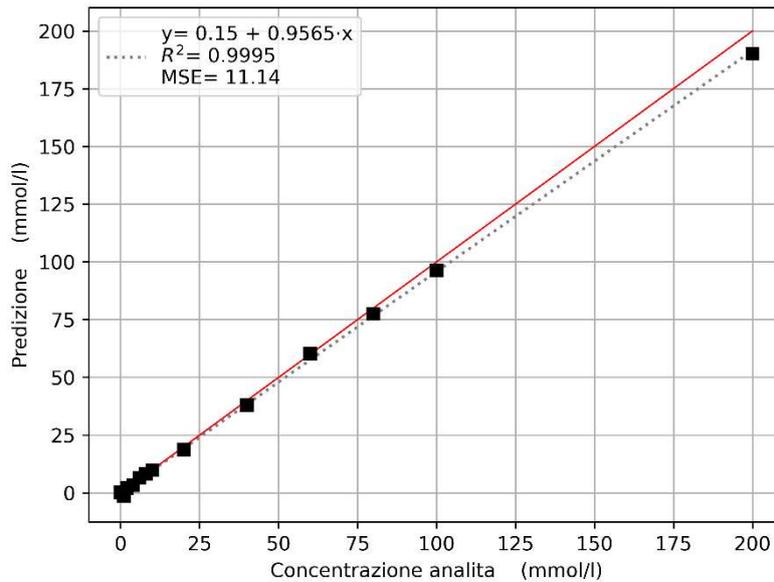


**Figura 4.31:** Diagrammi biplot del set di glucosio a concentrazione variabile in presenza di interferenti.



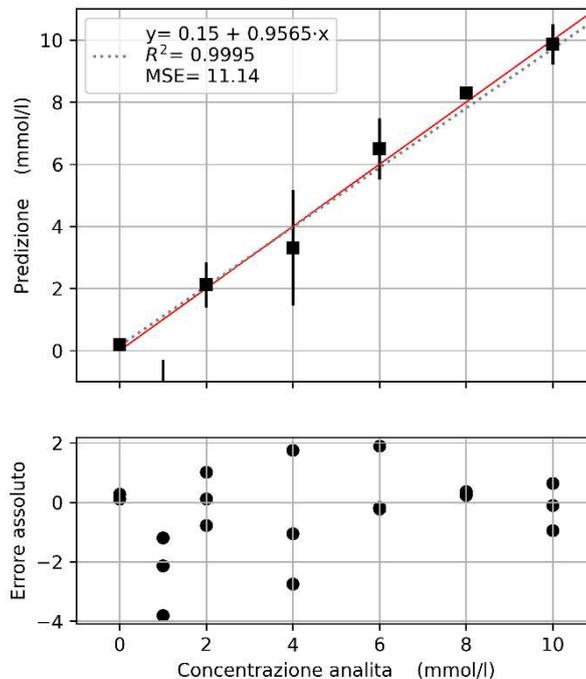
**Figura 4.32:** Rappresentazione tridimensionale dei diagrammi di *clustering* del glucosio in presenza di interferenti.

È stato possibile identificare il campione come quello di concentrazione 1 mM, sia perché corrisponde alle tre ripetizioni con più elevato RMSD dal diagramma degli *outliers*, sia perché provando ad eliminare questo campione dagli spettri considerati, i picchi anomali non risultano più nel diagramma cumulativo degli spettri Raman. Dai diagrammi di *clustering* è stato inoltre confermato che queste tre prove non sono anomalie gravi per la costruzione della regressione in quanto non sono distaccate dalle altre, ma le misure simili sono tutte raggruppate tra loro e presentano anche una certa linearità su questi grafici. Sono quindi state mantenute nel set per la regressione con PCR (Figura 4.33 e Figura 4.34) e PLSR (Figura 4.35) sebbene ci si aspetta che la predizione per queste prove possa presentare l'errore assoluto maggiore rispetto agli altri campioni.



**Figura 4.33:** PCR effettuata su soluzioni di glucosio con interferenti a concentrazione crescente.

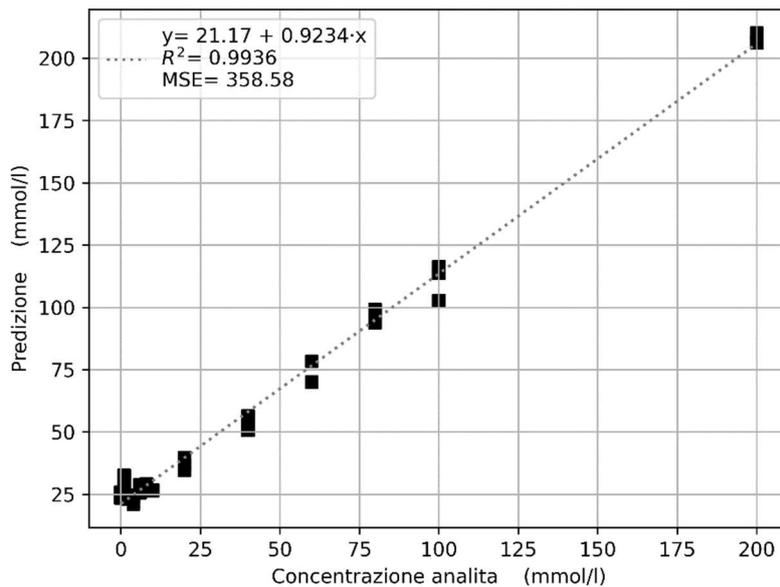
Nella PCR, si è ottenuta una buona linearità, soprattutto per quanto riguarda il coefficiente di determinazione  $R^2$  molto vicino all'unità. Anche il valore dell'intercetta e della pendenza della retta indicano una buona correlazione tra le predizioni del modello e i valori reali. Lo scarto quadratico medio e la forma del grafico sottolineano una leggera perdita di linearità per i campioni più concentrati, a partire da circa 80 mM. Questo comportamento è dovuto alla pendenza non unitaria della retta interpolante, anche se considerando i livelli fisiologici del glucosio nel sudore già discussi ci si accorge che non rappresenta un limite per l'applicabilità del metodo, in quanto neanche in case di grave iperglicemia si raggiungerebbero questi valori.



**Figura 4.34:** Ingrandimento della PCR sul set di glucosio con interferenti ed errore assoluto.

Analizzando l'ingrandimento della PCR nella zona da 0 a 10 mM si nota che la linearità è buona anche per i campioni più diluiti, con l'unica eccezione delle ripetizioni del campione 1 mM per i motivi trattati sopra. A questo campione corrisponde infatti l'errore assoluto massimo di 4 mM su una singola ripetizione.

La regressione ottenuta con PLSR ha invece presentato risultati peggiori rispetto alla PCR.



**Figura 4.35:** PLSR effettuata sul set di glucosio con interferenti e a concentrazione crescente.

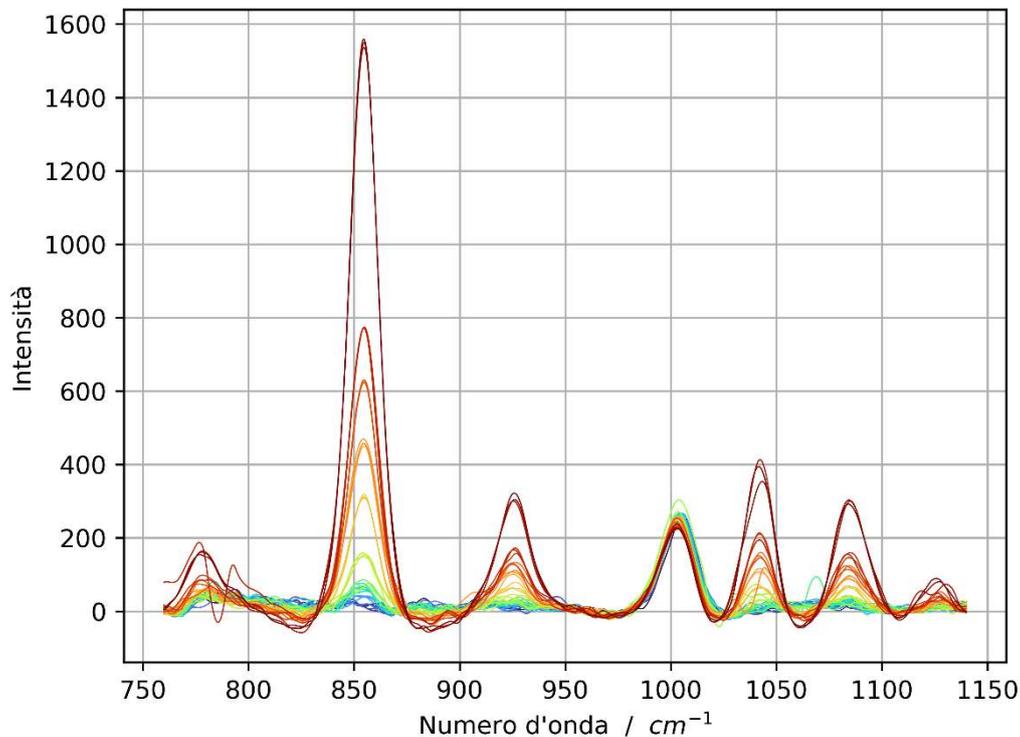
Sebbene i valori di  $R^2$  e pendenza siano buoni, segno di una buona linearità della predizione, lo scarto quadratico medio ed il valore dell'intercetta segnalano che il modello non riconosce direttamente il valore di concentrazione delle misure. In particolare, ogni campione viene maggiorato dall' algoritmo di una quantità pari a 21,17 mM. Questo significa che il modello è in grado di riconoscere con buona sensibilità la differenza di concentrazione tra due campioni, ma non è in grado di determinare il valore in assoluto a meno che non venga utilizzato un fattore correttivo, numericamente pari all'opposto dell' intercetta. Infatti, se a tutte le predizioni del modello venisse sottratto il valore 21,17 i risultati sarebbero in linea con i valori di concentrazione reali delle soluzioni in esame. L'MSE pari a 358,58 è anche dato da questo effetto, ma, se si applicasse il fattore correttivo, rientrerebbe in un range nuovamente accettabile.

In generale si può comunque dire che utilizzando la PLSR si hanno predizioni molto meno accurate in quanto, essendo il modello troppo adattato al set di riferimento, minime variazioni negli spettri incogniti portano a grandi variazioni nel risultato finale e quindi l'algoritmo si comporta come ipotizzato, andando a sottolineare differenze anche minime fra campioni molto simili.

#### 4.2.2. Analisi del lattato con interferenti

Per questa analisi il lattato era presente in quantità variabile, mentre glucosio e urea erano fissi rispettivamente a 1 mM e 20 mM.

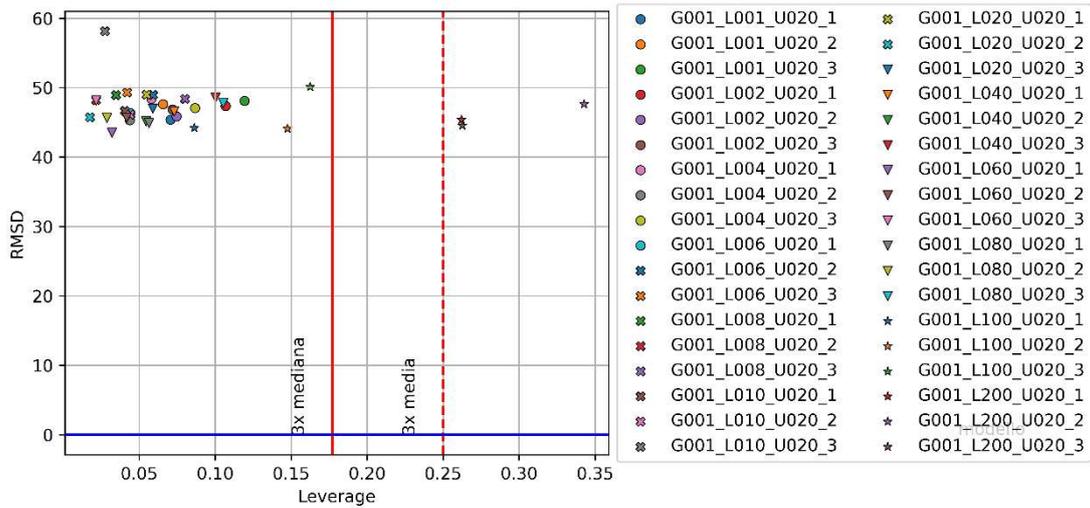
Visualizzando gli spettri non si notano particolari problematiche delle misure, se non un solo campione ad alta concentrazione il cui picco a  $775\text{ cm}^{-1}$  assume un andamento anomalo (Figura 4.36).



**Figura 4.36:** Spettri Raman del lattato in soluzioni con interferenti. La scala cromatica va dal viola, corrispondente al campione 0 mM, fino al rosso scuro, rappresentante i campioni 200 mM.

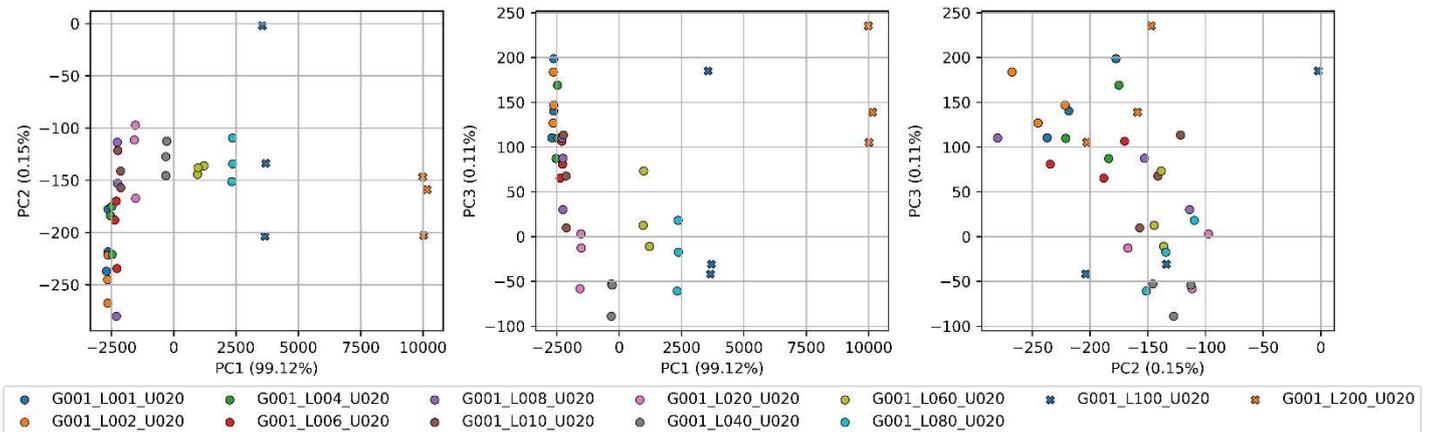
Si nota il picco dell'urea a  $1000\text{ cm}^{-1}$ , di intensità paragonabile a tutti i picchi secondari dello spettro classico del lattato. L'influenza del glucosio è praticamente nulla, in quanto questo è presente in concentrazione 1 mM, ovvero molto bassa. A questa concentrazione i suoi picchi hanno intensità assolutamente trascurabile rispetto a quelli degli altri due analiti. Possiamo dunque dire che il glucosio in concentrazione fisiologica non è un interferente nella misura di lattato e urea.

Lo studio degli *outliers* (Figura 4.37) ha mostrato che non vi sono misure anomale, sottolineando inoltre valori di RMSD piuttosto contenuti e ravvicinati, segno di una buona ricostruzione degli spettri nello spazio vettoriale degli autovettori.

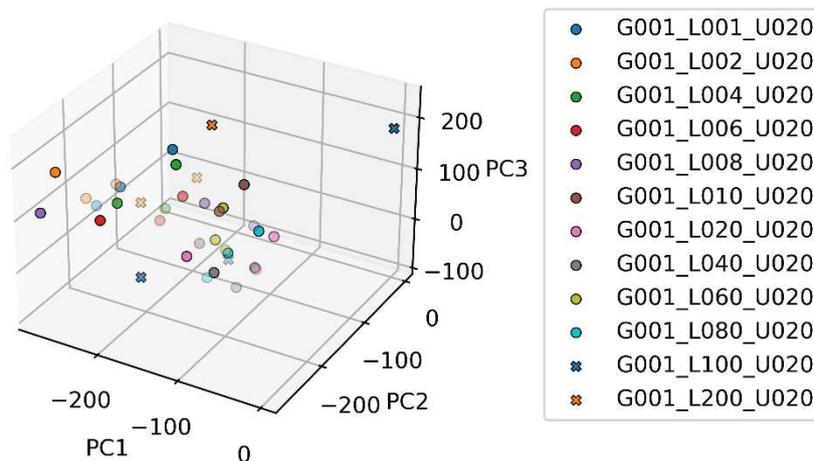


**Figura 4.37:** Determinazione degli *outliers* per il set di lattato a concentrazione variabile con interferenti.

Anche i diagrammi di clustering, bidimensionali (Figura 4.38) e tridimensionali (Figura 4.39), non mostrano prove da riconsiderare, è presente solo una lieve ripartizione delle prove 100 mM, probabilmente data dal campione discusso a valle della presentazione degli spettri cumulativi.

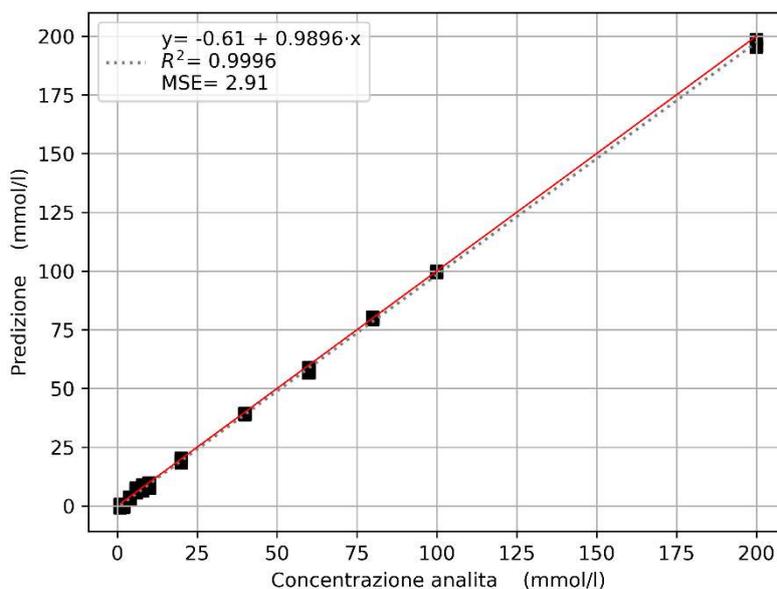


**Figura 4.38:** Diagrammi biplot del set di lattato a concentrazione variabile in presenza di interferenti.



**Figura 4.39:** Rappresentazione tridimensionale dei diagrammi di *clustering* del lattato in presenza di interferenti.

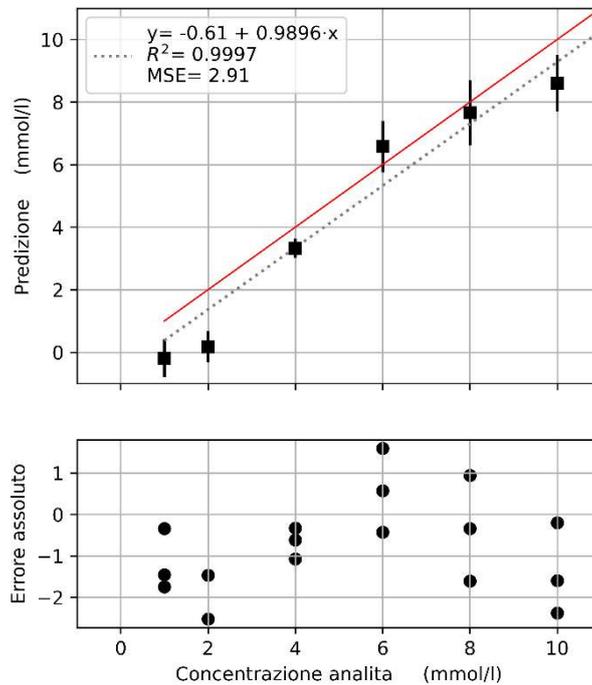
Per quanto riguarda la PCR, è stata ottenuta una buona linearità, con valori di coefficiente di determinazione e pendenza molto vicini all'unità. Lo scarto quadratico medio indica una buona aderenza delle predizioni del modello ai reali valori di concentrazione delle soluzioni (Figura 4.40).



**Figura 4.40:** PCR effettuata su soluzioni di lattato con interferenti a concentrazione crescente.

L'unico dato fonte potenziale di problemi è l'intercetta della retta di interpolazione, pari a  $-0,61$  mM. Questo valore non ha particolare influenza nella lettura dei campioni concentrati, ma può rappresentare una significativa sottostima per campioni molto diluiti, come visualizzato nell'ingrandimento nella zona tra 0 e 10 mM (Figura 4.41). Va comunque considerato che i livelli di lattato nel sudore sono mediamente molto superiori alle concentrazioni minori prese in esame, per cui solo in condizioni di grave

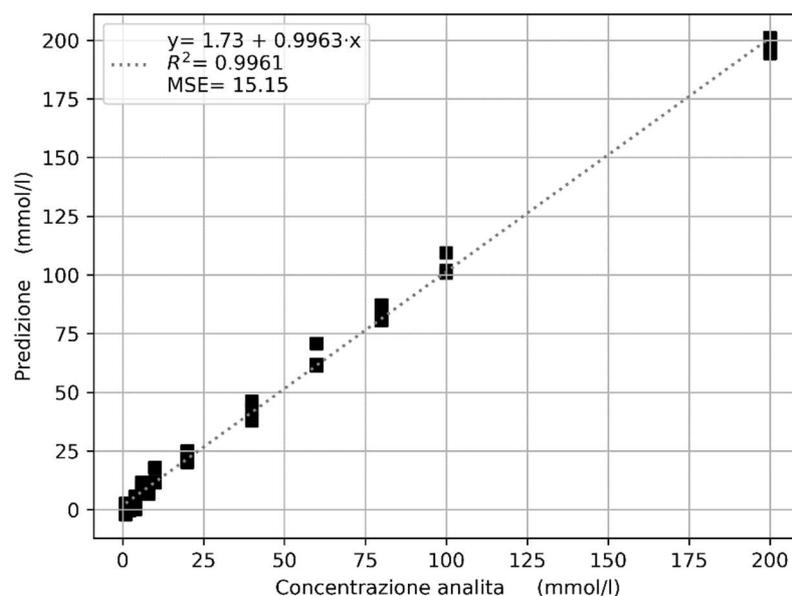
carenza si raggiungerebbero questi range. Inoltre, si potrebbe prevedere anche in questo caso un fattore correttivo, uguale ed opposto al valore dell'intercetta, da inserire nell'algoritmo dell'eventuale sensore di monitoraggio.



**Figura 4.41:** Ingrandimento della PCR sul set di lattato con interferenti ed errore assoluto.

Come discusso sopra, si nota un distacco tra la retta tratteggiata del modello e la diagonale teorica riportata in rosso. Gli errori assoluti sono comunque piuttosto contenuti, con variazioni massime di 2 mM per singole prove. In ultima analisi, è positiva la mancanza di perdita di linearità per alte concentrazioni, visto che il lattato è l'analita presente in concentrazioni maggiori nel sudore umano e, quindi, quello per cui si desidera una maggior precisione nell'intervallo di concentrazioni medio-alto, in maniera da avere buone predizioni anche in caso di accumulo eccessivo dell'analita.

Nella regressione eseguita con PLSR (Figura 4.42) i valori di  $R^2$  e pendenza indicano buona linearità e sensibilità del modello. Come nel caso precedente, i valori di intercetta e MSE sono più elevati rispetto alla PCR, il che si riflette maggiormente nella determinazione della concentrazione in soluzioni più diluite.



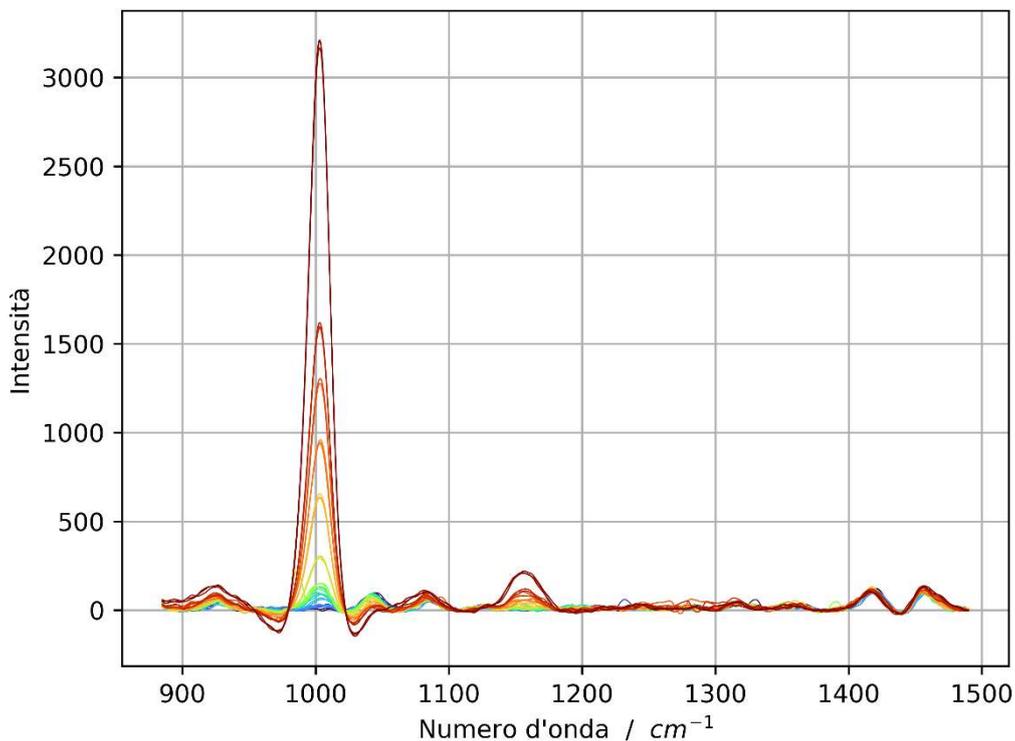
**Figura 4.42:** PLSR effettuata sul set di lattato con interferenti e a concentrazione crescente.

Anche in questo caso, l'andamento dei risultati della PLSR è da attribuire all'*overfitting* del modello di riferimento utilizzato per queste matrici.

#### 4.2.3. Analisi dell'urea con interferenti

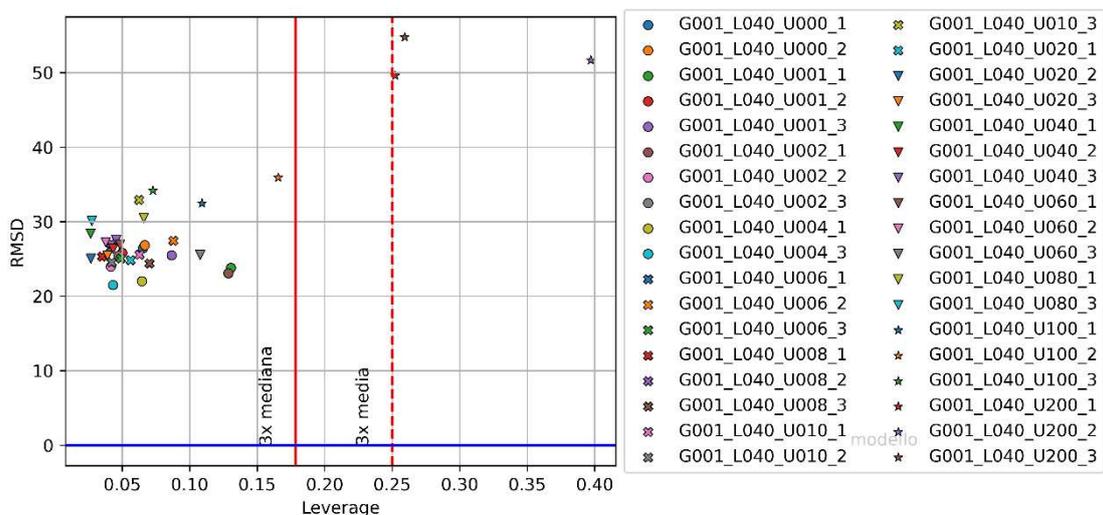
Si riportano i risultati dell'analisi eseguita sul set a concentrazione variabile di urea in soluzioni con glucosio 1 mM e lattato 40 mM come interferenti. Anche in questo caso, il glucosio è presente in concentrazione così bassa da non rappresentare un effetto visibile nella modifica dei picchi e sulla conseguente misura nella concentrazione dell'urea.

Dall'analisi degli spettri Raman acquisiti (Figura 4.43), si notano le influenze del lattato, visibili nei picchi a 1040 e 1080  $\text{cm}^{-1}$ . Questi picchi si sovrappongono in parte col picco principale dell'urea a 1000  $\text{cm}^{-1}$ , e va considerato che a questo numero d'onda corrisponde anche un terzo picco secondario del lattato, il quale viene coperto interamente dal principale, probabilmente modificandone in minima parte l'intensità. Vi sono poi anche due picchi a 1420 e 1450  $\text{cm}^{-1}$  circa, tipici del lattato ma non interferenti con i picchi dell'urea e dunque non importanti per la misura.



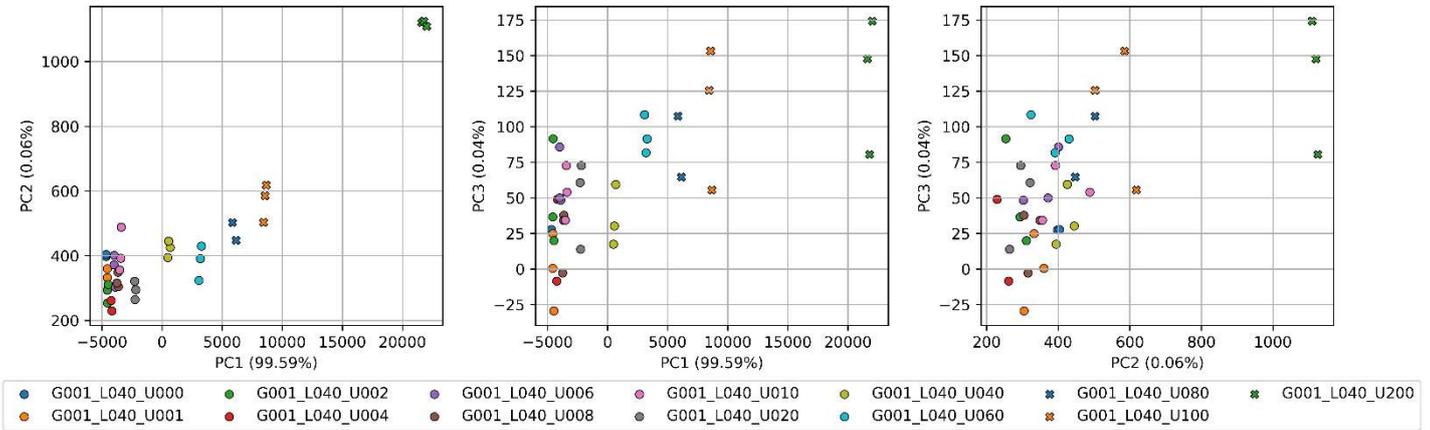
**Figura 4.43:** Spettri Raman dell'urea in soluzioni con interferenti. La scala cromatica va dal viola, corrispondente al campione 0 mM, fino al rosso scuro, rappresentante i campioni 200 mM.

L'analisi degli *outliers* non ha sottolineato alcuna anomalia nelle misure eseguite, presentando valori piuttosto bassi di deviazione media quadratica e di *leverage*, con solo i tre campioni a 200 mM presenti oltre le due soglie tipiche (Figura 4.44).

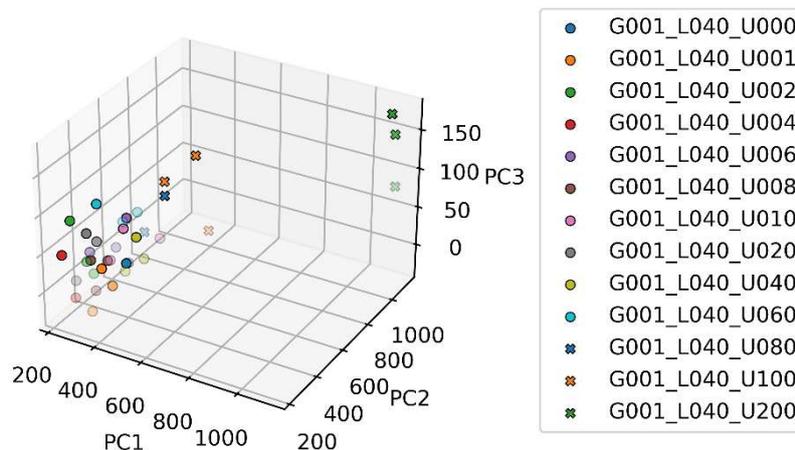


**Figura 4.44:** Determinazione degli *outliers* per il set di urea a concentrazione variabile con interferenti.

Questo risultato è confermato anche dai diagrammi di clustering bidimensionali e tridimensionali, i quali sottolineano un'ottima correlazione fra le prove di pari concentrazione (Figura 4.45 e Figura 4.46).

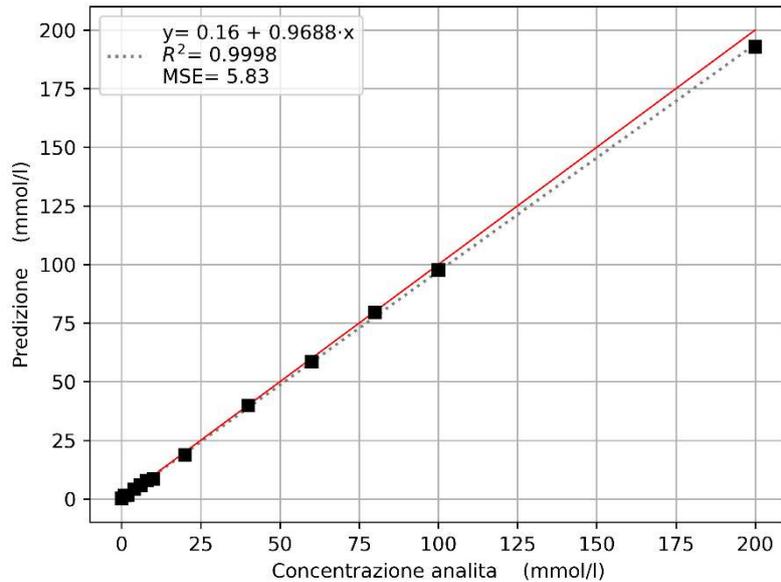


**Figura 4.45:** Diagrammi biplot del set di urea a concentrazione variabile in presenza di interferenti.



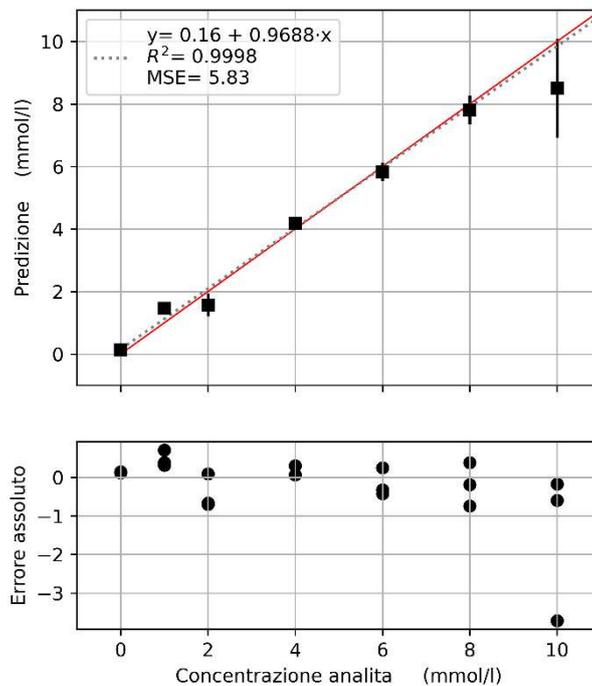
**Figura 4.46:** Rappresentazione tridimensionale dei diagrammi di *clustering* dell'urea in presenza di interferenti.

Il coefficiente di determinazione relativo alla regressione con PCR indica una buona linearità. Inoltre, i valori di pendenza e intercetta della retta interpolante indicano che il modello possiede buona sensibilità fra campioni a concentrazioni diverse ed è in grado di rilevare anche basse concentrazioni di urea in soluzione (Figura 4.47).



**Figura 4.47:** PCR effettuata su soluzioni di urea con interferenti a concentrazione crescente.

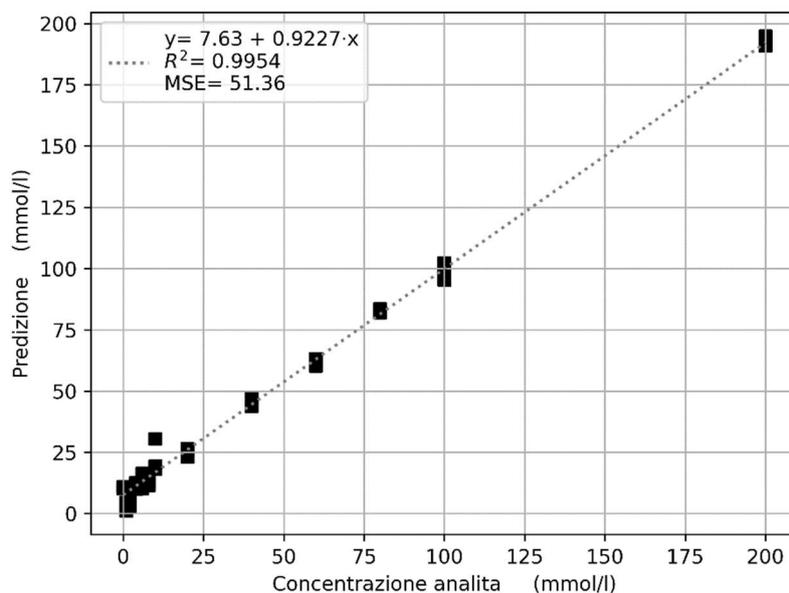
Si può notare l'effetto della pendenza non unitaria più marcato per alte concentrazioni, ma nuovamente queste concentrazioni non sono raggiungibili nemmeno in casi di grave insufficienza renale o per malattie epatiche acute, che, come noto, portano all'accumulo di urea. Lo scarto quadratico medio non è molto elevato, soprattutto se si considera che in buona parte il suo valore è dato dalla perdita di linearità sopra discussa, che, come detto, non è un vero problema per l'applicabilità del modello ad un caso reale. Si riporta l'ingrandimento della zona sotto i 10 mM (Figura 4.48).



**Figura 4.48:** Ingrandimento della PCR sul set di urea con interferenti ed errore assoluto.

Si evidenzia una sola ripetizione del campione a 10 mM con un errore assoluto pari a 3 mM, ma al di fuori di questa gli errori sono tutti molto contenuti rispetto ai casi precedenti. La linearità è comunque buona anche in questo campo e la lettura del bianco, ovvero la concentrazione nulla, presenta uno scarto minimo, per cui anche in questo caso non risulta necessario l'utilizzo di fattori correttivi.

La PLSR non ha portato a risultati altrettanto buoni (Figura 4.49).



**Figura 4.49:** PLSR effettuata sul set di urea con interferenti e a concentrazione crescente.

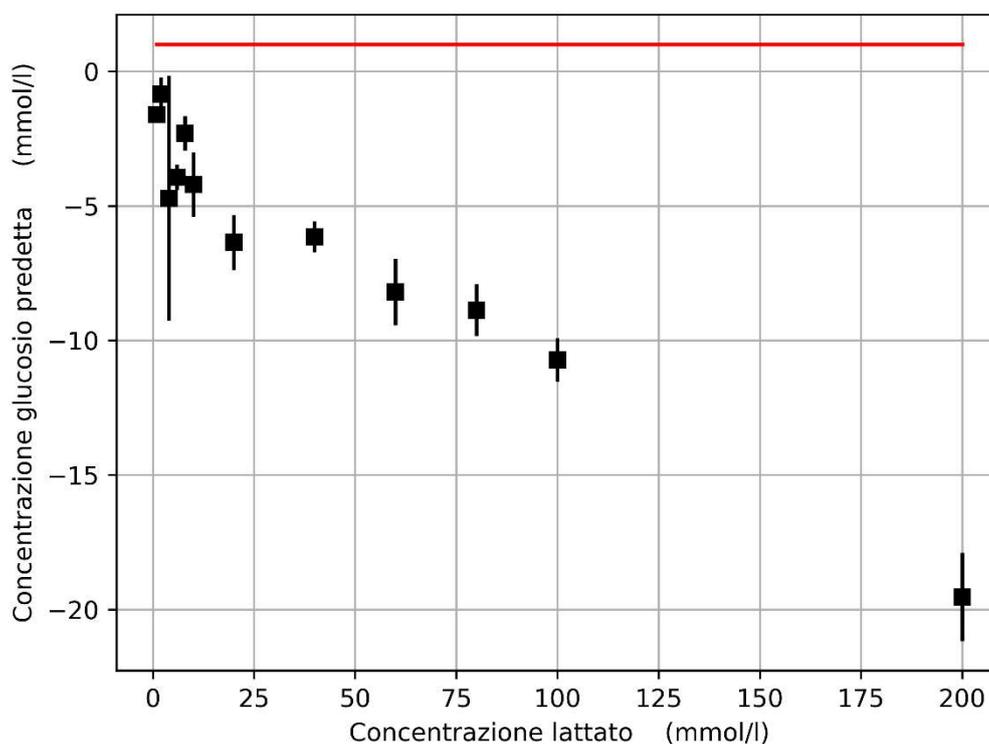
Il grado di linearità è comunque soddisfacente, sia per quanto riguarda il valore di  $R^2$  che per quanto riguarda la pendenza della retta interpolante, ma il reale problema è rappresentato dallo scarto quadratico medio e dal valore dell'intercetta, entrambi piuttosto alti. Questo è dato dalla dispersione dei dati per concentrazioni minori di 20 mM. Per questi campioni, il modello risulta troppo adattato al set di base e la predizione presenta errori piuttosto elevati. Il valore dell'intercetta indica che il modello richiederebbe un fattore correttivo elevato all'inizio della retta, ma se si osservano le predizioni associate ai campioni a 200 mM si nota che essi sono minori di questo valore. Questo è dovuto al valore della pendenza della retta, più lontana dall'unità rispetto ai casi di lattato e glucosio. Questo fa sì che i valori ad alta concentrazione necessitino fattori di correzione positivi, al contrario dei campioni diluiti, per cui anche questo parametro andrebbe adattato tenendo in considerazione la pendenza della retta.

### 4.3. Misura degli interferenti al variare dell'analita principale

In queste analisi, i set di dati di lattato e urea in soluzioni con interferenti trattati nelle sottosezioni 4.2.2 e 4.2.3 vengono rianalizzati con il modello di ricerca del glucosio. Per eseguire questo processo, vanno inseriti come input del modello sia l'informazione di quale sia il componente a concentrazione variabile, che assume la funzione di interferente, che l'informazione di quale analita si voglia ricercare, ovvero il glucosio. I risultati sono ottenuti mediante la retta di regressione ottenuta con PCR, in quanto, come mostrato sopra, risulta essere il metodo più efficace per le analisi su queste soluzioni.

#### 4.3.1. Analisi del glucosio in soluzioni di lattato a concentrazione variabile

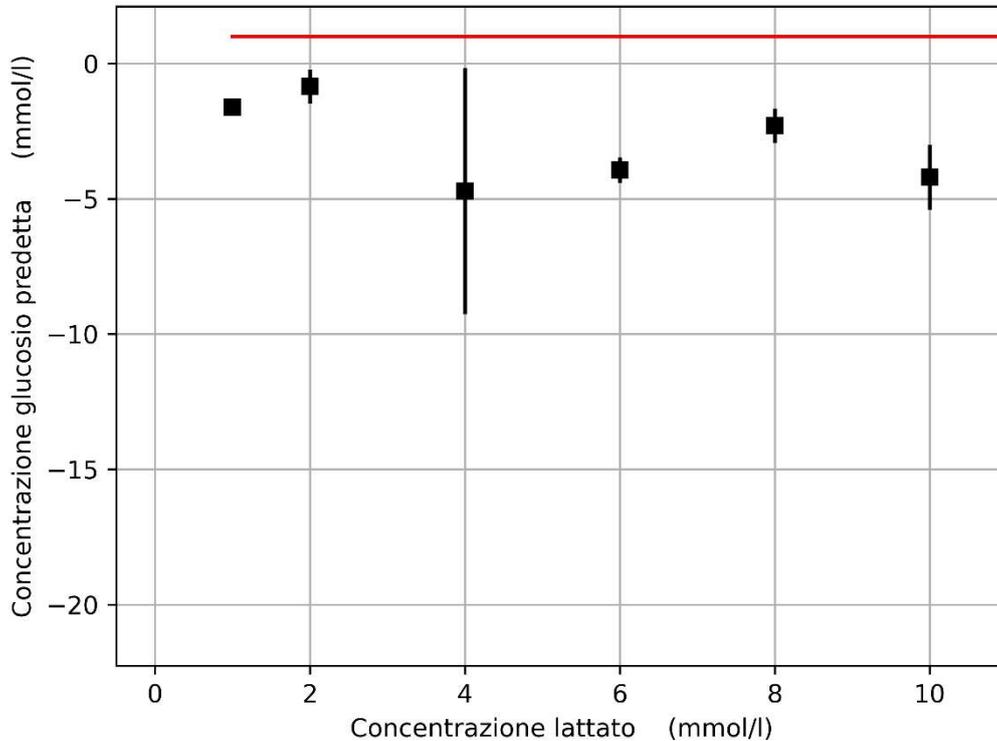
I risultati sono stati riportati in un grafico in cui per ogni campione viene indicata sia la concentrazione di lattato, in ascissa, che la concentrazione di glucosio, riportata in ordinata (Figura 4.50).



**Figura 4.50:** Determinazione del glucosio in soluzioni a concentrazione crescente di lattato e concentrazione fissa di urea. La linea rossa indica il valore reale di concentrazione del glucosio.

Il grafico in Figura 4.50 mostra che al crescere della concentrazione di lattato in soluzione, la previsione è sempre meno accurata. In particolare, si identifica una decrescita piuttosto lineare all'aumentare della quantità di lattato. Questo andamento è imputabile alla sempre maggior interferenza che il lattato crea nel calcolo della

concentrazione di glucosio, anche vista la notevole differenza tra la quantità dei due analiti. In più, vi è anche l'effetto dell'interferenza data dall'urea, fissa a 20 mM. Se si ingrandisce la zona tra 0 mM e 10 mM, entrando quindi nel campo di concentrazioni più classiche del glucosio, anche in condizioni di iperglicemia, si ottengono risultati interessanti (Figura 4.51).

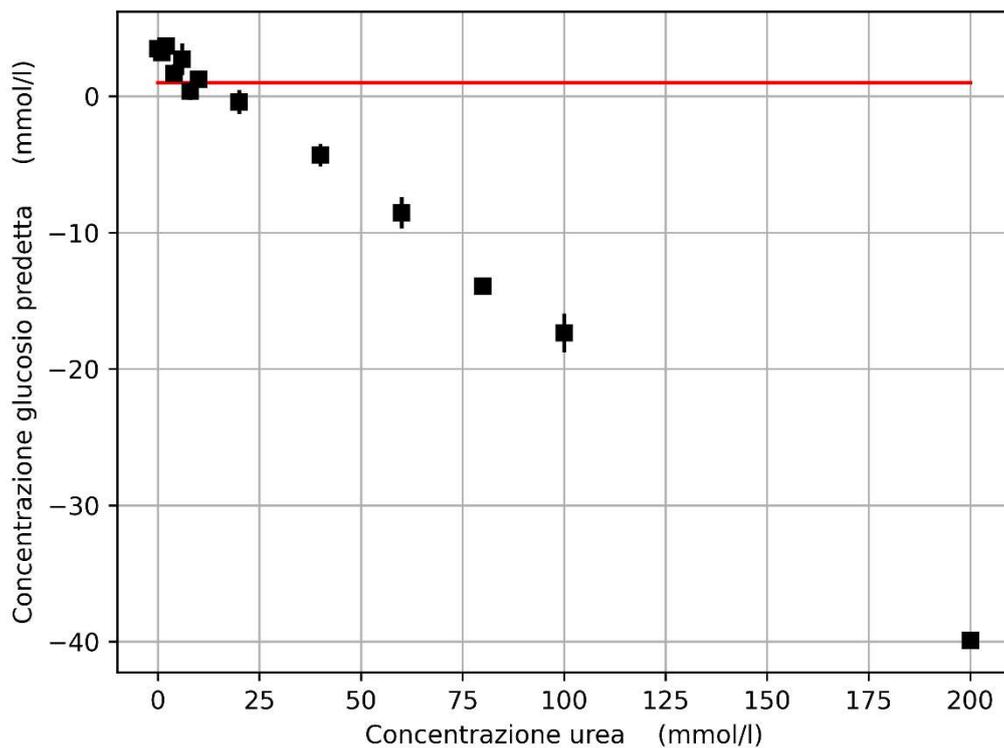


**Figura 4.51:** Ingrandimento della determinazione del glucosio in soluzioni a concentrazione crescente di lattato e concentrazione fissa di urea. La linea rossa indica il valore reale di concentrazione del glucosio.

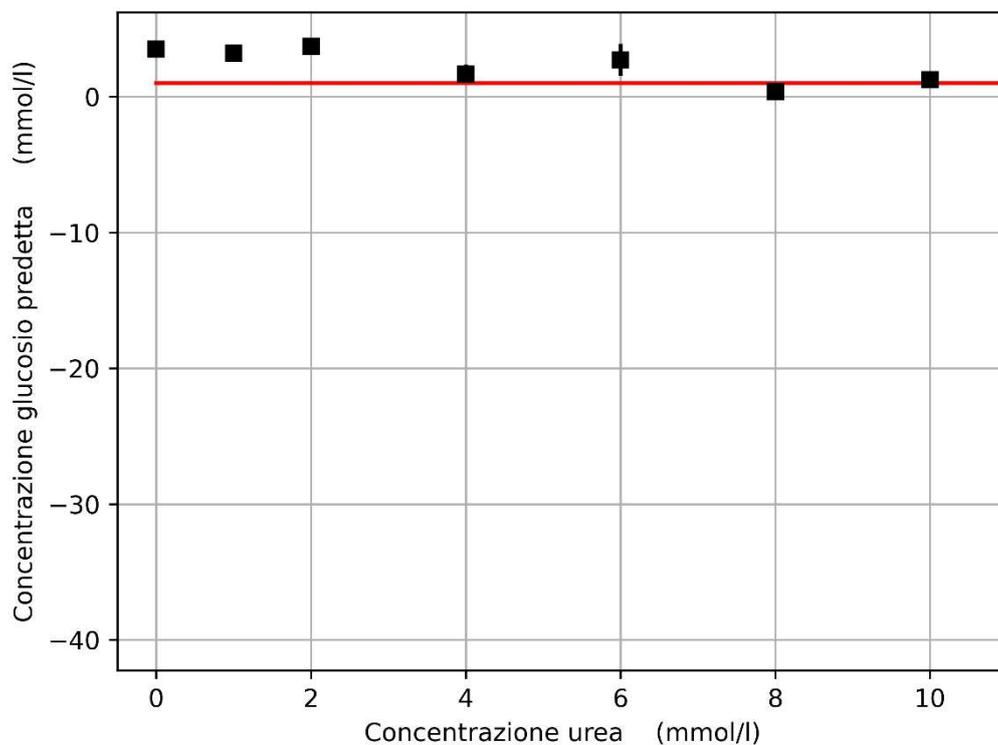
In questa zona le predizioni sono ancora poco accurate, ma con errori molto minori di quelli dei campioni a concentrazioni maggiori. In media, si ha una sottostima delle predizioni rispetto al valore reale di circa 3,5 mM. Possiamo dunque affermare che in questa analisi vi sono due tipologie di interferenze, la prima di carattere costante, dovuta all'interferenza nella lettura fornita dalla presenza di un quantitativo fisso di urea, e uno variabile, dato dall'aumento della concentrazione, e quindi dell'influenza, del lattato.

#### 4.3.2. Analisi del glucosio in soluzioni di urea a concentrazione variabile

Sono riportati i risultati di questa analisi sul set intero (Figura 4.52) e l'ingrandimento nel range di concentrazioni minori di 10 mM (Figura 4.53).



**Figura 4.52:** Determinazione del glucosio in soluzioni a concentrazione crescente di urea e concentrazione fissa di lattato. La linea rossa indica il valore reale di concentrazione del glucosio.

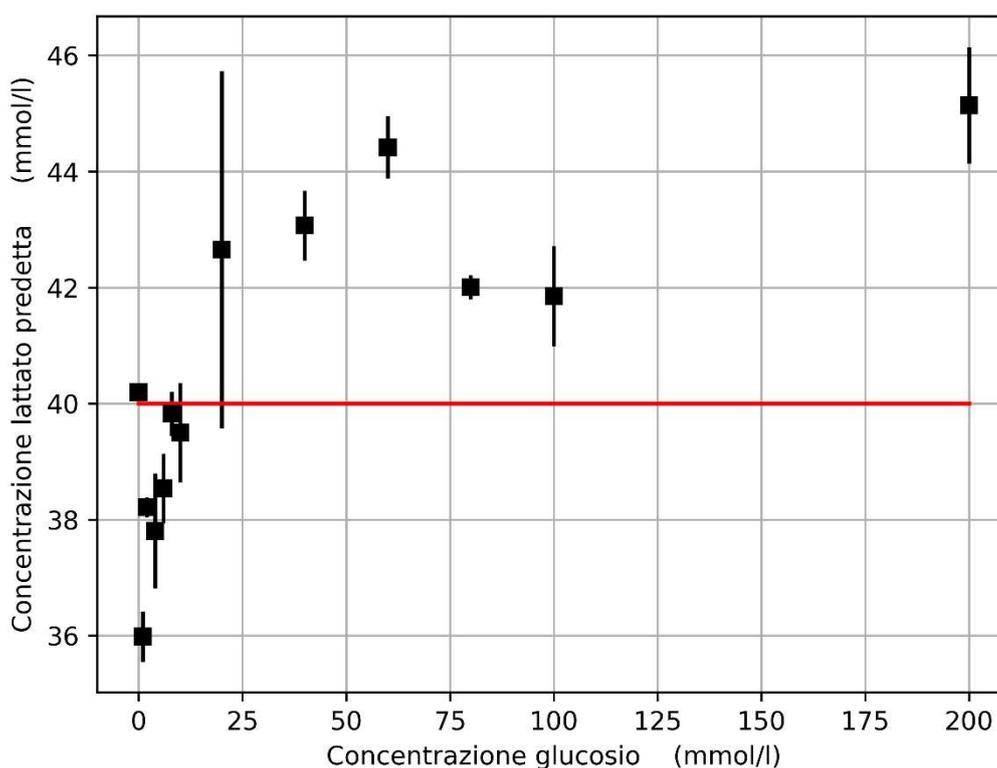


**Figura 4.53:** Ingrandimento della determinazione del glucosio in soluzioni a concentrazione crescente di urea e concentrazione fissa di lattato. La linea rossa indica il valore reale di concentrazione del glucosio.

Anche in questo caso si notano due effetti di interferenza di genere differente, uno fisso, dovuto alla presenza costante di 40 mM di lattato, e uno variabile creato dall'aumento della concentrazione di urea. Si nota comunque che rispetto al caso precedente, l'errore variabile dovuto all'aumento di concentrazione dell'urea è più marcato. Questo aspetto è in accordo con quanto visualizzato in precedenza, ovvero che, a parità di concentrazione, l'urea presenta picchi più intensi del lattato e dunque fornisce un'interferenza maggiore e, di conseguenza, un errore sempre maggiore. In maniera opposta, l'errore di predizione costante dovuto alla presenza del lattato è inferiore al caso precedente, conclusione che deriva però dallo stesso ragionamento appena eseguito. Analizzando la zona ingrandita, si nota infatti che l'errore massimo di lettura è di circa 2 mM, quasi la metà rispetto all'analisi precedente.

#### 4.3.3. Analisi del lattato in soluzioni di glucosio a concentrazione variabile

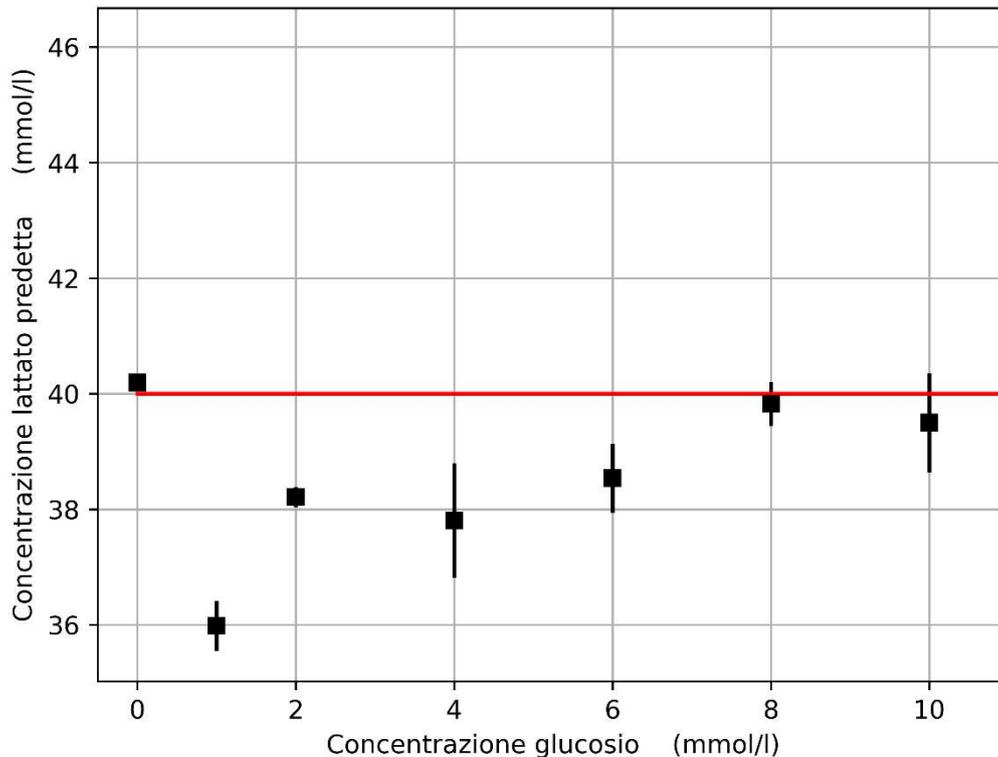
In questa analisi è ricercata la concentrazione del lattato nelle soluzioni a concentrazione variabile di glucosio usate nella sottosezione 4.2.1. Si riportano i risultati ottenuti sul set intero (Figura 4.54) e nell'intervallo tra 0 mM e 10 mM di glucosio (Figura 4.55).



**Figura 4.54:** Determinazione del lattato in soluzioni a concentrazione crescente di glucosio e concentrazione fissa di urea. La linea rossa indica il valore reale di concentrazione del lattato.

In questa analisi i risultati sono molto meno lineari rispetto ai rilevamenti di glucosio nelle due analisi precedenti. Le predizioni si dispongono in maniera migliore attorno alla retta della concentrazione reale, indicando quindi una minore interferenza

del glucosio all'aumentare della sua concentrazione. Nonostante questo, si nota che gli errori assoluti sono piuttosto alti, con variazioni fino a 6 mM su un valore reale di 40 mM e un errore relativo medio del 7%.

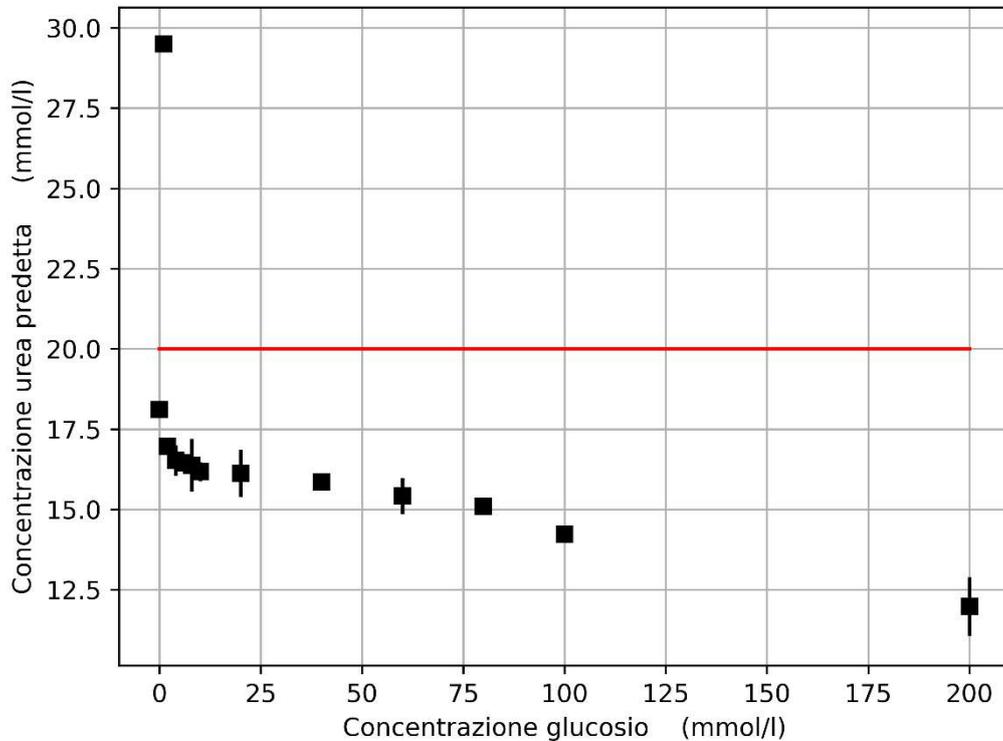


**Figura 4.55:** Ingrandimento della determinazione del lattato in soluzioni a concentrazione crescente di glucosio e concentrazione fissa di urea. La linea rossa indica il valore reale di concentrazione del lattato.

Nell'ingrandimento della zona tra 0 mM e 10 mM si sono ottenuti errori inferiori, con una sola prova oltre le 4 mM di differenza. Si ottiene un errore relativo medio del 3%. Il principale risultato per queste misure è il fatto che all'aumentare della concentrazione di glucosio, l'errore sulla misura del lattato non aumenta linearmente, ma rimane all'incirca costante. Questo sottolinea che l'effetto di interferenza del glucosio nei confronti del lattato è relativamente bassa, nonostante le visibili oscillazioni nell'errore di misura. Tale effetto è ricollegabile all'alto numero di picchi del lattato nel proprio spettro, che rendono possibile la sua misura non solo tramite l'utilizzo del picco principale, che potrebbe essere sovrapposto a quelli del glucosio. La presenza di altri picchi utilizzabili per la misura fa sì che l'errore dovuto alla sovrapposizione di un solo picco del lattato venga diminuito dalla presenza degli altri picchi non sovrapposti.

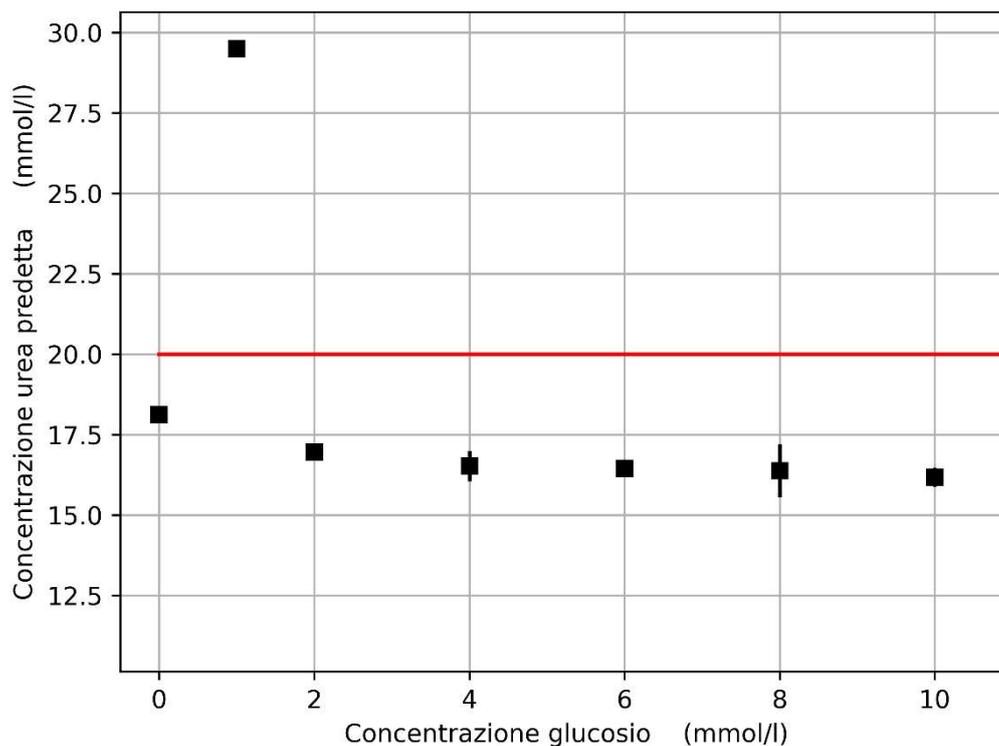
#### 4.3.4. Analisi dell'urea in soluzioni di glucosio a concentrazione variabile

Si è analizzata la concentrazione dell'urea nel set di soluzioni a concentrazione di glucosio variabile usato nella sezione precedente (Figura 4.56).



**Figura 4.56:** Determinazione dell'urea in soluzioni a concentrazione crescente di glucosio e concentrazione fissa di lattato. La linea rossa indica il valore reale di concentrazione dell'urea.

In questo caso si ha l'effetto opposto rispetto alla determinazione del lattato in questo set di soluzioni. Infatti, si nota che all'aumentare della concentrazione di glucosio in soluzione, l'errore nel calcolo della concentrazione di urea aumenta con andamento quasi lineare. Questo effetto è visibile, anche se in minor parte, nell'ingrandimento dell'intervallo tra 0 mM e 10 mM di glucosio (Figura 4.57).



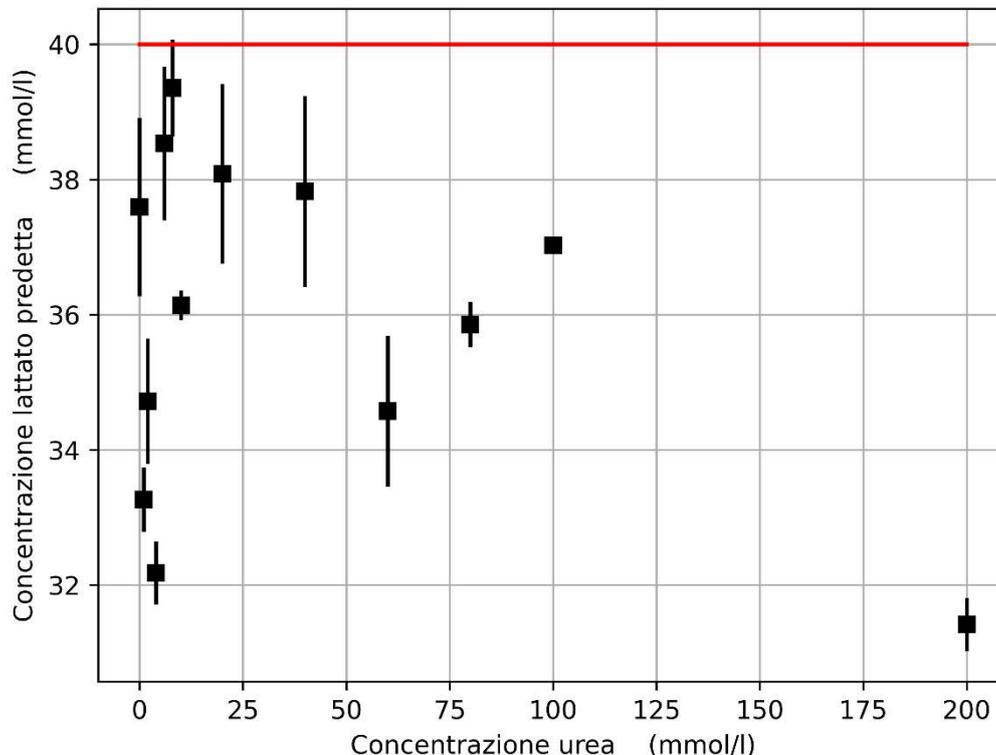
**Figura 4.57:** Ingrandimento della determinazione dell'urea in soluzioni a concentrazione crescente di glucosio e concentrazione fissa di lattato. La linea rossa indica il valore reale di concentrazione dell'urea.

Si nota un solo problema nella misura di uno dei campioni più diluiti. Come riportato anche nella sezione 4.2.1, il campione 1 mM di questo set presentava un picco anomalo a circa  $1000\text{ cm}^{-1}$  dovuto ad una concentrazione doppia di urea. Questa analisi è un'ulteriore conferma di questo errore nella preparazione delle soluzioni, in quanto il campione 1 mM è l'unico a distaccarsi dall'andamento tenuto da tutte le altre prove e presenta anche un errore relativo doppio rispetto a quello atteso se fosse rispettata tale linearità. Se si trascura tale prova, si nota comunque un andamento piuttosto lineare, con errori relativi massimi del 15%, valori ancora piuttosto alti. Il distacco medio è comunque minore rispetto al caso del lattato, a riprova ancora del fatto che l'urea, avendo un picco più intenso di quello principale del lattato, presenta un'interferenza maggiore nei confronti del glucosio.

Inoltre, a differenza del caso precedente, l'allontanarsi della predizione dal valore reale è imputabile alla presenza di un solo picco principale dell'urea, ed alla conseguente assenza di picchi secondari che possano diminuire l'effetto della sovrapposizione del picco principale, comportamento opposto al caso del lattato.

#### 4.3.5. Analisi del lattato in soluzioni di urea a concentrazione variabile

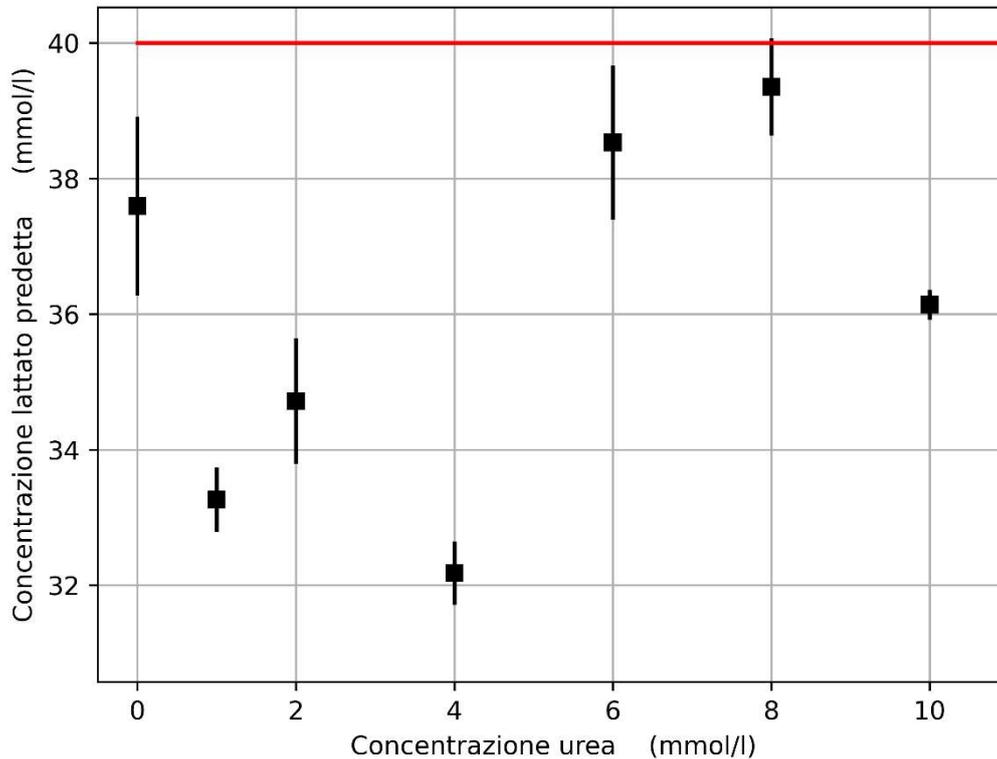
In questa analisi si è ricercata la concentrazione del lattato nel set di soluzioni utilizzate nella sottosezione 4.2.3. I risultati sul set intero sono riportati di seguito (Figura 4.58).



**Figura 4.58:** Determinazione del lattato in soluzioni a concentrazione crescente di urea e concentrazione fissa di glucosio. La linea rossa indica il valore reale di concentrazione del lattato.

I risultati indicano che le predizioni presentano andamento non lineare. Infatti, l'urea provoca interferenza, all'aumentare della concentrazione, solo nell'intorno dei  $1000\text{ cm}^{-1}$ , creando sovrapposizioni solo con picchi secondari del lattato. Il picco principale non viene influenzato, per cui anche per alti valori di concentrazione di urea l'errore nella predizione non aumenta di conseguenza.

Nonostante ciò, gli errori sono piuttosto elevati, con errore relativo medio di 12,5% ed errore assoluto massimo di quasi 9 mM (quasi il 25%). Ingrandendo l'intervallo 0-10 mM si ottengono i seguenti risultati (Figura 4.59).



**Figura 4.59:** Ingrandimento della determinazione del lattato in soluzioni a concentrazione crescente di urea e concentrazione fissa di glucosio. La linea rossa indica il valore reale di concentrazione del lattato.

Anche in questo campo di concentrazioni si notano errori elevati, come nel caso del campione 4 mM che presenta un errore molto vicino a quello del campione 200 mM, ad ulteriore riprova dell'assenza di un aumento progressivo dell'errore con l'aumento della concentrazione.

Rispetto ai casi precedenti si nota anche una maggior dispersione delle misure a pari concentrazione, come visibile dalle barre d'errore nei grafici, più ampie e frequenti che nelle altre analisi. Questo è indice di una precisione minore del modello nella ricerca del lattato

#### 4.3.6. Analisi dell'urea in soluzioni di lattato a concentrazione variabile

Per questa analisi è stato utilizzato il set di soluzioni della sezione 4.2.2. i risultati sono riportati per il set intero (Figura 4.60) e ingrandendo la zona tra 0 e 10 mM (Figura 4.61).



In questo caso si torna ad avere un andamento decrescente nella predizione all'aumentare della concentrazione di lattato. Ancora una volta, questo effetto è attribuibile alla maggiore interferenza del lattato nei confronti dell'urea, piuttosto che nel caso contrario. Visto che il lattato presenta picchi in più zone differenti, all'aumentare della sua concentrazione corrisponde una modifica di altrettante parti dello spettro e quindi ad una più probabile sovrapposizione con il picco dell'urea. Nell'ingrandimento si nota una certa linearità nella predizione, con un errore assoluto medio pari a 4 mM, ovvero il 20% in termini relativi. In generale gli errori sono piuttosto alti, nonostante gli alti livelli di concentrazione considerati e dunque il forte effetto di interferenza creato.



## 5. Conclusioni e sviluppi futuri

In conclusione, si è costruito un modello in grado di calcolare la concentrazione di glucosio, lattato ed urea a partire dagli spettri Raman della soluzione incognita che li contiene. Si è presentato il lavoro svolto per l'addestramento del modello, mediante un set di riferimento degli analiti disciolti singolarmente in acqua deionizzata. È seguita la fase di misura con soluzioni simulanti il sudore umano, contenenti simultaneamente le tre componenti con solo una presente in concentrazione variabile. Infine, è stata discussa l'ultima fase di test in cui si è predetta la quantità degli interferenti costanti nel set precedente.

La prima analisi mediante PCA ha permesso di processare una grande quantità di dati e di individuare ed eliminare prove rumorose o anomale che avrebbero potuto creare problemi soprattutto nella calibrazione del modello. I risultati sono stati ottenuti analizzando non gli spettri originali, ma la loro derivata prima, tecnica frequente per analisi simili. La fase di misura su soluzioni di sudore artificiale ha portato a risultati positivi, mostrando, dopo la regressione con PCR e PLSR, coefficienti di determinazione  $R^2$  sempre maggiori di 0,99 i quali indicano un'ottima linearità. Le rette interpolanti hanno dimostrato che la linearità è mantenuta fino ad alte concentrazioni, sicuramente al di fuori del campo di concentrazioni che ci si aspetta di trovare in campioni reali di questo genere, dimostrando quindi l'applicabilità del modello anche per concentrazioni patologiche dei tre componenti al fine di individuare possibili malattie metaboliche. Inoltre, sono stati ottenuti valori dell'intercetta prossimi allo zero, dimostrando quindi anche una buona sensibilità di lettura per soluzioni diluite, che nel caso di lattato e urea possono essere ricollegate a patologie risultanti nella loro carenza. Per quanto riguarda il glucosio, il campo di soluzioni a concentrazione minore di 10 mM è quello di interesse primario che indica i corretti livelli fisiologici, per cui in questo caso una buona linearità in questo intervallo era a maggior ragione desiderata.

In generale, la PCR ha fornito risultati migliori rispetto alla PLSR. In contrapposizione alle previsioni, il confronto fra i due metodi ha visto la PCR costantemente più performante, sebbene nella maggior parte dei casi la differenza nella predizione è dell'ordine delle micromoli al litro con errori dello 0,1%. Tale differenza di prestazioni è parzialmente spiegata dalla relativa semplicità dei data set acquisiti, aspetto che porta nel caso della PLSR all'*overfitting* del modello, ovvero ad una eccessiva aderenza e correlazione dei dati di addestramento al modello stesso. Questo effetto si nota soprattutto sui set di analisi dell'urea, proprio perché lo spettro Raman relativo a questa molecola è più semplice rispetto al glucosio e lattato. Se si considera, però, che le analisi su pazienti reali avrebbe una matrice di analisi molto più complessa e varia, comprendendo acido urico, creatinina, acido piruvico, cloruro di sodio e molti solfati e fosfati inorganici, si arriva alla conclusione che l'accuratezza della PLSR aumenterà rispetto alla PCR.

Nell'ultima fase di questo lavoro, è stata eseguita la misura della concentrazione del glucosio nei set in cui questo compariva come interferente. I risultati sono stati soddisfacenti, considerate le basse concentrazioni considerate. Si è verificato che, a parità di concentrazione, l'urea presenta picchi più intensi rispetto al lattato e questo ha

influenza soprattutto nelle analisi dei campioni più concentrati. Al contrario, sebbene i suoi picchi siano meno intensi, il lattato crea interferenze più importanti soprattutto per le soluzioni diluite, in quanto, essendo chimicamente e strutturalmente più simile al glucosio, i suoi picchi si trovano in corrispondenza di un maggior numero di picchi del glucosio stesso.

La misura della concentrazione di lattato ed urea al variare della concentrazione del glucosio ha portato anch'essa a risultati positivi, indicando una bassa interferenza da parte del glucosio nei confronti dei due analiti. L'effetto dell'interferenza risulta più marcato nel caso dell'urea in quanto questa presenta un solo picco di determinazione, mentre nel caso del lattato la presenza di più picchi diminuisce l'effetto della sovrapposizione di uno solo di essi con uno del glucosio.

Infine, sono state misurate la concentrazione di urea in soluzioni a concentrazione di lattato variabile e, viceversa, la concentrazione di lattato in soluzioni a concentrazione di urea variabile. Queste analisi hanno portato a risultati con discostamenti maggiori rispetto alle analisi precedenti, con errori relativi medi oscillanti tra 10% e 20%.

In generale, queste prove effettuate su soluzioni a concentrazioni di interferenti variabili presentano errori maggiori sia per la presenza di una terza componente a concentrazione fissa, sia per la relativa influenza della componente a concentrazione variabile. L'efficienza di queste misure potrebbe essere incrementata implementando fattori correttivi nell'ottenimento della predizione

Il passo successivo nello studio di questa tecnica per la determinazione e monitoraggio del glucosio metabolico sarebbe l'analisi di campioni di sudore eccrino di origine umana, e non più di soluzioni simulanti. Questo porterebbe ad un miglioramento delle prestazioni della PLSR. Un altro accorgimento per incrementare l'efficacia del modello potrebbe essere l'implementazione di una differente metodologia di *cross-validation* considerata nella tecnica, ovvero separando il dataset in due sottogruppi, uno di addestramento e uno di prova. In tale maniera si costruisce la calibrazione sul primo set mentre gli spettri presenti nel secondo gruppo vengono utilizzati come dati per una "prova incrociata" volta a valutare le prestazioni del modello. Nel complesso, la tecnica risulta molto promettente, ma necessita di ulteriori analisi, cambiando la matrice in esame e i metodi di consolidazione del modello.

## Bibliografia

- [1] Egi M., Bellomo R., Stachowski E., French C.J., Hart G.K., Hegarty C., Bailey M., 2008, Blood glucose concentration and outcome of critical illness: The impact of diabetes, *Critical Care Medicine*. 36 (8), 2249-2255.
- [2] Kaul K., Tarr J.M., Ahmad S.I., Kohner E.M., Chibber R., 2013, Introduction to Diabetes Mellitus, in 'Diabetes. Advances in Experimental Medicine and Biology', Chap. 1., vol 771. Springer, New York, NY.
- [3] Maraschin J.d.F., 2013, Classification of Diabetes, 'Diabetes. Advances in Experimental Medicine and Biology', Chap. 2., vol 771. Springer, New York, NY.
- [4] Winer N., 2004, Epidemiology of diabetes. *The Journal of Clinical Pharmacology*. 44 (4), 397-405.
- [5] Centers for Disease Control and Prevention. National Diabetes Statistics Report website e American Diabetes Association. Disponibile online presso <https://www.cdc.gov/diabetes/data/statistics-report/index.html>
- [6] Gonzales W.V., Mobashsher A.T., Abbosh A., 2019, The progress of glucose monitoring – a review of invasive to minimally and non-invasive techniques, devices and sensors, *Sensors*. 19 (4), 800.
- [7] Vashit S.K., 2012, Non-invasive glucose monitoring technology in diabetes management: a review, *Analytica chimica Acta*. 750, 16-27.
- [8] Bruen D., Delaney C., Florea L., Diamond D., 2017, Glucose Sensing for Diabetes Monitoring: Recent Developments, *Sensors*. 17 (8), 1866
- [9] Moyer J., Wilson D., Finkelshtein I., Wong B., Potts R., 2012, Correlation Between Sweat Glucose and Blood Glucose in Subjects with Diabetes, *Diabetes Technology & Therapeutics*. 14 (5), 398-402.
- [10] Baker L.B., Wolfe A.S., 2020, Physiological mechanisms determining eccrine sweat composition, *European Journal of Applied Physiology*. 120, 719-752.
- [11] Vaquer A., Baron E., de la Rica R., 2021, Detection of low glucose levels in sweat with colorimetric wearable biosensors, *Analyst*. 146, 3273.
- [12] Lin P.H., Sheu S.C., Chen C.W., Huang S.C., Li B.R., 2022, Wearable hydrogel patch with noninvasive, electrochemical glucose sensor for natural sweat detection, *Talanta*. 241, 123187.

- [13] Li Y., Guo Y., Chen H., Xiao X., Long F., Zhong H., Wang K., Guo Z., Zhuang Z., Liu Z., 2024, Flexible Wearable Plasmonic Paper-Based Microfluidics with Expandable Channel and Adjustable Flow Rate for Portable Surface-Enhanced Raman Scattering Sweat Sensing, *ACS Photonics*. 11 (2), 613-625.
- [14] Golparvar A., Thenot L., Boukhayama A., Carrara S., 2023, Soft Epidermal Paperfluidics for Sweat Analysis by Ratiometric Raman Spectroscopy, *Biosensors*. 14 (1), 12.
- [15] Patterson M.J., Galloway S.D.R., Nimmo M.A., 2001, Variations in regional sweat composition in normal human males, *Experimental Physiology*. 85 (6), 869-875.
- [16] Verde T., Shephard R.J., Corey P., Moore R., 1982, Sweat composition in exercise and in heat, *Journal of applied physiology*. 53 (6), 1540-1545.
- [17] Zafar H., Channa A., Jeoti V., Stojanovic G.M., 2022, Comprehensive Review on Wearable Sweat-Glucose Sensors for Continuous Glucose Monitoring, *Sensors*. 22, 638.
- [18] Baker L.B., 2019, Physiology of sweat gland function: The roles of sweating and sweat composition in human health, *Temperature*. 6 (3), 211-259.
- [19] Karpova E.V., Laptev A.I., Andreev E.A., Karyakina E.E., Karyakin A.A., 2020, Relationship between sweat and blood lactate levels during exhaustive physical exercise, *ChemElectroChem*. 7, 191.
- [20] Siddiqui S.A., Zhang Y., Lloret J., Song H., Obradovic Z., 2018, Pain free blood glucose monitoring using wearable sensors: recent advancements and future prospects, *IEEE Reviews in Biomedical Engineering*. 11, 21-35.
- [21] Xu J., Fang Y., Chen J., 2021, Wearable biosensors for non-invasive Sweat Diagnostics, *Biosensors*. 11 (8) 245.
- [22] Todaro B., Begarani F., Sartori F., Luin S., 2022, Is Raman the best strategy towards the development of non-invasive continuous glucose monitoring devices for diabetes management?. *Frontiers in Chemistry*. 10.
- [23] Zhao Z., Li Q., Dong Y., Gong J., Li Z., Zhang J., 2021, Core-shell structured gold nanorods on thread-embroidered fabric-based microfluidic device for Ex Situ detection of glucose and lactate in sweat, *Sensors and Actuators B: Chemical*. 353.
- [24] Koh E.H., Lee W.C., Choi Y.J., Moon J.I., Jang J., Park S.G., Choo J., Kim D.H., Jung H.S., 2021, A wearable surface-enhanced Raman scattering sensors for label-free molecular detection, *ACS Applied Materials Interfaces*. 13 (2), 3024-3032.
- [25] Iannucci L., 2021, Chemometrics for Data Interpretation: Application of Principal Components Analysis (PCA) to Multivariate Spectroscopic Measurements, *IEEE Instrumentation & Measurement Magazine*. 24 (4), 42-48.

- [26] Vigneau E., Devaux M.F., Qannari E.M., Robert P., 1997, Principal component regression, ridge regression and ridge principal component regression in spectroscopy calibration, *Journal of Chemometrics*. 11, 239-249.
- [27] Eilers P.H.C., Boelens H.F.M., 2003, A perfect smoother, *Analytical Chem.* 75, 3631-3636.
- [28] Savitzky A., Golay M.J.E., 1964, Smoothing and differentiation of data by simplified least square producers, *Analytical Chem.* 36, 1627-1639.
- [29] Golparvar A., Boukhayma A., Loayza T., Caizzone A., Enz C., Carrara S., 2021, Very selective detection of low physiological glucose levels by spontaneous Raman spectroscopy with univariate data analysis, *BioNanoSci.* 11, 871-877.
- [30] Golparvar A., Kim J., Boukhayma A., Briand D., Carrara S., 2023, Higly accurate multimodal monitoring of lactate and urea in sweat by soft epidermal optofluidics with single-band Raman scattering, *Sensors and Actuators B: Chemical*. 387.



## Ringraziamenti

Questo lavoro è la conclusione di un lungo percorso e di una complessa parte della mia vita. Devo ringraziare in primis coloro che hanno permesso la sua realizzazione, in particolare il Professor Leonardo Iannucci, per l'aiuto nel lavoro pratico svolto in laboratorio e soprattutto nella disponibilità e cortesia di questi mesi anche a distanza.

Un ringraziamento alla Professoressa Sabrina Grassini, per la proposta di questo lavoro e per il supporto di alcune semplici parole, estremamente apprezzate.

Un doveroso ringraziamento anche alla Professoressa Leila Es Sebar, per il suo interesse verso questo lavoro e per avermi proposto punti di vista nuovi e precisi per l'utilizzo di una metodologia che non conoscevo approfonditamente, senza il suo contributo la realizzazione avrebbe richiesto molto più tempo.

Ringrazio anche il Dottor Ata Golparvar, per le esperte osservazioni e per i consigli in materia.

Grazie ai miei amici, in particolar modo ai miei compagni di studio del gruppo N.C., che hanno faticato con me per lo stesso scopo (e mannaggia solo in due mi avete aspettato). Anche voi siete parte fondamentale di questo momento, per l'aiuto che ci siamo sempre scambiati in ogni momento di questi cinque anni.

Un grazie speciale a Fabio, hai rappresentato il collegamento tra il mondo torinese e quello cuneese, in molte serate di discorsi, scleri e platini vari.

Un grande ringraziamento ad amiche ritrovate, Roberta e Chiara, con voi ho nuovamente un punto di riferimento anche a Borgo, che mi mancava davvero dopo l'inizio di questo percorso e che non so perché fosse venuto meno.

Infine, alla mia famiglia, in cui ognuno mi ha portato un pezzo diverso, mia sorella Maria il supporto sugli aspetti più specifici di quest'esperienza, mia madre e mio padre per le cose che non sempre vengono notate, ed incessantemente le fate comunque, e anche un po' a Nerone per due feste la sera al ritorno, che comunque fanno.

Grazie!