



UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI INGEGNERIA INDUSTRIALE

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

**MODELLAZIONE STATISTICA MULTIVARIATA DI UN
PROCESSO CONTINUO DI GRANULAZIONE
PER L'INDUSTRIA FARMACEUTICA**

Relatore: Prof. Fabrizio Bezzo
Correlatori: Ing. Emanuele Tomba
Dr. Simeone Zomer

Laureanda: MARIALUISA DE MARTIN

ANNO ACCADEMICO 2011 – 2012

Riassunto

In questa Tesi viene illustrato come metodi chemiometrici, quali l'analisi delle componenti principali (*principal component analysis*) e la proiezione su strutture latenti (*partial least squares regression*), possano essere utilizzati a supporto dello sviluppo di una linea industriale per la produzione in continuo di compresse a base di paracetamolo. La linea comprende una fase di granulazione e una successiva fase di compattazione del granulato.

I dati disponibili, relativi ad esperimenti effettuati sul processo industriale con materie prime di diverse caratteristiche e a diverse condizioni operative al granulatore e alla comprimitrice, sono stati analizzati al fine di comprendere le principali fonti di variabilità agenti sul sistema e aventi un effetto sulla qualità del prodotto finale. È stato quindi possibile ordinare queste forze motrici secondo la loro importanza, costituendo un valido strumento al fine di assicurare il monitoraggio delle operazioni, in modo da ottenere una qualità riproducibile del prodotto anche a fronte di possibili cambiamenti nelle caratteristiche del materiale iniziale o delle condizioni operative.

Nella Tesi vengono inoltre mostrate alcune potenziali applicazioni dei metodi chemiometrici nella realtà produttiva come strumento per il monitoraggio di processo, al fine di rilevare in tempo reale eventuali anomalie di marcia, o per ottenere stime in tempo reale delle caratteristiche di prodotti intermedi o della qualità del prodotto finale.

I risultati ottenuti dimostrano come le metodologie applicate possano essere agevolmente impiegate sia in attività di sviluppo per la definizione del *design space* del processo produttivo, sia in ambito produttivo per assicurare il controllo della qualità e il miglioramento continuo delle operazioni.

Nota.

La Tesi è stata svolta in collaborazione con l'azienda GlaxoSmithKline (GSK; Harlow, U.K.), che desidero ringraziare per la disponibilità mostrata durante lo sviluppo del progetto e per aver messo a disposizione i dati da analizzare e la conoscenza sul processo.

Indice

INTRODUZIONE.....	1
CAPITOLO 1 – Generalità.....	3
1.1 PROCESSO DI PRODUZIONE DI COMPRESSE.....	3
1.2 L’APPROCCIO “QUALITY BY DESIGN”.....	7
1.3 METODI STATISTICI MULTIVARIATI.....	8
1.4 PROCEDURE PRELIMINARI DI TRATTAMENTO DEI DATI.....	9
1.5 ANALISI DELLE COMPONENTI PRINCIPALI.....	9
1.5.1. Convalida incrociata.....	11
1.5.2. Srotolamento (unfolding).....	12
1.6. PROIEZIONI SU STRUTTURE LATENTI.....	13
1.6.1. Multiblock partial least squares.....	15
1.6.2. Limiti di controllo.....	16
CAPITOLO 2 – Il processo di granulazione e compattazione.....	19
2.1 STRUTTURA DEI DATI RELATIVI AL PARACETAMOLO.....	19
2.1.1 Condizioni iniziali.....	20
2.1.2 Misure on-line al granulatore.....	24
2.1.3 Analisi all’uscita del granulatore.....	27
2.1.4 Analisi all’uscita del mulino.....	28
2.1.5 Analisi alla comprimitrice.....	29
2.2 STRUTTURA DEI DATI RELATIVI A UN SECONDO GRANULATO.....	31
CAPITOLO 3 – Analisi statistica multivariata del processo.....	33
3.1 ANALISI DEI DATI RELATIVI AL PARACETAMOLO.....	33
3.1.1 Analisi delle condizioni iniziali (Z).....	35
3.1.2 Analisi delle misure al granulatore (X₁).....	36
3.1.3 Analisi dei dati relativi alla corrente all’uscita del mulino (Y₁).....	39
3.1.4 Analisi dei dati relativi alla fase di compattazione (Y₂).....	40
3.1.5 Analisi congiunta dei dati relativi alle condizioni iniziali, all’uscita del granulatore e all’uscita del mulino.....	44
3.1.5.1 Relazione tra due sistemi per la misura della distribuzione granulometrica.....	48

3.1.6 Analisi congiunta dei dati relativi alle condizioni iniziali, al granulatore, all'uscita del granulatore e all'uscita del mulino.....	50
3.1.7 Analisi congiunta dei dati relativi alle condizioni iniziali, all'uscita del granulatore, all'uscita del mulino e di compressione.....	51
3.2 MODELLI PER LA STIMA DELLE CARATTERISTICHE DEI GRANULI ALL'USCITA DEL MULINO.....	53
3.2.1 Analisi delle relazioni tra variabili operative.....	53
3.2.2 Analisi multiblock del processo.....	55
3.2.3 Determinazione dell'importanza di ciascun blocco nella predizione delle caratteristiche dei granuli all'uscita del granulatore.....	58
3.3 ANALISI DEI DATI RELATIVI AL SECONDO GRANULATO.....	61
CAPITOLO 4 – Sviluppo di sensori virtuali.....	65
4.1 STIMA DELLA RESISTENZA A TRAZIONE DELLE COMPRESSE DAL SET DI DATI DEL PARACETAMOLO.....	65
4.2 STIMA DELLA RESISTENZA A TRAZIONE DEL SECONDO GRANULATO.....	69
4.3 STIMA DELLE CARATTERISTICHE DEL GRANULATO ALL'USCITA DEL MULINO.....	72
4.3.1 Stime per il paracetamolo.....	72
4.3.2 Stime per il secondo granulato.....	72
CAPITOLO 5 –Monitoraggio del processo produttivo.....	75
5.1 INDIVIDUAZIONE DI ANOMALIE DURANTE LA GRANULAZIONE.....	75
5.1.1 Granulazione del lotto {2}.....	76
5.1.2 Granulazione del lotto {5}.....	78
5.2 INDIVIDUAZIONE DI ANOMALIE DURANTE LA FASE DI GRANULAZIONE E STIMA DELLE CARATTERISTICHE DEL PRODOTTO ALL'USCITA DEL MULINO.....	82
5.2.1 Granulazione del lotto {2}.....	83
5.2.2 Granulazione del lotto {5}.....	84
CONCLUSIONI.....	89
NOMENCLATURA.....	93
Appendice 1 – Rappresentazione delle matrici dei dati.....	97
Appendice 2 – Figure e codici di calcolo.....	101
RIFERIMENTI BIBLIOGRAFICI.....	105

Introduzione

La produzione farmaceutica è stata per decenni basata su processi discontinui (*batch*). Recentemente, l'interesse si sta però sempre più orientando verso processi continui, che consentono di collegare tra loro diverse operazioni unitarie (come reazione, filtrazione, distillazione, cristallizzazione, granulazione, essiccazione e macinazione) con il vantaggio di ridurre le dimensioni delle apparecchiature, diminuire gli scarti generati, ridurre l'energia consumata, minimizzare i problemi di gestione e sicurezza dell'impianto, e quindi ridurre l'impatto ambientale (Jiménez-González *et al.*, 2011). Secondo Woodcock (2011), nei prossimi venticinque anni il sistema produttivo farmaceutico cambierà in favore di processi di tipo continuo più puliti ("verdi"), flessibili ed efficienti rispetto a quelli discontinui.

Parallelamente, incentivata dalle linee guida proposte dagli enti regolatori (*Food and Drug Administration*, FDA e *European Medicines Agency*, EMA), l'industria farmaceutica sta attraversando una fase di cambiamento nel modo di affrontare le attività di sviluppo di nuovi prodotti e processi, volta a migliorare la comprensione del processo produttivo al fine di assicurare elevati standard di qualità dei prodotti. Secondo questo approccio (che viene definito *Quality-by-Design*, QbD), la qualità non va verificata sul prodotto ottenuto *alla fine* del processo produttivo, ma deve essere una conseguenza naturale del processo ossia automaticamente *incorporata all'interno (embedded)* dello stesso attraverso la comprensione delle relazioni tra le condizioni iniziali del materiale, quelle operative del processo e la qualità finale del prodotto. La combinazione *multidimensionale* di condizioni iniziali e parametri di processo, che porta alla qualità desiderata del prodotto finale, definisce lo spazio di progetto (*design space*) del processo (FDA, 2004 e 2006). Il QbD farmaceutico quindi è un approccio sistematico, scientifico e globale, allo sviluppo farmaceutico, la cui implementazione può essere riassunta nelle fasi seguenti (Yu, 2008):

- definizione del profilo di qualità del prodotto;
- progettazione e sviluppo del processo produttivo e del prodotto;
- identificazione delle proprietà critiche del prodotto, dei parametri operativi di processo e delle fonti di variabilità;
- monitoraggio continuo del processo per assicurare una qualità costante del prodotto finale.

In questa Tesi è dimostrato come metodi chemiometrici come la *principal component analysis* (PCA) e la *partial least square regression* (PLS) possono essere utilizzati per supportare lo sviluppo di processi farmaceutici continui, ottenendo una maggiore comprensione delle forze motrici del sistema e costituendo un valido strumento per la definizione del *design space* del processo.

In particolare, questo approccio è utilizzato a supporto dello sviluppo di una linea di produzione in continuo di compresse, mediante granulazione e successiva compattazione, nella quale l'interesse è capire come diverse condizioni iniziali (tecniche di isolamento o di additivazione del principio attivo) e parametri di processo (condizioni operative) agiscano sugli attributi di qualità del prodotto finale.

L'obiettivo di questa Tesi è pertanto quello di comprendere quali siano le principali fonti di variabilità che influenzano il processo, al fine di assicurarne il controllo e di ottimizzare la produzione, consentendo di ottenere una qualità costante del prodotto anche a fronte di eventuali cambiamenti del materiale iniziale o delle condizioni operative.

A questo proposito, sono utilizzati metodi di analisi statistica multivariata come PCA e PLS, che consentono di individuare e riordinare le fonti di variabilità del sistema e sono impiegati anche come strumento per monitorare il processo produttivo rivelandone in tempo reale eventuali anomalie. Inoltre, nella Tesi i metodi PLS sono utilizzati per ottenere una stima in tempo reale delle caratteristiche intermedie del materiale e della qualità finale del prodotto.

La Tesi è suddivisa in Cinque capitoli nei quali sono descritti il processo, le analisi statistiche effettuate e i risultati ottenuti. Nel primo Capitolo è descritto il processo di granulazione continua e la successiva compattazione, e sono fornite le generalità sui metodi statistici multivariati utilizzati. Nel secondo Capitolo sono elencati i dati raccolti ed è descritta la loro organizzazione in matrici. Nel terzo Capitolo sono presentati i risultati dell'analisi statistica multivariata, identificando e ordinando le principali forze motrici del processo. Nel quarto Capitolo è descritto lo sviluppo di sensori virtuali per il monitoraggio della fase di compattazione. Infine, nel quinto Capitolo i modelli creati sono impiegati per il monitoraggio del processo produttivo, permettendo di rilevare in tempo reale anomalie di funzionamento, e favorendo quindi il controllo della qualità del prodotto. Una sezione finale descrive le principali conclusioni dello studio.

Capitolo 1

Generalità

1.1. Processo di produzione di compresse

L'industria farmaceutica per decenni si è dedicata alla produzione di farmaci mediante processi discontinui (*batch*) cioè processi di durata finita caratterizzati da una fase di carico, una di reazione, e una di scarico. Tale approccio è stato abbandonato da altre industrie (come quella petrolchimica, chimica e per certi versi anche alimentare) che hanno sviluppato un approccio continuo, riducendo i costi e l'impatto ambientale e migliorando la qualità del prodotto. I motivi principali del passaggio da un approccio *batch* ad un approccio continuo sono la mancanza di scalabilità dell'approccio discontinuo alla crescita del processo, il desiderio di minimizzare le dimensioni dell'impianto e di aumentarne l'efficienza. Recentemente, anche l'industria farmaceutica ha iniziato a rivedere i propri processi produttivi impiegando operazioni unitarie in continuo per sfruttare l'efficienza di questo approccio.

La produzione o di un principio attivo o di un farmaco, prevedono una serie di fasi (*step*) e quindi di operazioni unitarie, come ad esempio reazione, filtrazione, distillazione, cristallizzazione, granulazione, essiccazione e macinazione, che possono essere svolte in continuo con minori problemi di gestione. Nell'industria farmaceutica le compresse sono la forma solida di dosaggio più popolare e maggiormente utilizzata, portando ad un sempre maggiore interesse per l'ingegnerizzazione delle particelle e delle polveri farmaceutiche. La progettazione delle particelle prevede quindi di migliorare l'efficienza dei sistemi produttivi e di dare una maggiore funzionalità al farmaco e agli eccipienti.

I processi che portano all'aumento delle dimensioni delle particelle (*size enlargement*), sono stati definiti come trasformazioni nelle quali piccole particelle vengono agglomerate, compattate o comunque unite in particelle più grandi, di massa relativamente permanente, nella quale le particelle iniziali possono essere distinte (Perry e Green, 1997). Il processo di *size enlargement* permette di migliorare le proprietà di scorrimento, di creare miscele di polveri non segreganti, di controllare la porosità e il rapporto superficie-volume, e di migliorare l'aspetto del prodotto. L'agglomerazione può avvenire mediante diverse operazioni unitarie, come agitazione e miscelazione. L'insieme di tutte le operazioni unitarie che porta alla formazione degli agglomerati è chiamato processo di granulazione.

Il processo di granulazione può avvenire a secco (cioè senza l'ausilio di solventi, con minori problemi di alterazioni chimico fisiche), o a umido (cioè con una fase liquida e successivo

essiccamento del prodotto). Le operazioni che costituiscono il processo di granulazione a umido e successiva compattazione possono essere così riassunte (Figura 1.1):

- pretrattamento per uniformare le caratteristiche dei materiali di partenza attraverso setacciatura e macinazione;
- miscelazione delle polveri;
- miscelazione dei campioni solidi con il fluido di granulazione (“bagnatura”);
- formazione di granuli;
- essiccamento a temperatura controllata per evitare processi degradativi;
- setacciatura e, talvolta, ulteriore macinazione;
- compattazione;
- riciclo del materiale non utilizzato.

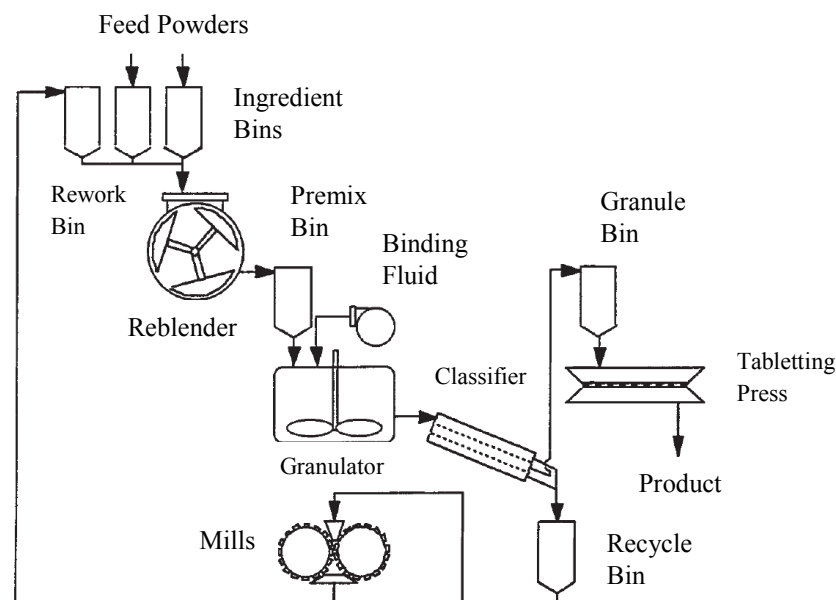


Figura 1.1 Esempio di processo di granulazione a umido e successiva compattazione (Perry e Green, 1997; p. 20-20)

La granulazione è suddivisa in quattro fasi: miscelazione solido-liquido, coalescenza o crescita, consolidamento e rottura (Figura 1.2). La miscelazione solido-liquido è fortemente influenzata dalla modalità con la quale il liquido viene distribuito per favorire una maggiore contatto. Nella fase di coalescenza e accrescimento, le particelle, parzialmente bagnatesi, fondono per formare granuli di dimensioni maggiori. Segue una fase di consolidamento nella quale i nuclei accrescono e aumentano le forze di aggregazione. Questa fase è molto importante perché influenza la porosità del granulo nonché la resistenza meccanica e la dispersione del prodotto finale. Alla fine ci può essere una fase di rottura se i granuli formati hanno dei difetti o le forze interne sono deboli. Le proprietà finali dell'agglomerato dipendono dalle dimensioni e dalla porosità, che possono essere modificate aggiustando le variabili operative o la formulazione iniziale. La scelta delle variabili operative, del sistema di

agglomerazione e delle sezioni secondarie, sommata alla scelta del materiale e quindi degli ingredienti piuttosto che della viscosità, della tensione superficiale, e della distribuzione, definiscono il meccanismo cinetico di miscelazione solido-liquido, coalescenza, consolidamento e rottura. Pertanto, se ci sono dei problemi connessi a una di queste fasi della granulazione, è necessario modificare la struttura del processo o la formulazione del prodotto (Perry e Green, 1997).

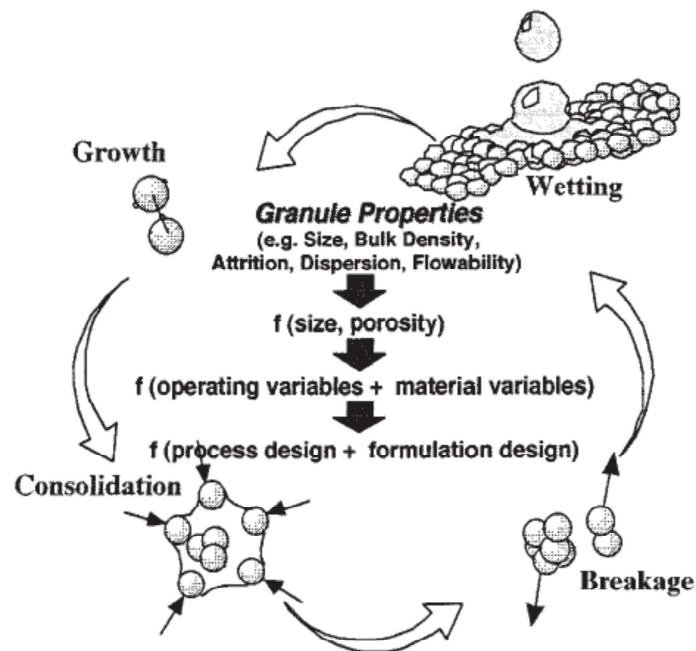


Figura 1.2 Fasi differenti della granulazione (Perry e Green, 1997, p. 20-20)

Il processo studiato in questa tesi è una linea produttiva di compresse mediante granulazione e successiva compattazione (*tableting*). L'azienda farmaceutica GlaxoSmithKline (GSK) sta sviluppando processi di *tableting* continuo per capire quali possono essere gli sviluppi e i vantaggi commerciali rispetto all'approccio *batch*. L'operazione in continuo permette di integrare tutte le operazioni unitarie, dalle materie prime al prodotto finale, in un'unica linea produttiva per ottimizzare il processo, mantenendo la qualità del prodotto finale. Per raggiungere lo scopo, è necessario capire quali sono le relazioni tra le proprietà delle materie prime, le variabili di processo e le caratteristiche del prodotto finale.

Una delle unità più importanti del processo produttivo esaminato, che porta alla formazione di compresse, è la granulazione in continuo, ottenuta impiegando un granulatore a doppia vite rotante (Parikh, 1997). GSK ha effettuato alcuni studi sperimentali per sviluppare tale tecnologia da un punto di vista meccanico e per spiegare problematiche relative a composti sviluppati per studi clinici. Tuttavia rimangono oggetto di ricerca alcuni aspetti della conoscenza scientifica quantitativa e qualitativa del processo, le complesse interazioni tra i parametri e lo sviluppo di un sistema di controllo.

La sezione Particle Generation Control and Engineering (PGCE) ha condotto degli esperimenti per studiare come la variazione di modalità di formulazione e di metodo di miscelazione tra principio attivo (*active pharmaceutical ingredient*, API) e additivi possa influire su varie fasi del processo come la granulazione in continuo, l'essiccamento, la macinazione e la compressione. La granulazione è stata eseguita a umido e necessita una fase di essiccamento per far evaporare il solvente, che può portare all'aumento delle porosità e della fragilità del materiale condizionando le fasi successive e la qualità del prodotto finale (Appelgren e Eskilson, 1990; Bergman, *et al.*, 1998).

Il processo in esame è schematizzato in Figura 1.3 per aumentare la comprensione delle unità e indicare in quali sezioni siano state effettuate le analisi. Le correnti sono state numerate per facilitare la descrizione del processo:

- corrente 1: ingresso al processo del materiale pretrattato;
- corrente 2: ingresso del materiale solido al granulatore;
- corrente 3: ingresso della corrente liquida al granulatore;
- corrente 4: uscita dal granulatore;
- corrente 5: uscita dall'essiccatore;
- corrente 6: uscita dal mulino;
- corrente 7: uscita dalla comprimitrice.

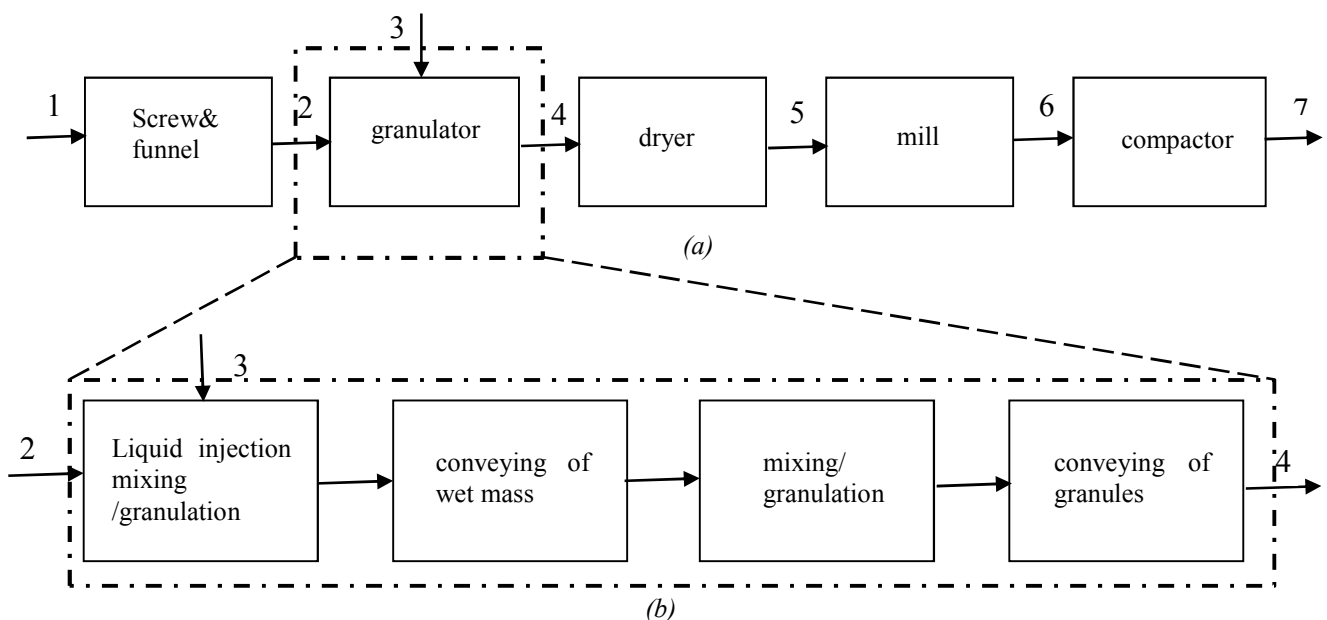


Figura 1.3 (a) Diagramma a blocchi del processo di granulazione in continuo del paracetamolo. (b) Dettaglio del granulatore.

Il processo GSK che porta alla formazione di compresse a base di paracetamolo può essere suddiviso nelle cinque fasi illustrate in Figura 1.3a:

- alimentazione dei campioni in polvere (corrente 1) mediante una coclea a doppia vite rotante; movimentazione della miscela al granulatore (corrente 2) per gravità grazie ad una tramoggia;
- miscelazione della polvere nel granulatore con diverse quantità di acqua purificata *granulating liquid* (corrente 3);
- essiccamento dei campioni in uscita mediante Aeromatic Strea-1 ($T_{in} = 60^{\circ}\text{C}$; $T_{out} = 35^{\circ}\text{C}$);
- macinazione del prodotto mediante un mulino Quadro CoMill 197 con “frusta” rotonda (un campione non è stato macinato);
- compattazione mediante comprimitrice munita di punzone.

Il processo è stato monitorato attraverso misure delle proprietà di flusso, misure di distribuzione delle particelle, misure on-line al granulatore e misure di compressione. Le condizioni iniziali sono state inserite nell’analisi perché hanno un ruolo significativo sul processo e sulle proprietà finali del prodotto (Soh *et al.*, 2008).

1.2. L’approccio “Quality-by-Design”

Per migliorare il ciclo produttivo di un farmaco rendendo il processo più rapido minimizzando i costi è necessario applicare in modo sistematico i metodi *Process Systems Engineering* (PSE). In questo contesto, la raccomandazione da parte di FDA (2004) ad impiegare *Process Analytical Technology* (PAT) ha chiarito la posizione degli enti regolatori in favore di metodi produttivi più efficienti che garantiscano elevati standard di qualità del prodotto finale. La guida PAT presentata da FDA (2004) promuove la ricerca di un *design space*, cioè di quella combinazione multidimensionale di condizioni iniziali e parametri operativi di processo che permette di ottenere le caratteristiche desiderate del prodotto finale. Cambiamenti al processo, che muovono lo stesso all’interno del *design space* non necessitano di essere sottoposti ad autorizzazione da parte degli organi regolatori. La guida PAT promuove inoltre la ricerca di condizioni operative “critiche”, cioè variabili che hanno un’impatto significativo sulla qualità “critica” e quindi devono essere monitorate e controllate per assicurare la qualità desiderata. Una *Critical Quality Attribute* (CQA) viene definita come una caratteristica del prodotto intermedio che deve essere all’interno di appropriati limiti per assicurare la qualità desiderata del prodotto finale.

L’analisi statistica multivariata dei processi farmaceutici permette di comprendere la combinazione di proprietà misurabili del prodotto intermedio o finale che vengono considerate “critiche” per stabilire la purezza, efficienza e la sicurezza desiderate. Gli strumenti necessari per raggiungere tali obiettivi sono:

- l’acquisizione di dati multivariati;
- analisi di processo mediante metodi chemiometrici;

- sviluppo di sistemi di controllo di processo;
- uso di strumenti per il miglioramento continuo e la gestione delle conoscenze.

PSE ha un ruolo importante nella definizione del *design space* per comprendere processi farmaceutici continui e nell'uso di PAT per il monitoraggio e controllo di processo (Figura 1.4). I tre problemi chiave di determinazione del *design space*, progettazione del processo produttivo e monitoraggio dello stesso devono essere affrontati in modo sinergico usando modelli meccanicistici o basati su dati per ottenere la comprensione del processo e un agevole trasferimento di informazioni tra le diverse scale di processo (laboratorio, impianto pilota, produzione) (Gernaey *et al.*, 2012).

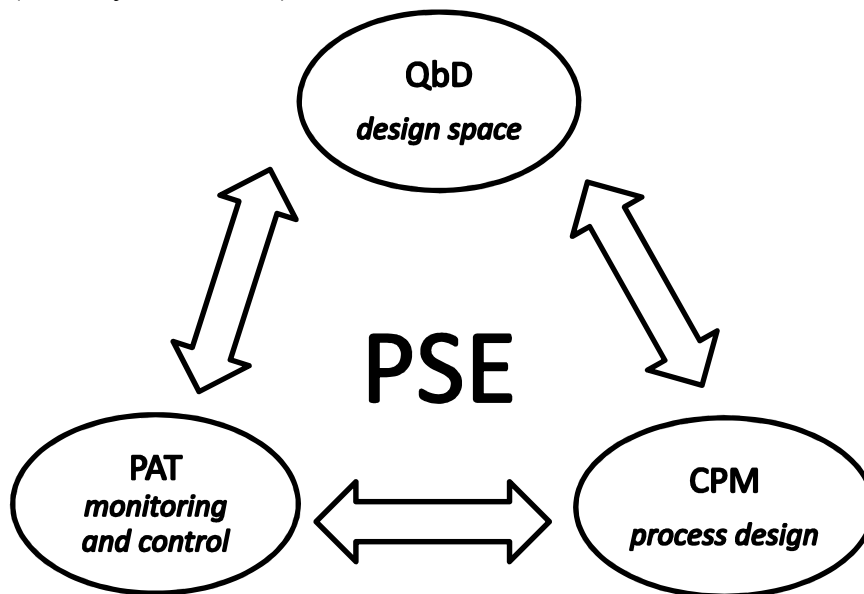


Figura 1.4 Il ruolo centrale della PSE nell'affrontare in modo sinergico i tre problemi chiave dell'industria farmaceutica (Gernaey *et al.*, 2012).

1.3. Metodi statistici multivariati

Nell'industria chimica, ogni processo viene monitorato attraverso la raccolta in poco tempo di moltissimi dati relativi a variabili diverse che rendono difficile il controllo e quindi il raggiungimento della qualità del prodotto desiderata. Con il termine *analisi multivariata* si indica quell'insieme di metodi statistici utilizzati per analizzare simultaneamente più fonti di variabilità. L'esistenza di molte variabili interagenti l'una con l'altra complica il sistema rispetto al caso monovariato, ma ridurre la dimensione eliminando alcune variabili potrebbe portare alla perdita di informazione preziosa per l'analisi del processo (Facco, 2005).

1.4. Procedure preliminari di trattamento dei dati

I dati raccolti durante il monitoraggio del processo hanno unità di misura diverse e possono avere pesi diversi. Per questo motivo, è necessario uniformare i dati e facilitare i calcoli successivi mediante il bilanciamento al valor medio (*mean centering*) e la riduzione della scala (*scaling*).

Secondo Geladi e Kowalski (1986) ci sono tre modi per trattare le variabili in questa fase:

- non scalare i dati quando le variabili hanno la stessa unità di misura;
- applicare una riduzione di scala basata sulla varianza (*variance scaling*) se all'interno di ciascun blocco di dati ci sono variabili di unità di misura diverse in modo tale che tutte abbiano varianza unitaria;
- decidere che alcune variabili sono meno importanti di altre e non influenzano il modello, e da dare loro un peso inferiore.

Il secondo metodo è il più utilizzato e associa la centratura rispetto alla media con la riduzione di scala sulla varianza e viene comunemente detto *autoscaling*.

Durante il processo vengono acquisite delle misure che sono state raccolte in una generica matrice \mathbf{X} del processo di dimensioni $(I \times J)$ dove I è il numero di campioni processati, J è il numero di variabili misurate e \mathbf{x}_i è il vettore della riga i -esima di \mathbf{X} corrispondente alle variabili misurate per campione i -esimo.

L'operazione di *mean centering* viene effettuata calcolando la media per ogni variabile del set di calibrazione:

$$\bar{x}_j = \frac{\sum_{i=1}^I x_{i,j}}{I} \quad (1.1)$$

dove $x_{i,j}$ è l'elemento della matrice \mathbf{X} $(I \times J)$ e sottraendo a ciascun elemento la propria media. In questo modo è possibile confrontare agevolmente le fonti di variabilità senza che ci siano effetti sulle inter-correlazioni fra le variabili (Kourti, 2003).

L'operazione di *scaling* prevede di dividere i valori di una variabile per la loro deviazione standard, formule (1.2) e (1.3), in modo che ciascuna variabile abbia varianza unitaria.

$$var(\mathbf{x}_j) = \frac{\sum_{i=1}^I (x_{i,j} - \bar{x}_j)^2}{I} \quad , \quad (1.2)$$

$$\sigma_{std} = \sqrt{var(\mathbf{x}_j)} \quad . \quad (1.3)$$

1.5. Analisi delle componenti principali

L'analisi delle componenti principali (PCA, *principal component analysis*) è uno degli strumenti più efficaci in chemometria per ridurre la dimensione dei dati e facilitare la

comprensione del processo (Jackson, 1991).

L'analisi delle componenti principali consiste nella decomposizione delle variabili di processo in autovalori della matrice di covarianza delle misure, chiamata anche matrice di correlazione se \mathbf{X} è stata autoscalata (Wise e Gallagher, 1996):

$$\text{cov}(\mathbf{X}) = \frac{\mathbf{X}'\mathbf{X}}{I-1} \quad (1.4)$$

La PCA decompone la matrice \mathbf{X} nella somma del prodotto tra i vettori \mathbf{t}_m e \mathbf{p}_m più una matrice dei residui \mathbf{E} :

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}'_1 + \mathbf{t}_2\mathbf{p}'_2 + \dots + \mathbf{t}_A\mathbf{p}'_A + \mathbf{E} \quad (1.5)$$

dove k è inferiore o uguale alla dimensione minima di \mathbf{X} , $A \leq \{i, j\}$.

I vettori \mathbf{t}_m sono gli *scores* e contengono l'informazione relativa alle relazioni tra i campioni mentre \mathbf{p}_m sono i vettori contenenti gli autovettori della matrice di covarianza:

$$\text{cov}(\mathbf{X})\mathbf{p}_m = \lambda_m\mathbf{p}_m \quad (1.6)$$

dove λ_m è l'autovalore associato all'autovettore \mathbf{p}_m . I vettori \mathbf{p}_m sono i *loadings* e contengono l'informazione relativa alle relazioni tra le variabili. Le caratteristiche di questi vettori sono:

- l'ortogonalità degli *scores*, $\mathbf{t}'_m\mathbf{t}_m = 0$ per $m \neq h$ (1.7)
- l'ortonormalità dei *loadings*,

$$\mathbf{p}'_m\mathbf{p}_h = 0 \text{ for } m \neq h \quad (1.8)$$

$$\mathbf{p}'_m\mathbf{p}_h = 1 \text{ for } m = h \quad (1.9)$$

- gli *scores* sono combinazioni lineari dei dati originali definiti da \mathbf{p}_m :

$$\mathbf{X}\mathbf{p}_m = \mathbf{t}_m \quad (1.10)$$

Le coppie $(\mathbf{t}_m, \mathbf{p}_m)$ sono disposte secondo l'ordine decrescente degli autovalori corrispondenti, che sono misure della varianza spiegata dalla coppia $(\mathbf{t}_m, \mathbf{p}_m)$ cioè della quantità di informazione contenuta nel modello. Con questa struttura la prima coppia $(\mathbf{t}_m, \mathbf{p}_m)$ cattura la più grande quantità di varianza/informazione rispetto a tutte le altre coppie.

I vettori degli *scores* vengono raggruppati nella matrice \mathbf{T} di dimensioni $(I \times A)$, dove I è il numero di campioni e A il numero di componenti principali che descrivono il sistema, mentre i vettori dei *loadings* vengono raggruppati nella matrice \mathbf{P} di dimensioni $(J \times A)$, dove J è il numero di variabili originarie e A il numero di componenti principali. La prima componente principale è una combinazione lineare di J variabili misurate, dà la direzione di massima variabilità dei dati ed è ortogonale alle altre componenti principali.

La Figura 1.5 rappresenta lo spazio tridimensionale dove si dispongono i dati originali che vengono descritti da due componenti principali del modello. L'obiettivo dell'analisi è proprio quello di ridurre le dimensioni del sistema in modo tale che poche componenti principali descrivano lo spazio dei dati. Quando il sistema possiede un numero elevato di variabili altamente correlate si può comunque spiegare la maggior parte della varianza della matrice \mathbf{X} impiegando solamente le prime componenti. Ciò che non viene spiegato dall'iperpiano di componenti principali scelte sarà racchiuso nella matrice dei residui \mathbf{E} della (1.5).

Tale matrice deve essere minimizzata in modo tale da non perdere informazioni rilevanti per il processo e graficamente è proporzionale alla distanza euclidea delle osservazioni dal piano delle variabili latenti.

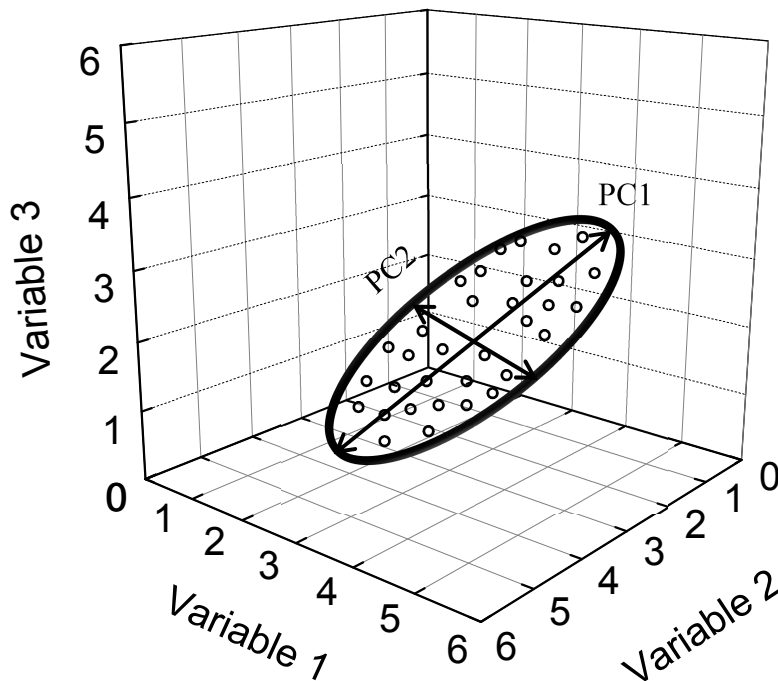


Figura 1.5 Modello PCA di un dataset tridimensionale che giace su un unico piano.

1.5.1. Convalida incrociata

Per la selezione della dimensione dello spazio delle componenti principali ci sono molti metodi:

- scegliere il numero di PC in base alla varianza spiegata complessiva, stabilendo una soglia oltre la quale la varianza associata alla PC è tanto piccola da poter essere trascurata;
- usare l'autovalore associato a ciascuna PC, scartando le PC con λ inferiore a 1;
- usare la convalida incrociata riportata da Jackson (1991, p.353).

La convalida incrociata (*cross-validation*) consiste nel suddividere la matrice dei dati in segmenti, costruire un modello PCA per ciascun segmento e proiettare ogni volta sull'iperpiano descritto dal modello il *set* di dati non utilizzato. L'errore di ricostruzione viene diagrammato in funzione delle componenti principali utilizzate per la costruzione del modello in termini di RMSECV, (*root mean square error in cross validation*) che dipende dal valore di PRESS (*prediction error of sum of squares*) cioè l'errore di predizione sulla somma dei quadrati:

$$RMSECV_j = \sqrt{\frac{PRESS_j}{I}}, \quad (1.11)$$

$$PRESS_j = \sum_{i=1}^I (x_{i,j} - \hat{x}_{i,j})^2 \quad . \quad (1.12)$$

L'aggiunta di nuove componenti principali al modello dovrebbe far decrescere l'errore finché esse non descriveranno solo rumore, facendolo aumentare. Quindi ricercando il minimo del RMSECV sarà possibile ottenere il numero ottimale di variabili latenti da utilizzare nel modello.

1.5.2. Srotolamento (Unfolding)

Il monitoraggio di un processo produttivo prevede l'acquisizione di dati in modo sequenziale durante le varie fasi di lavorazione. I dati sperimentali sono disposti in una matrice tridimensionale ($I \times J \times S$), rappresentata in Figura 1.6, dove I rappresenta il numero dei lotti, J il numero delle variabili e S gli istanti temporali. Ciò significa che ogni sezione orizzontale è una matrice ($J \times S$) che rappresenta i profili temporali di tutte le variabili per un singolo lotto i di produzione, mentre le sezioni verticali ($I \times J$) rappresentano i valori misurati di ciascuna variabile in un istante di tempo s .

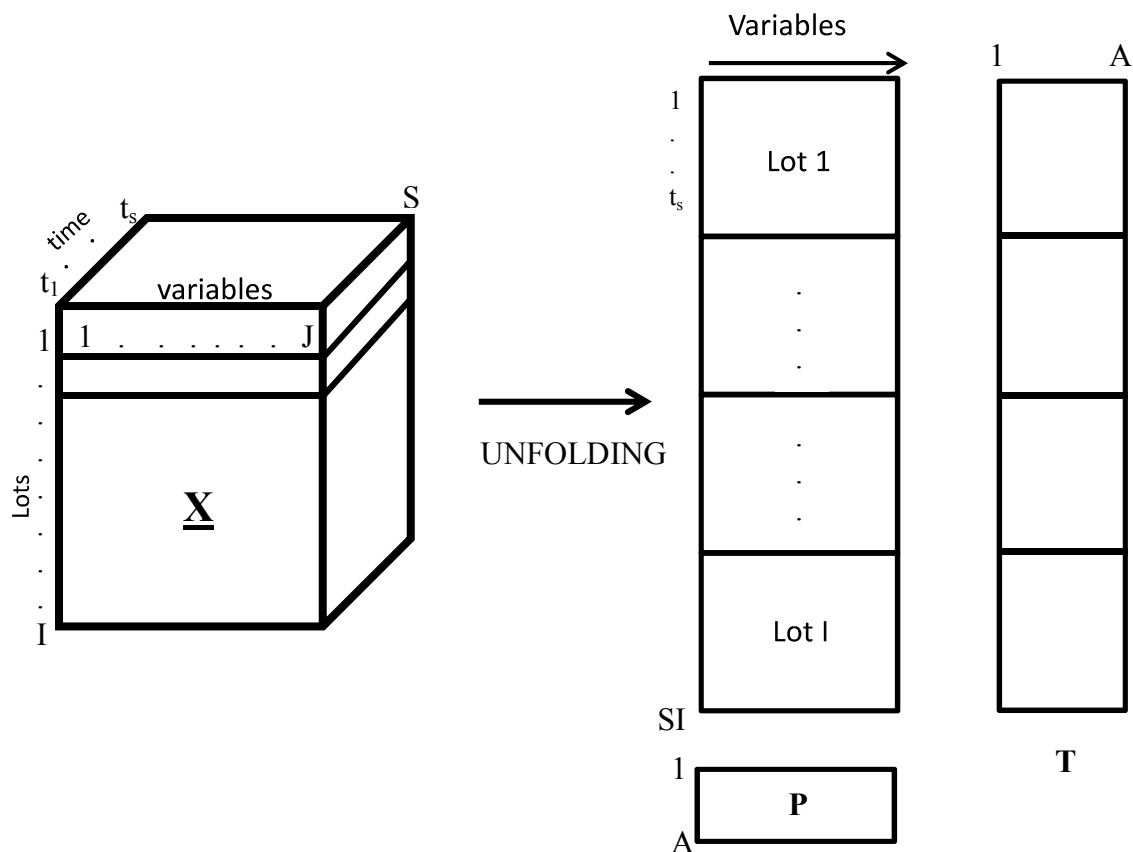


Figura 1.6 Riarrangiamento della struttura della matrice \underline{X} da tridimensionale a bidimensionale, mediante variable wise unfolding.

Per tener conto dell'andamento temporale delle variabili è possibile adattare PCA ad una matrice bidimensionale "riarrangiata" mediante srotolamento (*unfolding*). Ci sono due modi differenti di riarrangiare la matrice tridimensionale, che corrispondono a modi differenti di osservare la variabilità del sistema (Kourti, 2003).

Un processo continuo può essere analizzato mediante PCA utilizzando uno srotolamento nel senso delle variabili per poter capire le somiglianze tra vari lotti di materiale durante la lavorazione. Questa procedura consiste nel disporre ogni sezione ($J \times S$) della matrice \mathbf{X} sotto quella precedente per formare una matrice bidimensionale ($IS \times J$) in cui le prime s righe riguardano il primo lotto processato, come si vede in Figura 1.6.

1.6. Proiezione su strutture latenti

Nell'analisi di un processo è possibile avere a disposizione *set* di variabili misurate in linea in punti diversi dell'impianto. Il metodo della proiezione su strutture latenti (PLS, *partial least squares regression*) cerca di correlare due o più *set* di dati massimizzando la covarianza tra le matrici, non solo cercando le direzioni di massima varianza del processo ma usando le variabili indipendenti, come le condizioni iniziali o le variabili di processo, per predire quelle dipendenti, come la qualità del prodotto.

Tale metodo permette non solo di ridurre le dimensioni delle matrici, del processo \mathbf{X} e della qualità del prodotto \mathbf{Y} , ma anche di trovare lo spazio di \mathbf{X} più predittivo per \mathbf{Y} .

La matrice \mathbf{Y} è una matrice bidimensionale che contiene le caratteristiche del prodotto e ha dimensione ($I \times L$) dove I è il numero di campioni processati e L è il numero di variabili misurate che definiscono il prodotto.

Il modello PLS è composto da due relazioni, una esterna (relativa alle matrici prese singolarmente) e una interna (di collegamento tra le matrici). Le relazioni esterne per la matrice di processo \mathbf{X} e per la matrice delle variabili di qualità \mathbf{Y} sono uguali a quella per la decomposizione mediante PCA:

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E} = \sum_{a=1}^A t_a p'_a + \mathbf{E} \quad (1.13)$$

$$\mathbf{Y} = \mathbf{UQ}' + \mathbf{F} = \sum_{a=1}^A u_a q'_a + \mathbf{F}. \quad (1.14)$$

Il modello impone di minimizzare $\|\mathbf{E}\|$ e $\|\mathbf{F}\|$.

La relazione interna può essere costruita guardando i grafici degli *score* della matrice \mathbf{X} , t e della matrice \mathbf{Y} , u , ed è semplicemente una relazione lineare pari a:

$$u_m = b_m t_m \quad (1.15)$$

in cui i coefficienti di regressione sono:

$$b_m = \frac{u'_m t_m}{t'_m t_m}. \quad (1.16)$$

Il modello creato calcolando le componenti principali separatamente crea una debole relazione tra le matrici mentre sarebbe meglio dare a ciascuna le informazioni dell'altra per ruotare le componenti lungo la linea di regressione. Generalmente viene utilizzato un algoritmo di NIPALS nel quale *score* e *loading* sono calcolati utilizzando una serie di "pesi", w_m che servono a mantenere gli *score* ortogonali. Con questo metodo l'inversa di \mathbf{X} risulta essere:

$$\mathbf{X}^{-1} = \mathbf{W}(\mathbf{P}'\mathbf{W})^{-1}(\mathbf{T}'\mathbf{T})^{-1}\mathbf{T} , \quad (1.17)$$

dove gli *score* e i *loading* sono ruotati rispetto a quelli calcolati con PCA per predire meglio le variabili di qualità del prodotto. Per utilizzare il modello PLS è utile conoscere le proprietà di *score*, *loading* e pesi:

- i *loading* p'_m e q'_m hanno lunghezza unitaria $\|p'_m\| = \|q'_m\| = 1$;
- gli *score* sono centrati sullo zero e ortogonali;
- i pesi sono ortonormali $w'_i w_j = \delta_{ij} \|w'_i\|^2$ dove δ_{ij} è il delta di Kronecker.

Ogni modello di regressione può essere utilizzato per predire la matrice dipendente (cioè quella relativa alla qualità del prodotto) da quella indipendente (cioè quella relativa al processo). Questo può essere fatto decomponendo la matrice \mathbf{X} e ricostruendo la matrice \mathbf{Y} . Dopo aver costruito il modello PLS dallo storico del processo, viene utilizzata una nuova matrice \mathbf{X} con un numero di campioni diverso da quello utilizzato nella matrice precedente. La procedura prevede di decomporre \mathbf{X} :

$$\hat{\mathbf{t}}_m = \mathbf{E}_{m-1} \mathbf{w}_h \quad (1.18)$$

$$\mathbf{E}_m = \mathbf{E}_{m-1} - \hat{\mathbf{t}}_m \mathbf{p}'_m \quad (1.19)$$

e di ricostruire \mathbf{Y} :

$$\mathbf{Y} = \mathbf{F}_m = \sum_{m=1}^M b_m \hat{\mathbf{t}}_m \mathbf{q}'_m . \quad (1.20)$$

Per valutare le prestazioni di un modello nella stima della matrice \mathbf{Y} è importante valutare la varianza nella fase di calibrazione del modello (SS_c) e in quella di convalida incrociata (SS_{cv}), e disponendo di un nuovo dataset è possibile valutare la varianza in predizione (SS_p). Il calcolo della varianza della matrice \mathbf{Y} può essere effettuato con la relazione seguente:

$$SS = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (1.21)$$

dove y_i è la generica riga i -esima della matrice \mathbf{Y} , \hat{y}_i è la ricostruzione della riga i -esima e \bar{y} è il valore medio.

Inoltre è importante valutare l'errore commesso in termini di RMSE (*root mean square error*) in calibrazione, in convalida incrociata e in predizione.

$$RMSE = \sqrt{\frac{\sum_{i=1}^I (\hat{y}_i - y_i)^2}{n}} \quad (1.21)$$

dove y_i è la generica riga i -esima della matrice \mathbf{Y} , \hat{y}_i è la ricostruzione della riga i -esima.

Dopo aver costruito il modello è possibile calcolare l'importanza di ciascuna variabile indipendente nella proiezione delle variabili di qualità del prodotto finale (Chong e Jun, 2005). Il VIP (*variable importance in projection*) viene calcolato per ciascuna variabile di qualità e indica quali variabili di processo sono maggiormente correlate in predizione alla j -esima variabile di qualità:

$$VIP_j = \sqrt{\frac{p \sum_{a=1}^A (SS(b_a \mathbf{t}_a) (\mathbf{w}_{jk} / \|\mathbf{w}_k\|)^2)}{\sum_{a=1}^A (SS(b_a \mathbf{t}_a))}} \quad (1.22)$$

$$\text{dove } SS(b_a \mathbf{t}_a) = b_a^2 \mathbf{t}_a' \mathbf{t}_a, \quad (1.23)$$

p è il numero di variabili predittrici e \mathbf{w} è il peso assegnato a ciascuna variabile di processo.

1.6.1. *Multiblock partial least squares*

Generalmente un processo produttivo è suddiviso in varie sezioni che vengono monitorate separatamente ottenendo più set di dati relativi a ciascuno *step*. Il processo può essere valutato mediante l'analisi dei vari dataset separatamente, per comprendere la struttura di correlazione in ciascun blocco, o tutti insieme per avere una visione globale. Il modello *multiblock partial least squares* (MBPLS) (Kourti e MacGregor, 1994) è utile nell'analisi globale del sistema per poter combinare l'informazione relativa alle condizioni iniziali con quella del processo senza perdere la struttura del sistema definita dai vari blocchi. Infatti la matrice \mathbf{X} è composta da tutti i blocchi relativi ai vari step del processo e viene utilizzata nel modello per predire i dati relativi alla qualità del prodotto finale, matrice \mathbf{Y} . Prima di procedere con la costruzione del modello MBPLS ad ogni blocco della matrice \mathbf{X} viene dato un peso per poter suddividere le informazioni:

$$\mathbf{X} = [\mathbf{X}_1 / m_{X_1}^{1/2}, \dots, \mathbf{X}_B / m_{X_B}^{1/2}] \quad (1.23)$$

Inoltre è possibile creare, per ogni blocco, un sottospazio degli *score* che identifica l'influenza di ciascuno *step* sulla qualità del prodotto finale:

$$\mathbf{w}_b = \frac{\mathbf{X}_b' \mathbf{u}}{\mathbf{u}' \mathbf{u}} \quad (1.24)$$

$$\mathbf{t}_b = \frac{\mathbf{X}_b \mathbf{w}_b'}{m_{X_b}^{1/2}} \quad (1.25)$$

$$\mathbf{T} = [\mathbf{t}_1, \dots, \mathbf{t}_B] \quad (1.26)$$

$$\mathbf{w}_T = \frac{\mathbf{T}'\mathbf{u}}{\mathbf{u}'\mathbf{u}} \quad (1.27)$$

La formula (1.26) permette di calcolare il pesi di ciascun blocco, per variabile, nella rotazione dello spazio di \mathbf{X} per massimizzare la correlazione in predizione, mentre (1.28) definisce un indice sintetico che rappresenta il peso di ciascun blocco nella rotazione dello spazio chiamato anche *super weight*.

Per comprendere l'importanza di ciascun blocco nella predizione della qualità in un modello MBPLS è possibile utilizzare la seguente formula:

$$BIP_b = \sqrt{\frac{b \sum_{a=1}^A \left(SS_y SS_x \left(\mathbf{w}_T \mathbf{T}_b / \|\mathbf{w}_T \mathbf{T}_b\| \right)^2 \right)}{\sum_{a=1}^A (SS_y \times SS_x)}} \quad (1.29)$$

dove SS_y è la varianza associata alla matrice di qualità, SS_{xb} è la varianza associata a ciascun blocco della matrice \mathbf{X} , \mathbf{w}_T sono i *super weight*, B è il numero dei blocchi predittori e A il numero di variabili latenti utilizzate nella costruzione del modello e b è il generico blocco.

Questa formula è stata ricavata dall'idea di Yacoub (2011) di utilizzare un indice che tenga conto della varianza associata alla qualità e dei pesi per blocco, *super weight*, per determinare l'importanza di ciascun blocco della matrice predittrice nella predizione della qualità.

1.6.2. Limiti di controllo

È possibile quantificare la distanza di un campione dalla media del sistema e quindi quanto la sua proiezione si distanzia dall'origine nello spazio descritto dalle componenti principali utilizzando la somma degli *score* normalizzati al quadrato chiamata Hotelling T^2 :

$$T_i^2 = \mathbf{t}_i \boldsymbol{\lambda}^{-1} \mathbf{t}_i' = \mathbf{x}_i \mathbf{P} \boldsymbol{\lambda}^{-1} \mathbf{P}' \mathbf{x}_i' \quad (1.30)$$

dove \mathbf{t}_i si riferisce alla i -esima riga di \mathbf{T}_a , la matrice contenente gli A vettori degli *scores* del modello. La matrice $\boldsymbol{\lambda}^{-1}$ è una matrice diagonale contenente l'inverso degli autovalori corrispondenti agli autovettori a -esimi considerati nel modello.

Wise e Gallagher (1996) affermano che gli *scores* delle variabili latenti siano più normalmente distribuiti delle variabili originarie. Solitamente i modelli statistici multivariati vengono creati sul campionamento di una grande quantità di variabili tale da poter assumere che i campioni abbiano distribuzione normale. Dopo aver creato il modello è necessario calcolare i limiti di confidenza del sistema cioè i limiti che il modello ha nella rappresentazione di nuovi campioni.

I limiti di confidenza per il T^2 definiscono un'ellisse nel piano sul quale è proiettato il punto di funzionamento normale. Secondo Jackson (1991, p.343-344) in una rappresentazione bidimensionale del diagramma degli *score*, i semiassi maggiore s_M e minore s_m dell'ellisse di confidenza sono calcolate con:

$$s_M = \sqrt{\lambda_1 T_{A,k,\alpha}^2} \quad (1.31)$$

e

$$s_m = \sqrt{\lambda_2 T_{A,k,\alpha}^2} \quad (1.32)$$

dove λ_1 e λ_2 sono gli autovalori associati alle prime due componenti principali e $T_{A,k,\alpha}^2$ è calcolato dalla distribuzione F come:

$$T_{A,k,\alpha}^2 = \frac{A(k-1)}{k-A} F_{A,k-1,\alpha} \quad (1.33)$$

dove k è il numero di campioni e A è il numero di componenti principali usati per creare il modello statistico multivariato e $F_{A,k-1,\alpha}$ dipende non solo da k e A ma anche dal limite di confidenza applicato $(1-\alpha)$.

Inoltre è possibile calcolare il contributo di ciascuna variabile al T^2 per determinare quale variabile è maggiormente responsabile della variazione dei campioni dalla media. Tale contributo viene calcolato con la formula seguente:

$$T_{con} = \mathbf{t}_i \boldsymbol{\lambda}^{-1} \mathbf{p}_i' \quad (1.34)$$

Inoltre è possibile calcolare la mancanza di accuratezza statistica del modello di regredire i dati, attraverso la somma dei quadrati di ciascuna riga della matrice \mathbf{E} , cioè per ciascun campione, chiamato *SPE (Squared Prediction Error)*:

$$SPE_i = \mathbf{e}_i \mathbf{e}_i' = \mathbf{x}_i (\mathbf{I} - \mathbf{P}_k \mathbf{P}_k') \mathbf{x}_i' = \sum_{j=1}^J (x_{i,j} - \hat{x}_{i,j})^2 \quad (1.35)$$

dove \mathbf{e}_i è un vettore riga della matrice dei residui \mathbf{E} , \mathbf{P}_k è la matrice contenete i vettori dei *loading* delle prime componenti principali e \mathbf{I} è matrice identità.

Secondo Wise e Gallagher (1996), la statistica SPE indica quanto un campione si identifica nel modello ed è una misura della varianza non catturata dalle k componenti principali usate nel modello.

I limiti di confidenza per i residui, SPE, vengono calcolate con (Wise e Gallagher, 1996):

$$SPE_\alpha = \Theta_1 \left[\frac{c_\alpha \sqrt{2\Theta_2 h_0^2}}{\Theta_1} + \frac{\Theta_2 h_0 (h_0 - 1)}{\Theta_1^2} \right]^{1/h_0} \quad (1.35)$$

dove

$$\Theta_i = \sum_{j=A+1}^M \lambda_j^i \quad \text{per } i = 1-3 \quad (1.36)$$

e

$$h_0 = 1 - \frac{2\Theta_1 \Theta_2}{3\Theta_2^2} \quad (1.37)$$

Nella (1.35) c_α è la deviazione standard normale che corrisponde al percentile maggiore $(1-\alpha)$. Nella (1.36) A sono le componeti principali usate nel modello e r è il numero totale di componeti principali pari al minimo tra I e J , cioè il rango della matrice \mathbf{X} analizzata. Il limite

di SPE è definito come la distanza al di fuori del piano che viene considerata inusuale per condizioni operative normali.

Inoltre se alcuni campioni hanno SPE che esce dai limiti descritti dal modello è possibile stabilire quali variabili siano responsabili di questa anomalia calcolando i residui, cioè quella parte della matrice che non viene descritta dal modello, come differenza tra i valori veri e quelli stimati dal modello:

$$SPE_{con} = (y - \hat{y}) \quad (1.38)$$

Ipotizzando che i dati siano normalmente distribuiti e che i contributi al T^2 e a SPE siano normalmente distribuiti è possibile calcolare i limiti di confidenza nel modo seguente:

$$x \pm z_{\alpha/2} \hat{\sigma}_x \quad (1.39)$$

dove x è un generico contributo medio, α è il livello di confidenza e di solito è compreso tra 0.01 e 0.05, cioè un intervallo di confidenza del 99% o 95%, $z_{\alpha/2}$ è la funzione calcolata utilizzando la distribuzione t-Student e σ_x è la deviazione standard dei contributi (Conlin *et al.*, 2000).

Capitolo 2

Il processo di granulazione e compattazione

Il processo di granulazione deve essere monitorato per capire quali siano le correlazioni tra la polvere in ingresso, il materiale intermedio e il prodotto finale. Inoltre è importante conoscere quali misure effettuare e in quale sezione del processo siano più efficienti. A questo scopo i dati disponibili sono stati riorganizzati per facilitare la comprensione del processo e le analisi successive.

2.1. Struttura dei dati relativi al paracetamolo

In Figura 2.1 è rappresentato un diagramma a blocchi semplificato del processo continuo di granulazione del paracetamolo. I dati, relativi alle misure eseguite sul materiale durante il processo di granulazione di Figura 2.1 sono stati riorganizzati in matrici, come indicato in figura.

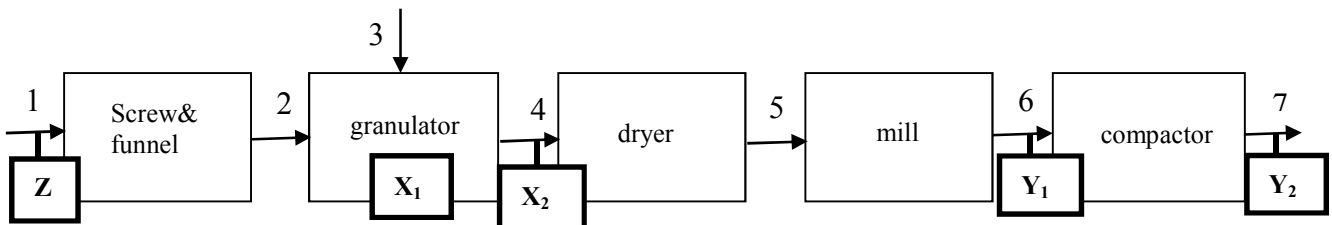


Figura 2.1 Diagramma a blocchi del processo di granulazione in continuo del paracetamolo. Le lettere maiuscole rappresentano i simboli impiegati per la descrizione dei dati sperimentali

Sono disponibili dati relativi alle condizioni iniziali del materiale in ingresso al processo, le misure on-line al granulatore e analisi di compressione. I dati sono stati riorganizzati per facilitare le analisi successive e permettere un confronto più agevole tra le varie sezioni del processo, suddividendoli nelle quattro matrici indicate in Figura 2.2 (Kourti e MacGregor, 1995; Kourti, 2003):

- condizioni iniziali **Z**;
- misure on-line al granulatore **X₁**;
- analisi in uscita al granulatore **X₂**;

- analisi in uscita al mulino Y_1 ;
- analisi alla comprimitrice Y_2 .

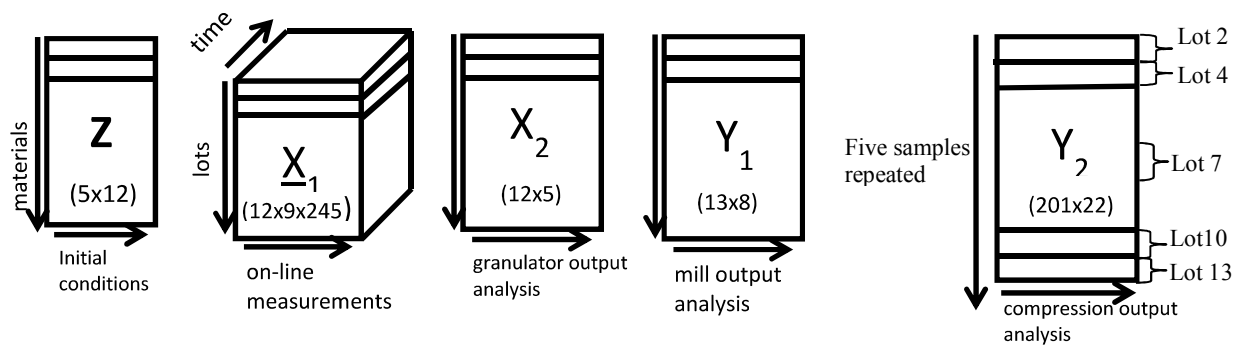


Figura 2.2 Organizzazione in matrici dei dati disponibili e relativa simbologia impiegata

Per facilitare la discussione dei dati è stata utilizzata la seguente notazione:

- [] per indicare una variabile;
- { } per indicare materiali, lotti, istanti di campionamento o campioni.

In ingresso al processo entrano cinque materiali che vengono addizionati al granulatore con diverse quantità d'acqua. Le diverse combinazioni di materiale e acqua sono state denominate lotti. Ogni lotto di materiale è stato campionato all'uscita di alcune unità del processo, come granulazione e compattazione, per valutare le proprietà del prodotto.

Il termine *lotto* identifica una parte di prodotto suddivisa per unità di tempo o di quantità che possiede caratteristiche uniformi (Chatterjee, 2012).

2.1.1. Condizioni iniziali

Nella Tabella 2.1 è riportata la matrice Z dove sono stati raggruppati i dati relativi alle condizioni iniziali di ciascun materiale in ingresso al processo (corrente 1 di Figura 2.1).

Si identificano:

- variabili booleane che identificano i pretrattamenti;
- proprietà di flusso, come densità aerata [5] e battuta [6], rapporto di Hausner [7], indice di Carr e *feed factor* [12];
- misure di distribuzione della dimensione delle particelle ([8], [9], [10]).

I materiali in ingresso al processo sono cinque, come riportato in Tabella 2.1 e sono stati caratterizzati con dodici variabili, anche se ci sono dati mancanti relativi a misure di densità.

Le misure mancanti sono state identificate con NaN (*Not a Number*) in Tabella 2.1.

Le prime quattro variabili della matrice Z in Tabella 2.1 descrivono i pretrattamenti che il materiale ha subito e le modalità con le quali il principio attivo è stato additivato.

Tabella 2.1 Matrice **Z** delle condizioni iniziali

		Variables											
		[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
		Z ₁	Z ₂	Z ₃	Z ₄	Z ₅	Z ₆	Z ₇	Z ₈	Z ₉	Z ₁₀	Z ₁₁	Z ₁₂
		Aldrich material different isolation	weight (g)	(1) “+ excipients”/ (0) “API alone”	(1) Wet milled/ (0) microfluidised	ρ aerated “raw”	ρ tapped “raw”	Hausner Ratio “raw”	x10 “raw”	x50 “raw”	x90 “raw”	span “raw”	feed factor at charge (kg/h)
Materials	{M1}	1	702	1	0	NaN	NaN	NaN	4.09	13.6	80.5	5.62	13.38
	{M2}	0	1213	1	1	0.42	0.73	1.74	7.93	28.1	92.5	3.01	16.03
	{M3}	0	922	1	0	0.35	0.6	1.74	3.86	12.3	81.5	6.31	12.8
	{M4}	0	710	0	1	0.39	0.62	1.6	9.11	33.8	102	2.75	14.12
	{M5}	0	500	0	0	0.36	0.59	1.61	6.67	21.1	137	6.18	13.89

I pretrattamenti, le modalità di miscelazione e le modalità di isolamento sono riportati in Tabella 2.2.

Tabella 2.2 Condizioni iniziali relative alla matrice **Z**

Number	Sample Number	Description	Powder blend reference	Mass (g)	Isolation means
M1	R18138/185/1	Microfluidised (Aldrich material) + Excipients		702	Agitated Filter Dryer
M2	R17950/113/4	Wetmilled API + Excipients		1213	Centrifuge and conical dryer
M3	R17939/110/4	Microfluidised API + Excipients		922	Centrifuge and conical dryer
M4	R17950/111/3	WetMilled API alone	108353-106	710	Centrifuge and conical dryer
M5	R17950/108/4	Microfluidised API alone	108353-107	500	Centrifuge and conical dryer

I materiali in ingresso al processo (corrente 1 di Figura 2.1) sono distinti in Tabella 2.2 in base alla modalità di miscelazione API-eccipiente:

- “+ Excipients” significa che il principio attivo è stato additivato al punto di isolamento;
- “API alone” significa che il principio attivo è stato additivato in un recipiente di 15 L, Pharmatec IBC, a 17 rpm per 15 min.

Al termine dei pretrattamenti i materiali corrispondono tutti alla medesima formulazione, che corrisponde a quella in Tabella 2.3.

Tabella 2.3 Formulazione del granulato farmaceutico

Component Name	Wt % dry basis	Function
Paracetamol API	75	API
Microcrystalline cellulose	9	Compression Aid/Filler/Absorbent
Hydroxypropylmethylcellulose	5	Binder
Lactose Monohydrate	9.5	Filler
Crosscarmellose sodium Ac-Di-Sol	1.5	Super-disintegrant

Ciascun materiale in ingresso è stato caratterizzato con misure di densità, aerata (ρ_p) e battuta (ρ_t), per valutare le proprietà di flusso. La densità è un parametro importante per valutare la quantità di polvere che può essere introdotta in un miscelatore, in una tramoggia o in una comprimitrice (Yihong *et al.*, 2009).

Dalle misure di densità aerata e battuta è possibile ricavare alcuni parametri, caratteristici della polvere, come il rapporto di Hausner e l'indice di Carr (Schulze, 2007). Il rapporto di Hausner (HR) è un numero correlato alla fluidità della polvere ed è calcolato con la formula $HR = \rho_t / \rho_p$. Questo indice dipende dalla modalità con la quale vengono calcolate le variabili e permette di valutare la variazione di densità in relazione al cambiamento di forma delle particelle. L'indice di Carr (CI) è un'indicazione della compressibilità della polvere ed è una misura relativa delle interazioni intermolecolari. Tale indice permette di caratterizzare il materiale sapendo che le due densità variano molto se le particelle sono coesive, mentre sono simili se le particelle sono scorrevoli:

$$CI(\%) = \left(\frac{\rho_t - \rho_p}{\rho_t} \right) \times 100 \quad (2.1)$$

Tabella 2.4 Confronto tra l'indice di Carr e rapporto di Hausner (Yihong *et al.*, 2009)

Flow character	CI	HR
Excellent	≤ 10	1.00-1.11
Good	11-15	1.12-1.18
Fair	16-20	1.19-1.25
Passable	21-25	1.26-1.34
Poor	26-31	1.35-1.45
Very poor	32-37	1.46-1.59
Very, very poor	> 38	> 1.60

Non ci sono differenze qualitative sostanziali tra i due indici, e minori essi sono maggiori sono le proprietà di scorrimento, come riportato in Tabella 2.4.

Un altro parametro importante misurato per la caratterizzazione del materiale in ingresso è il *feed factor*, cioè la portata alimentata (in kg/h) alla massima velocità della vite. Tale fattore è

fortemente correlato alla densità del materiale, perché a parità di volume libero della vite, maggiore è la densità maggiore è la portata ponderale che può essere alimentata. I materiali in ingresso al processo sono stati caratterizzati anche con misure di distribuzione della dimensione delle particelle. Questa misura permette di capire come varia la popolazione delle particelle in ogni unità del processo e di comprendere anche le fasi della granulazione, dalla nucleazione alla rottura dei granuli. In Figura 2.3 si può osservare come varia la distribuzione delle dimensioni delle particelle tra l'ingresso e l'uscita di un granulatore.

Infatti la dimensione dei granuli dipende dalle condizioni iniziali della polvere, dalla quantità d'acqua al granulatore, dalle modalità di granulazione, di essiccamento e successiva macinazione e dalle modalità di compattazione.

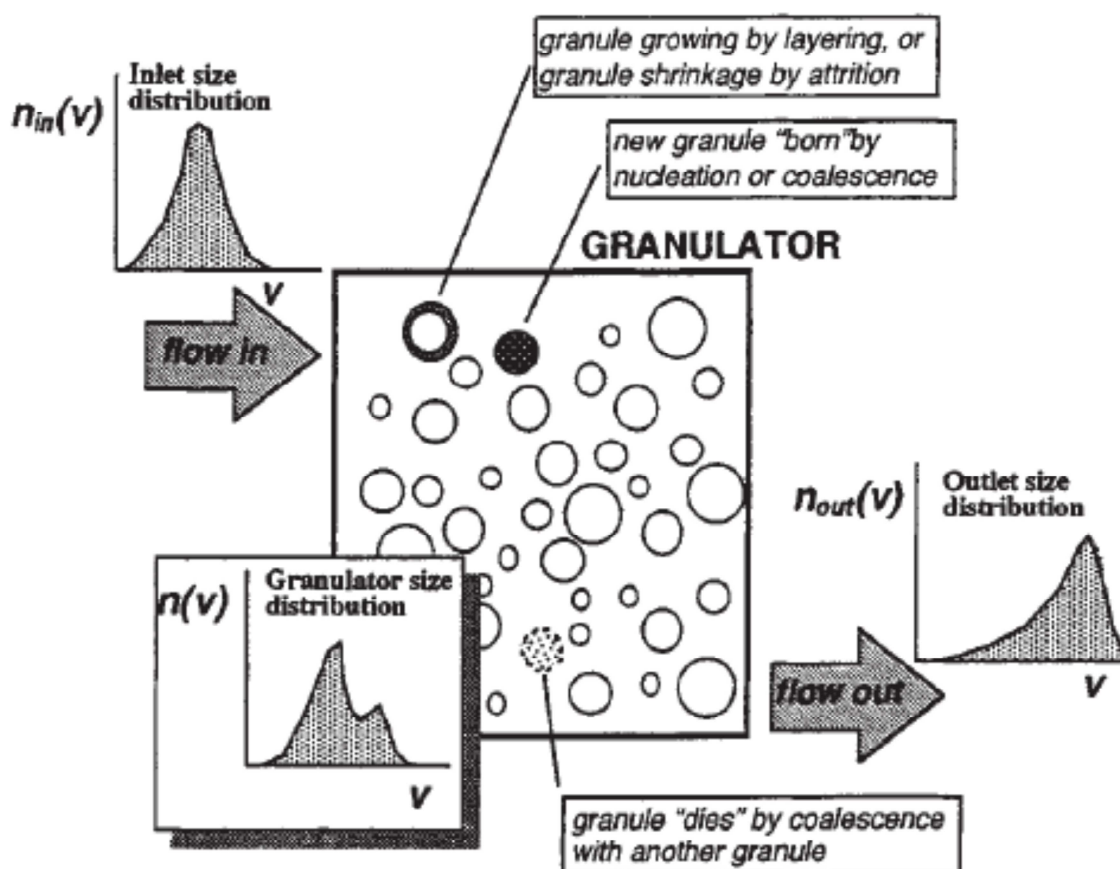


Figura 2.3 Distribuzione della dimensione dei granuli misurata in punti diversi del granulatore per comprendere il processo (Perry e Green, 1997, 20-85)

Le misure sono state effettuate utilizzando un diffrattometro laser (Synpatec, Helos/Gradis) che permette di ottenere la mediana (d_{50}), il decimo (d_{10}) e il novantesimo (d_{90}) percentile (in μm) per poter valutare la distribuzione delle dimensioni delle particelle. Inoltre è stato calcolato lo *span*, cioè l'ampiezza della distribuzione:

$$\text{Span} = \frac{d(x_{90}) - d(x_{10})}{d(x_{50})} \quad (2.2)$$

Esso non ha nessuna relazione con il diametro medio delle particelle, ma permette di capire l'uniformità delle particelle in ingresso al processo.

I dati disponibili sono stati riorganizzati nella matrice Z di Tabella 2.1 per facilitare le analisi successive eliminando le variabili ridondanti, come l'indice Carr (tale indice è infatti proporzionale al rapporto di Hausner).

2.1.2. Misure on-line al granulatore

I materiali in ingresso al granulatore (corrente 3 di Figura 2.1) sono monitorati in continuo durante la fase di granulazione. In passato i sistemi di granulazione sono stati monitorati valutando informazioni relative a proprietà meccaniche della massa bagnata utilizzando parametri come il consumo di materiale o i segnali della coppia (Kristensen e Schaefer, 1987; Faure *et al.*, 2001).

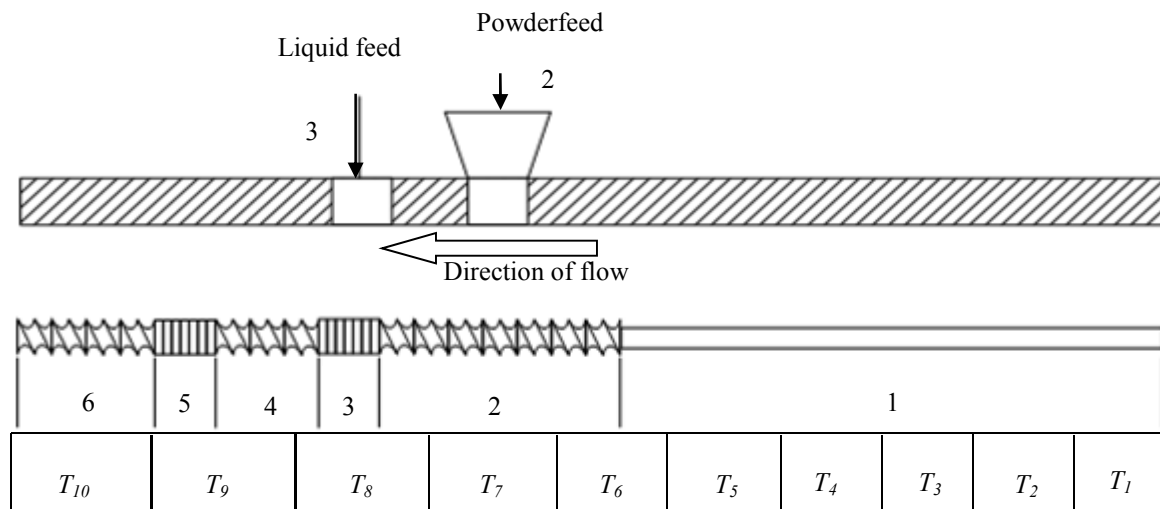


Figura 2.4 Particolare del granulatore. Le correnti 2 e 3 sono gli ingressi al granulatore, i numeri da 1 a 6 rappresentano le suddivisioni della vite del granulatore e la lettera T indica le sezioni di misura della temperatura.

In questo studio il granulatore è stato monitorato in continuo mediante misure di temperatura in varie sezioni del granulatore stesso, misure relative alla coppia e misure di portata.

Lo *step* di granulazione del materiale solido con l'aggiunta di un liquido (correnti 2 e 3 di Figura 2.1) è stato suddiviso in più fasi:

- miscelazione del solido con l'acqua fino al raggiungimento di uno stato stazionario;
- stato stazionario per un periodo variabile tra 300 e 900 secondi;
- fase transitoria di miscelazione del solido con una quantità maggiore di acqua;
- stato stazionario della nuova miscela;
- fase di spegnimento del granulatore.

Il granulatore schematizzato in Figura 2.4 è stato suddiviso in sei sezioni per misurare le varie fasi di granulazione, e in dieci sezioni per acquisire il profilo di temperatura lungo la coclea. La matrice X_1 di Figura 2.2 contiene i dati relativi alle misure al granulatore ed è formata da nove variabili:

- “*ExtTorqServoNM*”, cioè la coppia misurata al motore in Nm;
- “*ExtTorqServoPV*”, cioè la coppia massima in percentuale misurata al motore;
- “*ExtTorqTransd*”, cioè la coppia misurata al trasduttore in Nm;
- “ T_7, T_8, T_9, T_{10} ”, cioè le temperature misurate nelle ultime quattro zone del granulatore in °C;
- “*ExtSpdAV*”, cioè la velocità del motore misurata in rpm;
- “*FeederRate*”, cioè la portata alimentata al granulatore in kg/h;

Sono state alimentate al granulatore portate d’acqua diverse per capire come varino le caratteristiche del prodotto finale. Le diverse miscele di solido e acqua sono state monitorate durante la fase stazionaria per un periodo variabile tra i 5 e gli 11 min. In Tabella 2.5 è riportata la durata di monitoraggio di ciascun lotto e le problematiche riscontrate durante la fase di granulazione.

Tabella 2.5 Caratteristiche dei dati on-line al granulatore (matrice X_1)

Lot number	Time (s)	Time (min)	Time in	Time out	Problems
1	481	8	15:59:00	16:07:00	
2	421	7	16:08:00	16:15:00	motor torque decreases after 16.13 for lack of the material and feeder rate goes to 1.86
3	601	10	10:37:00	10:47:00	
4	661	11	10:48:00	10:59:00	
5	661	11	11:00:00	11:11:00	motor torque and feeder rate go to zero at 11.05 because motor has been stopped and restarted
6	361	6	14:38:00	14:44:00	
7	481	8	14:45:00	14:53:00	
8	301	5	14:54:00	14:59:00	At 14.58.10 motor torque goes down
9	421	7	14:25:10	14:32:10	“ <i>ExtTorqTransd</i> ” starts from zero and the other motor torque a low value
10	901	15	14:43:10	14:58:10	At 14.50.50 motor torque and feeder rate go to zero
12	481	8	15:19:10	15:27:10	“ <i>ExtTorqTransd</i> ” starts from zero and the other motor torque a low value
13	301	5	15:28:10	15:33:10	At 15.32.14 motor torque goes down

Il quarto materiale alimentato al granulatore è stato additivato con due quantità d'acqua (15% e 17.5%); quindi le miscele solido-liquido monitorate nella fase stazionaria sono state chiamate lotto {9} e lotto {10}. Sono stati riscontrati dei problemi relativi al processo di granulazione del lotto {10}, come si vede in Figura 2.5.

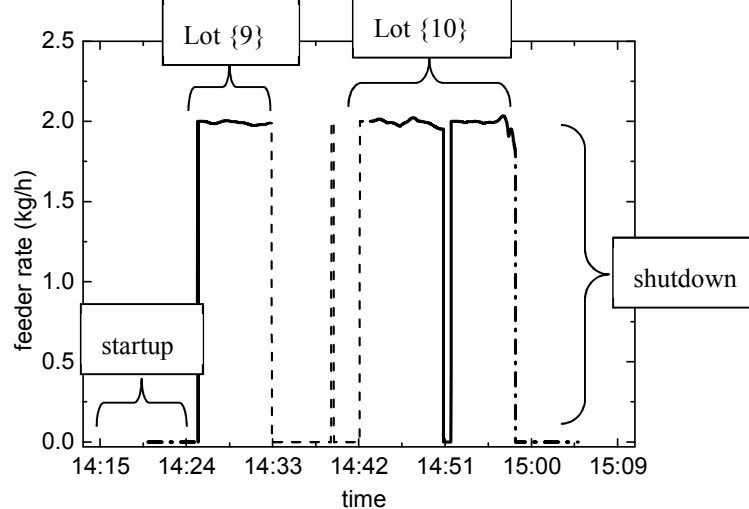


Figura 2.5 Profilo della portata entrante al granulatore relativa al quarto materiale alimentato, lotti 9 e 10. Le linee tratteggiate rappresentano i transitori mentre le linee continue gli stazionari.

La misura della portata di solido in ingresso al granulatore in Figura 2.5 è stata suddivisa nei vari *step* dall'accensione all'arresto e durante le misure del secondo stato stazionario, cioè del lotto {10}, è stata riscontrata un'anomalia, cioè il probabile spegnimento del motore che porta a zero l'alimentazione.

I dati sono stati quindi riorganizzati nella matrice \underline{X}_1 riportata in Figura 2.6 considerando solamente 245 istanti di campionamento per ciascun lotto nelle condizioni di processo normale.

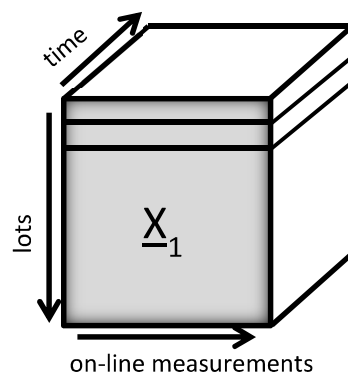


Figura 2.6 I dati relative al monitoraggio al granulatore sono stati raggruppati nella matrice \underline{X}_1

Nella Tabella 2.6 è riportata la struttura della matrice \underline{X}_1 per il primo istante di campionamento. La matrice contiene dodici lotti e non compaiono dati relativi al lotto {11} poiché in questa fase del processo ha le stesse caratteristiche del lotto {10}.

Le variabili sono otto e la portata d'acqua alimentata al granulatore è l'unica variabile operativa.

Tabella 2.6 Sezione della matrice \underline{X}_1 relativa al primo istante di campionamento

		[47]	[48]	[49]	[50]	[51]	[52]	[53]	[54]	[55]
		X1 ₁	X1 ₂	X1 ₃	X1 ₄	X1 ₅	X1 ₆	X1 ₇	X1 ₈	X1 ₉
		ExtTorq Servo NM	ExtTorq Servo PV	Ext Torq Transd	zone7T	zone8T	zone9T	zone10T	ExtSpd AV	Feeder Rate
lots	{1}	3.333	14.035	2.643	20.17	22.17	25.56	23.09	199.31	1.99

	{13}	2.892	12.213	2.377	20.24	23.02	27.43	27.00	199.14	2.002

2.1.3. Analisi all'uscita del granulatore

La corrente in uscita al granulatore (corrente 4 di Figura 2.1) è stata campionata per determinare la distribuzione delle particelle con uno strumento PIMS, *Particle Imaging Measurement System*, realizzato dall'azienda. Le misure sono dello stesso tipo di quelle effettuate sul materiale in ingresso al processo (corrente 1 di Figura 2.1) poiché sono stati misurati il decimo, il cinquantesimo e il novantesimo percentile. Tale misura, effettuata in linea, è preferibile rispetto ad un'analisi off-line perché è più rapida e meno costosa. I valori di distribuzione riportati nella matrice \underline{X}_2 sono la media delle prove eseguite in linea all'uscita del granulatore.

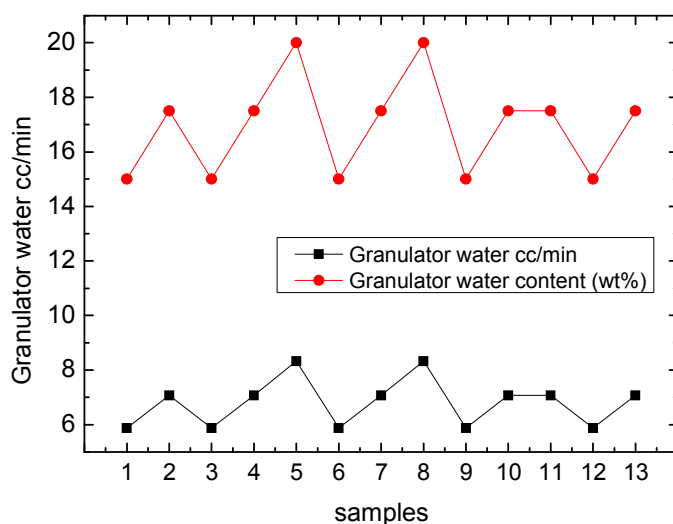


Figura 2.7 Confronto tra la portata d'acqua alimentata al granulatore e il contenuto d'acqua della miscela in uscita al granulatore

Nella Figura 2.7 la portata d'acqua alimentata al granulatore e il contenuto d'acqua dopo la granulazione, hanno lo stesso andamento.

Infatti il contenuto d'acqua della miscela all'uscita del granulatore è stato calcolato con la relazione:

$$\text{Granulator water content (wt\%)} = \frac{w_{\text{water}}}{w_{\text{total}}} \times 100 = \frac{\text{liquid feed}}{(\text{powder feed} + \text{liquid feed})} \quad (2.3)$$

Le variabili della matrice \mathbf{X}_2 sono state utilizzate per descrivere la corrente all'uscita del granulatore e sono riportate in Tabella 2.7.

Tabella 2.7 *Analisi effettuate all'uscita del granulatore (matrice \mathbf{X}_2)*

		Variables				
		[13]	[14]	[15]	[16]	[17]
		X2 ₁	X2 ₂	X2 ₃	X2 ₄	X2 ₅
		granulator water content (%)	L10 PIMS	L50 PIMS	L90 PIMS	Span PIMS
Lots	1	15	62.9	213.75	697.56	2.97
	2	17.5	94.3	272.63	899.73	2.95
	3	15	NaN	NaN	NaN	NaN
	4	17.5	NaN	NaN	NaN	NaN
	5	20	NaN	NaN	NaN	NaN
	6	15	62.9	153.49	567.2	3.29
	7	17.5	76.36	220.59	773.94	3.16
	8	20	94.3	298.75	1047.75	3.19
	9	15	94.32	297.72	897.97	2.7
	10	17.5	125.78	419.1	1139.7	2.42
	12	15	88.62	232.46	726.62	2.74
	13	17.5	88.02	273.94	831.6	2.71

2.1.4. *Analisi all'uscita del mulino*

La corrente in uscita al mulino (corrente 6 di Figura 2.1) è stata analizzata per determinarne le proprietà di flusso. Sono state effettuate misure di densità e di distribuzione delle particelle con le stesse tecniche utilizzate per le analisi del materiale in ingresso del processo per poter confrontare le varie fasi di granulazione e capire quali variabili siano rilevanti per la qualità finale del prodotto. Una parte del materiale in uscita all'essiccatore, corrispondente al lotto {10}, non è stata macinata andando a formare il lotto {11}.

Le variabili della matrice Y_1 sono state utilizzate per descrivere la corrente all'uscita al mulino e sono riportate in Tabella 2.8.

Tabella 2.8 Analisi effettuate all'uscita del mulino (matrice Y_1)

		Variables							
		[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]
		Y_{1_1}	Y_{1_2}	Y_{1_3}	Y_{1_4}	Y_{1_5}	Y_{1_6}	Y_{1_7}	Y_{1_8}
		not milled	ρ aerated	ρ tapped	Hausner Ratio	x10	x50	x90	span
Lots	1	0	0.49	0.65	1.33	33.2	306	798	2.5
	2	0	0.49	0.61	1.25	43.6	395	934	2.25
	3	0	0.51	0.68	1.34	34.9	313	1087	3.36
	4	0	0.52	0.66	1.28	46.8	546	1357	2.4
	5	0	0.5	0.67	1.34	36.1	378	859	2.18
	6	0	0.46	0.62	1.35	21.7	209	630	2.91
	7	0	0.48	0.65	1.34	32.2	334	958	2.77
	8	0	0.46	0.61	1.32	55.2	592	1273	2.06
	9	0	0.51	0.66	1.31	50.5	361	841	2.19
	10	0	0.51	0.65	1.26	59.6	438	952	2.04
	11	1	0.48	0.56	1.19	361	1749	2821	1.41
	12	0	0.48	0.63	1.3	36	348	1055	2.93
	13	0	0.49	0.64	1.3	45.3	377	922	2.33

2.1.5. Analisi alla comprimitrice

Dopo la macinazione (corrente 6 di Figura 2.1), le miscele con quantità d'acqua intermedia, lotti {2,4,7,10,13}, sono state sottoposte a compressione per ottenere il prodotto finale (corrente 7 di Figura 2.1).

Il materiale in ingresso alla comprimitrice (corrente 6) è stato sottoposto a diverse prove di compressione variando il valore della distanza del punzone, la profondità di riempimento e mantenendo costante la durata della prova (prime tre variabili della Tabella 2.8).

Le compresse sono state poi caratterizzate con misure di dimensione, proprietà di flusso e di resistenza a trazione come riportato in Tabella 2.8.

Ogni materiale in uscita al mulino è stato campionato circa quaranta volte, e metà di queste prove hanno nove misure mancanti. Le prove di compattazione permettono di capire la resistenza del materiale alla compressione relativa alla frazione solida. Solitamente i granulati farmaceutici sono fragili ed è preferibile usare prove di durezza per stabilire la resistenza a trazione (Yihong *et al.*, 2009). Il test permette di ottenere valori di durezza, cioè di resistenza

del materiale a deformazioni plastiche, correlabili alla resistenza a trazione (*tensile strength* variabile [46] della Tabella 2.9).

Tabella 2.9 Variabili di compressione (matrice Y_2)

number	Symbol	Variable		Description		formula
[26]	l	minimum punch tip separation distance	mm	operating parameter	assigned	-
		duration of profile	s	operating parameter (constant)	assigned	-
[27]	fl	Fill depth	mm	operating parameter	assigned	-
[28]	w	Tablet weight	mg	tablet property	measured	-
[29]	t	tablet thickness	mm	tablet property	measured	-
[30]	D	tablet diameter	mm	tablet property	measured	-
[31]	H	Hardness	kp	tablet property	measured	-
[32]	ρ	Density	g/cm ³	tablet property	calculated	$\rho = w / (\pi(D / 2)^2 \times t)$
[33]	ρ_r	Relative density	[]	tablet property	calculated	$\rho_r = \rho / 1.35$
[34]	ϕ	Porosity	[]	tablet property	calculated	$\phi = 1 - \rho_r$
[35]	E_{tot}	Total energy	J		calculated	
[36]	E_{rec}	Recovered energy	J		calculated	
[37]	E_{irr}	Irrecoverable energy	J		calculated	$E_{irr} = E_{tot} - E_{rec}$
[38]	P	Plasticity ratio	[]	tablet property	calculated	$P = E_{irr} / E_{tot}$
[39]	F_{UpP}	Max Upper Punch Force	kN	machine property	measured	-
[40]	F_{LoP}	Max Lower Punch Force	kN	machine property	measured	-
[41]	F_{Ej}	Max Ejection Force	kN	machine property	measured	-
[42]	S_{UpP}	Max Upper Punch Stress	MPa	machine property	calculated	$S_{UpP} = F_{UpP} / (\pi(10/2)^2)$
[43]	S_{LoP}	Max Lower Punch Stress	MPa	machine property	calculated	$S_{LoP} = F_{LoP} / (\pi(10/2)^2)$
[44]	S_{Ej}	Max Ejection Stress	MPa	machine property	calculated	$S_{Ej} = F_{Ej} / (\pi 10t)$
[45]	S_{tr}	Stress Transmission Ratio	[]	machine property	calculated	$S_{tr} = S_{LoP} / S_{UpP}$
[46]	σ	Tensile strength	MPa	tablet property	calculated	$\sigma = 2P / \pi Dt$

2.2. Struttura dei dati relativi ad un secondo granulato

Sono disponibili dati relativi ad un materiale diverso (secondo granulato) che è stato sottoposto allo stesso processo di granulazione.

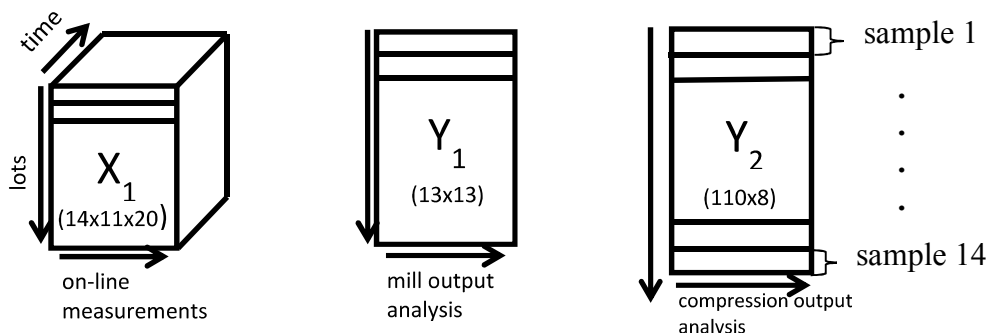


Figura 2.8 Riorganizzazione in matrici dei dati relativi al granulato B

I dati sono stati suddivisi in matrici (Figura 2.8) come nel caso precedente:

- misure on-line al granulatore, matrice X_1 ;
- analisi all'uscita del mulino, matrice Y_1 ;
- analisi alla comprimitrice, matrice Y_2 .

Le prove sono state effettuate variando la portata alimentata al granulatore, la velocità del motore e la portata di polvere in ingresso al granulatore. Il granulatore è stato monitorato con misure di temperatura, di portata e di variabili relative alla coclea, e le misure disponibili sono riportate in Tabella 2.10.

Tabella 2.10 Variabili relative al monitoraggio del granulatore (matrice X_1) per il secondo granulato

		Variables										
		[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]
		$X1_1$	$X1_2$	$X1_3$	$X1_4$	$X1_5$	$X1_6$	$X1_7$	$X1_8$	$X1_9$	$X1_{10}$	$X1_{11}$
		Water addition % w/w	Screw Speed rpm	Mass flow rate Kg/h	ExtTorq Servo NM	ExtTorq ServoPV	zone7T	zone8T	zone9T	zone10T	ExtSpd AV	Feeder Rate
Lots	{1}	25	150	2	9.03	8.15	24.07	23.01	24.37	32.86	149.89	1.99

	{14}	15	300	2	5.28	4.46	23.45	26.186	28.99	38.96	300.38	2.00

I campioni in uscita al mulino sono stati caratterizzati con misure di densità aerata, battuta, dinamica e misure di distribuzione della dimensione delle particelle mediante setaccio. Tali variabili sono riportate nella Tabella 2.11 e costituiscono la matrice Y_1 .

Tabella 2.11 Variabili relative al monitoraggio del granulatore (matrice Y_1) per il granulato B

Variables		Description
ρ aerated (initial)	g/ml	measured
ρ tapped	g/ml	measured
compaction compressibility	%	measured
ρ dynamic	g/ml	measured
Hausner ratio	[]	calculated
total cumulative undersize	%	measured
total weight retained	g	measured
percentage recovery	%	measured
median particle size	μm	measured
particle size at 10% undersize	μm	measured
particle size at 90% undersize	μm	measured
uniformity number		calculated
uniformity index	%	calculated

Inoltre sono state eseguite diverse prove di compattazione su tutti i materiali in uscita al mulino. In questo caso sono state misurate nove variabili (Tabella 2.12) che descrivono le proprietà delle compresse e non sono state misurate quelle variabili che descrivono l'energia e lo sforzo necessari durante la compattazione.

Tabella 2.12 variabili all'uscita della comprimitrice (matrice Y_2) per il secondo granulato

Variables		Description
weight	mg	measured
thickness	mm	measured
hardness	kp	measured
wall height	mm	measured
tensile strength	MPa	calculated
density	mg/mm^3	calculated
solid fraction	[]	calculated
compression pressure	MPa	measured
predicted TS	MPa	calculated

Nel capitolo successivo le matrici con numero di righe diverso da quello descritto nei paragrafi precedenti ma lo stesso numero di variabili sono state identificate con un apice come Z^* .

Capitolo 3

Analisi statistica multivariata del processo

Il processo di granulazione è stato analizzato mediante metodi statistici multivariati (PCA e PLS) per comprendere, a partire dai dati, le forze motrici del sistema. Ogni fase del processo (condizioni iniziali, al granulatore, al mulino o di compattazione) è stata analizzata singolarmente per capire le correlazioni tra le variabili all'interno di ogni blocco e poi in modo congiunto per capire le relazioni tra i vari blocchi. Sono state utilizzate tecniche PLS per stimare le caratteristiche dei granuli all'uscita del mulino dalle fasi precedenti del processo.

2.1. Analisi dei dati relativi al paracetamolo

Prima di procedere con l'analisi delle componenti principali (PCA), i dati disponibili (condizioni iniziali, misure rilevate al granulatore, all'uscita del mulino o di compattazione, sono stati autoscalati) cioè centrati rispetto alla media e ridotti a varianza unitaria, perché hanno unità di misura e ordini di grandezza diversi e questa procedura permette di uniformare il peso delle variabili. Dopo aver pretrattato i dati, è necessario determinare il numero di componenti principali (PC) da considerare nel modello. Tale scelta è stata effettuata valutando l'autovalore associato a ciascuna PC, la varianza spiegata da ciascuna PC e l'errore in convalida incrociata (*cross validation*).

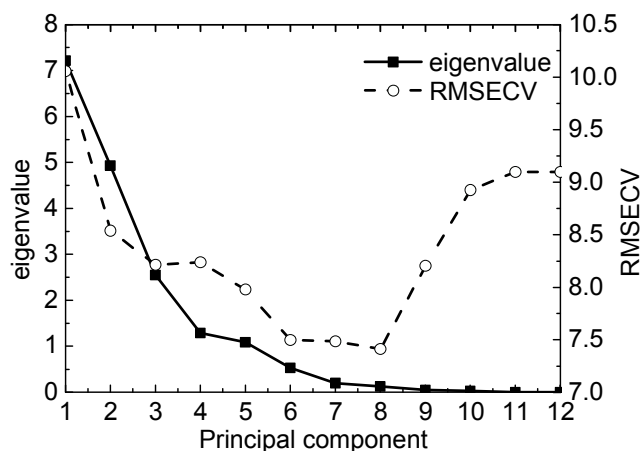


Figura 3.1 Andamento di autovalori e RMSECV per ciascuna PC nel modello della matrice B, dei dati relativi alle condizioni iniziali, all'uscita del granulatore e all'uscita del mulino.

L'entità degli autovalori diminuisce all'aumentare del numero di componenti principali impiegate, e per la determinazione del numero ottimale di PC si valuta l'andamento degli autovalori al variare del numero di PC. Il punto in cui la curva diminuisce sensibilmente la propria pendenza ("il ginocchio") corrisponde in genere al numero di PC da utilizzare (Wise *et al.*, 2006).

Dalla Figura 3.1 si nota come quattro o cinque componenti principali possano essere selezionate. Un criterio alternativo che può essere considerato per la determinazione del numero di PC è valutare l'autovalore associato a ogni componente principale. Ciascun autovalore può essere interpretato come il numero di variabili della matrice originale descritte dalla corrispondente componente principale (Wise *et al.*, 2006); per esempio, nella Figura 3.1 la prima PC cattura la varianza di circa 7 delle variabili originali. Non è opportuno utilizzare componenti principali con autovalore inferiore a 1, perché non descrivono varianza sistematica.

È stato inoltre valutato il minimo dell'errore quadratico medio in convalida incrociata (*root mean squared error of cross validation*, RMSECV). Come si vede in Figura 3.1, la curva dell'errore presenta un minimo tra la sesta e l'ottava PC. La convalida incrociata consiste nel dividere la matrice in segmenti uguali (di una o più righe) selezionate in maniera casuale, e nel creare un modello PCA sulla matrice senza un segmento, per poi proiettare sul modello appena creato i dati del segmento non considerato (Wise e Ricker, 1991), valutando il corrispondente errore di ricostruzione. Ripetendo l'analisi per un numero predefinito di volte è possibile ottenere un valore medio dell'errore di ricostruzione (RMSECV).

Tabella 3.1 Autovalori, varianza spiegata da ciascuna componente principale e varianza cumulata per il modello PCA creato sulla matrice **B**

Number of PC	Eigenvalues	Variance	Cumulative variance
1	7.21	40.07	40.07
2	4.93	27.39	67.46
3	2.55	14.16	81.62
4	1.29	7.17	88.79
5	1.09	6.04	94.83
6	0.53	2.94	97.77
7	0.19	1.09	98.86
8	0.13	0.70	99.56
9	0.05	0.27	99.83
10	0.03	0.15	99.98
11	0.002	0.01	99.99
12	3.23E-4	0.002	99.992

L'errore di ricostruzione decresce all'aumentare del numero delle componenti principali impiegate, finché la PC descrive solo rumore per cui RMSECV ricomincia a crescere come si vede in Figura 3.1. Sulla base dei risultati ottenuti, si è ritenuto opportuno selezionare cinque PC per la costruzione del modello sulla matrice **B** riportata in Appendice.

Ogni successiva analisi statistica è stata effettuata utilizzando queste stesse regole per la scelta del numero di componenti principali, valutando caso per caso quale fosse il numero di PC più opportuno.

Sono stati condotti due tipi di modellazione: uno su ciascun dataset relativo al processo, e uno sulla matrice generata dall'insieme di tutti i dataset.

Tale suddivisione è stata effettuata per due motivi:

- comprendere eventuali interazioni tra le variabili all'interno di ciascun blocco (crf. §2.1) e identificare somiglianze tra i campioni (i cinque materiali in ingresso al processo, i tredici lotti in uscita al granulatore, i dati relativi alle misure in continuo al granulatore e le diverse misure di compattazione effettuate);
- comprendere eventuali interazioni tra le variabili di un blocco e quelle di altri blocchi.

3.1.1. Analisi delle condizioni iniziali (**Z**)

Le condizioni iniziali (matrice **Z**) riportate in Tabella 2.1 di cinque materiali in ingresso al granulatore sono state analizzate costruendo un modello con tre componenti principali per comprendere le fonti di variabilità dei diversi materiali e le correlazioni tra le variabili. I risultati del modello PCA sono riportati in Figura 3.2 solo per le prime due PC. La Figura 3.2a riporta il digramma degli *scores*, mentre in Figura 3.2b è riportato il diagramma dei *loadings*. Ci sono dati mancanti per alcune variabili del materiale {M1}, che vengono stimate dall'algoritmo per poter stimare i parametri del modello PCA.

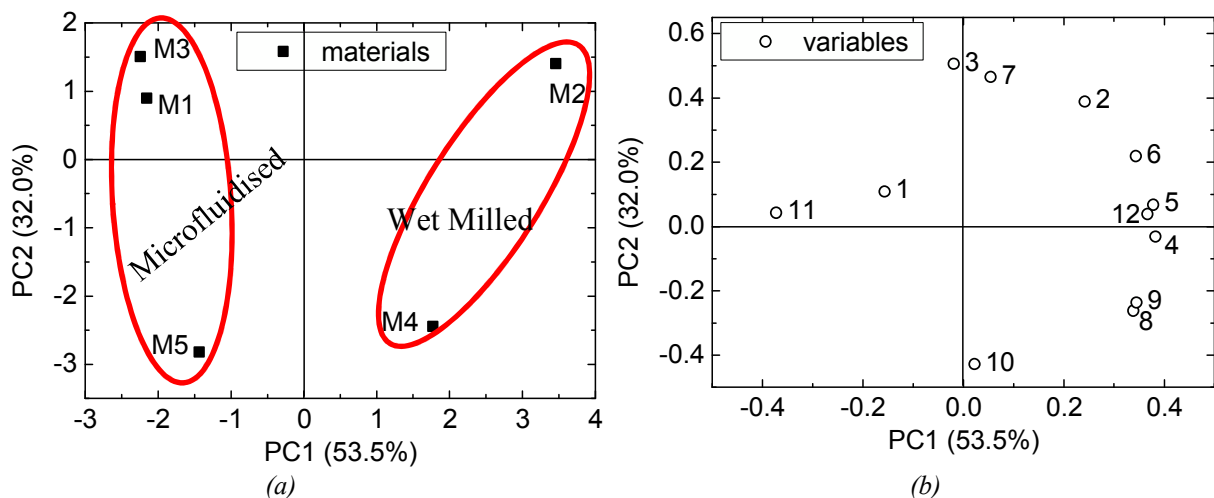


Figura 3.2 Modello PCA sulle condizioni iniziali, matrice **Z**. (a) Scores plot; (b) loading plot.

I cinque punti proiettati nel piano degli *scores* sulle prime due PC di Figura 3.2(a) rappresentano le cinque diverse tipologie di materiale in ingresso al processo (corrente 1 di Figura 2.1). Sulla prima componente principale, a cui è associata una varianza spiegata del 53.5%, sono distinti nello *scores plot* di Figura 3.2a i materiali *microfluidised* {M1,M3,M5} e i materiali *wet milled* {M2,M4}. Nel *loading plot* (Figura 3.2b) è possibile osservare che i materiali *microfluidised* hanno un elevato *span* [11], mentre quelli *wet milled* sono caratterizzati da valori elevati di *feed factor* [12], densità [5,6] e mediana [8] e cinquantesimo percentile della distribuzione [9].

La seconda fonte di variabilità, lungo PC2 (32%), è dovuta alla differenza tra i materiali “+ *excipient*” {M1,M2,M3}, caratterizzati da *Hausner ratio* [7] elevate e i materiali “API alone” {M4,M5}, che risultano essere caratterizzati dal novantesimo percentile [10] elevato, e quindi da distribuzioni dimensionali più allungate.

3.1.2. Analisi delle misure al granulatore (X_1)

I dati relativi alle misure in continuo al granulatore sono stati suddivisi in quattro fasi: rampa iniziale, stazionario, transitorio e arresto. Durante la fase stazionaria di lavorazione di alcuni lotti di materiale sono state riscontrate alcune anomalie di misura, come lo spegnimento del motore o la riduzione della portata di solido al granulatore. Dopo aver eliminato gli istanti di campionamento relativi alle anomalie riscontrate (di esse si parlerà nel Capitolo 5), si è scelto di dare a tutti i lotti la stessa lunghezza temporale, pari a quella del campione più corto (245 istanti di campionamento), affinché tutti i lotti abbiano lo stesso peso nell’analisi.

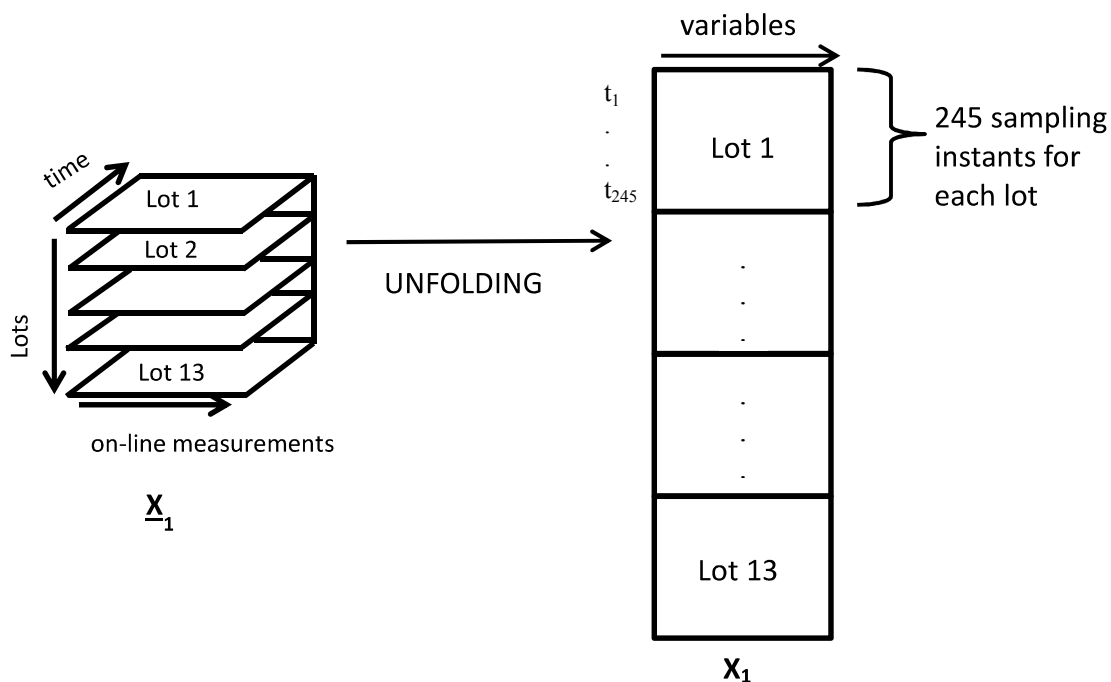


Figura 3.3 Riarrangiamento della struttura della matrice X_1 da tridimensionale a bidimensionale, mediante *variable wise unfolding*.

Per facilitare la comprensione della fase di granulazione, la matrice tridimensionale \mathbf{X}_1 è stata riarrangiata in una matrice bidimensionale mediante *variable wise unfolding* (Kourti, 2003). In tal modo, gli istanti di campionamento sono disposti uno sotto l'altro come si vede in Figura 3.3.

Gli scopi dell'analisi condotta sulla matrice \mathbf{X}_1 sono:

- comprendere le correlazioni tra le variabili monitorate in linea durante il processo;
- capire se tali variabili vengono influenzate dalle condizioni iniziali del materiale in ingresso al granulatore;
- conoscere eventuali somiglianze tra i diversi lotti sottoposti alla granulazione.

Le prime due componenti principali del modello sono riportate in Figura 3.4. Nello *scores plot* (Figura 3.4a) ciascun lotto è rappresentato con simboli diversi per facilitare la comprensione del diagramma e identificare eventuali cluster.

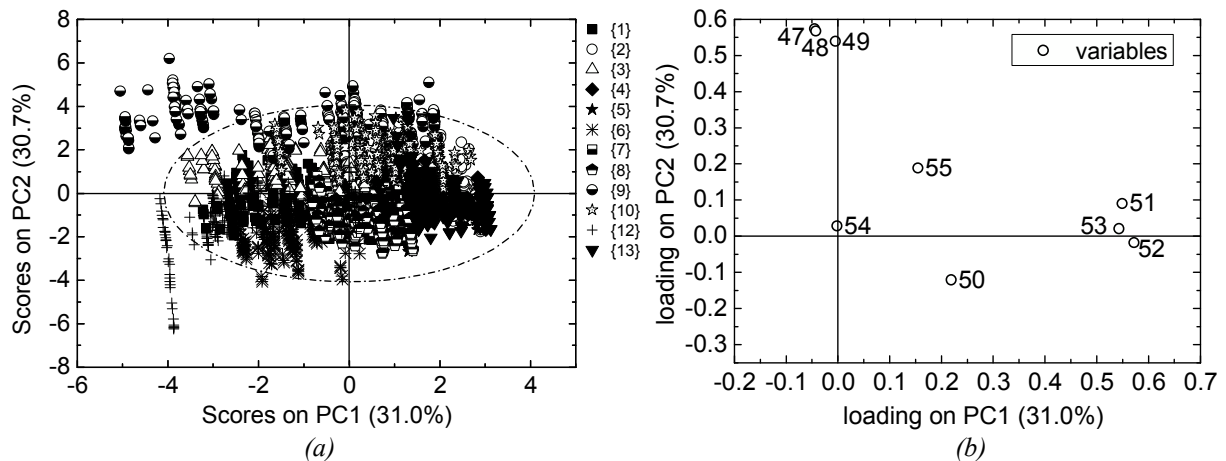


Figura 3.4 Modello PCA relativo ai dati di monitoraggio al granulatore (a) score plot (b) loading plot. I diversi lotti di materiale sono indicati con simboli diversi.

I risultati di quest'analisi sono:

- i dati relativi ai lotti {9} e {12} escono dai limiti di confidenza, ossia vengono riconosciuti come “diversi” rispetto alla media degli altri dati poiché:
 - il lotto {9} ha valori mediamente più elevati di *motor torque* [47,48,49];
 - il lotto {12} ha valori di *motor torque* [47,48,49] mediamente più bassi.

Le variabili relative al *motor torque* sono un indice delle forze coesive del materiale processato, e probabilmente il lotto {9} ha forze di coesione più elevate del lotto {12}. Inoltre, si può ipotizzare che tali forze siano strettamente correlate alle variabili relative alla distribuzione delle dimensioni delle particelle dei materiali in ingresso al processo. La mediana della distribuzione iniziale è elevata per il lotto {9} e bassa per il lotto {12}. Inoltre il lotto {9} ha una distribuzione iniziale molto stretta e quindi valori di *span* bassi, mentre il lotto {12} ha una distribuzione allungata e *span* elevati.

Quindi è possibile ipotizzare che materiali caratterizzati da dimensioni delle particelle elevate possano necessitare un maggiore momento torcente durante la granulazione;

- i dati relativi al campionamento di lotti con portata d'acqua intermedia o elevata (17.5% o 20%) hanno PC1 positiva. Dal diagramma dei *loading* si evince che le temperature sono responsabili della divisione dei campioni su PC1. Le temperature misurate sono direttamente correlate alla portata d'acqua, e avendo PC1 positiva spostano i campioni verso destra in Figura 3.4b;
- le variabili relative al momento torcente [47,48,49] sono strettamente correlate e sono responsabili della discriminazione dei lotti lungo PC2;
- la variabilità descritta da PC1 (31.0%) è quantitativamente uguale a quella descritta da PC2 (30.7%);
- non c'è una distinzione netta tra i vari tipi di materiale, cioè la variabilità dovuta al materiale non è associata alla variabilità delle condizioni al granulatore;
- le proiezioni del diagramma degli score non presentano traiettorie temporali;
- le temperature [51,52,53] sono molto correlate tra loro;
- dal diagramma dei *loading* (Figura 3.4b), la velocità del motore [54] non sembra essere importante per discriminare i lotti;
- la temperatura relativa alla sezione di ingresso del solido [50] è poco correlata alle altre e ha poco peso su PC2;

Come accennato, il modello creato sui dati relativi allo stato stazionario della granulazione non identifica sul diagramma degli *scores* traiettorie temporali. È stata perciò eseguita un'ulteriore analisi PCA *sulla media* di ciascuna misura lungo tutti gli istanti di campionamento, per comprendere eventuali somiglianze tra i lotti sottoposti alla granulazione. In quest'analisi è stata inserita anche la variabile relativa alla percentuale d'acqua presente in ciascun lotto alla fine della granulazione [13] che, essendo proporzionale alla portata d'acqua in ingresso al granulatore, può essere correlata alle variabili monitorate durante la granulazione.

Le prime due componenti principali del modello sono state presentate nello *scores plot* di Figura 3.5a ed è molto più facile capire le somiglianze tra i vari lotti rispetto alle proiezioni in Figura 3.4a. La prima componente principale descrive le differenze tra i lotti {2,4,5,7,8,13} dai lotti {1,3,6,9,12}; queste differenze sono legate alle temperature rilevate [50,51,52,53] e/o alle diverse quantità d'acqua presenti nei vari lotti [13] all'uscita del granulatore.

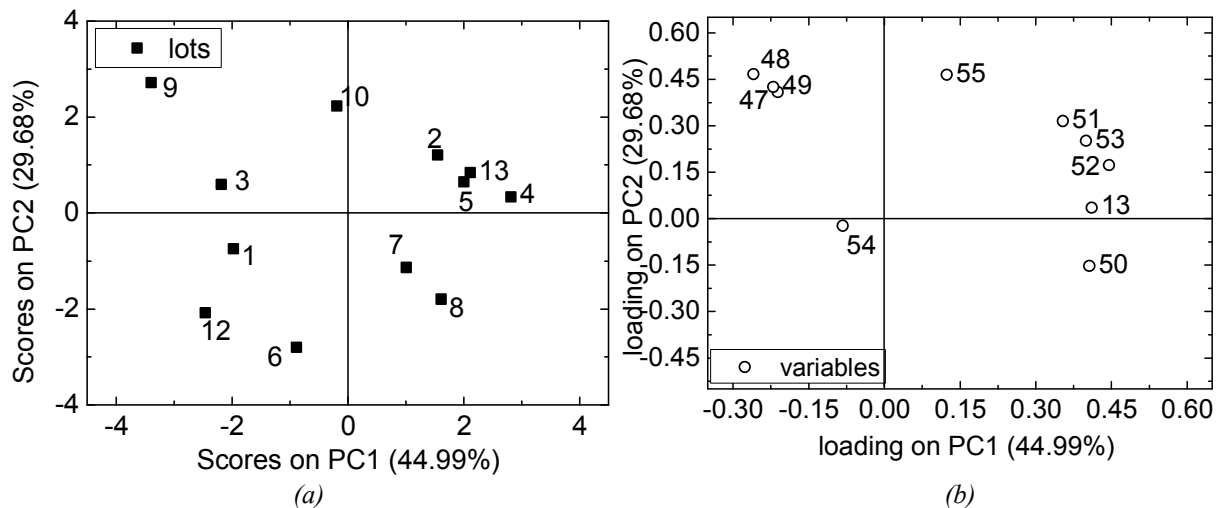


Figura 3.5 Modello PCA relativo ai dati medi delle variabili monitorate al granulatore. (a) Score plot; (b) loading plot.

Infatti, nel *loading plot* Figura 3.5b le variabili relative alle temperature [50,51,52,53] hanno PC1 elevata e sono strettamente correlate alla portata d'acqua e quindi alla quantità d'acqua presente in ogni lotto all'uscita del granulatore [13], come era stato ipotizzato nell'analisi precedente.

La seconda componente principale spiega le differenze tra i lotti {2,9,10} con *motor torque* e portata entrante di solido elevata, e i lotti {6,8,12} con *motor torque* basso. Nel *loading plot* le variabili relative al *motor torque* [47,48,49] hanno elevata PC2 e sono strettamente correlate, come nell'analisi precedente.

3.1.3. Analisi dei dati relativi alla corrente all'uscita del mulino (Y_1)

Le misure di qualità all'uscita del mulino (corrente 6 di Figura 2.1) sono state analizzate per comprendere le correlazioni tra le variabili in uscita e capire se le differenze o somiglianze tra i lotti vengono mantenute. Il lotto {11} non è stato macinato e quindi non è incluso in questo tipo di analisi.

Le variabili di qualità dei dodici lotti macinati della Tabella 2.7 sono state analizzate utilizzando tre componenti principali, e le prime due sono riportate in Figura 3.6. L'analisi delle variabili di qualità all'uscita del mulino permette di osservare che non ci sono dei cluster ben definiti nello *score plot* di Figura 3.6a, e quindi la variabilità dovuta alla granulazione porta ad una diversa posizione dei lotti nel piano rispetto all'analisi dei materiali di Figura 3.2a, perché i parametri al mulino sono costanti per tutti i lotti. Lo *step* di granulazione modifica quindi la sorgente dominante di variabilità rilevata nelle materie prime.

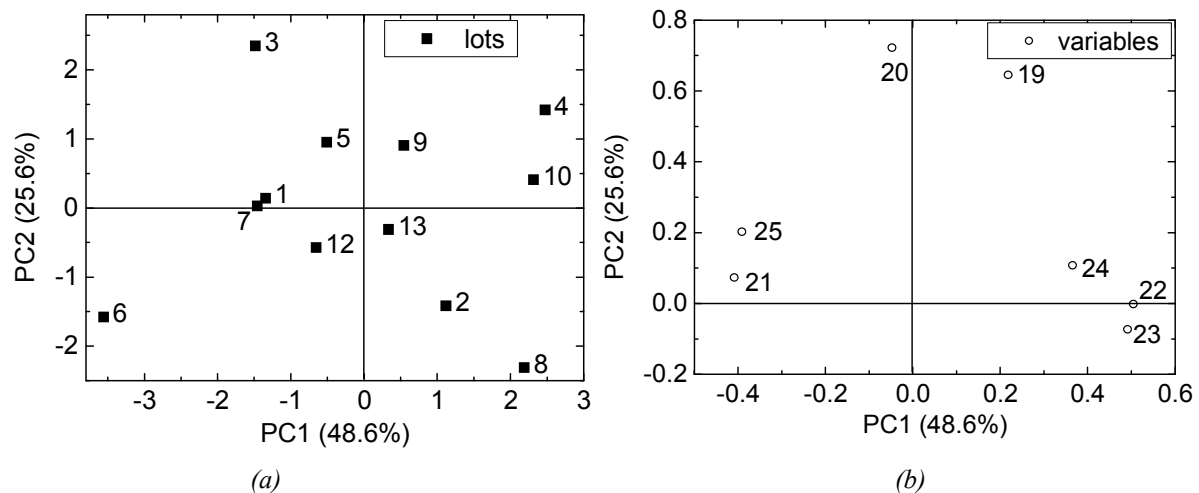


Figura 3.6 Modello PCA sui dati relativi alla qualità in uscita al mulino, matrice Y_1 . (a) Scores plot; (b) loading plot.

I lotti di uno stesso materiale sono tra loro distanziati; per esempio, i lotti {9} e {10} hanno condizioni iniziali pari a quelle del materiale {M4}, ma la diversa quantità d'acqua con la quale {M4} viene processato fa in modo che le proiezioni dei lotti {9} e {10} siano distanziate in Figura 3.6a.

Inoltre è possibile distinguere le differenze dovute alle condizioni iniziali lungo la seconda componente principale; infatti, i lotti *wet milled* {3,4,5,9,10} hanno PC2 elevata mentre quelli *microfluidised* {1,2,6,7,8,12,13} hanno PC2 negativa.

Il *loading plot* di Figura 3.6b permette di osservare lungo la prima componente principale che le variabili relative alla distribuzione del diametro delle particelle [22,23,24] sono anticorrelate al rapporto di Hausner e allo *span* [21,25]. La seconda componente principale discrimina i lotti sulla base della loro densità [19,20].

Le condizioni iniziali influenzano la qualità all'uscita del mulino; infatti, i lotti *wet milled* {3,4,5,9,10} mantengono densità elevate mentre quelli *microfluidised* hanno densità basse. Inoltre i lotti *wet milled* {4,10} hanno distribuzioni elevate [22,23,24] e *span* [25] bassi, come i lotti *microfluidised* {2,8}. Questi lotti {2,4,8,10} hanno tutti una quantità d'acqua all'uscita del granulatore pari al 17.5%, e quindi è possibile ipotizzare che materiali sottoposti a granulazione con portate d'acqua elevate possano portare alla formazione di granuli più grandi e con distribuzioni delle particelle più strette.

3.1.4. Analisi dei dati relativi alla fase di compattazione (Y_2)

I dati di compressione (matrice Y_2 di Tabella 2.8) sono stati analizzati mediante PCA per cercare le correlazioni tra le variabili e capire meglio lo *step* finale del processo.

Sono state eseguite più analisi per comprendere tutte le correlazioni:

- analisi della matrice Y_2 ;

- analisi dei singoli cluster trovati nell'analisi della matrice Y_2 ;
- analisi dei lotti della matrice Y_2 provenienti da un singolo materiale.

La matrice Y_2 contiene le misure relative alle prove di compressione effettuate su cinque lotti dei tredici in uscita al mulino (corrente 6 di Figura 2.1) aventi la stessa quantità d'acqua (17.5%). La durata delle prove è costante e le variabili operative sono la distanza del punzone [26] e la profondità di riempimento [27]. In Figura 3.7a è stato riportato il diagramma degli *scores* sulle prime due componenti principali, mentre in Figura 3.7b quello dei *loading*.

L'analisi ha evidenziato alcuni cluster lungo la diagonale dello *scores plot*, primo-terzo quadrante (Figura 3.7a). Ogni cluster è rappresentato con un simbolo diverso e corrisponde a prove effettuate con la stessa distanza del punzone (da 3 a 4.2 mm). Per esempio, i campioni del cluster cerchiato in nero sono stati creati con distanza del punzone pari a 3 mm.

I risultati dell'analisi effettuata senza le variabili [26] e [27] sono gli stessi. Ciò indica che le condizioni operative nella fase di compattazione sono molto importanti e determinano le proprietà delle compresse introducendo specifiche correlazioni tra le variabili.

Dal diagramma dei *loading* si vede che la distanza del punzone [26] è correlata allo spessore della compressa [29] e al grado di plasticità [38], cioè al rapporto tra l'energia non recuperata e l'energia totale, mentre la profondità di riempimento [27] è correlata al peso della compressa [28].

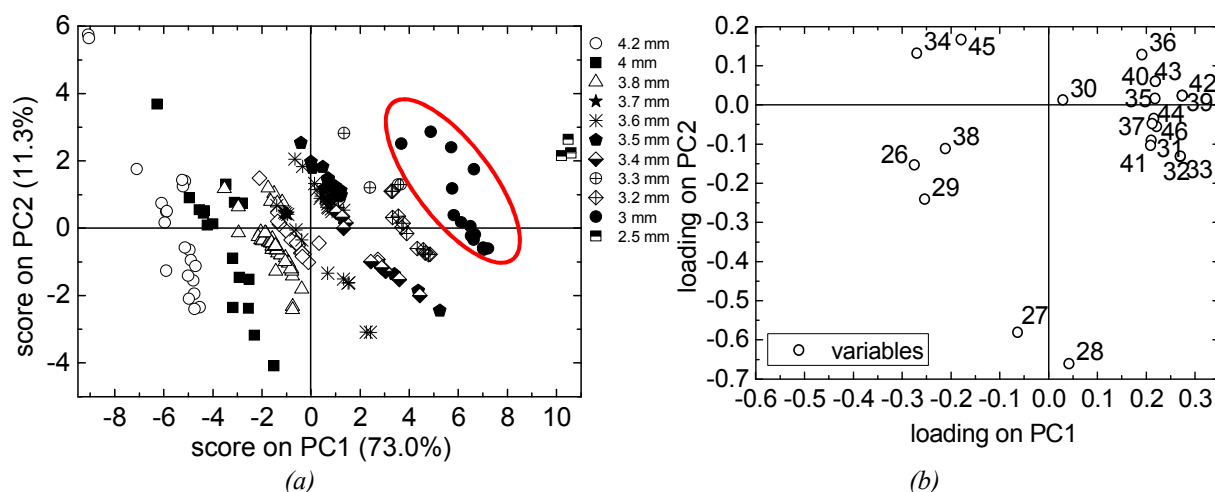


Figura 3.7 Modello PCA dei dati relativi alla compattazione, matrice Y_2 . (a) Score plot per 201 campioni; (b) loading plot per 21 variabili. Le diverse distanze del punzone sono contrassegnate con simboli diversi.

Le prove effettuate con una distanza del punzone elevata portano alla formazione di compresse ad elevato spessore e con un elevato grado di plasticità. Inoltre nel *loading plot* si osserva che lungo la PC1 la plasticità [38] è anticorrelata alla densità [32] e alla durezza [31], mentre lungo la PC2 il peso della compressa [28] è anticorrelato alla porosità [35]. Questo significa che ci si aspetta che compresse più dense e più compatte siano anche meno plastiche e che compresse più pesanti siano anche meno porose.

Tutte le altre variabili sono strettamente correlate alla densità [32] come si vede in Figura 3.7b e ciò determina la discriminazione dei campioni lungo PC1.

Per comprendere le fonti di variabilità del sistema, prescindendo dalle variabili manipolate durante le prove di compattazione, è stata eseguita un'analisi sul cluster a distanza minore del punzone, pari a 3 mm (cerchi neri di Figura 3.7). I risultati dell'analisi sono presentati in Figura 3.8, dove sono riportate le prime due componenti principali dello *scores plot* (Figura 3.8a) e del *loading plot* (Figura 3.8b). La numerazione dei campioni in Figura 3.8a si riferisce al numero del campione nella matrice contenete i dati di compattazione Y_2 .

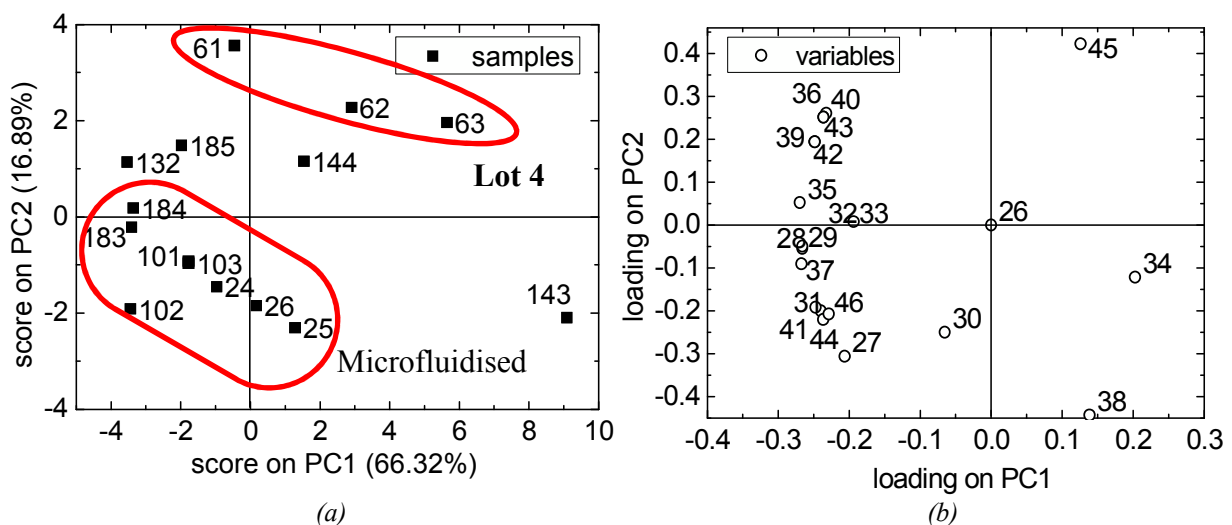


Figura 3.8 Modello PCA del cluster di dati della matrice Y_2 con distanza del punzone pari a 3 mm. (a) Score plot per 15 campioni; (b) loading plot.

È possibile osservare che:

- la distanza del punzone [26] non ha effetto sulla variabilità; ciò è ragionevole, visto che le prove (benché provenienti da lotti diversi), sono state effettuate alla stessa distanza del punzone, pari a 3 mm;
- nel diagramma degli *scores*, i campioni sono divisi per materiale, anche se le variabili analizzate sono solo di compressione e non includono proprietà dei materiali;
- le misure effettuate sullo stesso lotto, come per esempio i campioni {24,25,26} del lotto {1}, dovrebbero essere identiche e le proiezioni sul diagramma degli *score* dovrebbero essere vicine, perché la distanza del punzone e la profondità di riempimento sono costanti. Quindi la differenza che si nota nello *score plot* tra questi campioni è dovuta a una fonte di variabilità esistente, ma non descritta dal modello (perché ne manca una misura rappresentativa);
- i campioni {61,62,63} del lotto {4} hanno proprietà diverse e proiezioni diverse nello *score plot* perché sono stati effettuati a differenti profondità di riempimento [27];

- i campioni di uno stesso lotto sono disposti lungo la diagonale secondo-quarto quadrate secondo valori di porosità [34] e di plasticità [38] crescenti. Infatti tali variabili hanno PC2 negativa e PC1 positiva in Figura 3.8b.
- si osserva una certa distinzione tra le misure effettuate su lotti “*wet milled*” o “*microfluidised*”, anche se le misure sono esclusivamente di compressione e non ci sono quindi variabili che identifichino tale differenza. Inoltre, si può osservare che le compresse ottenute da lotti *microfluidised* sono più resistenti, poiché hanno valori di *hardness* [31] e resistenza a trazione [46] elevati.

In Figura 3.9 sono riportati lo *score plot* e il *loading plot* per la seconda e la terza componente principale.

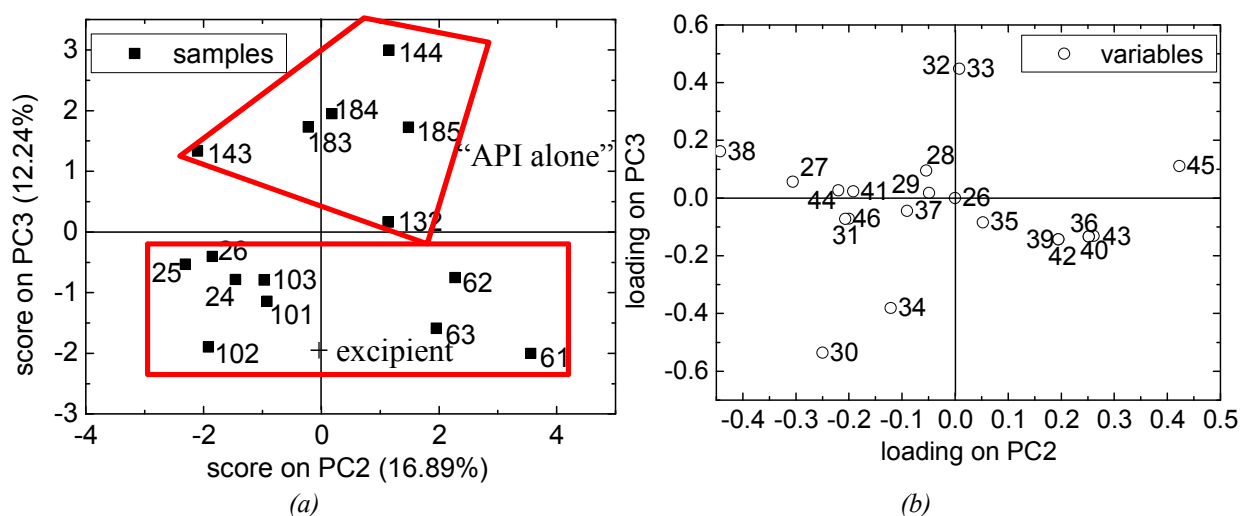


Figura 3.9 Modello PCA del cluster di dati della matrice Y_2 con distanza del punzone pari a 3 mm. (a) Score plot PC3-PC2; (b) loading plot PC3-PC2.

Nel diagramma degli *scores* di Figura 3.9a, i campioni {132,143,144} e i campioni {183,184,185} hanno PC3 positiva e sono tutti “*API alone*”, mentre gli altri sono “+ *excipient*”. I campioni “*API alone*” sono caratterizzati da densità elevata [32] e diametri delle compresse [30] inferiori alla media. Nel *loading plot* si osserva che quasi tutte le variabili non hanno peso su PC3 e che la terza componente principale descrive le differenze dovute alle proprietà delle compresse come diametro [30], porosità [34], densità [32] e densità relativa [33]. Alcune proprietà delle compresse come diametro [30], densità [32] e porosità [34] sono dovute quasi esclusivamente alla modalità con cui il principio attivo è stato additivato.

È stata successivamente effettuata un’analisi sulle misure di compressione realizzate su un unico lotto in uscita al mulino, il lotto {1}, e le prime due componenti principali del diagramma degli *scores* e di quello dei *loading* sono riportate in Figura 3.10. Ogni campione è stato rappresentato con un simbolo diverso in base alla distanza del punzone con la quale è stata effettuata la misura.

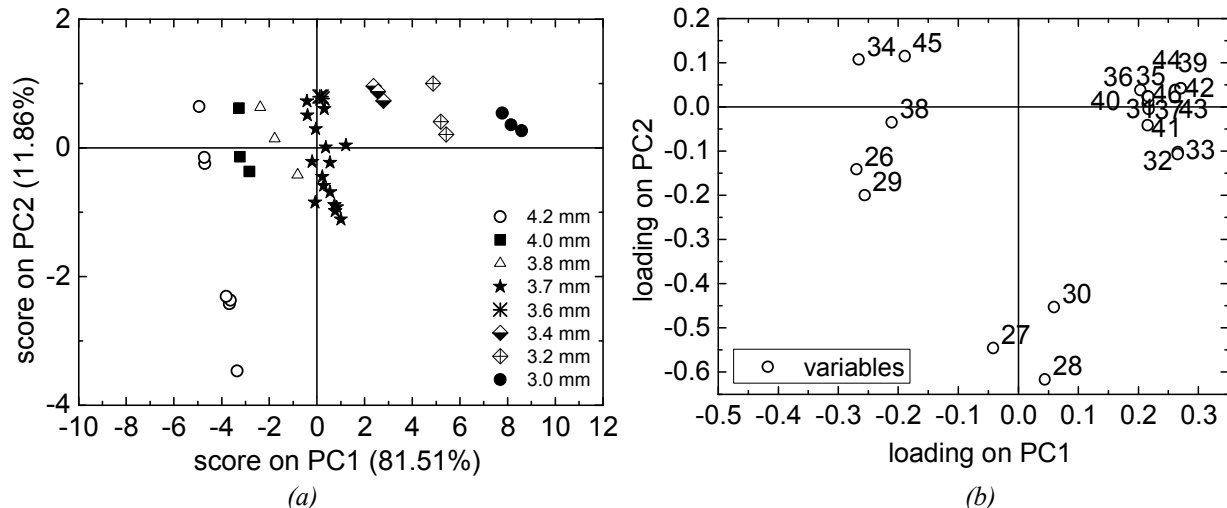


Figura 3.10 Modello PCA sulle misure di compattazione effettuate sul lotto {1} in ingresso alla comprimitrice. (a) Score plot; (b) loading plot.

I risultati di quest'analisi sono:

- in Figura 3.10a le misure effettuate con la stessa distanza del punzone sono molto vicine lungo la prima componente principale, come è stato riscontrato nell'analisi di tutti i lotti di Figura 3.8a.
- i campioni discriminati lungo PC2 dello *score plot* sono stati effettuati a diverse profondità di riempimento. Infatti i campioni con PC2 negativa hanno alta profondità di riempimento [27];
- nel *loading plot*, la profondità di riempimento [27] è strettamente correlata al peso [28] e al diametro [30] della compressa;
- la distanza del punzone [26] è correlata allo spessore della compressa [29], al grado di plasticità [38] e in modo minore alla porosità [34] e allo *stress transmission ratio* [45]. La variabile [45] è il rapporto tra le misure di sforzo alla massima e minima distanza del punzone;
- la PC1 descrive l'anticorrelazione tra le variabili [26,29,34,38,45] e tutte le altre. Le compresse create con distanze del punzone elevate presentano spessori, porosità, plasticità e sforzo di trasmissione elevati.

3.1.5. Analisi congiunta dei dati relativi alle condizioni iniziali, all'uscita del granulatore e all'uscita del mulino

Le condizioni iniziali (matrice Z^* di Tabella 2.1), i dati all'uscita del granulatore (matrice X_2 di Tabella 2.6) e i dati all'uscita del mulino (matrice Y_1 di Tabella 2.7) sono stati raggruppati nella matrice B in Figura 3.11 per analizzare la correlazione tra le variabili relative alle diverse tipologie di materiale, la qualità dei prodotti all'uscita del granulatore e all'uscita del mulino. Nella matrice B , ad ogni lotto all'uscita del granulatore è stata associata la propria

condizione iniziale, matrice Z^* , che differisce dalla matrice Z perché ha tredici righe (lotti) e non cinque (materiali). I tredici lotti considerati si distinguono per le cinque tipologie di materiale di partenza e le tre diverse quantità d'acqua addizionate al granulatore.

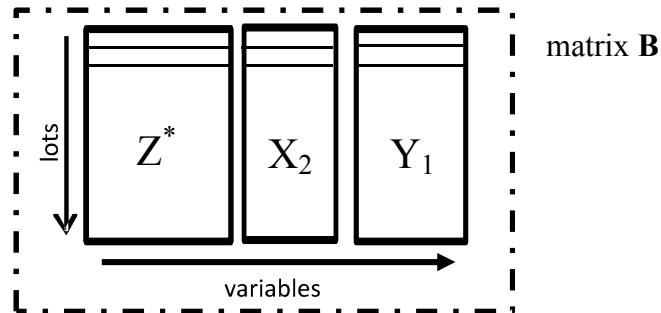


Figura 3.11 Riorganizzazione dei dati nella matrice B (13x25)

La matrice B è stata analizzata con l'algoritmo NIPALS perché permette di studiare matrici con *missing data* senza stimare i dati mancanti.

I risultati di quest'analisi relativi alle prime due componenti principali sono riportati in Figura 3.12 e possono essere così riassunti:

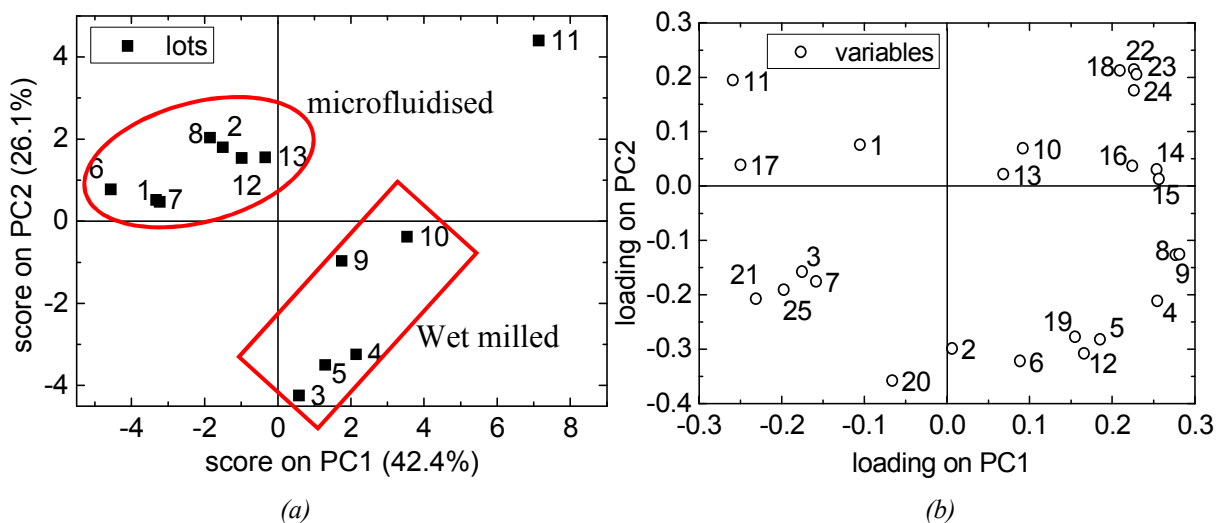


Figura 3.12 Modello PCA della matrice B . (a) Score plot; (b) loading plot.

- nello *score plot* (Figura 3.12a), i lotti sono divisi in due cluster: “*wet milled*” {3,4,5,9,10,11} e “*microfluidised*” {1,2,6,7,8,12,13}. Complessivamente, i lotti sembrano quindi comportarsi in modo diverso principalmente a seconda del metodo di isolamento del principio attivo;
- i lotti creati a partire da materiali additivati al punto di isolamento (da {1} a {8}) sono distribuiti in entrambi i cluster. La modalità con la quale l'API viene miscelato agli additivi sembra quindi meno importante delle modalità di isolamento (cioè “*wet milled*” piuttosto che “*micorfluidised*”);
- il cluster di lotti “*wet milled*” è diviso in due sotto gruppi:

- i lotti {3,4,5} sono stati additivati al punto di isolamento [2], hanno densità elevata [5,6,19,20] e un elevato *feed factor* [12]. Quindi se un materiale *wet milled* viene additivato al punto di isolamento, questo mediamente si associa ad un aumento della densità e quindi della portata massima entrante al granulatore;
- i lotti {9,10} sono stati additivati in una fase successiva (“API alone”) e sono caratterizzati da elevata distribuzione iniziale [8,9,10] e all’uscita del mulino [14,15,16]. Quindi se un materiale viene additivato alla fine dei trattamenti iniziali, di solito è caratterizzato da granuli di dimensioni elevate;
- i lotti “*microfluidised*” {1,2,6,7,8} sono caratterizzati, lungo PC1, da span elevati [11,17] e *Hausner ratio* elevati [7,21];
- i lotti {3, 4, 5, 9, 10} sono tutti “*wet milled*” e sono caratterizzati da valori elevati di distribuzione iniziale [8, 9], di densità iniziale e all’uscita al mulino [5,6,19, 20] e di *feed factor* [12], come risulta nel *loading plot*;
- i lotti provenienti dallo stesso materiale di partenza sono molto vicini. Per esempio, i lotti {3,4,5} sono caratterizzati dallo stesso materiale di partenza;
- i lotti {12,13} sono più vicini all’origine, segno che sono molto vicini alle condizioni “medie” del sistema;
- il lotto {11} si trova molto distanziato dagli altri lotti nello *score plot*, poiché è l’unico che non è stato sottoposto a macinazione. Tale lotto presenta valori di distribuzione [22, 23, 24] in uscita al mulino molto più elevati. Infatti, nel *loading plot* di Figura 3.12b tali variabili si trovano nella posizione corrispondente a quella del campione nello *score plot*. Inoltre, il lotto presenta bassi valori di *Hausner ratio* [21] e di *span* della distribuzione in uscita al mulino [25], che nel *loading plot* si trovano dalla parte opposta alle variabili relative alla distribuzione;
- nei cluster si osserva che i lotti con PC elevate hanno distribuzioni elevate [14,15,16,22,23,24]. Nel *loading plot* queste variabili hanno valori positivi di PC1 e PC2;
- nel *loading plot*, il quantitativo d’acqua [13] non sembra essere molto importante sulle prime due PC nella discriminazione dei lotti poiché si trova abbastanza vicino all’origine;
- la densità misurata all’inizio del processo [5,6] è correlata a quella misurata all’uscita del mulino [19,20]; ciò significa che il processo sostanzialmente non modifica questa proprietà del materiale;
- il *feed factor* [12] è strettamente correlato alla densità aerata in ingresso e all’uscita del mulino [5,19]. Tale variabile definisce una proprietà di scorrimento del materiale, ed è lecito aspettarsi che sia correlata alle densità del materiale in ingresso e in uscita al processo;

- nel *loading plot*, su PC1 si nota l'anticorrelazione tra due cluster di variabili, che risultano molto distanti fra loro. Il primo gruppo è formato da: “*raw*” *span* [11], PIMS *span* [17], *output Hausner ratio* [21] e *output span* [25], mentre il secondo gruppo dalla distribuzione delle particelle iniziale, all'uscita del granulatore e all'uscita al mulino [8,9,14,15,16,22,23,24] e dalla caratteristica “*wet milled*” [4]. L'aumento della differenza tra i valori delle variabili di ciascun gruppo dovrebbe portare all'aumento della separazione dei due cluster di lotti lungo la PC1 nello *score plot*;
- nel *loading plot*, su PC2 si nota l'anticorrelazione tra la densità battuta [20] e la distribuzione [22,23,24] in uscita al mulino. Quindi si può ipotizzare che lotti con dimensioni elevate delle particelle abbiano anche distribuzioni “strette” e più omogenee;
- la massa del materiale iniziale [2] ha un peso negativo lungo la seconda componente principale, determinando uno spostamento verso il basso dei lotti con massa elevata {3,4,5} come si vede in Figura 3.12a.
- il rapporto di Hausner in uscita al mulino [21] è correlato allo *span* della in uscita al mulino [25].

I risultati dell'analisi di PC2 e PC3 sono (Figura 3.13):

- la terza componente principale spiega la differenza tra i lotti {9,10,12,13} e gli altri cioè tra i diversi metodi di miscelare API con gli eccipienti. La variabile “+ *excipient*” [3] ha una PC3 negativa molto bassa, quindi è importante nella discriminazione dei lotti. Questa variabilità descritta dalla terza componente principale è però solo circa il 14% di quella complessiva;
- i lotti dello stesso materiale si differenziano lungo la PC3 per il contenuto d'acqua [13]. Infatti [13] risulta essere piuttosto importante su PC3 come si nota dal *loading plot* di Figura 3.13b. Per esempio i lotti {6,7,8} hanno quantità d'acqua crescenti;
- i lotti con PC3 negativa, tutti “+ *excipient*” tranne {11}, si dividono in due gruppi {3,4,5} e {1,2,6,7,8}, e tale suddivisione è dovuta al diverso metodo di isolamento del principio attivo. Infatti, i lotti {3,4,5} sono “+ *excipient*” e “*wet milled*”, mentre gli altri sono “+ *excipient*” e “*microfluidised*”. I lotti {3,4,5} hanno un valore di massa [2], densità [5,6,7] e “*feed factor*” [12] elevati, e nel *loading plot* queste variabili hanno valori di PC2 negativi;

- i lotti {9,10,12,13} “API alone” hanno un’elevata distribuzione iniziale [8,9,10] e un’elevata densità in uscita al mulino [19,20,21].

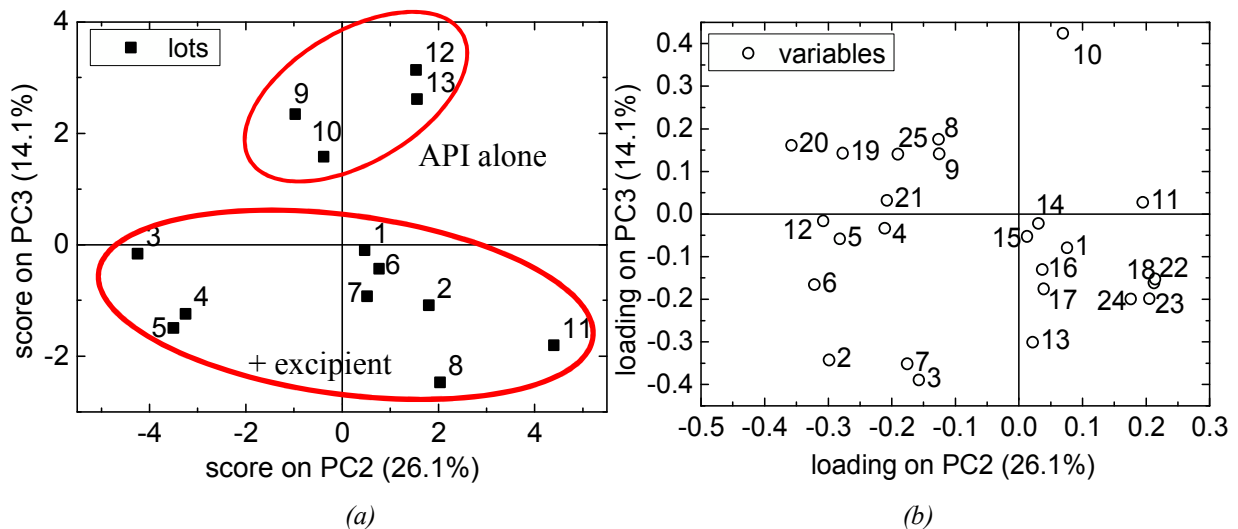


Figura 3.13 Modello PCA della matrice **B**. (a) Score plot PC3 vs PC2 per 13 lotti; (b) loading plot PC3 vs PC2.

3.1.5.1. Relazione tra due sistemi per la misura della distribuzione granulometrica

I dati della matrice **B** non sono completi, perché per alcuni dei lotti considerati mancano misure di densità in ingresso e di distribuzione delle particelle in uscita al granulatore.

Sono state eseguite tre analisi per confermare la stretta correlazione tra le distribuzioni misurate in punti diversi del processo eliminando il terzo, il quarto e il quinto lotto che presentano dati mancanti in corrispondenza della distribuzione in uscita al granulatore:

- utilizzando tutte le variabili ($\mathbf{B}_{3,4,5}$);
- eliminando le variabili relative alla distribuzione in uscita al granulatore ($\mathbf{BP}_{3,4,5}$);
- eliminando le variabili relative alla distribuzione in uscita al mulino ($\mathbf{BS}_{3,4,5}$).

Questo approccio permette di comprendere la correlazione tra la distribuzione misurata all’uscita del granulatore e quella misurata all’uscita del mulino mediante il confronto dei tre *score plot*. Se le proiezioni dei lotti si trovano nella stessa posizione in tutte le analisi significa che le misure di distribuzione eseguite all’uscita del granulatore aggiungono la stessa informazione delle misure di distribuzione eseguite all’uscita del mulino.

Le tre analisi sono state eseguite per confermare la forte correlazione tra le misure di distribuzione all’uscita del mulino e quelle eseguite all’uscita del granulatore, eliminando i lotti {3,4,5} che hanno dati mancanti. In questo modo l’analisi considera lo stesso numero di lotti “+ excipient” e “API alone” dando maggiore importanza alla variabilità dovuta alla modalità di additivazione piuttosto che alla variabilità dovuta al metodo di isolamento.

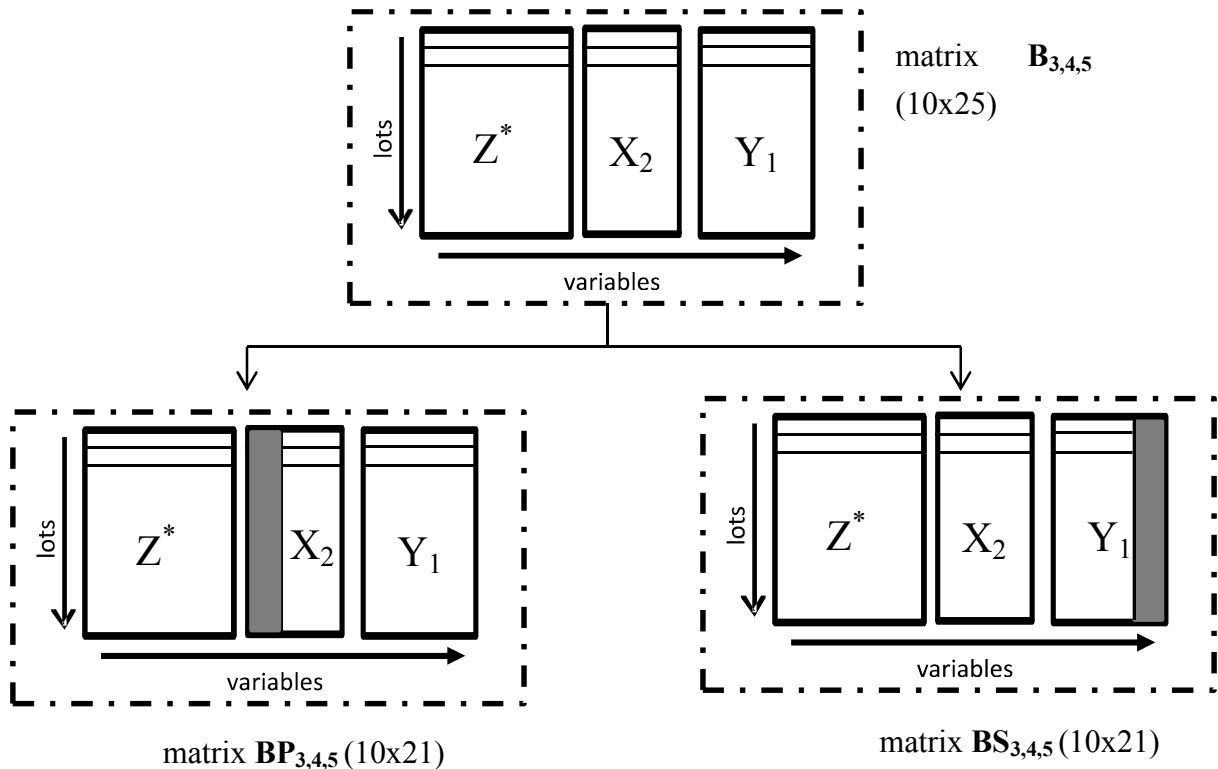


Figura 3.14 Matrici utilizzate nell'analisi: matrice $B_{3,4,5}$ è il caso generale con tutte le variabili, matrice $BP_{3,4,5}$ non contiene la distribuzione fornita dal PIMS e matrice $BS_{3,4,5}$ non contiene la distribuzione in uscita al mulino.

Le tre analisi confermano che le misure delle distribuzioni sono strettamente correlate; infatti, la posizione dei lotti nello *score plot* è la stessa nei tre casi e viene riportato solo il risultato della prima analisi ($B_{3,4,5}$) in Figura 3.15. Nella Figura 3.15, relativa al caso generale, si osserva che i lotti sono suddivisi in base alla modalità con la quale il principio attivo è additivato. I lotti {1,2,6,7,8} sono “+ excipient” mentre i lotti {9,10,12,13} sono “API alone”.

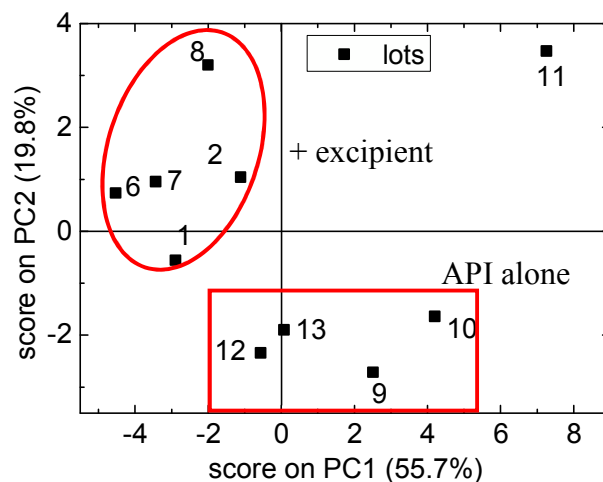


Figura 3.15 score plot PC2 vs PC1 dell'analisi della matrice $B_{3,4,5}$

In questo caso non è possibile identificare la principale forza motrice che caratterizzava le analisi precedenti, cioè la differenza tra lotti *microfluidised* e lotti *wet milled* probabilmente perché il numero di lotti *wet milled*, è molto più basso, riducendo la variabilità del sistema dovuta ai differenti metodi di isolamento.

3.1.6. Analisi congiunta dei dati relativi alle condizioni iniziali, al granulatore, all'uscita del granulatore e all'uscita del mulino

I dati relativi alle misure stazionarie analizzate nel paragrafo §3.1.2 sono stati mediati per poter effettuare un'analisi congiunta con i dati relativi agli altri *step* del processo. L'analisi delle condizioni iniziali (matrice Z^*), delle misure al granulatore (matrice X_1), all'uscita del granulatore (matrice X_2) e all'uscita del mulino (matrice Y_1) è stata effettuata tramite l'analisi delle componenti principali affiancando le matrici come si vede in Figura 3.16.

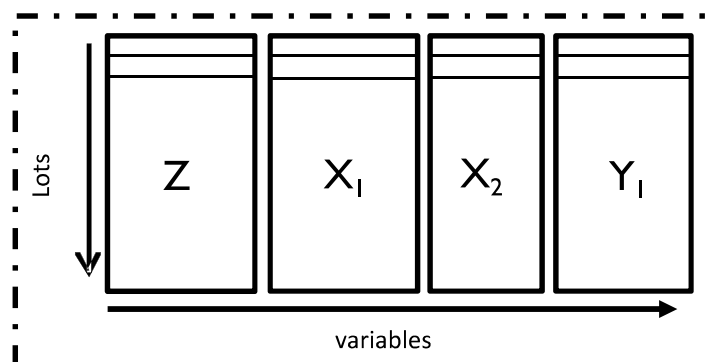


Figura 3.16 Matrici utilizzate nell'analisi, formata da quattro matrici affiancate

Lo scopo dell'analisi è capire le correlazioni tra tutte variabili misurate durante il processo, eventuali relazioni tra materie prime, processo e attributi del prodotto, e comprendere somiglianze tra i campioni dopo l'aggiunta dei dati inerenti il processo. I risultati dell'analisi sono presentati mediante *score plot* e *loading plot* in Figura 3.17 e 3.18 per le prime tre componenti principali. In particolare, in Figura 3.17 sono stati riportati i risultati sulle prime due PC e in Figura 3.18 i risultati su PC2 e PC3. Nei *loading plot* le variabili relative al granulatore [47,...,55] sono rappresentate con il simbolo ● per capire meglio le strutture di correlazione con le altre variabili.

Sulle prime due componenti principali (che descrivono circa il 55% della variabilità totale dei dati) si nota ancora una distinzione tra i lotti dovuta al metodo di isolamento della materia prima. Infatti nella Figura 3.16a si possono distinguere due cluster: i materiali *microfluidised* e *wet milled*. Nei due *cluster*, i lotti spostati verso PC1 positive e PC2 negative, in Figura 3.16a hanno distribuzione elevate all'uscita del granulatore [14,15,16] e all'uscita del mulino [22,23,24]; inoltre, sono caratterizzati da quantità d'acqua elevate [13] e temperature elevate [52,53]. Quindi la quantità d'acqua addizionata al granulatore influenza la qualità del prodotto aumentando la dimensione dei granuli.

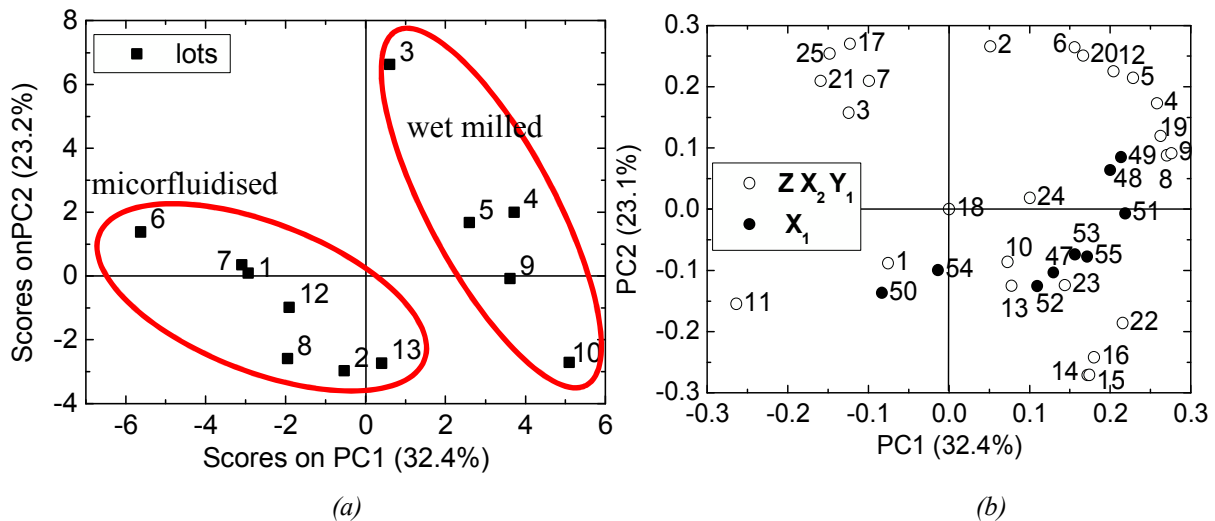


Figura 3.17 Modello PCA sul processo di granulazione, matrici Z, X_1, X_2, Y_1 (a) score plot PC2 vs PC1 (b) loading plot

La terza componente principale (che descrive circa il 20% della variabilità totale dei dati) spiega il contributo dovuto allo *step* di granulazione. In Figura 3.17a si osserva che i lotti sono suddivisi in base alla quantità d’acqua [13], che è strettamente correlata alle temperature [50,51,52,53] probabilmente perché l’aumento del quantitativo d’acqua aumenta il volume di miscela aumentando l’attrito alle pareti e quindi la temperatura.

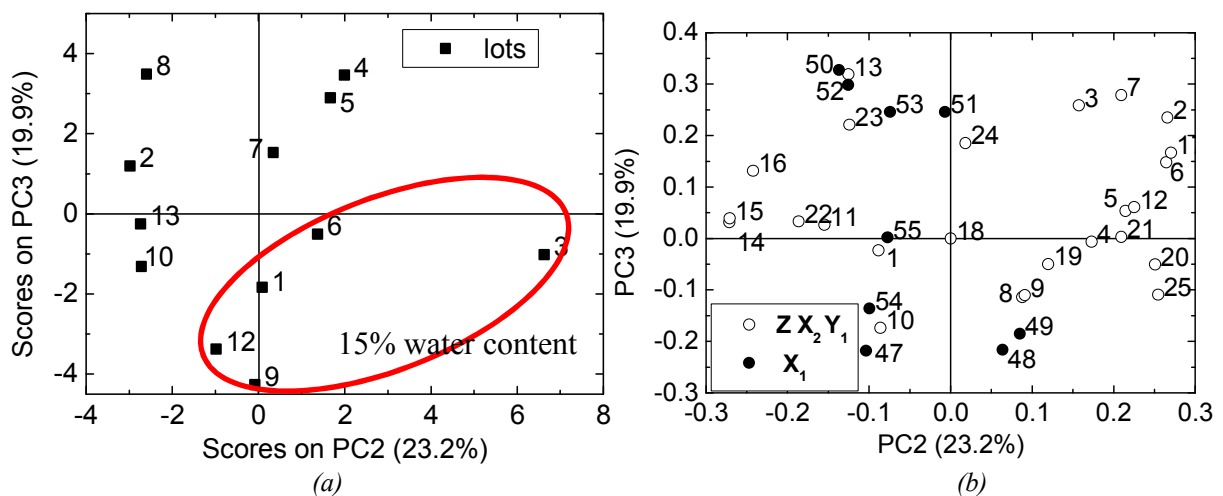


Figura 3.18 Modello PCA sul processo di granulazione, matrici Z, X_1, X_2, Y_1 (a) score plot PC3 vs PC2 (b) loading plot PC3 vs PC2

3.1.7. Analisi congiunta dei dati relativi alle condizioni iniziali, all’uscita del granulatore, all’uscita del mulino e di compressione

Le condizioni iniziali (matrice Z^* di Tabella 2.1), le analisi al granulatore (matrice X_2^* di Tabella 2.6), e le analisi all’uscita del granulatore (matrice Y_1^* di Tabella 2.7) che formano la matrice B sono state allineate con i dati di compressione (matrice Y_2 di Tabella 2.8), formando la matrice C Figura 3.19.

La matrice è composta da 201 campioni, corrispondenti alle diverse prove di compressione effettuate, e 44 variabili. Ad ogni campione della matrice Y_2 sono state associate le condizioni iniziali e quelle di processo, e le variabili utilizzate in quest'analisi sono riportate nelle Tabelle 2.1, 2.6, 2.7, 2.8 del Capitolo precedente.

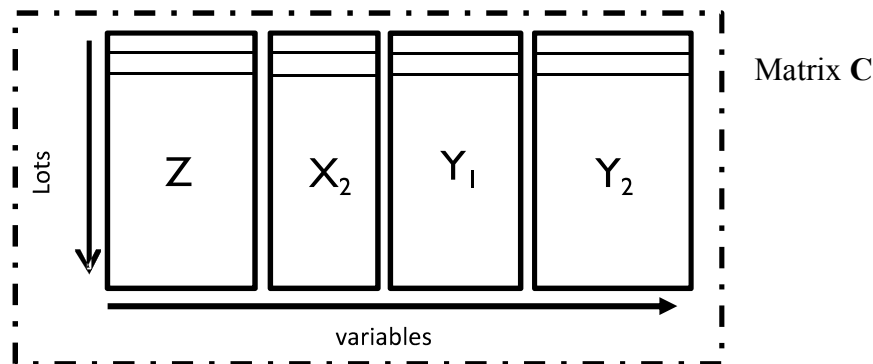


Figura 3.19 Matrice utilizzata nell'analisi, formata da quattro matrici affiancate

La matrice C è costituita da dati relativi a lotti con lo stesso quantitativo d'acqua della miscela all'uscita del granulatore (pari al 17.5%) e sono state eliminate le variabili "quantità d'acqua" [13] e "not milled" [18] della matrice Y_1 e "duration of profile" della matrice Y_2 perché costanti. Il modello PCA è stato costruito per comprendere la struttura di correlazione tra le variabili disponibili senza considerare la granulazione. Inoltre è importante comprendere come la fase di compattazione venga influenzata dalle fasi precedenti.

Nella Figura 3.20 sono riportati lo *scores plot* e il *loading plot* per le prime due componenti principali e i campioni di lotti diversi sono rappresentati con simboli diversi.

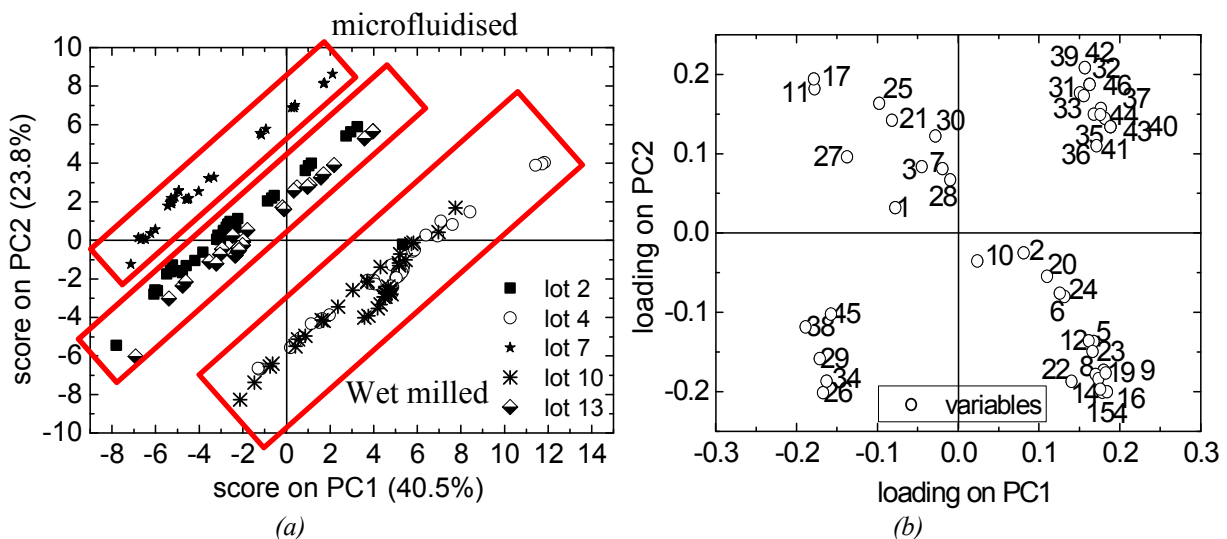


Figura 3.20 Modello PCA sulla matrice C (a) score plot PC2 vs PC1 per 201 campioni, (b) loading plot PC2 vs PC1 per 44 variabili

I risultati di quest'analisi possono essere riassunti:

- i campioni risultano suddivisi in tre cluster lungo la diagonale primo e terzo quadrante di Figura 3.20a:
 - i campioni eseguiti sul lotto {7}, caratterizzati da *span* delle distribuzioni in ingresso, uscita al granulatore e uscita al mulino [11,17,25] elevati, quindi da granuli piccoli ma distribuzione allargata;
 - i campioni eseguiti sui lotti {2} e {13};
 - i campioni eseguiti sui lotti {4} e {10}, caratterizzati da distribuzioni in ingresso, uscita al granulatore e uscita al mulino [8,9,14,15,16,22,23,24] elevate e quindi granuli di dimensioni elevate rispetto agli altri lotti;
- è possibile distinguere due cluster suddivisi in base al tipo di isolamento, “*microfluidised*” {2,7,13} o “*wet milled*” {4,10};

Dall’esame del *loading plot* di Figura 3.20b si nota che:

- le differenze tra i campioni lungo la diagonale tra il secondo e il quarto quadrante è dovuta alle variazioni di condizioni iniziali (matrice \mathbf{Z}), all’uscita del granulatore (matrice \mathbf{X}_2), e all’uscita del mulino (matrice \mathbf{Y}_1). Infatti in Figura 3.20b le variabili di tali matrici si trovano su questa diagonale;
- le differenze tra i campioni all’interno di un cluster, lungo la diagonale tra il primo e il terzo quadrante, è dovuta alla variazione delle variabili di compressione (matrice \mathbf{Y}_2).

3.2. Modelli per la stima delle caratteristiche dei granuli all’uscita del mulino

In questo paragrafo si descrive un approccio di tipo PLS e MBPLS per la stima delle caratteristiche dei granuli all’uscita del mulino, che sono state considerate come indice di qualità del prodotto dopo la macinazione. Nella costruzione di un modello PLS è molto importante scegliere il numero di variabili latenti ottimale. Li *et al.* (2002) hanno descritto vari metodi che possono essere utilizzati nella scelta del numero di variabili latenti. Poiché il numero di lotti considerati in questa analisi è molto piccolo, il numero di variabili latenti per ciascun modello è stato scelto valutando il minimo dell’errore quadratico medio in convalida incrociata calcolato secondo il metodo *leave one out*.

3.2.1. Analisi delle relazioni tra variabili operative

Le condizioni iniziali (matrice \mathbf{Z}^*) riportate in Tabella 2.1 e l’unica variabile operativa dello *step* di granulazione, cioè la portata d’acqua in ingresso al granulatore, sono state utilizzate nella predizione della qualità in uscita al mulino (matrice \mathbf{Y}_1) mediante PLS. In Figura 3.21 è riportata la struttura delle matrici utilizzate nell’analisi.

In questo caso non è stato considerato il lotto {11} perché non è stato macinato e presenta le stesse caratteristiche del lotto {10} prima del mulino, e sarebbe molto difficile per il modello predire la qualità all'uscita del mulino.

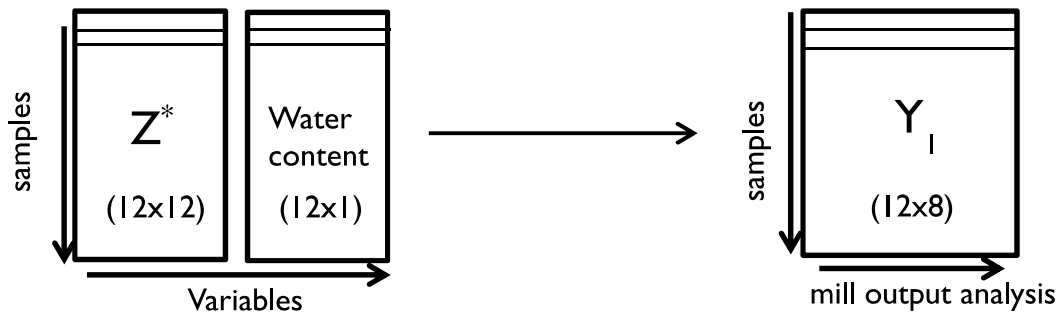


Figura 3.21 Struttura delle matrici analizzate con PLS

Le varianze spiegate medie del modello in calibrazione e in convalida sono pari a 0.62 e 0.227. Tali valori sono bassi, sottolineando le scarse prestazioni del modello soprattutto in convalida incrociata forse per la scarsa quantità di dati con cui è stato costruito il modello.

È stato effettuato uno studio per comprendere il peso di ciascuna variabile manipolata nella predizione della qualità finale del prodotto. Per raggiungere tale scopo, è stato utilizzato un indice complessivo dell'importanza delle variabili in proiezione (*variable importance in projection*, VIP), unico per la matrice Y_1 e distinto per ciascuna variabile predittrice:

$$VIP_{C_a} = \sqrt{p \times \frac{\sum_{i=1}^I (SS_x \times SS_y \times w / \|w\|)^2}{\sum_{i=1}^I (SS_x \times SS_y)}} \quad (3.1)$$

Dove a indica una generica variabile predittrice, p è il numero di variabili di qualità, I è il numero di campioni, SS_x e SS_y sono le varianze spiegate per le due matrici (Z^*w e Y_1), analizzate, e w sono i pesi relativi a ciascuna variabile predittrice nel modello PLS (crf. §1.4.3).

Da Figura 3.22, le variabili più importanti nella predizione della qualità un uscita al mulino sembrano essere la quantità d'acqua [13], il *feed factor* [12], il decimo percentile [8], la mediana [9] e lo *span* [11] della distribuzione iniziale, la densità aerata [5] e la differenza *wet milled – microfluidised* [4].

Nella Figura 3.22, le variabili “altro metodo di isolamento e altro materiale” [1], peso [2], differenza “+ *excipient*”/“*API alone*” [3] e il novantesimo percentile [10] sembrano essere meno importanti nella predizione della qualità all'uscita del mulino. Il novantesimo percentile [10] è la variabile con VIP più basso, anche se nei modelli precedenti era correlato alla distribuzione all'uscita del mulino.

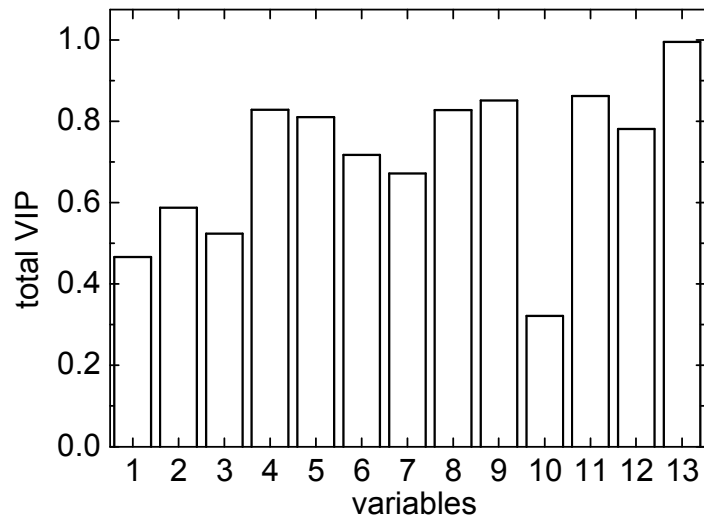


Figura 3.22 VIP complessivo per il modello PLS costruito a partire dalle variabili operative e dalle condizioni iniziali

3.2.2. Analisi multiblock del processo

Per riuscire a determinare gli effetti sulla qualità finale del prodotto, ogni fase dell'intero processo di granulazione deve essere compresa. Quindi, dopo aver determinato che nella predizione della qualità all'uscita del mulino tutte le variabili, condizioni iniziali o operative, devono essere prese in considerazione, è stata effettuata un'ulteriore analisi per comprendere il peso di ciascun blocco nella predizione.

Le condizioni iniziali (matrice Z^*) e i dati medi della fase stazionaria durante la fase di granulazione (matrice X_1) sono stati utilizzati per predire le caratteristiche dei granuli all'uscita al mulino. Nella matrice X_1 c'è anche la variabile relativa alla quantità d'acqua presente nel materiale all'uscita del granulatore, perché è correlata alla portata d'acqua in ingresso.

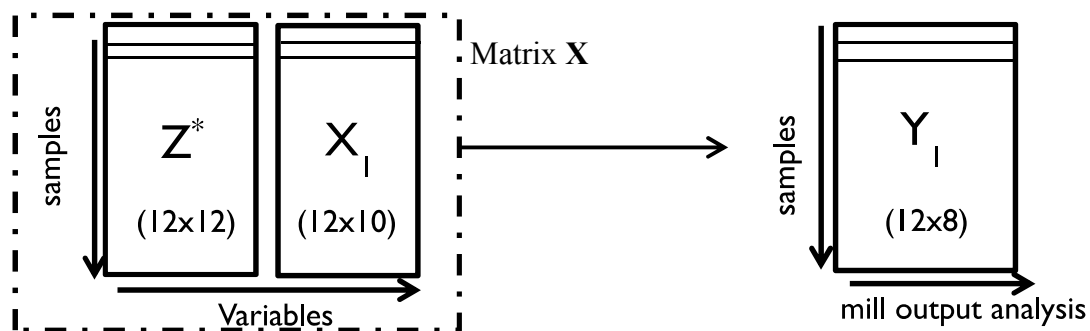


Figura 3.23 Struttura delle matrici analizzate con MBPLS

In Figura 3.23 è rappresentata la struttura del modello MBPLS e non sono state utilizzate le misure (X_2) effettuate all'uscita del granulatore perché le condizioni all'essiccatore e al mulino sono costanti per tutti i lotti e l'alta correlazione tra X_2 e Y_1 , riscontrata nelle analisi

precedenti (§3.1.5), ridurrebbe il peso delle fasi precedenti (condizioni iniziali e granulazione) nella predizione.

Quindi, il modello costruito includendo anche i dati all'uscita del granulatore, matrice (\mathbf{X}_2), sarebbe fortemente influenzato dall'alta correlazione \mathbf{X}_2 - \mathbf{Y}_1 , che farebbe apparire meno importanti gli altri blocchi, \mathbf{Z}^* e \mathbf{X}_1 , mascherandone il reale effetto.

Secondo Westerhuis *et al.* (1998), l'uso di un metodo PLS *multiblock*, mantenendo le variabili divise in blocchi, è utile per aumentare l'interpretabilità del modello e per questa ragione un approccio di tipo *multiblock* è stato utilizzato per questo tipo di analisi.

I blocchi, dopo essere stati sottoposti ad *autoscaling*, vengono divisi per il numero di variabili:

$$\mathbf{X} = [\mathbf{X}_1/m_{x_1}^{1/2}, \dots, \mathbf{X}_B/m_{x_B}^{1/2}], \quad (3.2)$$

dove B è il numero di blocchi e m è il numero di variabili del blocco \mathbf{X}_b . Lo scopo è quello di suddividere la matrice \mathbf{X} (Figura 3.23) in blocchi di ugual peso affinché siano confrontabili, per capire quale blocco e quali variabili di ciascun blocco influenzino la qualità all'uscita del mulino (Liu *et al.*, 2011).

Dopo aver calibrato il modello, è stato calcolato un indice sintetico del peso di ciascun blocco nella predizione di \mathbf{Y}_1 utilizzando l'idea presentata da Yacoub *et al.* (2011) di correlare i pesi di ciascun blocco alla varianza spiegata della matrice predetta.

Il *BIP* (*block importance in projection*) è stato calcolato con la relazione:

$$BIP_b = \sqrt{\frac{\left(B \times \sum_{a=1}^A (SS_y \times SS_b \times WT / \|WT\|)^2 \right)}{\sum_{a=1}^A SS_y \times SS_b}} \quad (3.3)$$

dove b indica il generico blocco, SS è la varianza spiegata per ciascun blocco del processo (b) e della qualità (y), e WT sono i "super weight", cioè i pesi per blocco per variabile latente (1.29), A è il numero di variabili latenti e B è il numero di blocchi della matrice \mathbf{X} .

I *BIP* per blocco riportati in Tabella 3.2 sono molto simili e indicano che le matrici hanno peso simile nella predizione della qualità all'uscita del mulino.

Tabella 3.2 *BIP, block importance in projection, per i due blocchi utilizzati nell'analisi MBPLS*

Block	BIP
Initial condition (\mathbf{Z})	0.707
Sep of granulation (\mathbf{X}_1)	0.757

Tuttavia, è possibile che le variabili monitorate durante la granulazione siano influenzate dalle condizioni iniziali del materiale, facendo sì che la correlazione tra i due blocchi e le variabili misurate all'uscita del mulino sia simile.

Il modello massimizza la correlazione tra le variabili predittrici e quelle predette utilizzando quattro variabili latenti per ottenere la varianza spiegata riportata in Tabella 3.3.

Tabella 3.3 Varianza spiegata dal modello MBPLS per ciascun blocco

Block	Variance explained
Initial condition (Z)	0.837
Sep of granulation (X_1)	0.850
Quality out of mill (Y_1)	0.676

Il vantaggio di utilizzare un approccio *multiblock* sta nel fatto che è possibile ottenere un modello separato per ciascun blocco, ma al tempo stesso legato agli altri blocchi tramite il modello *multiblock* globale. Inoltre tale modello può essere utilizzato per ricostruire il “percorso” di un materiale dalle condizioni iniziali, alla granulazione alla qualità all'uscita del mulino. Lo spazio di ciascun blocco, descritto dalle prime variabili latenti, può essere usato per comprendere come un lotto possa modificare le proprie caratteristiche durante il processo. In Figura 3.24 sono riportati i tre piani degli *scores* descritti dalle prime due variabili latenti relativi alle condizioni iniziali (Figura 3.24a), delle condizioni al granulatore (Figura 3.24b), e della qualità in uscita al mulino (Figura 3.24c).

Come esempio, si segue la traiettoria dei lotti $\{9,10\}$, che sono costituiti da identico materiale:

- in Figura 3.24a i due lotti sono molto vicini, perché sono costituiti dallo stesso materiale e quindi sono entrambi *wet milled* e *API alone*, hanno distribuzioni elevate, densità e *feed factor* abbastanza elevati;
- in Figura 3.24b hanno LV1 negativa e LV2 positiva, e si differenziano per la quantità d'acqua [13] che sposta il lotto $\{10\}$ verso LV2 inferiori. Inoltre, il lotto $\{10\}$ ha momento torcente più elevato e temperature più elevate nella fase di granulazione;
- in Figura 3.24c hanno LV1 e LV2 negativa e sono caratterizzati da densità abbastanza elevate e distribuzioni elevate rispetto agli altri lotti, ma molto diverse tra loro per la diversa quantità d'acqua.

Questo modello potrebbe essere utilizzato, in prospettiva, per stabilire quali devono essere le condizioni al granulatore affinché un lotto raggiunga la qualità desiderata all'uscita del processo, conoscendo le condizioni iniziali. Inoltre, potrebbe essere possibile rilevare anomalie nel processo poiché, conoscendo le condizioni operative normali di processo sarebbe possibile identificare agevolmente anomalie nei valori monitorati durante la fase di

granulazione, utilizzando il modello per scopi di monitoraggio del processo in linea (MacGregor *et al.*, 1994; Choi e Lee, 2005).

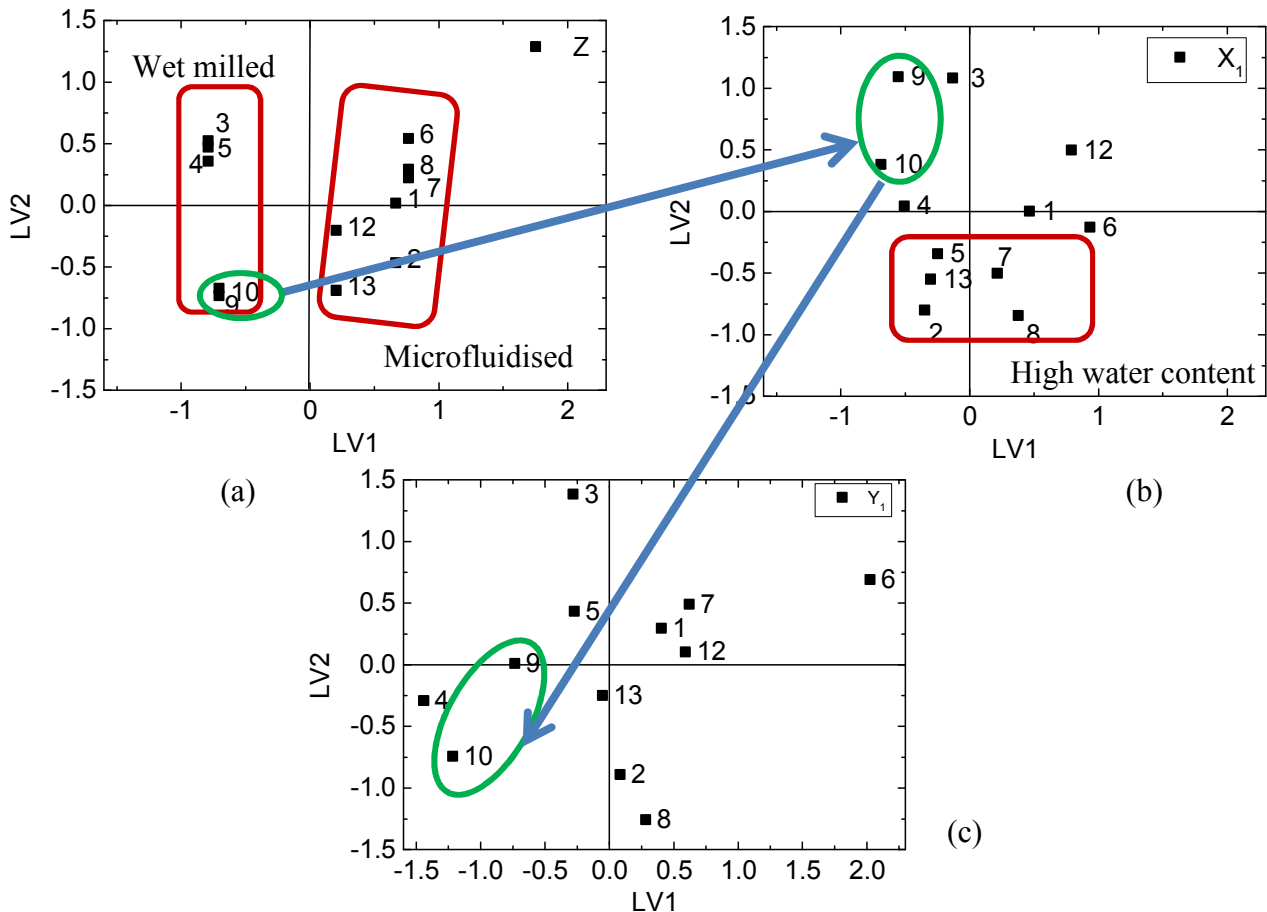


Figura 3.24 Modello MBPLS: (a) score plot delle condizioni iniziali; (b) score plot dello step di granulazione; (c) score plot della qualità all'uscita del processo.

3.2.3. Determinazione dell'importanza di ciascun blocco nella predizione delle caratteristiche dei granuli all'uscita del mulino

La fase di granulazione è influenzata dalle condizioni iniziali del materiale in ingresso, e quindi le variabili monitorate durante la granulazione contengono una fonte di variabilità dovuta al processo e una fonte di variabilità relativa al materiale. Per determinare il reale peso delle condizioni iniziali e della fase di granulazione nella predizione della qualità in uscita al mulino è necessario eliminare il contributo delle condizioni iniziali sul processo di granulazione.

La procedura per ottenere il contributo di ciascun blocco separatamente è stata creata grazie all'idea sviluppata da Duchesne e MacGregor (2004) di separare i contributi delle variabili manipolate nella predizione della qualità finale del prodotto.

Il procedimento prevede:

1. di creare un modello PLS per predire il valor medio delle variabili monitorate al granulatore a partire dalle condizioni iniziali (Z^*);
2. di ottenere la matrice $X_{1(\text{no}Z)}$ dei dati relativi alle misure al granulatore senza il contributo delle condizioni iniziali (matrice Z^*);
3. di aggiungere alla matrice $X_{1(\text{no}Z)}$ la variabile “quantità d’acqua”, relativa alla portata d’acqua in ingresso al granulatore;
4. di utilizzare le condizioni iniziali e la matrice così creata per predire la qualità all’uscita del mulino mediante un modello MBPLS;
5. di stimare i valori di BIP.

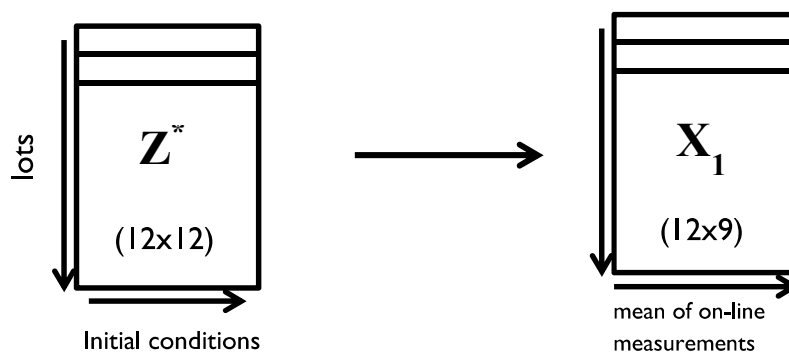


Figura 3.25 Modello PLS per predire le variabili monitorate al granulatore, matrice X_1 , dalle condizioni iniziali, matrice Z^* .

Il modello PLS, Figura 3.26, per predire le variabili misurate durante la fase di granulazione dalle condizioni iniziali permette di ottenere la ricostruzione della matrice predetta. Le prestazioni di questo modello sono scarse; infatti, la varianza media in convalida incrociata SS_{cv} è pari 0.227 mentre quella in calibrazione SS_c è pari a 0.620.

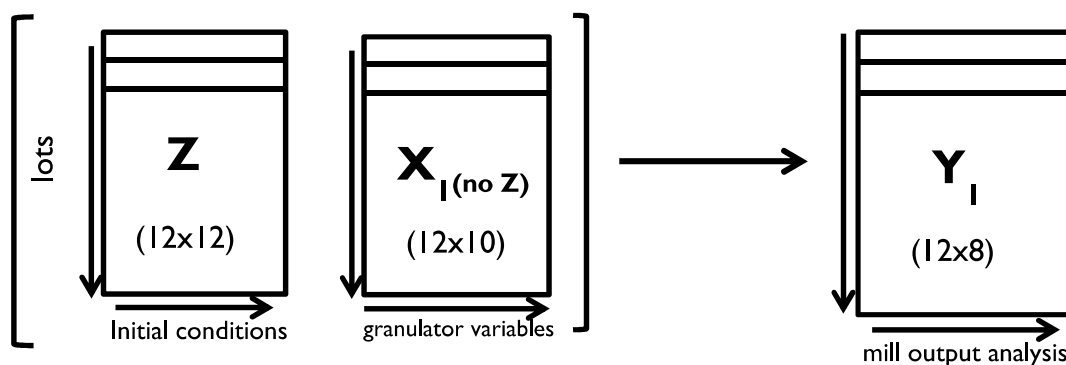


Figura 3.26 Modello MBPLS per predire la qualità all’uscita del mulino, matrice Y_1 , dalle condizioni iniziali, matrice Z^* , e dalla matrice $X_{1(\text{no}Z)}$ che contiene le variabili monitorate senza il contributo di Z^* e l’unica variabile manipolata.

Alcune variabili di X_1 sono influenzate dalle condizioni iniziali che hanno un maggiore potere predittivo delle stesse, quindi la matrice ricostruita dal modello rappresenta il contributo di Z^* su X_1 . L’errore commesso nella predizione, cioè la differenza tra la matrice X_1 reale e quella

ricostruita, è influenzato principalmente dal processo e dall'unica variabile manipolata, la portata d'acqua alimentata al granulatore sulle variabili monitorate.

L'errore di ricostruzione delle variabili monitorate durante la granulazione e l'unica variabile operativa, cioè la quantità d'acqua presente in ciascun lotto, sono state raggruppati nella matrice $\mathbf{X}_{1(\text{no } \mathbf{Z})}$ di Figura 3.26 per essere utilizzati nel modello *multiblock* PLS per predire la qualità all'uscita del granulatore.

Il modello MBPLS creato (Figura 3.26) permette di stimare i valori di BIP riportati in Tabella 3.4. Tali valori sono simili e confermano che sia le condizioni iniziali che il processo di granulazione influenzano la qualità all'uscita del mulino.

Tabella 3.4 BIP (block importance in projection) per i due blocchi utilizzati nell'analisi MBPLS

Block	BIP
Initial condition (\mathbf{Z}^*)	0.693
Sep of granulation (\mathbf{X}_1) without \mathbf{Z}^*	0.779

È stata inoltre eseguita un'analisi PCA sulla matrice $\mathbf{X}_{1(\text{no } \mathbf{Z})}$ per capire se le correlazioni tra le variabili siano simili a quelle rilevate dall'analisi della matrice reale (§3.1.2). I risultati per le prime due componenti principali sono riportati in Figura 3.27.

Tale modello permette di osservare che la struttura di correlazione tra le variabili relative alla granulazione viene mantenuta; infatti la posizione delle proiezioni dei lotti nello *score plot* (Figura 3.27a) è molto simile a quello del modello PCA eseguito sui dati reali in Figura 3.5, mentre la correlazione tra le variabili cambia.

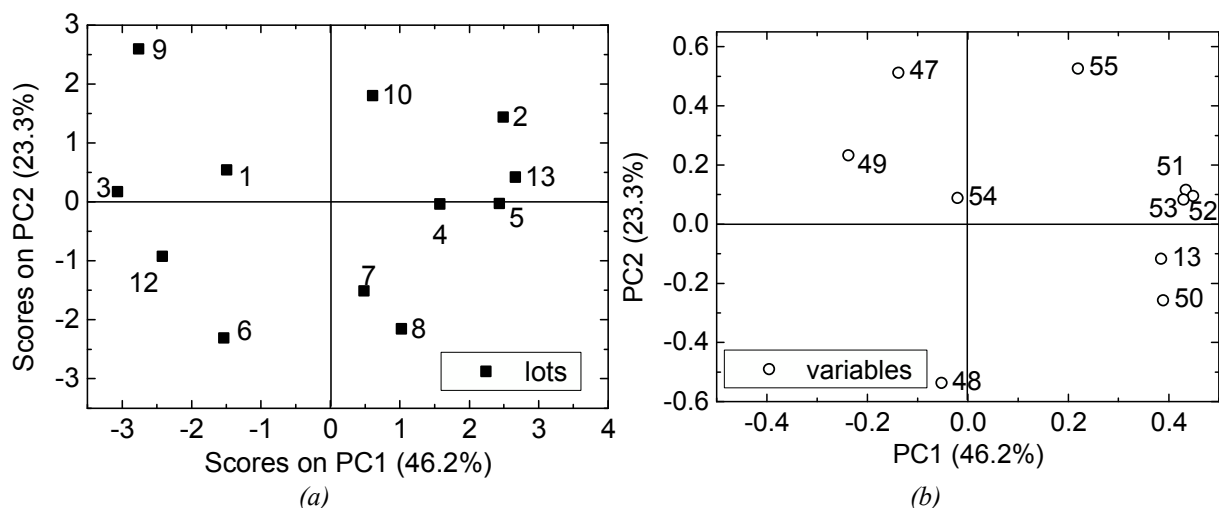


Figura 3.27 Modello PCA della matrice \mathbf{X}_1 senza il contributo di \mathbf{Z}^* . (a) Score plot; (b) loading plot.

Il modello conferma che le variabili relative al momento torcente [47,48,49] sono ortogonali alle quantità d'acqua [13] presente nei vari lotti e alle temperature misurate nelle varie sezioni [50,51,52,53].

Nel *loading plot* (Figura 3.27b) si può però osservare che le variabili [47,48], relative al momento torcente sono anticorrelate mentre nel modello eseguito sui valori reali (Figura 3.5b) c'è una stretta correlazione tra le tre variabili relative al momento torcente [47,48,49].

Il risultato di questa analisi conferma l'ipotesi che il materiale abbia un effetto sul momento torcente, poiché queste tre variabili devono essere intrinsecamente correlate. Infatti, il momento torcente descrive le proprietà reologiche del materiale perché varia al variare delle forze coesive e in §3.1.2 si era ipotizzato che tali variabili fossero correlate alle condizioni iniziali. Inoltre nelle Figure 3.16b e 3.17b del §3.1.6, le variabili relative al momento torcente sono risultate strettamente correlate alle variabili relative alla distribuzione iniziale [8,9,10]. Per esempio, i lotti {9,10} presentano valori elevati di distribuzione iniziale delle dimensioni delle particelle e quindi granuli di dimensioni elevate, hanno densità iniziali elevate, e in Figura 3.27b si osserva che hanno momento torcente elevato. L'analisi PCA (Figura 3.25) conferma che le condizioni iniziali influenzano le variabili relative al momento torcente e la procedura permette capire il vero contributo dei due blocchi nella predizione della qualità all'uscita del mulino con il modello MBPLS. In Tabella 3.4 i valori dei contributi sono molto simili confermando che tutte le variabili, siano esse condizioni iniziali o di processo, sono ugualmente importanti nella predizione.

3.3. Analisi dei dati relativi al secondo granulato

I dati relativi al secondo granulato, diverso da quello considerato fin qui, sono stati analizzati per comprendere le correlazioni tra le variabili all'interno di ogni blocco (dai dati relativi al monitoraggio dello *step* di granulazione alla qualità all'uscita del mulino) e per cercare di confrontare questo set di dati con quello precedente allo scopo di confermare o meno i risultati ottenuti per il paracetamolo e per dare generalità all'analisi fatta.

Dopo aver riorganizzato i dati, sono state eseguite tre analisi per comprendere la struttura delle variabili:

- analisi delle misure al granulatore;
- analisi della qualità all'uscita del mulino;
- analisi della qualità all'uscita della comprimitrice.

Le variabili misurate nella fase di granulazione, riportate in Tabella 2.9, sono le stesse del processo di granulazione del paracetamolo. Il materiale è unico e non ci sono variazioni nelle condizioni iniziali ma, diversamente dal processo precedente, sono state manipolate tre variabili per il granulatore:

- la quantità d'acqua [1]: da 12 a 25% su cinque livelli;

- la portata di solido entrante al granulatore [3] (corrente 2 di Figura 2.1): da 1 a 2 kg/h su tre livelli;
- la velocità del motore [2]: da 150 a 350 rpm su quattro livelli.

La prima analisi è stata eseguita sulla media dei valori relativi alle misure durante la fase stazionaria dello *step* di granulazione (matrice X_1) e in Figura 3.28 sono riportati lo *score plot* e del *loading plot* sulle prime due componenti principali.

Nella Figura 3.28b si può osservare la stessa struttura di correlazione del modello riportato in Figura 3.5b relativo ai dati del paracetamolo; infatti, c'è una forte correlazione tra le temperature [7,8,9] e in modo minore con la temperatura misurata nella zona d'entrata del solido [6], e c'è una forte correlazione anche tra le variabili relative alla coppia del motore e del trasduttore [4,5].

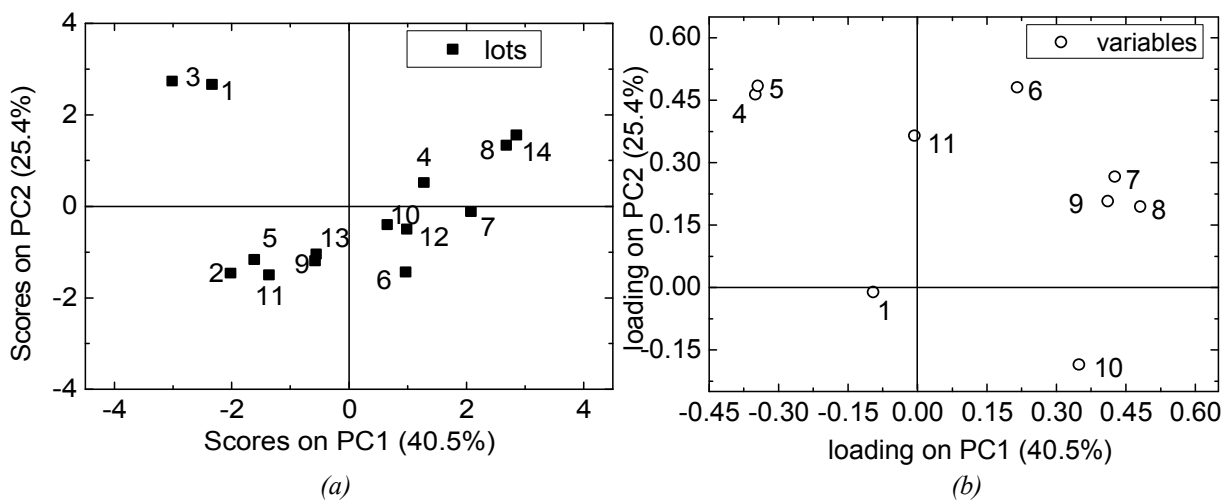


Figura 3.28 Modello PCA sui misure al granulatore per il secondo set di dati. (a) Score plot; (b) loading plot.

In questo set di dati, la portata di solido alimentato [11] non è costante per i vari lotti (anche se ha un peso solo sulla seconda componente principale), mentre la velocità del motore [10] ha un peso rilevante nel discriminare i campioni sulle prime due componenti principali.

Nello *score plot* di Figura 3.28 i lotti con PC2 positiva {1,3,4,8,14} hanno la stessa portata entrante di solido [11], pari a 2 kg/h, e hanno valori elevati di temperatura [6] misurata nella zona di ingresso del solido al granulatore. I lotti {1,3} sono gli unici ad avere valori elevati di *motor torque* [4,5], senza un'apparente motivazione. Essi sono stati lavorati con velocità del motore basse [10] (pari a 150 rpm) come i lotti {2,9}, ma quest'ultimi hanno valori di *motor torque* [4,5] bassi.

I lotti con PC1 positiva in Figura 3.27 hanno valori di temperature elevate [7,8,9] e sono stati processati alla stessa velocità del motore [10], tranne i lotti {10} e {12}.

In Figura 3.29, il lotto {9} esce dai limiti del SPE perché ha temperature elevate rispetto ai lotti processati con lo stesso quantitativo d'acqua e la stessa velocità del motore.

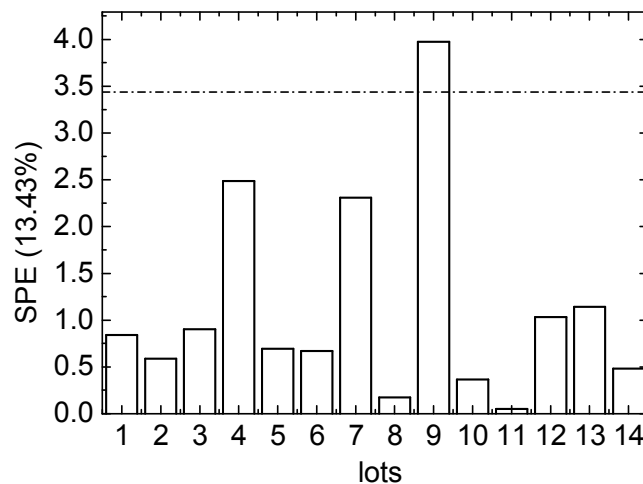


Figura 3.29 SPE relativi al modello PCA sui dati di processo, matrice X_1

È stata eseguita anche un'analisi sui dati relativi all'ultima fase del processo, cioè alla compattazione (matrice Y_2). Le variabili misurate durante questa fase sono solo quelle relative alle proprietà finali della compressa come il peso, la durezza, lo spessore e la resistenza a compressione.

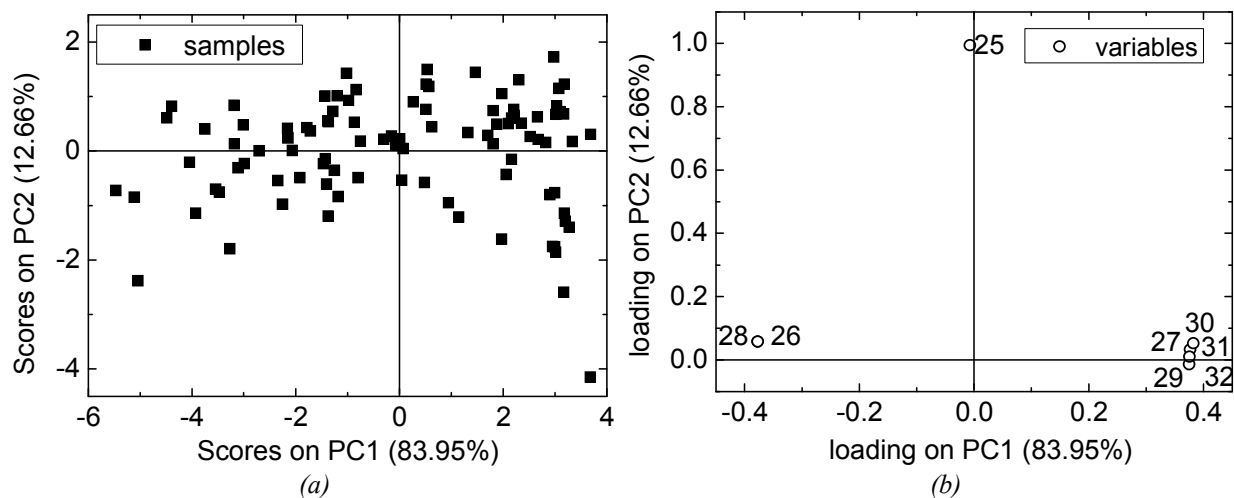


Figura 3.30 Modello PCA sui dati relativi alla fase di compattazione per il secondo set di dati (a) scores plot (b) loading plot

L'analisi, in Figura 3.30, conferma quanto visto al §3.1.4, cioè che le variabili spessore delle compresse [26,28] e durezza [27] sono sempre anticorrelate, e che durezza [27], resistenza alla compattazione [29], densità [30] e frazione solida [31] sono strettamente correlate. Inoltre tali variabili sono ortogonali al peso delle compresse [25]. Quest'analisi è utile per comprendere non solo la struttura di correlazione, ma anche se ci possono essere delle analogie tra i due set di dati.

È stata poi eseguita un'analisi dei due set di dati utilizzando le variabili comuni monitorate durante lo *step* di granulazione, riportate in Tabella 3.5. Tale analisi è stata eseguita per capire se è possibile costruire un unico modello considerando i dataset relativi ai diversi materiali.

Tabella 3.5 Variabili relative alle misure al granulatore utilizzate nel confronto tra i due set di dati.

Number	Variables
[13]	Water content (% w/w)
[47]	ExtTorqServo (NM)
[49]	ExtTorqServoPV
[50]	T_7
[51]	T_8
[52]	T_9
[53]	T_{10}
[54]	ExtSpdAV (rpm)
[55]	Feeder Rate (kg/h)

Per comprendere se la struttura di correlazione tra le variabili dei due dataset è la medesima, è stato prima creato un modello PCA con i dati relativi al secondo materiale e poi i dati medi relativi al paracetamolo sono stati proiettati cercando similitudini tra le due strutture di correlazione.

L'analisi dimostra che i due set di dati non hanno la stessa struttura di correlazione. Infatti, nella Figura 3.31 i due set di dati hanno SPE molto diverse.

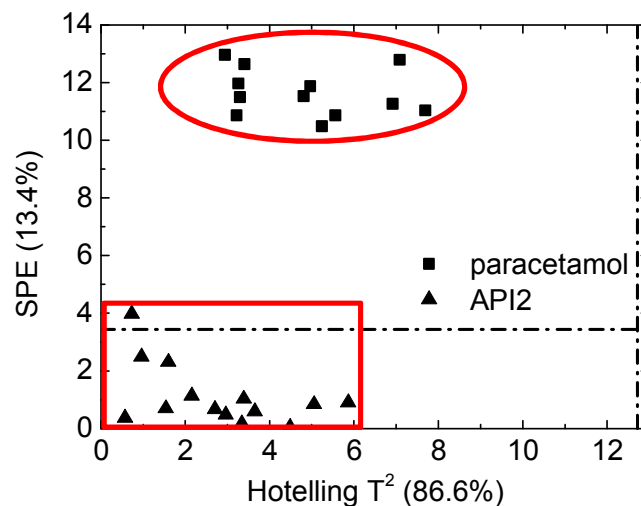


Figura 3.31 SPE vs Hotelling T^2 per i due set di dati

I dati relativi al paracetamolo, proiettati sul modello, hanno media simile a quella del secondo set di dati perché sono all'interno dei limiti del T^2 , ma hanno valori molto diversi di SPE residuo e quindi sono disposti in uno spazio diverso da quello descritto nel modello del secondo set di dati. Si conclude che è poco ragionevole costruire un modello unico considerando entrambi i dataset.

Capitolo 4

Sviluppo di sensori virtuali

Modelli statistici di regressione multivariata sono stati utilizzati per stimare variabili utili a valutare le prestazioni del processo (caratteristiche dei lotti all'uscita del mulino) e alcune caratteristiche delle compresse che possono essere difficili da misurare (Facco *et al.*, 2009). In particolare, è stato creato un modello PLS per stimare le caratteristiche dei granuli all'uscita del mulino dalle variabili operative. In un secondo caso è stato utilizzato un modello di regressione a variabili latenti per stimare la resistenza a trazione di compresse prodotte alla fine del ciclo produttivo sulla base di altre variabili misurate durante la compattazione.

4.1. Stima della resistenza a trazione delle compresse dal set di dati del paracetamolo

La qualità finale della compressa dopo la fase di compattazione viene definita da varie proprietà come le dimensioni, la densità, la plasticità e la resistenza a trazione. La resistenza a trazione (tensile strength) σ è la massima tensione che un materiale può sopportare prima della rottura e viene misurata come forza di contatto per unità di superficie (4.1), calcolata mediante la relazione seguente:

$$\sigma = \frac{2H}{\pi Dt} , \quad (4.1)$$

dove H è la durezza e D e t sono rispettivamente il diametro e lo spessore della compressa.

È stata inoltre fornita una relazione empirica per stimare tale parametro dalla porosità (adimensionale):

$$\sigma_{est} [MPa] = 7.4643 \times e^{(-11.03 \times porosity)} \quad (4.2)$$

L'analisi dei dati di compressione (Figura 3.7b) ha evidenziato che ci sono molte variabili correlate alla resistenza a trazione [46] come:

- *MaxEjection Stress* [44],
- *Irrecoverable energy* [37],
- *Hardness* [31],
- *Max Ejection Force* [41],
- *Total energy* [35],
- *Max Lower Punch Stress* [43],
- *Max Lower Punch Force* [40],

- *Density* [32],
- *Relative density* [33],
- *Max Upper Punch Force* [39],
- *Max Upper Punch Stress* [42].

Queste variabili sarebbero le più indicate per la stima della resistenza a trazione, ma alcune presentano dati mancanti.

È stato quindi utilizzato un modello PLS per stimare la resistenza a trazione utilizzando le dieci variabili della matrice Y_2 definite in Tabella 4.1, che non hanno dati mancanti.

Tabella 4.1 Variabili usate nel modello PLS per la stima della resistenza a trazione

Number	Variables
[26]	minimum punch separation distance
[27]	fill depth
[28]	tablet weight
[29]	tablet thickness
[30]	tablet diameter
[32]	density
[33]	relative density
[34]	porosity
[39]	max upper punch force
[42]	max upper punch stress

Il set di dati di compressione, matrice Y_2 , è stato suddiviso in tre parti:

- S_1 per la calibrazione del modello PLS (36% di Y_2), comprendente prove senza dati mancanti;
- S_2 per la convalida del modello (24% di Y_2), comprendente prove senza dati mancanti;
- S_3 comprendente misure in cui la resistenza a trazione non è stata misurata ma stimata in base alla porosità (4.2), per confrontare la stima fatta utilizzando la porosità con la stima fatta utilizzando il modello PLS.

Tabella 4.2 Varianza spiegata da ciascuna variabile latente per la matrice (X) predittrice e quella predetta (Y)

Number of LV	Variance X	Cumulative variance	Variance Y	Cumulative variance
1	65.94	65.94	92.11	92.11
2	13.30	79.25	2.393	94.50
3	9.345	88.59	0.977	95.48
4	6.55	95.15	0.29	95.77
5	4.71	99.86	0.24	96.01

Il modello è stato calibrato selezionando il sottoinsieme S_1 dall'insieme S_1+S_2 in modalità casuale, ripetendo tale procedura più volte affinché le prestazioni del modello non fossero influenzate dalla scelta del set di calibrazione.

In Tabella 4.2 sono riportate le varianze spiegate da ciascuna variabile latente del modello relative alla matrice predittrice (\mathbf{X}) e a quella predetta (\mathbf{Y}). Questi indici sono utili per capire quante variabili latenti devono essere utilizzate nel modello PLS. Il modello è stato costruito con tre variabili latenti perché la varianza spiegata dalla quarta LV è molto bassa.

I modelli creati permettono di ottenere in media un errore relativo in predizione del 16% (Tabella 4.3) considerando solo i campioni del *set* di convalida che presentano T^2 e SPE all'interno dei rispettivi limiti di confidenza del 95%. Infatti, se anche un solo campione ha valori di T^2 e SPE molto elevati, l'errore in predizione del modello diventa molto elevato.

Tabella 4.3 Indici statistici per stabilire la bontà del modello PLS nella stima della resistenza a trazione

SS_c	0.955
SS_{cv}	0.944
SS_p	0.941
RMSEC	0.266 Mpa
RMSECr	0.4591
RMSECV	0.323 Mpa
RMSEP	0.318 Mpa
RMSEPr	0.163

In Figura 4.1 è stato riportato il confronto tra la stima con il modello PLS e i valori reali in calibrazione e in convalida. Il modello PLS costruito con le dieci variabili relative alla sezione di compressione (Tabella 4.1) ha buone prestazioni per la predizione della resistenza a trazione utilizzando tre variabili latenti, ottenendo una varianza spiegata in predizione, SS_p del 94% (Tabella 4.3).

Il modello PLS così costruito è stato utilizzato per predire la resistenza a trazione nelle prove in cui tale variabile non è stata misurata, e la predizione PLS è stata confrontata con la stima effettuata con la relazione empirica (4.2). In Figura 4.2 è stato riportato il confronto tra i valori predetti dal modello e quelli calcolati con (4.2).

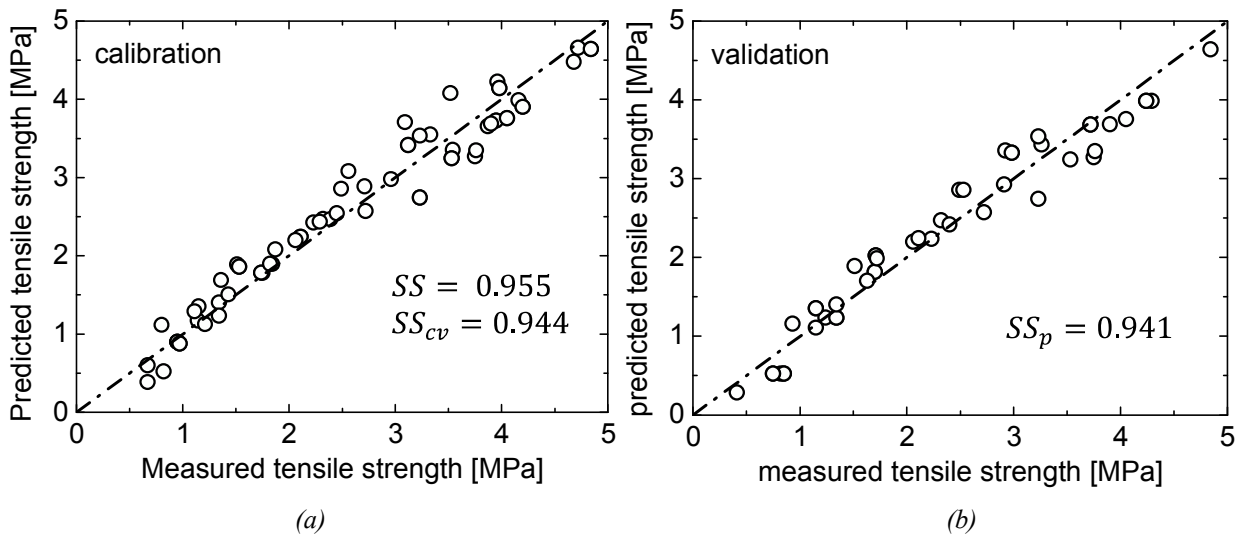


Figura 4.1 Confronto tra resistenza a compattazione (tensile strength) misurata e predetta con PLS (a) in calibrazione (b) in convalida

I dati mancanti sono relativi a prove effettuate all'interno dell'intervallo delle variabili operative di compattazione e il confronto con la stima effettuata con la porosità permette di osservare che le due stime sono abbastanza simili; infatti, la differenza, $RMSE_r$, tra le due stime è pari al 12%. Questa analisi suggerisce che, se ci sono dati mancanti relativi a proprietà della compressa utili alla determinazione della resistenza a trazione (come ad esempio la durezza), e quindi non è possibile impiegare la (4.1), è auspicabile utilizzare un modello che consideri più proprietà della compressa piuttosto che un modello pesato sulla sola porosità.

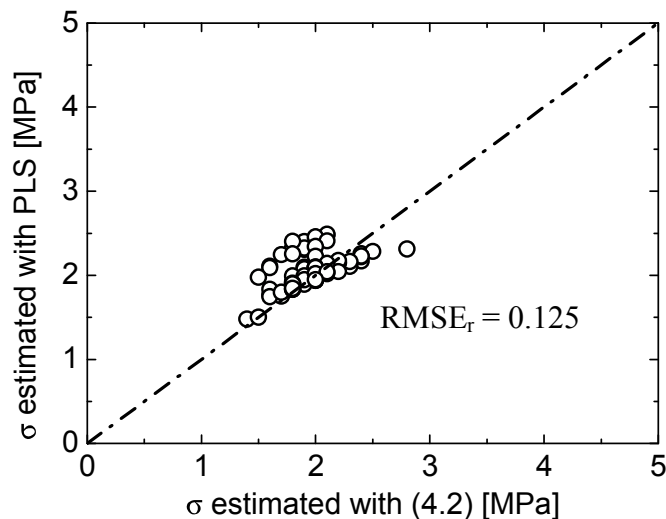


Figure 4.2 Confronto tra resistenza a trazione predetta con PLS e stimata con la formula empirica (4.2)

In Figura 4.3 è riportato il VIP della resistenza a trazione per comprendere quale variabile (di Tabella 4.1) è maggiormente predittiva.

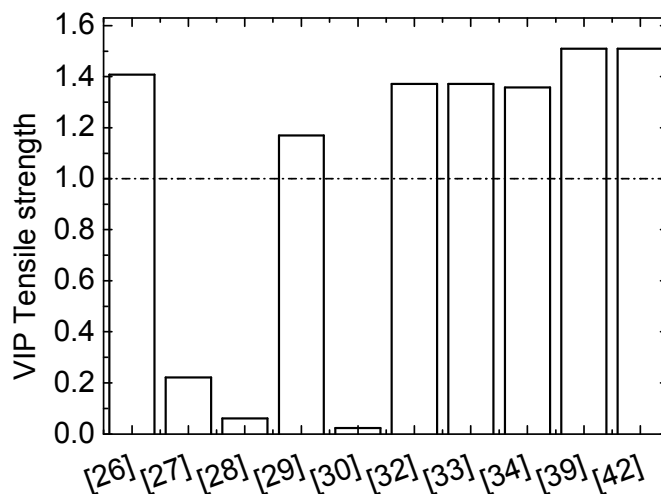


Figure 4.3 Importanza di ciascuna variabile utilizzata nel modello PLS nella predizione della resistenza a trazione

È possibile osservare che le variabili correlate a σ , come lo spessore della compressa [29], la densità [32], la densità relativa [33], *Max Upper Punch Force* [39] e *Max Upper Punch Stress* [42], sono molto importanti nella predizione della resistenza a trazione. Infatti, essa è correlata alla tensione di compattazione, cioè alla massima pressione esercitata sul punzone.

Anche la distanza del punzone [26] e la porosità [34] risultano importanti nella predizione della resistenza a trazione; infatti al § 3.1.4 era stato osservato che [26], essendo una variabile operativa, modifica le caratteristiche dei lotti in ingresso uniformando quelle del prodotto in uscita al compattatore.

Nel modello PCA creato in precedenza per i dati di compressione (Figura 3.7b) si è osservato che ci sono alcune variabili strettamente correlate alla resistenza a trazione, e tra queste non compare la porosità bensì la densità. Potrebbe quindi essere una valida alternativa utilizzare una relazione empirica basata sulla *densità* piuttosto che sulla porosità, se non è possibile usare strumenti multivariati come il PLS.

4.2. Stima della resistenza a trazione del secondo granulato

Nella descrizione del secondo granulato, crf. §2.2, sono stati elencate le variabili di compattazione misurate durante tale fase (Tabella 2.11) ed è stata fornita una stima della resistenza a trazione sulla base della porosità. Tale stima viene calcolata con parametri diversi per ogni lotto.

È stato creato un modello PLS per predire tale variabile usando in calibrazione il 70% dei dati disponibili, e il restante 30% in convalida. In questo modello non sono state utilizzate le misure di durezza, strettamente correlate alla resistenza a trazione, per verificare se è possibile stimare quest'ultima da variabili più facili da misurare come:

- weight

- thickness
- density
- solid fraction

Il numero delle variabili latenti è stato scelto valutando il minimo del RMSECV e valutando la varianza spiegata da ciascuna LV che è stata riportata in Tabella 4.4.

Tabella 4.4 Varianza spiegata da ciascuna variabile latente per la matrice (X)predittrice e quella predetta(Y)

Number of LV	Variance X	Cumulative variance	Variance Y	Cumulative variance
1	74.89	74.89	90.14	90.14
2	5.86	80.75	5.33	95.47
3	19.25	100	1.59	97.06

In Figura 4.4 è stato riportato il confronto tra la stima con il modello PLS e i valori reali in calibrazione e in convalida. Il modello PLS costruito con quattro variabili relative alla sezione di compattazione ha buone prestazioni per la predizione della resistenza a trazione.

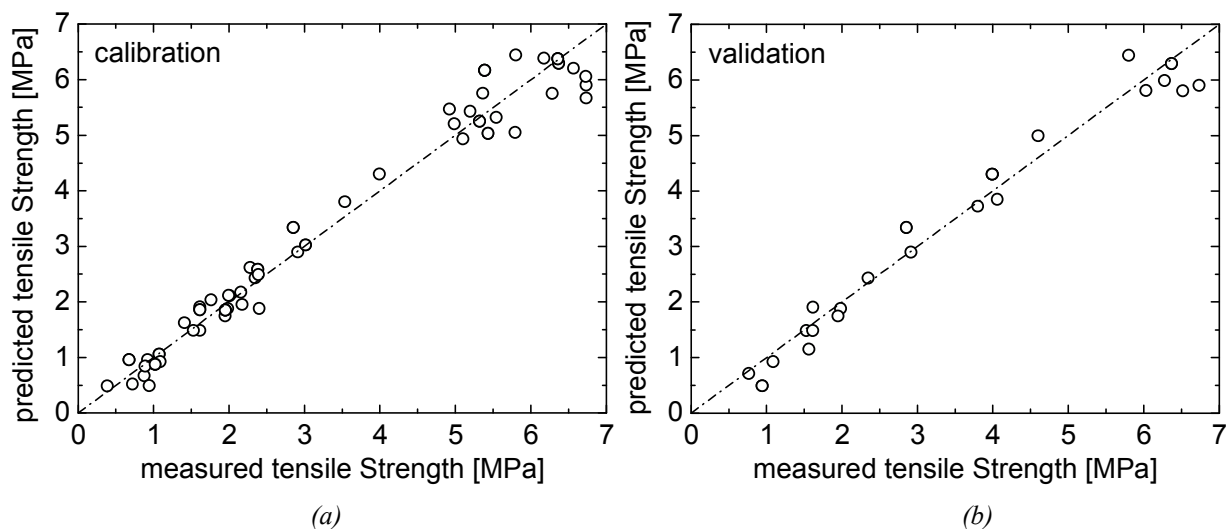


Figura 4.4 Confronto tra resistenza a compattazione (tensile strength) misurata e predetta con PLS (a) in calibrazione (b) in convalida

Utilizzando l'indice *VIP* è possibile determinare quali siano le variabili più importanti per la predizione della resistenza a trazione, e come si vede in Figura 4.5, il peso della compressa non ha influenza nella predizione, mentre le altre variabili sembrano essere similmente importanti.

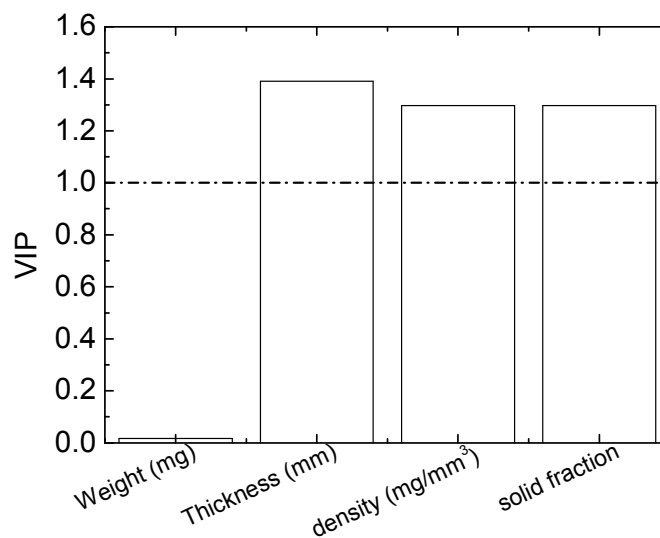


Figura 4.5 VIP per variabile nella predizione della resistenza a trazione

Il modello commette un errore relativo massimo del 16% nella predizione della resistenza a compressione, con valori di varianza in calibrazione (SS_c) e in predizione (SS_p) soddisfacenti, riportate in Tabella 4.5. Gli indici riportati in Tabella 4.5 sono stati descritti nel paragrafo §1.4.3.

Tabella 4.5 Indici statistici per stabilire la bontà del modello nella stima

SS_c	0.9706
SS_{cv}	0.9672
SS_p	0.9680
RMSEC	0.3608 Mpa
RMSECr	0.1412
RMSEP	0.3597 Mpa
RMSEPr	0.1656

L'analisi permette di dire che il modello PLS, essendo unico per tutti i lotti, potrebbe essere migliore delle stime fatte sulla base della porosità che utilizzano un modello diverso per ciascun lotto. Inoltre questo modello, che non impiega misure di durezza, ha varianza paragonabile al modello costruito con i dati relativi al paracetamolo e conferma l'elevata correlazione tra resistenza a trazione, spessore della compressa e densità. Infatti queste due variabili hanno un elevato VIP (Figura 4.4) ed sono risultate strettamente correlate in Figura 3.7b.

4.3. Stima delle caratteristiche del granulato all'uscita del mulino

4.3.1. Stime per il paracetamolo

Le condizioni iniziali (matrice \mathbf{Z}^*) e l'unica variabile manipolata, cioè la portata d'acqua al granulatore, sono state utilizzate per predire le caratteristiche dei lotti all'uscita del granulatore (matrice \mathbf{Y}_1) mediante un modello PLS. Per ciascuna delle variabili misurate all'uscita del mulino (\mathbf{Y}_1) è stato costruito un modello di stima di tipo PLS. In Tabella 4.6 sono riportati i valori di varianza in calibrazione (SS_c) e in convalida incrociata (SS_{cv}) per determinare la bontà di ciascun modello per la stima della relativa variabile.

Il numero di variabili latenti utilizzate per costruire ciascun modello è stato scelto valutando il minimo del RMSECV e la varianza spiegata da ciascuna LV. In Tabella 4.5 è riportato il numero di LV utilizzato per ciascun modello.

Tabella 4.5 SS in calibrazione e in convalida incrociata per ciascuna variabile di \mathbf{Y}_1 relativa al paracetamolo

Variables	No. of LV	SS_c	SS_{cv}
$\rho_{aerated}$ [19]	3	0.876	0.656
ρ_{tapped} [20]	1	0.660	0.412
HR [21]	2	0.487	0.005
x_{10} [22]	4	0.711	0.187
x_{50} [23]	3	0.500	0.0008
x_{90} [24]	1	0.164	0.051
span [25]	2	0.843	0.502

I valori riportati in Tabella 4.5 indicano che quasi tutti i modelli hanno prestazioni insoddisfacenti per avere una stima accurata delle variabili misurate all'uscita del mulino, forse perché la quantità di dati è troppo bassa per poter eseguire una predizione. Infatti, i valori in convalida incrociata eseguita in modalità *leave one out* sono accettabili solo per la prima [19] e l'ultima variabile [25].

4.3.2. Stime per il secondo granulato

Il secondo materiale farmaceutico è stato granulato variando la portata di solido e di liquido in ingresso al granulatore e la velocità del motore del granulatore. Queste tre variabili operative sono state utilizzate nella predizione delle variabili misurate all'uscita del mulino.

I risultati dell'analisi evidenziano le scarse prestazioni dei modelli nel predire le caratteristiche dei lotti all'uscita del mulino. Infatti, in Tabella 4.6 i valori di varianza in calibrazione (SS_c) e in convalida incrociata (SS_{cv}) sono molto bassi. Questo dimostra che i modelli non sono adeguati alla descrizione delle caratteristiche dei lotti all'uscita del mulino sulla base delle variabili di processo opportunamente manipolate. Ciò può essere dovuto al

fatto che si sta considerando modelli lineari, quando la relazione tra ingressi e uscite potrebbe essere non lineare, oppure che il piano sperimentale progettato non tiene conto di tutte le fonti di variabilità che agiscono sul processo e influenzano la qualità finale.

Tabella 4.6 *SS in calibrazione e in convalida incrociata per alcune variabili misurate all'uscita del mulino per il secondo set di dati*

Variables	No. of LV	SS_c	SS_{cv}
ρ_{aerated}	1	0.554	0.258
x_{50}	2	0.369	0.094
span	1	0.550	0.219

Capitolo 5

Monitoraggio del processo produttivo

I modelli statistici multivariati possono essere utilizzati per il monitoraggio del processo produttivo in linea allo scopo di identificare anomalie del sistema che, se non identificate per tempo, possono pregiudicare la qualità del prodotto finale (MacGregor *et al.*, 1994). I dati relativi alla fase di granulazione sono stati analizzati con un approccio di tipo PCA, in cui il modello è calibrato sullo storico del processo e poi utilizzato per il monitoraggio in tempo reale del sistema produttivo e l'individuazione di eventuali *fault*. Inoltre, è stato utilizzato un approccio di tipo PLS che permette sia di identificare in linea le anomalie del sistema, sia di fornire una stima delle caratteristiche del prodotto all'uscita del mulino ad ogni istante in cui vengono raccolti i dati del processo.

5.1. Individuazione di anomalie durante la granulazione

Modelli statistici multivariati come PCA e PLS possono essere utili per il monitoraggio del processo e l'individuazione di anomalie. In questo caso modelli di questo tipo sono stati utilizzati per rilevare in linea eventuali problematiche durante la fase di granulazione (riportate in Tabella 2.5).

Tabella 5.1 Autovalori, varianza spiegata da ciascuna componente principale e varianza cumulata per il modello PCA creato sulla matrice X_1

Number of PCs	Eigenvalues	Variance	Cumulative variance
1	2.79	31.03	31.03
2	2.77	30.74	61.77
3	1.00	11.14	72.91
4	0.92	10.20	83.12
5	0.89	9.84	92.96
6	0.28	3.07	96.03
7	0.19	2.15	98.17
8	0.10	1.13	99.30
9	0.06	0.70	100.00

Il modello PCA, descritto nel paragrafo §3.1.2, è stato utilizzato in questo caso allo scopo di monitorare la fase di granulazione e di rilevare anomalie nella lavorazione del lotto {2} e del lotto {5}. I dati relativi a queste anomalie, note a priori, verranno pertanto proiettati sul modello di monitoraggio, per valutare se quest'ultimo è in grado non solo di rilevare l'anomalia, ma anche di diagnosticarne le cause.

5.1.1. Granulazione del lotto {2}

Durante la lavorazione del lotto {2} si è esaurito il materiale in ingresso al granulatore, provocando la diminuzione della portata di solido al granulatore. L'obiettivo dell'analisi è quindi valutare se il modello PCA sia in grado di individuare prontamente l'anomalia. Per valutare le prestazioni del modello sono state proiettate le misure relative ad alcuni istanti di campionamento del lotto {2} (sul modello PCA descritto al §3.1.2).

In Figura 5.1 sono riportate le carte di monitoraggio di Hotelling T^2 (1.30) e del SPE (1.35) con i relativi limiti al 95% (cfr. §1.4.5). Si ricorda che la statistica T^2 rappresenta la vicinanza alla media del sistema, mentre SPE rappresenta la distanza delle proiezioni delle misure dal piano descritto dal modello.

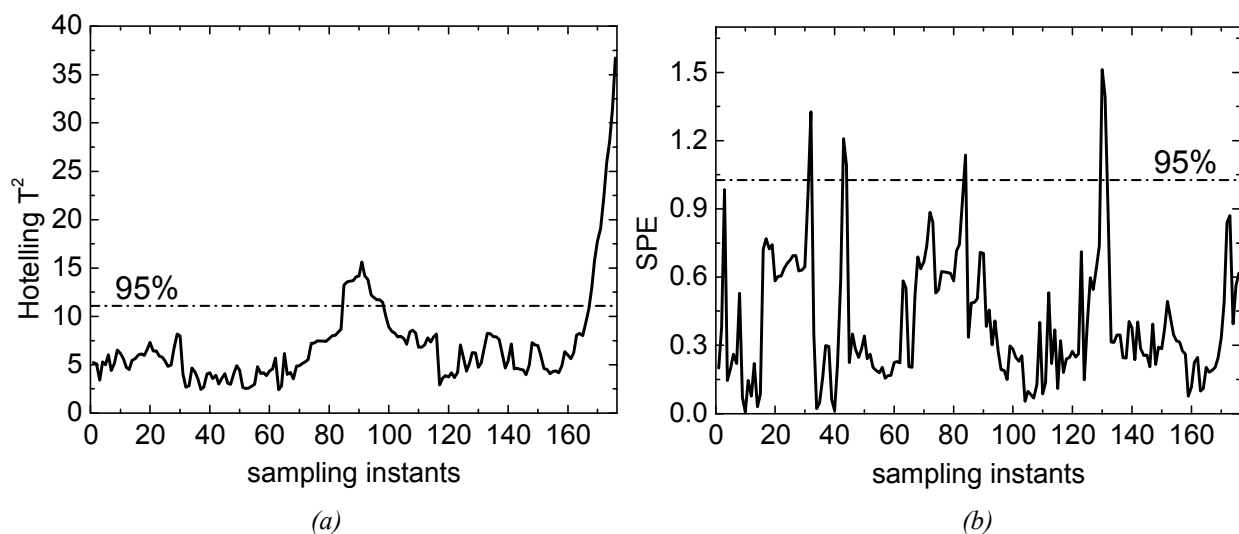


Figura 5.1 (a) Hotelling T^2 e (b) SPE per la determinazione del fault del lotto {2} con il modello PCA.

In Figura 5.1 si osserva che gli ultimi istanti di campionamento proiettati sul modello presentano valori di T^2 al di fuori del limite di confidenza del 95%, ma SPE all'interno. Questo significa che gli ultimi istanti di campionamento sono molto lontani dalla media del sistema, ma vengono comunque descritti bene dal modello. L'anomalia viene riscontrata allo stesso istante di campionamento, numero {168}, al quale è stata dichiarata avvenire sulla base dell'analisi del solo profilo di portata di solido in ingresso al granulatore.

Inoltre viene individuato un falso allarme in Figura 5.1 tra gli istanti {85} e {95}, dovuto ad una temporanea diminuzione dei valori di momento torcente.

Per comprendere quali variabili misurate risentano maggiormente dell'anomalia e quindi capire perché alcuni campioni hanno media diversa da quella dei dati con i quali è stato costruito il modello, è necessario calcolare il contributo al T^2 (cfr. §1.4.5).

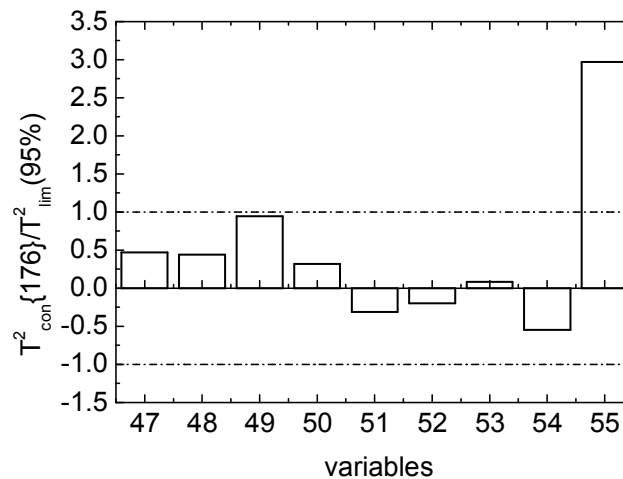


Figura 5.2 Rapporto tra i contributi al T^2 per variabile e i limiti di confidenza al 95% per l'ultimo istante di campionamento relativo al lotto {2} durante il malfunzionamento.

In Figura 5.2 è stato riportato il rapporto per ciascuna variabile di processo tra i contributi al T^2 e i rispettivi limiti di confidenza al 95% per il campione {176} (cfr. §1.4.5). Si può osservare che la portata di solido [55] al granulatore ha un valore molto elevato, fuori dai limiti di confidenza, modificando la media multivariata dei dati proiettati. L'analisi permette di osservare che la diminuzione del materiale in ingresso porta ad una variazione della portata di solido al granulatore [55], come si vede in Figura 5.2, senza modificare le altre variabili, che presentano contributi molto bassi al T^2 .

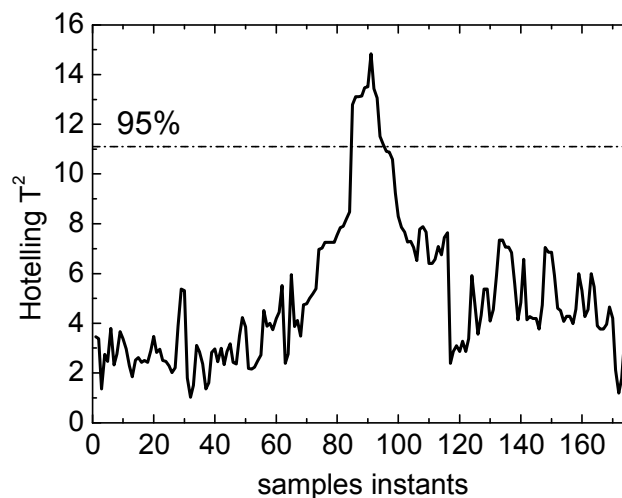


Figura 5.3 Hotelling T^2 per la determinazione del fault del lotto {2} con il modello PCA senza la variabile [55].

Per capire se l'identificazione del fault possa avvenire anche in assenza della misura [55] della portata del solido in ingresso al granulatore, è stato creato un modello PCA senza la variabile [55] e sono stati proiettate le misure relative agli stessi istanti di campionamento. Come si vede in Figura 5.3, questo modello non riesce a identificare il *fault*. Ciò significa che [55] è effettivamente l'unica variabile a subire l'effetto dell'anomalia.

5.1.2. Granulazione del lotto {5}

Durante la lavorazione del lotto {5} si è fermata la vite del granulatore, causando una diminuzione dei valori delle variabili relative al momento torcente e della portata di solido. Nelle Figure 5.4a e 5.4b è rappresentato l'andamento delle variabili misurate al granulatore ad ogni istante di campionamento. Le misure presentate nella figura sono state scalate rispetto alla media (divise rispetto alla media dei dati usati per la calibrazione del modello PCA) per poter osservare agevolmente l'andamento delle variabili nel tempo. In Figura 5.4 sono rappresentate le misure rilevate durante la lavorazione del lotto {5} e per comodità la misura {0} è stata posta 57 istanti prima del verificarsi del *fault*. Le misure dagli istanti {-245} a {-1} sono state utilizzate per la calibrazione del modello PCA (cfr. §3.1.2), mentre quelle da {0} a {245} sono le misure in prossimità degli istanti nei quali si è verificata l'anomalia. È possibile osservare che dopo l'istante {0} le variabili relative al momento torcente iniziano a decrescere prima che il *fault* venga dichiarato.

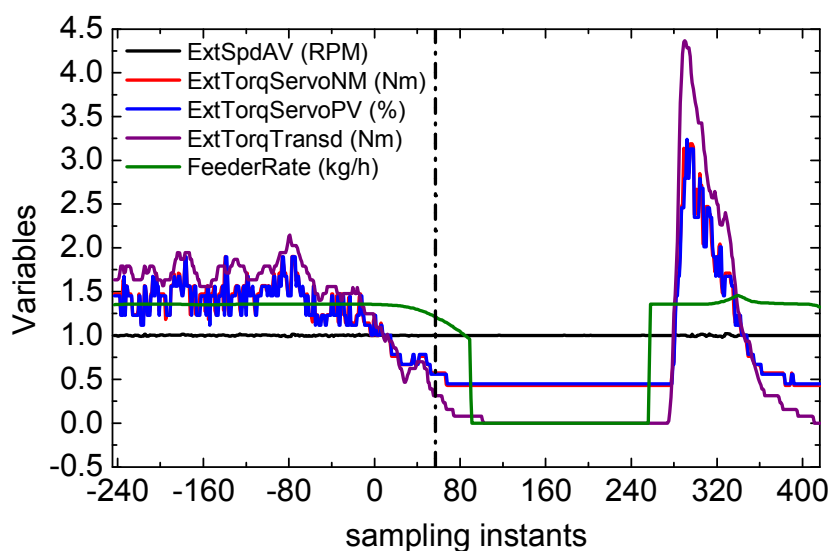


Figura 5.4 Profilo delle variabili misurate durante la fase di granulazione (a) in condizioni normali (b) durante il malfunzionamento. (Il *fault* è dichiarato all'istante {57})

Anche in questo caso per valutare le prestazioni del modello per il monitoraggio dell'operazione sono stati proiettati gli istanti relativi all'anomalia (da {0} a {245}) sul modello PCA calibrato sulle condizioni normali di tutti i lotti (da {-245} a {-1}).

In Figura 5.5 sono riportati i relativi diagrammi di Hotelling T^2 e di SPE. In Figura 5.5a si osserva che T^2 esce dai limiti di confidenza dopo l'istante numero {30}. Gli istanti successivi sono caratterizzati quindi da variabili con valori molto diversi da quelli medi del sistema. In Figura 5.5b si osserva che SPE esce dai limiti di confidenza dopo l'istante numero {65}; ciò significa che i campioni successivi hanno una struttura di correlazione totalmente diversa da quella del modello. Il modello ha buone prestazioni nell'individuazione delle anomalie perché i valori di T^2 permettono di individuare l'anomalia in anticipo (all'istante {30}), mentre il *fault* è stato accertato all'istante {57} di Figura 5.4b.

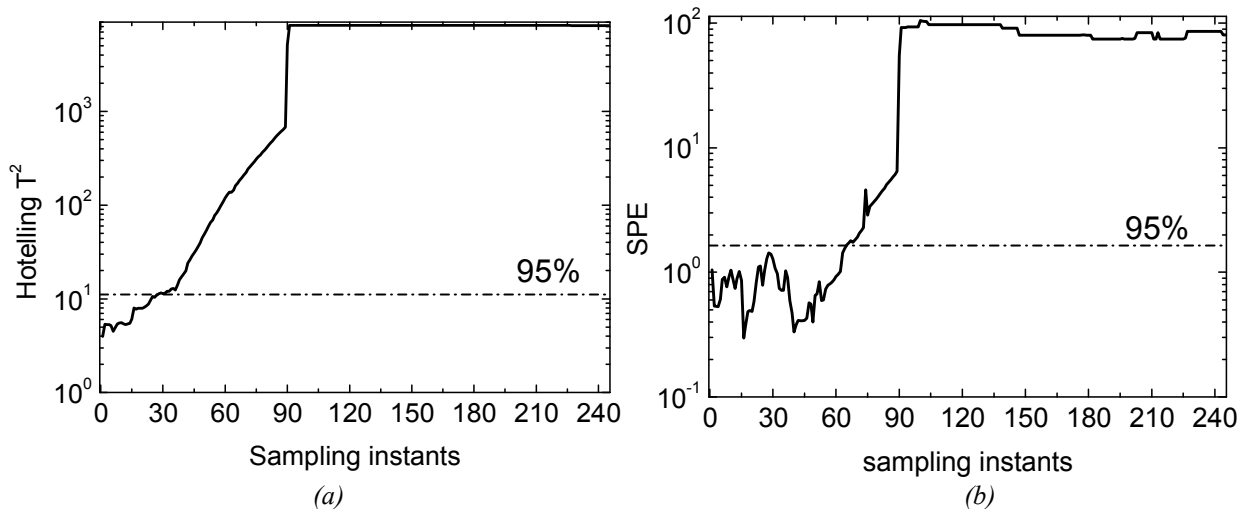


Figura 5.5 (a) Hotelling T^2 e (b) SPE residui per i dati relativi al *fault* durante la granulazione del lotto {5}.

In Figura 5.6 è riportato il rapporto tra i contributi al T^2 per variabile e i limiti di confidenza al 95% per l'istante di campionamento {30} di Figura 5.4b di seguito.

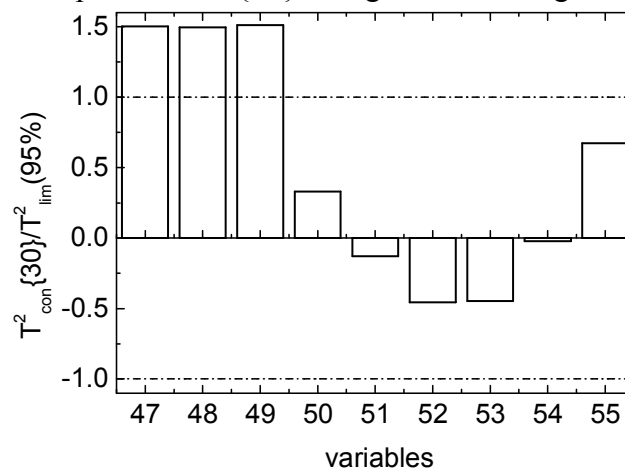


Figura 5.6 Rapporto tra i contributi al T^2 per variabile e i limiti di confidenza al 95% per l'istante {30} durante il malfunzionamento nella granulazione del lotto {5}

Si può osservare che le variabili relative al momento torcente [47, 48, 49] hanno contributi elevati al T^2 e quindi sono le prime variabili a rendere possibile l'identificazione

dell'anomalia. Infatti, le variabili relative al momento torcente diminuiscono poiché si ferma la vite del granulatore come si osserva in Figura 5.4b. Le misure negli istanti successivi, dal {30} al {65}, si trovano sul piano descritto dal modello, ma lontane dalla media.

Il rapporto tra i contributi al T^2 dell'istante {58} e i limiti di confidenza per ciascuna variabile di processo, riportati in Figura 5.7a, permettono di identificare le variabili che più contribuiscono all'individuazione dell'anomalia dopo l'istante in cui è stata dichiarata. In Figura 5.7a si osserva che solo le temperature misurate nelle ultime due sezioni del granulatore [52,53] e la velocità del motore [54] sono all'interno dei limiti di confidenza. Infatti l'arresto della vite provoca una diminuzione della portata di solido in ingresso al granulatore [55] e una variazione del momento torcente al trasduttore [47,48,49], come si osserva anche in Figura 5.4b.

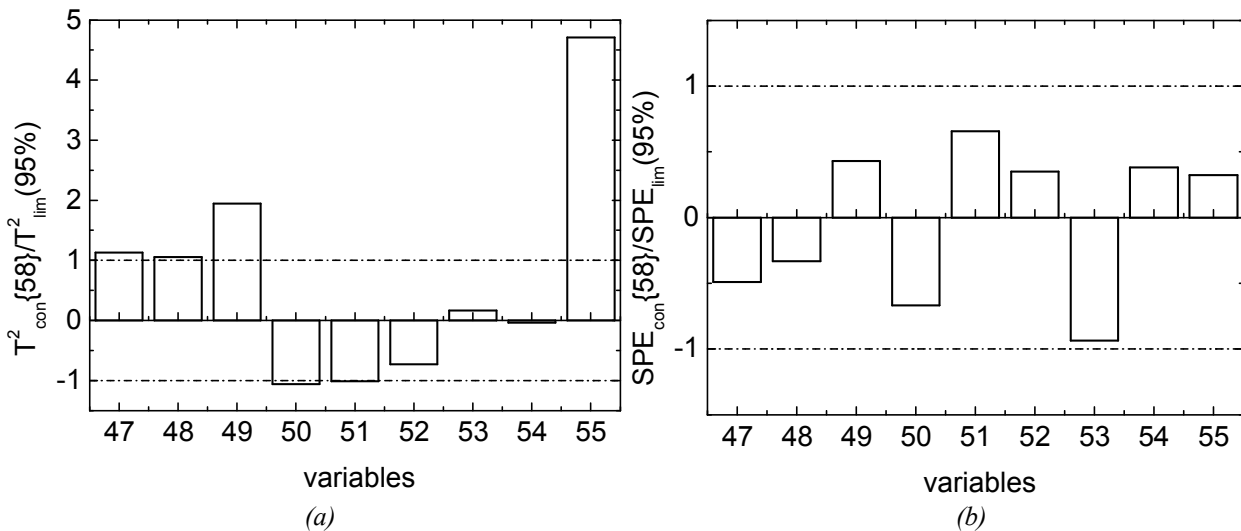


Figura 5.7 Malfunzionamento del lotto {5} (a) Rapporto tra i contributi al T^2 per variabile e i limiti di confidenza al 95% per l'istante di campionamento numero {58} (b) Rapporto tra i contributi a SPE per variabile e i limiti di confidenza al 95% per l'istante di campionamento numero {58}

In Figura 5.7b è riportato il rapporto tra i contributi a SPE per variabile e i limiti di confidenza al 95% per l'istante {58}. Questo istante ha un valore di SPE all'interno dei limiti di confidenza, come si osserva da Figura 5.5b.

In Figura 5.8 sono riportati gli stessi rapporti di Figura 5.7 ma per l'istante {90}. Tutte le variabili sono fuori dai limiti di confidenza, e quindi la struttura di correlazione è totalmente diversa da quella dei dati utilizzati per calibrare il modello.

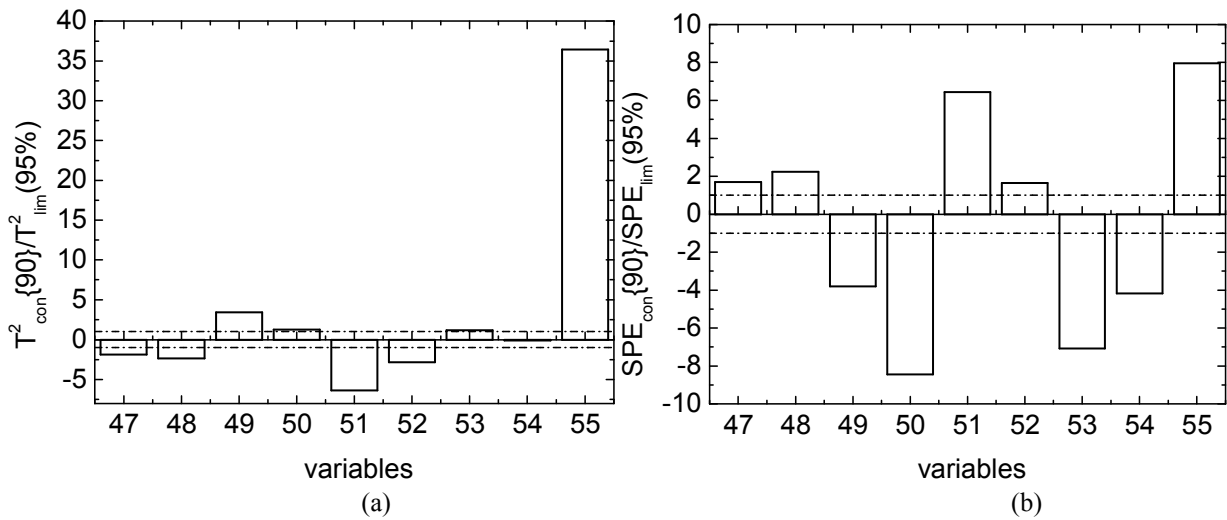


Figura 5.8 l'istante di campionamento numero {90} relativo al lotto {5} durante il malfunzionamento (a) Rapporto tra i contributi al T^2 per variabile e i limiti di confidenza al 95% (b) Rapporto tra i contributi a SPE per variabile e i limiti di confidenza al 95%.

In Figura 5.7a e in Figura 5.8a si osserva che la variabile [55], relativa alla portata di solido, ha un contributo molto elevato nell'individuare il campione fuori dai limiti del sistema. È stato quindi costruito un modello PCA senza questa variabile e sono stati proiettate le misure acquisite agli istanti di campionamento relativi al *fault* del lotto {5} per capire se l'anomalia sia ugualmente rilevabile. In Figura 5.9 sono riportati i relativi diagrammi di Hotelling T^2 e di SPE.

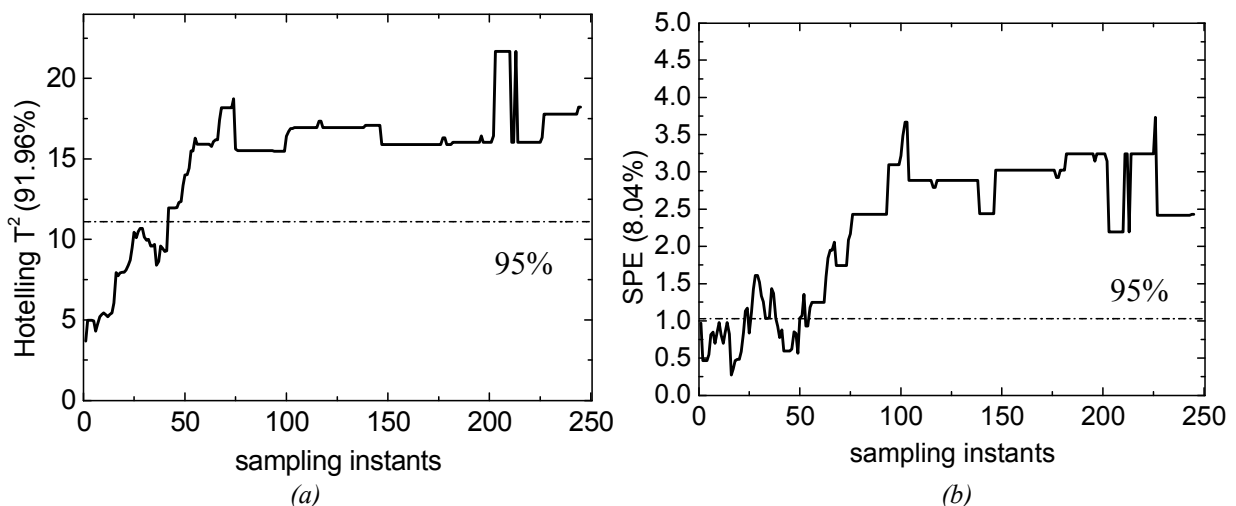


Figura 5.9 (a) Hotelling T^2 e (b) SPE per la determinazione del fault nel monitoraggio del lotto {5} per il modello PCA senza la variabile [55]

Il modello senza questa variabile riesce comunque a individuare l'anomalia. In Figura 5.9a i valori di T^2 escono dai limiti di confidenza all'istante {42}, con un lieve ritardo rispetto al modello con la variabile [55] in Figura 5.3a (dove veniva identificato all'istante {30}). In Figura 5.9b i valori di SPE escono dai limiti di confidenza all'istante {50}, in anticipo rispetto

al modello precedente in Figura 5.5b (dove veniva identificato all'istante {65}). L'anomalia viene comunque identificata dal modello perché rompe la struttura di correlazione tra le variabili di processo variando i valori di momento torcente e delle temperature misurate durante la granulazione, come era stato identificato in Figura 5.7-5.8.

5.2. Individuazione di anomalie durante la granulazione e stima delle caratteristiche del prodotto all'uscita del mulino

La proiezione su strutture latenti può essere utile per ottenere la stima delle caratteristiche del prodotto all'uscita del mulino e allo stesso tempo l'individuazione di anomalie durante il monitoraggio dello *step* di granulazione. A questo scopo, è stato calibrato un modello PLS utilizzando i dati relativi alla fase di granulazione (X_1) per stimare le caratteristiche del prodotto all'uscita del mulino (Y_1). La matrice Y_1 ha dodici righe, poiché è stata eseguita una sola misura per ciascun lotto all'uscita del mulino; pertanto, per avere un numero di righe uguali a quelle di X_1 , ogni riga di Y_1 è stata artificialmente ripetuta 245 ($2940/12=245$) volte (Figura 5.8). In questo modo si genera la matrice Y_1^* avente 2940 righe e 7 colonne.

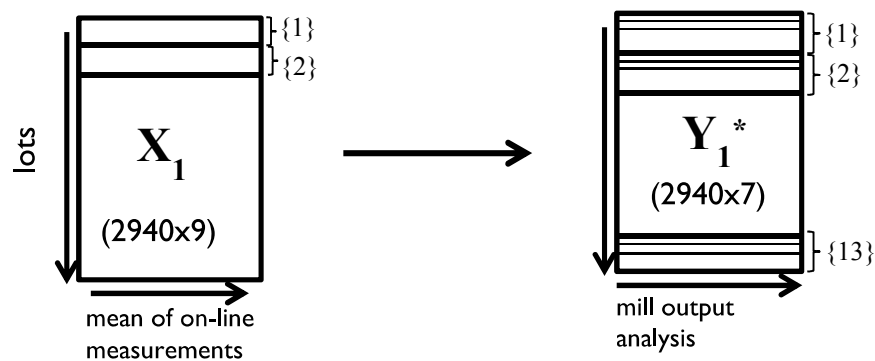


Figura 5.8 Modello PLS per predire i dati di qualità (matrice Y_1^*) dai dati relativi alla fase di granulazione (matrice X_1)

In Tabella 5.2 sono riportate le varianze spiegate dalle prime quattro variabili latenti scelte per costruire il modello.

Tabella 5.2 Varianza spiegata dalle prime 4 variabili latenti relativa alla fase di granulazione (X_1) e alle caratteristiche del prodotto all'uscita del mulino (Y_1^*)

Number of LVs	Variance X_1	Cumulative variance	Variance Y_1	Cumulative variance
1	30.74	30.74	20.4	20.47
2	30.53	61.27	4.50	24.98
3	7.905	69.18	4.86	29.84
4	5.516	74.70	4.58	34.43

In Tabella 5.3 sono riportate le prestazioni del modello in termini di varianza in calibrazione e in convalida incrociata per ciascuna variabile di qualità all'uscita del mulino. Le prestazioni sono scarse, indicando che le sole variabili misurate durante la granulazione non sono sufficienti per ottenere una stima adeguata. Inoltre, si dispone di poche misure all'uscita del mulino, e questo rende più difficile la calibrazione del modello.

Tabella 5.3 Varianza in calibrazione e in convalida incrociata per ciascuna variabile della matrice Y_1

Y_1 variables	SS_c	RMSEC	SS_{cv}	RMSECV
$\rho_{aerated}$ [g/mL]	0.52	0.013	0.52	0.013
ρ_{tapped} [g/mL]	0.34	0.018	0.33	0.018
Hausner ratio []	0.39	0.025	0.38	0.025
x_{10} [μm]	0.42	7.88	0.42	7.91
x_{50} [μm]	0.21	88.3	0.21	88.5
x_{90} [μm]	0.15	177.03	0.14	177.5
span []	0.38	0.31	0.38	0.31

5.2.1. Granulazione del lotto {2}

Le misure relative al *fault* del lotto {2} sono state proiettate sul modello per verificare se è possibile individuare il *fault* anche con questo modello PLS. In Figura 5.10 si può osservare che l'anomalia viene identificata da T^2 all'istante {167}, mentre non viene identificata da SPE.

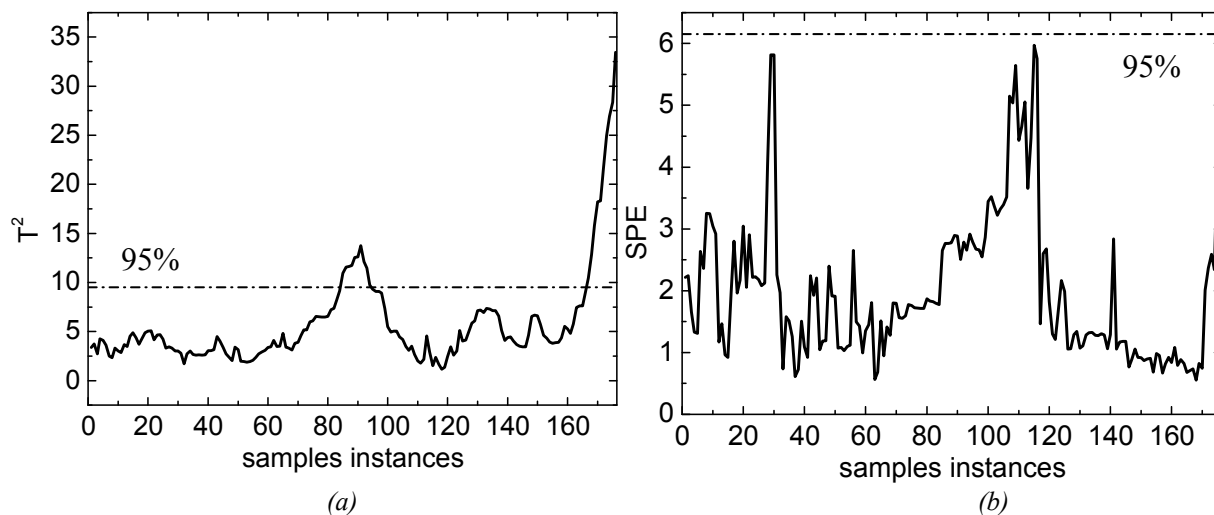


Figura 5.10 Modello PLS per determinare la qualità dai dati di processo. (a) Hotelling T^2 e (b) SPE per la determinazione del *fault* nel monitoraggio del lotto {2}

Inoltre si osserva che il T^2 esce dai limiti prima dell'annuncio del *fault* per dieci istanti di campionamento (istanti dal {85} al {95}), perché i valori delle variabili relative al momento torcente si abbassano.

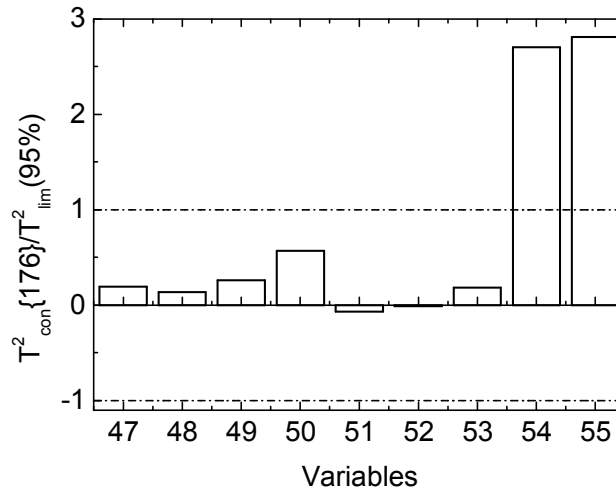


Figura 5.11 Rapporto tra i contributi al T^2 per variabile e i limiti di confidenza al 95% relativo alla proiezione dell'istante di campionamento numero {176} del lotto {2} sul modello PLS

In Figura 5.11 sono riportati i contributi al T^2 per variabile di processo per l'istante {176} ed è possibile osservare che le variabili relative alla portata di solido entrante al granulatore [55] e la velocità del motore [54] sentono gli effetti dell'anomalia. Diversamente da quanto riscontrato nel modello PCA (Figura 5.2), l'utilizzo del modello PLS porta ad individuare anche la variabile [54] come responsabile dell'anomalia. Questo è alquanto inaspettato dato che tale variabile ha una scarsa variabilità sia nel *set* di calibrazione sia nel *set* di dati proiettati, e di conseguenza presenta una bassa influenza (in termini di valori dei *loadings* qui non riportati) sul modello. Tuttavia si noti che il modello PLS è diverso dal modello PCA, dal momento che l'obiettivo è quello di descrivere la variabilità dei dati di \mathbf{X}_1 più predittiva nei confronti delle risposte (matrice \mathbf{Y}_1 di Figura 5.8). Infatti nel modello PLS i *loadings* sono ruotati e non sono indipendenti tra loro. La rotazione fa sì che [54] e [55] siano più correlate tra loro in relazione alle risposte che nel caso del modello PCA. Questo provoca l'alto contributo da parte della variabile [54].

5.2.2. Granulazione del lotto {5}

Sullo stesso modello sono state proiettate le misure relative al *fault* del lotto {5}, e in Figura 5.12 sono state riportate le relative carte di monitoraggio. Si osserva che il T^2 esce dai limiti all'istante numero {20} mentre SPE all'istante {57}, istante in cui viene annunciato il *fault*. L'anomalia viene dunque rilevata dall'analisi dei residui del modello PLS prima che dall'analisi dei residui del modello PCA di Figura 5.5b.

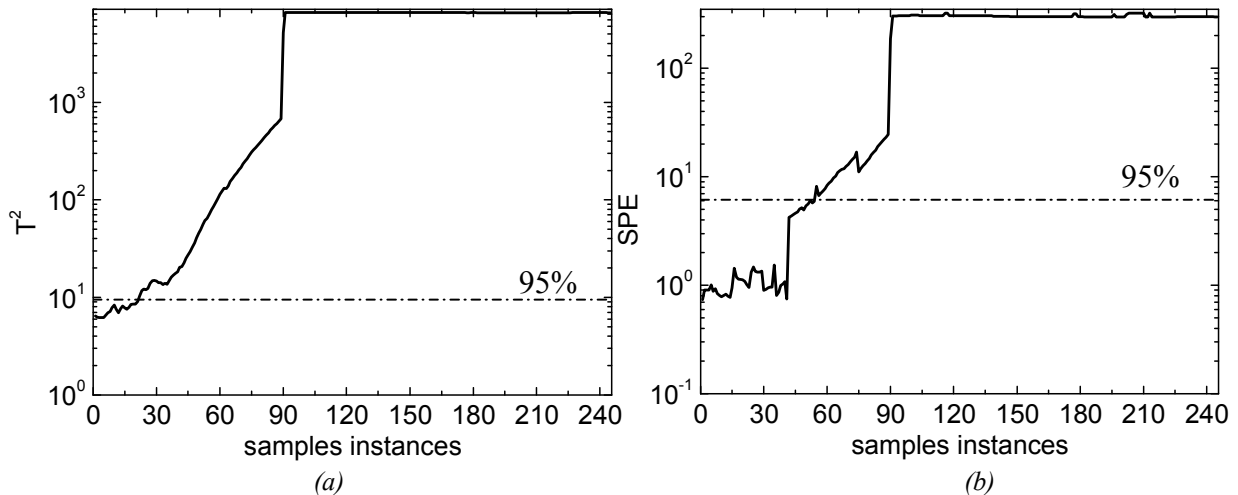


Figura 5.12 Modello PLS per determinare la qualità dai dati di processo. (a) Hotelling T^2 e (b) SPE per la determinazione del fault nel monitoraggio del lotto {5}

In Figura 5.13 sono riportati i rapporti tra i contributi al T^2 e a SPE e i limiti di confidenza al 95% per gli istanti {58} e {90}.

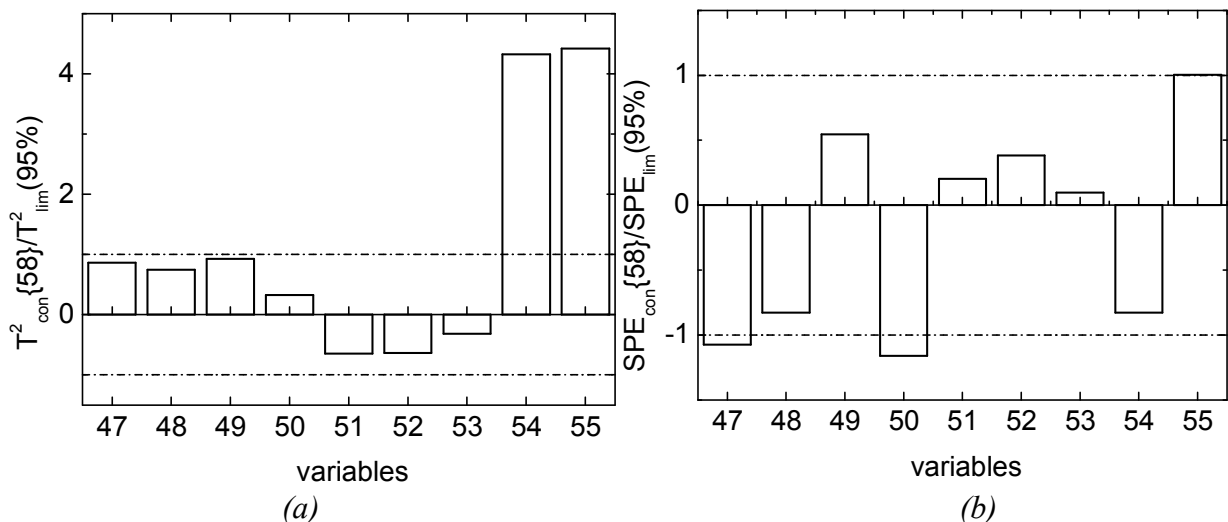


Figura 5.13 (a) Rapporto tra i contributi al T^2 per variabile e i limiti di confidenza al 95% (b) Rapporto tra i contributi a SPE per variabile e i limiti di confidenza al 95% per l'istante di campionamento {58} relativo al lotto {5} durante il malfunzionamento

All'istante {58} le variabili con contributo a T^2 elevato sono la portata di solido in ingresso al granulatore [55] (come in Figura 5.7a), la velocità del motore [54] e le variabili relative al momento torcente [47,49] e la temperatura rilevata all'uscita del granulatore [53]. All'istante {58} le variabili con contributo a SPE elevato sono il momento torcente [47], la temperatura rilevata nella prima sezione del granulatore [50] e in minima parte anche la portata di solido in ingresso al granulatore [55].

Nelle Figure 5.14a-5.14b tutte le variabili hanno contributi elevati di T^2 e SPE indica che la struttura di correlazione è totalmente diversa da quella del set di calibrazione.

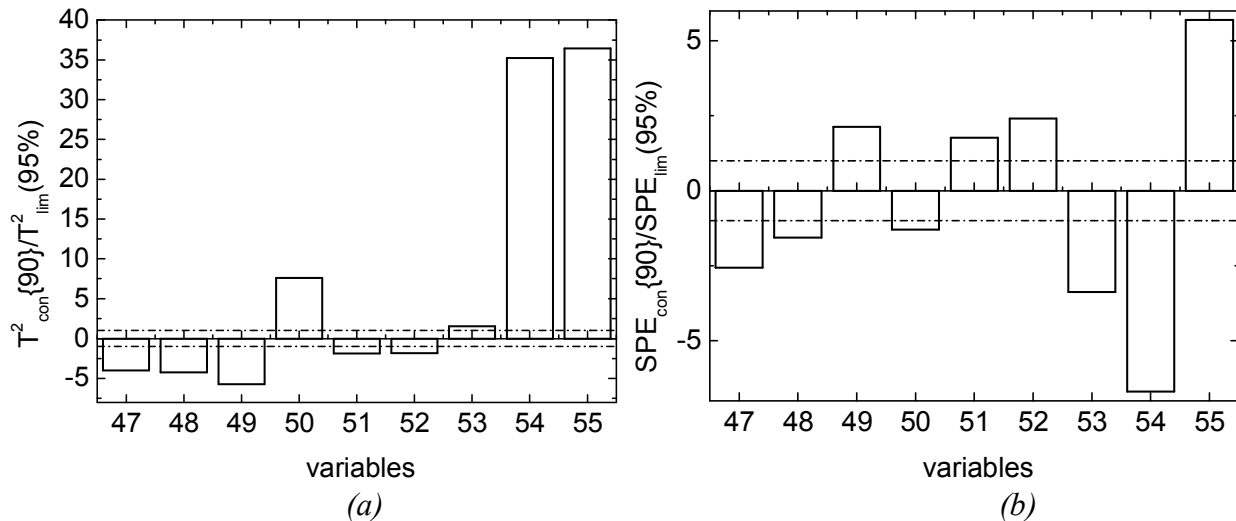


Figura 5.14 (a) Rapporto tra i contributi al T^2 per variabile e i limiti di confidenza al 95%
 (b) Rapporto tra i contributi a SPE per variabile e i limiti di confidenza al 95% per l'istante di campionamento {90} relativo al lotto {5} durante il malfunzionamento

È stato inoltre costruito un modello PLS senza considerare la variabile [55] cioè la portata di solido in ingresso al granulatore per valutare se i *fault* venivano comunque individuati. La varianza spiegata da ciascuna variabile latente è riportata in Tabella 5.4. Si noti che in questo caso sono state selezionate tramite convalida incrociata tre variabili latenti diversamente da quanto risultava per il modello costruito con tutte le variabili di \mathbf{X}_1 (4 LV, Tabella 5.2).

Tabella 5.4 Varianza spiegata dalle prime 3 variabili latenti relativa alla fase di granulazione (\mathbf{X}_1) e alle caratteristiche del prodotto all'uscita del mulino (\mathbf{Y}_1) per il modello PLS senza la variabile [55]

Number of LV	Variance X1	Variance cumulative	Variance Y1	Variance cumulative
1	33.21	33.21	19.50	19.50
2	34.58	67.80	4.257	23.76
3	3.938	71.74	7.670	31.43

In Tabella 5.5 sono riportate le prestazioni del modello in termini di varianza in calibrazione e in convalida incrociata per ciascuna variabile di qualità all'uscita del mulino. Le prestazioni sono scarse.

Tabella 5.3 Varianza in calibrazione e in convalida incrociata per ciascuna variabile della matrice Y_1

Y_1 variables	SS_c	RMSEC	SS_{cv}	RMSECV
$\rho_{aerated}$ [g/mL]	0.459	0.014	0.457	0.014
ρ_{tapped} [g/mL]	0.312	0.018	0.308	0.018
Hausner ratio []	0.355	0.025	0.351	0.025
x_{10} [μm]	0.420	7.9	0.417	7.92
x_{50} [μm]	0.198	89.07	0.195	89.21
x_{90} [μm]	0.102	181.6	0.099	181.9
span []	0.354	0.32	0.350	0.32

Sono state proiettate le misure relative al *fault* del lotto {5} sul modello PLS costruito senza la portata di solido [55] e in Figura 5.14 sono riportate le carte di monitoraggio.

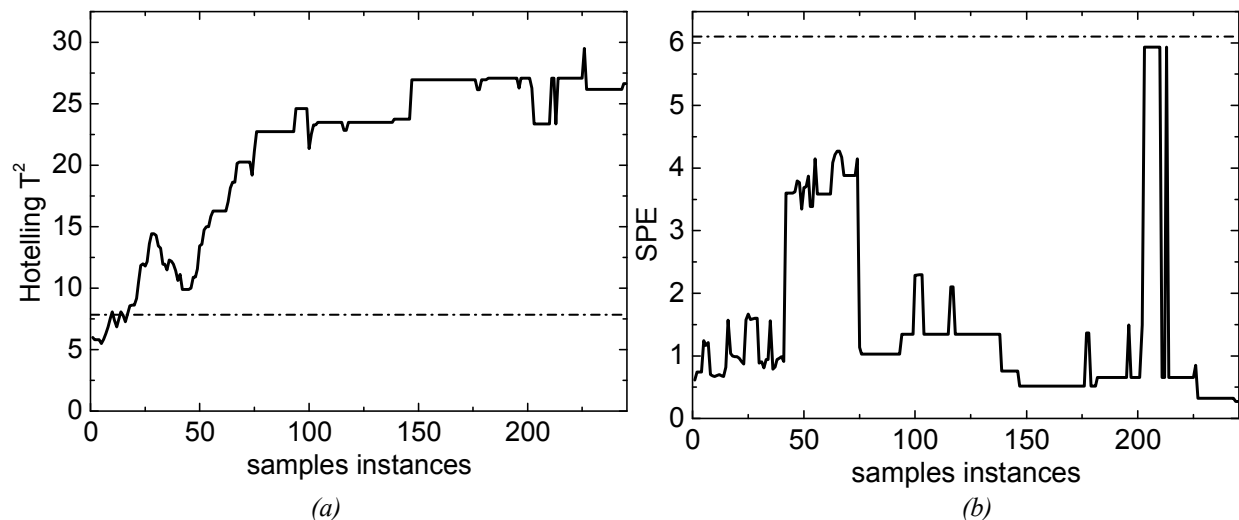


Figura 5.15 proiezione di 245 istanti di campionamento del lotto {5} sul modello PLS senza la variabile [55]

In Figura 5.15a si può osservare che il T^2 esce dai limiti di confidenza all'istante numero {20} come nel caso precedente (Figura 5.12a); la portata di solido entrante al granulatore non è quindi l'unica variabile che risente dell'anomalia, e l'anomalia non è dovuta ad un malfunzionamento strumentale.

Conclusioni

Il sistema produttivo farmaceutico, tradizionalmente basato su operazioni discontinue, sta spostando la propria attenzione verso l'utilizzo di processi manifatturieri continui, che consentono un più facile controllo delle operazioni, minimizzando i problemi di gestione e sicurezza degli impianti, con il vantaggio di favorire la riduzione delle dimensioni degli stessi, limitandone anche l'impatto ambientale. Allo stesso tempo, l'industria farmaceutica sta attraversando una fase di cambiamento per quanto riguarda le attività di sviluppo di nuovi processi e prodotti. In particolare, secondo il nuovo approccio, incentivato dalle agenzie di controllo come la *Food and Drug Administration* (FDA), lo sviluppo di nuovi prodotti e relativi processi produttivi deve passare attraverso la comprensione scientifica delle relazioni esistenti tra materie prime, processo produttivo e prodotto finale, al fine di progettare la qualità all'interno del prodotto stesso, piuttosto che verificarne l'accettabilità alla fine del processo produttivo.

Per queste ragioni, l'industria farmaceutica necessita di strumenti adeguati per accelerare le attività di sviluppo, attraverso l'organizzazione sistematica di esperimenti e la comprensione su base scientifica dei dati, al fine di comprendere le principali fonti di variabilità da controllare per ottimizzare e rendere robusta la produzione.

In questa Tesi sono stati utilizzati metodi chemiometrici, come PCA e PLS, come strumenti di supporto per lo sviluppo di una linea di produzione in continuo di compresse a base di paracetamolo. L'obiettivo è stato quello di identificare e ordinare le fonti di variabilità in ingresso al processo (proprietà delle materie prime, parametri di processo) secondo il loro effetto sulla qualità del prodotto finale, ai fini di assicurarne il controllo e di ottimizzare la produzione. A tale scopo sono stati raccolti e riorganizzati dati relativi alle condizioni iniziali, allo step di granulazione, all'uscita del granulatore, all'uscita del mulino e alla comprimitrice. L'analisi PCA dei dati relativi alle operazioni unitarie del processo ha evidenziato quali siano le principali forze motrici del processo. La principale fonte di variabilità dei lotti processati è risultata essere il diverso metodo di isolamento del materiale in ingresso al processo (*microfluidised* o *wet milled*). Questa fonte di variabilità influenza fortemente le caratteristiche del processo di granulazione e dei prodotti in uscita. Si è visto che i lotti *microfluidised* sono in generale caratterizzati da ampiezza (*span*) elevata della distribuzione granulometrica e granuli coesivi mentre i lotti *wet milled* sono caratterizzati da granuli di dimensioni elevate, densità elevate e portata massima di materiale in ingresso al granulatore elevata. La seconda fonte di variabilità è data dalle diverse modalità di additivazione, cioè il punto in cui il principio attivo è mescolato e additivato con gli eccipienti ("+" *excipient*" o *API alone*). Questa caratteristica della materia in ingresso condiziona, anche se in modo minore, il

processo: se un materiale *wet milled* viene additivato al punto di isolamento (+ *excipient*), questo mediamente porta ad un aumento della densità e quindi della portata massima entrante al granulatore e una diminuzione della dimensione dei granuli; mentre se viene additivato dopo l'isolamento (*API alone*) è caratterizzato da basse densità e granuli di dimensioni elevate. La terza forza motrice è la diversa portata d'acqua utilizzata nella granulazione ad umido dei vari materiali. Questa influenza le caratteristiche del prodotto all'uscita del granulatore, dato che a maggiori portate di acqua in ingresso corrispondono maggiori dimensioni dei granuli.

Le analisi dei dati a disposizione hanno permesso di identificare eventuali misure ridondanti che vengono effettuate sul processo e sul prodotto finale. Per esempio, si è visto che le misure di distribuzione all'uscita del granulatore (eseguite con PIMS) sono molto correlate a quelle eseguite all'uscita del mulino (con SYMPATEC). Quindi potrebbe essere sufficiente eseguire solo una delle due misure per ottenere la stessa informazione sul processo.

Oltre a ciò, è stato verificato che le condizioni operative della comprimitrice sono molto importanti e determinano le proprietà delle compresse introducendo specifiche correlazioni tra le variabili. La fase di compressione agisce sui lotti processati uniformando le differenze dovute ai diversi metodi di isolamento e alle diverse modalità di additivazione, e introducendo una variabilità dovuta solamente alle condizioni operative alla comprimitrice. In generale, prove effettuate con una distanza del punzone elevata portano alla formazione di compresse ad elevato spessore ed elevato grado di plasticità. Il modello PCA costruito sulle proprietà delle compresse ha permesso di capire che compresse più dense e più compatte sono anche meno plastiche, e che compresse più pesanti sono anche meno porose. Inoltre l'analisi PCA di dati relativi a compresse ottenute a pari distanza del punzone ha evidenziato che i diversi metodi di isolamento influenzano alcune proprietà delle compresse come durezza e resistenza a trazione mentre le diverse tipologie di additivazione influenzano altre proprietà come densità e porosità.

L'utilizzo di modelli PLS e MBPLS ha permesso di verificare che le condizioni iniziali e al granulatore hanno un'influenza analoga nella stima delle caratteristiche del prodotto all'uscita del mulino. Il modello MBPLS è inoltre utile dal momento che riunisce i dati delle singole operazioni unitarie fornendo un modello globale del processo. Ciò fornisce un potenziale strumento per lo sviluppo del *design space* del processo globale, attraverso l'unione dei *design spaces* delle singole operazioni unitarie.

I modelli PLS sono stati utilizzati anche per lo sviluppo di sensori virtuali per ottenere una stima di proprietà difficili da misurare, come la resistenza a trazione delle compresse, a partire da variabili misurate durante la fase di compressione.

Infine si è dimostrato come modelli PCA e PLS possono essere utilizzati per il monitoraggio in linea del granulatore al fine di rilevare eventuali anomalie, per mantenere un standard costante di qualità del prodotto finale. I modelli sono stati testati per la rilevazione di due

anomalie: l'esaurimento del materiale in ingresso al granulatore, e l'arresto della vite di movimentazione del materiale nell'apparecchiatura. In entrambi i casi i metodi usati hanno permesso di rilevare le anomalie con successo e puntualmente. In particolare, nel secondo caso (arresto della vite), i modelli sono riusciti a riscontrare l'anomalia in anticipo rispetto all'istante in cui questa è stata segnalata dagli operatori.

Questi risultati sono limitati al materiale a cui i dati analizzati si riferiscono (paracetamolo) o a materiali aventi caratteristiche simili. Le conclusioni a cui si è giunti in termini di comprensione del processo e costruzione di modelli di stima e monitoraggio non sono state confermate dall'analisi effettuata sui dati relativi un diverso granulato, probabilmente di caratteristiche diverse da quelle del paracetamolo. Inoltre lo scarso numero di dati analizzati e l'acquisizione non sistematica limita l'affidabilità dei modelli, soprattutto per quanto riguarda la stima delle proprietà del prodotto finale. Ciononostante, i metodi utilizzati mostrano elevate potenzialità anche da questo punto di vista, dal momento che possono essere utilizzati per guidare la sperimentazione, suggerendo quali possono essere i migliori esperimenti da effettuare per aumentare la conoscenza sul processo e sul suo *design space*. Questo offre importanti opportunità di sviluppare in futuro il lavoro svolto.

Nomenclatura

a	=	componente principale generica, pedice generico (-)
A	=	numero di componenti principali utilizzate in un modello PCA (-)
b	=	generico blocco (-)
b_m	=	coefficiente di regressione m-esimo del modello PLS (-)
B	=	numero complesso di blocchi (-)
\mathbf{B}	=	combinazione delle matrici Z , X_2 , Y_1
$\mathbf{B}_{3,4,5}$	=	matrice \mathbf{B} senza i lotti {3}, {4} e {5}
$\mathbf{BP}_{3,4,5}$	=	matrice $\mathbf{B}_{3,4,5}$ senza le variabili relative alla distribuzione granulometrica all'uscita del granulatore
$\mathbf{BS}_{3,4,5}$	=	matrice $\mathbf{B}_{3,4,5}$ senza le variabili relative alla distribuzione granulometrica all'uscita del mulino
BIP	=	<i>block importance in projection</i>
c_α	=	deviazione normale standard (-)
$cov(\mathbf{X})$	=	covarianza della matrice \mathbf{X} (-)
\mathbf{C}	=	combinazione delle matrici Z , X_2 , Y_1 , Y_2
CI	=	indice di Carr (%)
D	=	dimetro della compressa (mm)
\mathbf{e}_i	=	vettore riga della matrice dei residui (-)
\mathbf{E}	=	matrice dei residui della matrice \mathbf{X} (-)
E_{irr}	=	energia non recuperata (J)
E_{rec}	=	energia recuperata (J)
E_{tot}	=	energia totale (J)
$ExtTorqServoNM$	=	coppia misurata al motore (Nm)
$ExtTorqServoPV$	=	coppia massima in percentuale misurata al motore (%)
$ExtTorqTransd$	=	coppia misurata al trasduttore (Nm)
$ExtSpdAV$	=	velocità del motore misurata (rpm)
\mathbf{F}	=	matrice dei residui della matrice \mathbf{Y} (-)
$F_{A,k-1,\alpha}$	=	distribuzione statistica F (-)
fl	=	profondità di riempimento (mm)
F_{Ej}	=	<i>max ejection force</i> (kN)
F_{LoP}	=	<i>max lower punch force</i> (kN)
F_{UpP}	=	<i>max upper punch force</i> (kN)
h	=	pedice generico (-)
h_0	=	coefficiente dell'equazione di Jackson-Mudholkar (-)

H	=	durezza (kp)
HR	=	Hausner ratio (adm)
i	=	i -esimo componente della matrice \mathbf{X} / pedice generico (-)
I	=	numero delle righe della matrice \mathbf{X} (-)
J	=	numero delle variabili del processo/ di colonne della matrice \mathbf{X} (-)
k	=	generico indice di campionamento (-)
l	=	minimum punch-tip separation distance (mm)
L	=	numero di variabili della matrice \mathbf{Y} (-)
L10	=	decimo percentile della distribuzione granulometrica misurata con PIMS (μm)
L50	=	cinquantesimo percentile della distribuzione granulometrica misurata con PIMS (μm)
L90	=	novantesimo percentile della distribuzione granulometrica misurata con PIMS (μm)
m	=	pedice generico (-)
m_{x_i}	=	numero di variabili della matrice \mathbf{X}_1 (-)
M	=	numero di variabili latenti utilizzate nel modello PLS (-)
M_i	=	materiale i -esimo (-)
n	=	generico elemento (-)
N	=	numero di componenti (-)
NaN	=	<i>not a number</i>
p	=	numero di variabili della matrice (-)
\mathbf{p}_m	=	vettore dei <i>loadings</i> della matrice \mathbf{X} (-)
P	=	plasticità (-)
\mathbf{P}	=	matrice dei <i>loadings</i> della matrice \mathbf{X} (-)
PC_i	=	componente principale i -esima (-)
PIMS	=	<i>particle imaging measurement system</i>
PRESS	=	<i>prediction error of sum of squares</i>
q_m	=	vettore dei <i>loading</i> della matrice \mathbf{Y} (-)
\mathbf{Q}	=	matrice dei <i>loadings</i> della matrice \mathbf{Y} (-)
r	=	numero totale di componenti principali pari al minimo tra I e J (-)
RMSE	=	<i>root mean square error</i>
RMSECV	=	<i>root mean square error cross validation</i>
s	=	generico istante di tempo (-)
s_m	=	semiasse minore dell'ellisse di confidenza (-)
s_M	=	semiasse maggiore dell'ellisse di confidenza (-)
S	=	terza dimensione della matrice \mathbf{X} (-)
S_{Ej}	=	<i>max ejection stress</i> (MPa)

S_{LoP}	=	<i>max lower punch stress</i> (MPa)
Span	=	ampiezza della distribuzione
SPE	=	<i>squared prediction error</i>
SPE_{con}	=	contributo a SPE
SS_c	=	varianza in calibrazione (-)
SS_{cv}	=	varianza in convalida incrociata (-)
SS_p	=	varianza in predizione(-)
S_{tr}	=	<i>stress transmission ratio</i> (-)
S_{UpP}	=	<i>max upper punch stress</i> (MPa)
t	=	tablet thickness (mm)
\mathbf{t}_m	=	vettore degli <i>scores</i> della matrice \mathbf{X} (-)
\mathbf{T}	=	matrice degli <i>scores</i> della matrice \mathbf{X} (-)
T^2	=	Hotelling T^2
$T^2_{A,k,\alpha}$	=	limite di confidenza per il diagramma degli <i>scores</i> con limite di confidenza $(1 - \alpha)$ (-)
T_{con}	=	contributi al T^2 (-)
\mathbf{U}	=	matrice degli <i>scores</i> della matrice \mathbf{Y} (-)
$var(x_j)$	=	varianza del vettore x_j (-)
VIP	=	<i>variable importance in projection</i> (-)
$VIPC_a$	=	VIP complessivo, unico per tutte le variabili predette, della variabile predittrice a
x_{10}	=	decimo percentile della distribuzione granulometrica (μm)
x_{50}	=	cinquantesimo percentile della distribuzione granulometrica (μm)
x_{90}	=	novantesimo percentile della distribuzione granulometrica (μm)
\mathbf{x}_i	=	riga i -esima della matrice di processo \mathbf{X} (-)
$x_{i,j}$	=	elemento della matrice \mathbf{X} di dimensioni $(I \times J)$ (-)
\bar{x}_j	=	valore medio sugli elementi del vettore x_j (-)
\mathbf{X}	=	matrice di processo bidimensionale (-)
$\underline{\mathbf{X}}$	=	matrice di processo tridimensionale (-)
\mathbf{X}_1	=	matrice delle misure on-line al granulatore (-)
\mathbf{X}_2	=	matrice delle misure all'uscita del granulatore (-)
y_i	=	elemento generico della matrice \mathbf{Y} (-)
\hat{y}_i	=	i -esimo elemento ricostruito della matrice \mathbf{Y} (-)
\bar{y}	=	valore medio di ciascuna variabile della matrice \mathbf{Y} (-)
\mathbf{Y}	=	matrice delle caratteristiche del prodotto (-)
\mathbf{Y}_1	=	matrice delle misure all'uscita del mulino (-)
\mathbf{Y}_2	=	matrice delle misure alla comprimitrice (-)
T_i	=	temperatura della sezione i -esima del granulatore ($^{\circ}\text{C}$)

w	=	peso della compressa (mg)
w_m	=	peso m -esimo del modello PLS (-)
$w_t\%$	=	<i>granulator water content (%)</i>
\mathbf{W}	=	matrice dei pesi del modello PLS (-)
$z_{\alpha/2}$	=	funzione calcolata utilizzando la distribuzione t-Student con livello di confidenza α (-)
\mathbf{Z}	=	matrice delle condizioni iniziali (-)
[]	=	generica variabile
{ }	=	materiale/ lotto/ istante di campionamento/ campione

Lettere greche

α	=	grado di confidenza (-)
δ_{ij}	=	delta di Kronecker
Θ_i	=	coefficienti dell'equazione di Jackson-Mudholkar (-)
λ_m	=	autovalore associato al m -esimo vettore p_m (-)
ρ_t	=	densità battuta (g/ml)
ρ_p	=	densità aerata (g/ml)
ρ	=	densità (g/cm ³)
ρ_r	=	densità (adm)
σ	=	tensile strength/ resistenza a trazione (MPa)
σ_{est}	=	resistenza a trazione stimata con una relazione empirica (MPa)
σ_{std}	=	deviazione standard (-)
σ_x	=	deviazione standard dei contributi (-)
φ	=	porosità (adm)

Acronimi

CPM	=	continuous pharmaceutical manufacturing
CQA	=	critical quality attribute
MBPLS	=	multiblock partial least squares
NIPALS	=	non-linear iterative partial least squares
PAT	=	process analytical technology
PCA	=	principal component analysis
PLS	=	partial least squares
PSE	=	process systems engineering

Appendice 1

Rappresentazione delle matrici dei dati

Nell'Appendice sono riportate le matrici dei dati relativi alla linea di produzione di compresse a base di paracetamolo. In Figura A1.1 è rappresentato il diagramma del processo di granulazione e compattazione con il dettaglio del granulatore. In Figura A1.2 è riportata l'organizzazione in matrici dei dati disponibili e la relativa simbologia impiegata. In Tabella A1.0 è stata riportata la matrice \mathbf{B} , creata affiancando i dati contenuti nelle matrici \mathbf{Z} , \mathbf{X}_2 , \mathbf{Y}_1 di Figura A1.1. In Tabella A1.1 è riportata la matrice \mathbf{Y}_2 contenente i dati della fase di compressione. In Tabella A1.2 sono riportate le misure relative al primo istante di campionamento della matrice \mathbf{X}_1 relativa alle misure in linea al granulatore.

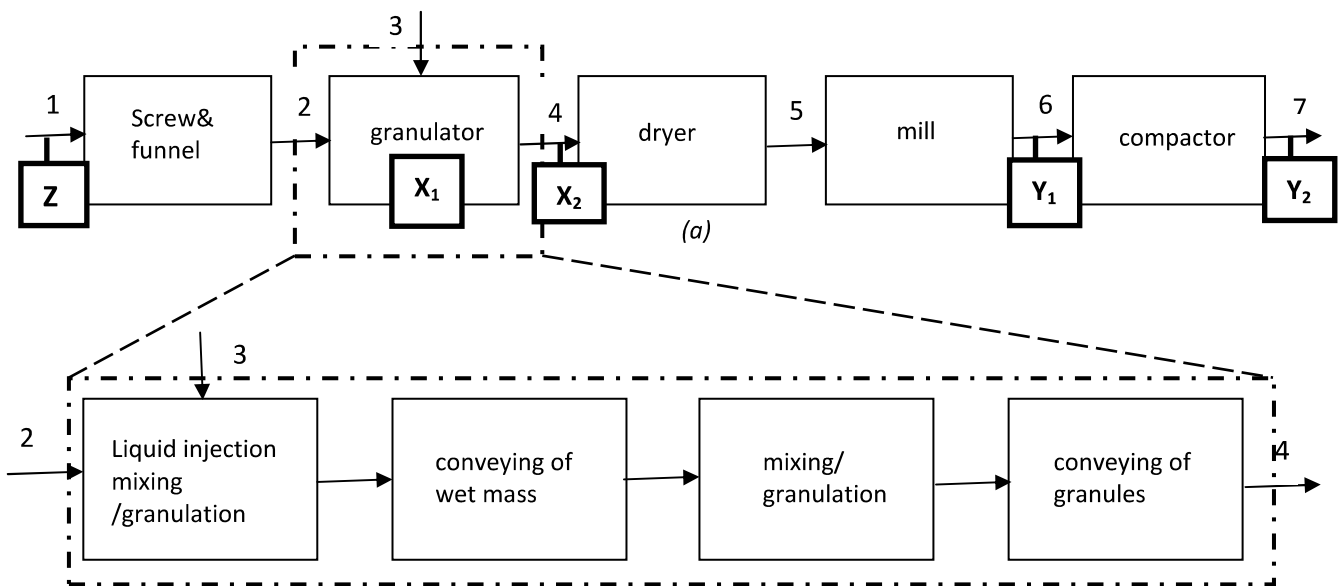


Figura A1.1 Diagramma a blocchi del processo di granulazione in continuo del paracetamolo e dettaglio del granulatore.

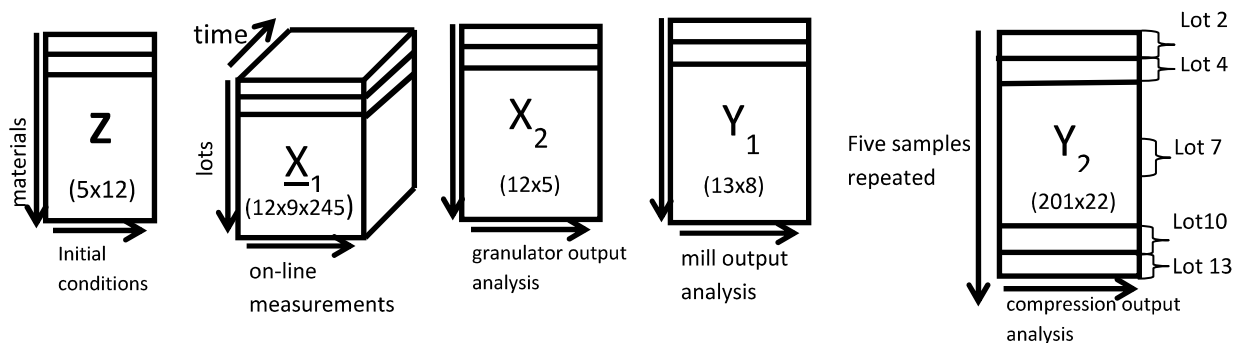


Figura A1.2 Organizzazione in matrici dei dati disponibili e relativa simbologia impiegata

Variables

[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]
Z ₁	Z ₂	Z ₃	Z ₄	Z ₅	Z ₆	Z ₇	Z ₈	Z ₉	Z ₁₀	Z ₁₁	Z ₁₂	X ₂₋₁	X ₂₋₂	X ₂₋₃	X ₂₋₄	X ₂₋₅	Y ₁₋₁	Y ₁₋₂	Y ₁₋₃	Y ₁₋₄	Y ₁₋₅	Y ₁₋₆	Y ₁₋₇	Y ₁₋₈
Aldrich material, different isolation	weight (g)	(1) "4 excipients"/ (0) "API alone"	(1) Wet milled/ (0) microfluidised	p aerated "raw"	p tapped "raw"	Hausner Ratio "raw"	x10 "raw"	x50 "raw"	x90 "raw"	span "raw"	feed factor at charge (kg/h)	granulator water content (%)	L10 PIMS	L50 PIMS	L90 PIMS	Span PIMS	not milled	p aerated	p tapped	Hausner Ratio	x10	x50	x90	span
{1}	108353-104A	1	0	NaN	NaN	NaN	4.09	13.6	80.5	5.62	13.38	15	62.9	213.75	697.56	2.97	0	0.49	0.65	1.33	33.2	306	798	2.5
{2}	108353-104B	1	0	NaN	NaN	NaN	4.09	13.6	80.5	5.62	13.38	17.5	94.3	272.63	899.73	2.95	0	0.49	0.61	1.25	43.6	395	934	2.25
{3}	108353-101A	1	1	0.42	0.73	1.74	7.93	28.1	92.5	3.01	16.03	15	NaN	NaN	NaN	NaN	0	0.51	0.68	1.34	34.9	313	1087	3.36
{4}	108353-101B	1	1	0.42	0.73	1.74	7.93	28.1	92.5	3.01	16.03	17.5	NaN	NaN	NaN	NaN	0	0.52	0.66	1.28	46.8	546	1357	2.4
{5}	108353-101C	1	1	0.42	0.73	1.74	7.93	28.1	92.5	3.01	16.03	20	NaN	NaN	NaN	NaN	0	0.5	0.67	1.34	36.1	378	859	2.18
{6}	108353-103A	0	0	0.35	0.6	1.74	3.86	12.3	81.5	6.31	12.8	15	62.9	153.49	567.2	3.29	0	0.46	0.62	1.35	21.7	209	630	2.77
{7}	108353-103B	0	0	0.35	0.6	1.74	3.86	12.3	81.5	6.31	12.8	17.5	76.36	220.59	773.94	3.16	0	0.48	0.65	1.34	32.2	334	958	2.91
{8}	108353-103C	0	0	0.35	0.6	1.74	3.86	12.3	81.5	6.31	12.8	20	94.3	298.75	1047.75	3.19	0	0.46	0.61	1.32	55.2	592	1273	2.06
{9}	108353-107A	0	1	0.39	0.62	1.6	9.11	33.8	102	2.75	14.12	15	94.32	297.72	897.97	2.7	0	0.51	0.66	1.31	50.5	361	841	2.19
{10}	108353-108	0	1	0.39	0.62	1.6	9.11	33.8	102	2.75	14.12	17.5	125.78	419.1	1139.7	2.42	0	0.51	0.65	1.26	59.6	438	952	2.04
{11}	108353-108A	0	1	0.39	0.62	1.6	9.11	33.8	102	2.75	14.12	17.5	125.78	419.1	1139.7	2.42	1	0.48	0.56	1.19	361	1749	2821	1.41
{12}	108353-109A	0	0	0.36	0.59	1.61	6.67	21.1	137	6.18	13.89	15	88.62	232.46	726.62	2.74	0	0.48	0.63	1.3	36	348	1055	2.93
{13}	108353-109B	0	0	0.36	0.59	1.61	6.67	21.1	137	6.18	13.89	17.5	88.02	273.94	831.6	2.71	0	0.49	0.64	1.3	45.3	377	922	2.33

Variables

[26]	[27]	[28]	[29]	[30]	[31]	[32]	[33]	[34]	[35]	[36]	[37]	[38]	[39]	[40]	[41]	[42]	[43]	[44]	[45]	[46]
Y ₂₋₁	Y ₂₋₂	Y ₂₋₃	Y ₂₋₄	Y ₂₋₅	Y ₂₋₆	Y ₂₋₇	Y ₂₋₈	Y ₂₋₉	Y ₂₋₁₀	Y ₂₋₁₁	Y ₂₋₁₂	Y ₂₋₁₃	Y ₂₋₁₄	Y ₂₋₁₅	Y ₂₋₁₆	Y ₂₋₁₇	Y ₂₋₁₈	Y ₂₋₁₉	Y ₂₋₂₀	Y ₂₋₂₁
minimum punch tip separation distance / mm	Fill depth / mm	Tablet weight / mg	tablet thickness / mm	Tablet diameter / mm	Hardness / KP	Density / g.cm-3	Relative density	Porosity	Total energy / J	Recovered Energy / J	Irrecoverable energy / J	Plasticity ratio	Max Upper Punch Force / kN	Max Lower Punch Force / kN	Max Ejection Force / kN	Max Upper Punch Stress / MPa	Max Lower Punch Stress / MPa	Max Ejection Stress / MPa	Stress Transmission Ratio	Tensile Stress / MPa
{1}	4.2	8.3	304.4	4.24	10.02	1.6	0.9104	0.33	1.27	0.06	1.21	0.95	2.01	2.19	0.173	25.59	27.88	1.30	1.09	0.24
{201}	3.8	9	366.3	4	10.02	NaN	1.1613	0.14	NaN	NaN	NaN	NaN	10.534	NaN	NaN	134.12	NaN	NaN	NaN	Estimated tensile stress

[47]	[48]	[49]	[50]	[51]	[52]	[53]	[54]	[55]
X ₁₋₁	X ₁₋₂	X ₁₋₃	X ₁₋₄	X ₁₋₅	X ₁₋₆	X ₁₋₇	X ₁₋₈	X ₁₋₉
ExtTorqServoNM	ExtTorqServoPV	ExtTorqTransd	zone7T	zone8T	zone9T	zone10T	ExtSpdAV	FeederRate
3.333	14.035	2.643	20.170	22.175	25.565	23.091	199.314	1.989
...
{13}	2.892	12.213	20.246	23.027	27.432	27.000	199.143	2.002

Processo di granulazione e compattazione del paracetamolo

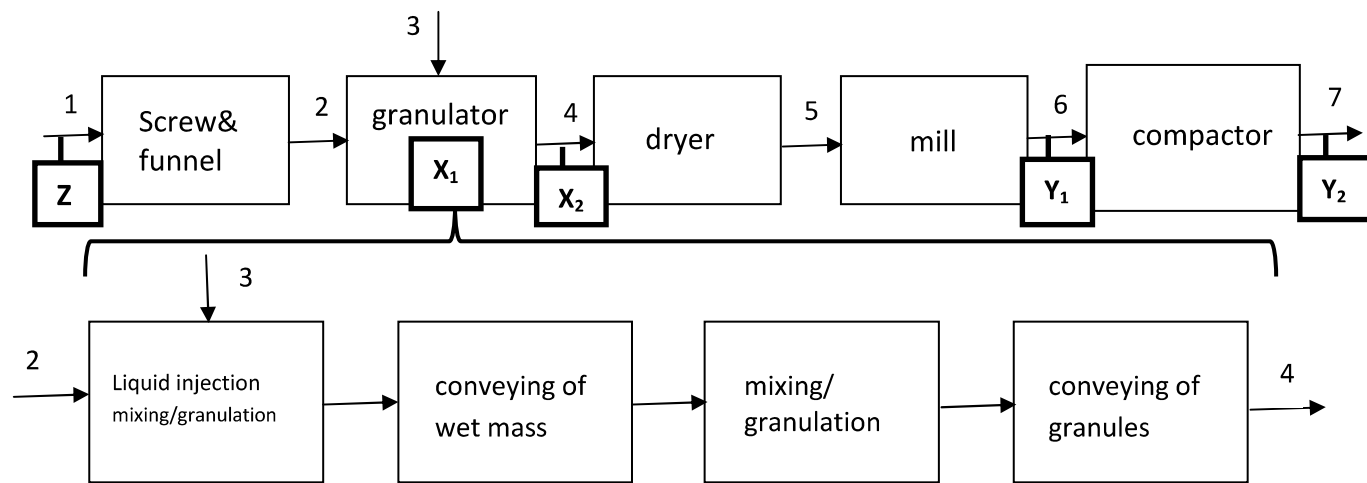


Figura A1.1 Diagramma a blocchi del processo di granulazione in continuo del paracetamolo e dettaglio del granulatore.

I dati contenuti nelle matrici Z , X_2 e Y_1 sono stati affiancati nella matrice B per capire le somiglianze tra lotti (Tabella A1.0).

Tabella A1.0 Condizioni iniziali, all'uscita del granulatore e all'uscita del mulino per 13 lotti (matrice B)

Lots		Variables																									
		[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]	
		Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}	Z_{11}	Z_{12}	X_{21}	X_{22}	X_{23}	X_{24}	X_{25}	Y_{11}	Y_{12}	Y_{13}	Y_{14}	Y_{15}	Y_{16}	Y_{17}	Y_{18}	
		Aldrich material, different isolation	weight (g)	"+" excipients"/ (0) "API alone"	(1) Wet milled/ (0) microfluidised	ρ aerated "raw"	ρ tapped "raw"	Hausner Ratio "raw"	x10 "raw"	x50 "raw"	x90 "raw"	span "raw"	feed factor at charge (kg/h)	granulator water content (%)	L10 PIMS	L50 PIMS	L90 PIMS	Span PIMS	not milled	ρ aerated	ρ tapped	Hausner Ratio	x10	x50	x90	span	
	{1}	108353-104A	1	702	1	0	NaN	NaN	NaN	4.09	13.6	80.5	5.62	13.38	15	62.9	213.75	697.56	2.97	0	0.49	0.65	1.33	33.2	306	798	2.5
	{2}	108353-104B	1	702	1	0	NaN	NaN	NaN	4.09	13.6	80.5	5.62	13.38	17.5	94.3	272.63	899.73	2.95	0	0.49	0.61	1.25	43.6	395	934	2.25
	{3}	108353-101A	0	1213	1	1	0.42	0.73	1.74	7.93	28.1	92.5	3.01	16.03	15	NaN	NaN	NaN	NaN	0	0.51	0.68	1.34	34.9	313	1087	3.36
	{4}	108353-101B	0	1213	1	1	0.42	0.73	1.74	7.93	28.1	92.5	3.01	16.03	17.5	NaN	NaN	NaN	NaN	0	0.52	0.66	1.28	46.8	546	1357	2.4
	{5}	108353-101C	0	1213	1	1	0.42	0.73	1.74	7.93	28.1	92.5	3.01	16.03	20	NaN	NaN	NaN	NaN	0	0.5	0.67	1.34	36.1	378	859	2.18
	{6}	108353-103A	0	922	1	0	0.35	0.6	1.74	3.86	12.3	81.5	6.31	12.8	15	62.9	153.49	567.2	3.29	0	0.46	0.62	1.35	21.7	209	630	2.91
	{7}	108353-103B	0	922	1	0	0.35	0.6	1.74	3.86	12.3	81.5	6.31	12.8	17.5	76.36	220.59	773.94	3.16	0	0.48	0.65	1.34	32.2	334	958	2.77
	{8}	108353-103C	0	922	1	0	0.35	0.6	1.74	3.86	12.3	81.5	6.31	12.8	20	94.3	298.75	1047.75	3.19	0	0.46	0.61	1.32	55.2	592	1273	2.06
	{9}	108353-107A	0	710	0	1	0.39	0.62	1.6	9.11	33.8	102	2.75	14.12	15	94.32	297.72	897.97	2.7	0	0.51	0.66	1.31	50.5	361	841	2.19
	{10}	108353-108	0	710	0	1	0.39	0.62	1.6	9.11	33.8	102	2.75	14.12	17.5	125.78	419.1	1139.7	2.42	0	0.51	0.65	1.26	59.6	438	952	2.04
	{11}	108353-108A	0	710	0	1	0.39	0.62	1.6	9.11	33.8	102	2.75	14.12	17.5	125.78	419.1	1139.7	2.42	1	0.48	0.56	1.19	361	1749	2821	1.41
	{12}	108353-109A	0	500	0	0	0.36	0.59	1.61	6.67	21.1	137	6.18	13.89	15	88.62	232.46	726.62	2.74	0	0.48	0.63	1.3	36	348	1055	2.93
	{13}	108353-109B	0	500	0	0	0.36	0.59	1.61	6.67	21.1	137	6.18	13.89	17.5	88.02	273.94	831.6	2.71	0	0.49	0.64	1.3	45.3	377	922	2.33

Tabella A1.1 Dati di compattazione (matrice Y_2)

Samples		Variables																					
		[26]	[27]	[28]	[29]	[30]	[31]	[32]	[33]	[34]	[35]	[36]	[37]	[38]	[39]	[40]	[41]	[42]	[43]	[44]	[45]	[46]	
		Y_{21}	Y_{22}	Y_{23}	Y_{24}	Y_{25}	Y_{26}	Y_{27}	Y_{28}	Y_{29}	Y_{210}	Y_{211}	Y_{212}	Y_{213}	Y_{214}	Y_{215}	Y_{216}	Y_{217}	Y_{218}	Y_{219}	Y_{220}	Y_{221}	
		minimum punch tip separation distance / mm	duration of profile / sec	Fill depth / mm	Tablet weight / mg	tablet thickness / mm	Tablet diameter / mm	Hardness / KP	Density / g.cm-3	Relative density	Porosity	Total energy / J	Recovered Energy / J	Irrecoverable energy / J	Plasticity ratio	Max Upper Punch Force / kN	Max Lower Punch Force / kN	Max Ejection Force / kN	Max Upper Punch Stress / MPa	Max Lower Punch Stress / MPa	Max Ejection Stress / MPa	Stress Transmission Ratio	Tensile Stress / MPa
	{1}	4.2	68	8.3	304.4	4.24	10.02	1.6	0.9104	0.6744	0.33	1.27	0.06	1.21	0.95	2.01	2.19	0.173	25.59	27.88	1.30	1.09	0.24
	{201}	3.8	68	9	366.3	4	10.02	NaN	1.1613	0.8602	0.14	NaN	NaN	NaN	NaN	10.534	NaN	NaN	134.12	NaN	NaN	NaN	Estimated tensile stress

DATI

I dati ricevuti sono stati divisi in cinque matrici (Figure A1.2): condizioni iniziali (Z), misure on-line al granulatore (X_1), all'uscita del granulatore (X_2), all'uscita del mulino (Y_1), dati di compattazione (Y_2).

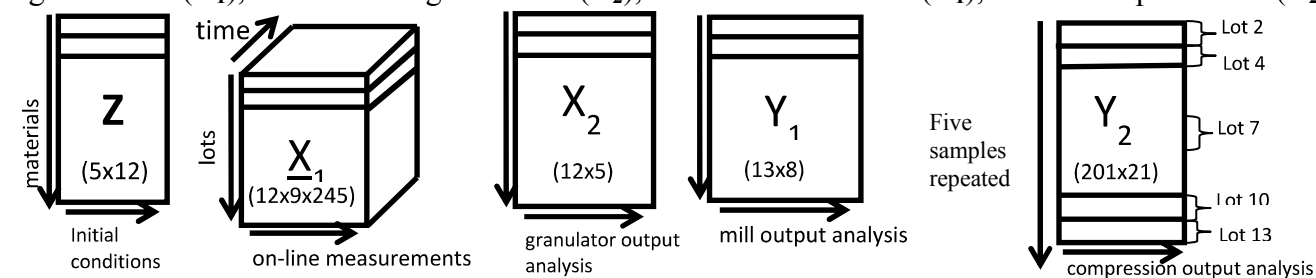


Figura A1.2 Organizzazione in matrici dei dati disponibili e relativa simbologia impiegata

Tabella A1.2 Misure on-line al granulatore per 12 lotti per il primo istante di campionamento (matrice X_1)

		[47]	[48]	[49]	[50]	[51]	[52]	[53]	[54]	[55]
		X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}
		ExtTorqServoNM	ExtTorqServoPV	ExtTorqTransd	zone7T	zone8T	zone9T	zone10T	ExtSpdAV	FeederRate
Lots	{1}	3.333	14.035	2.643	20.170	22.175	25.565	23.091	199.314	1.989

	{13}	2.892	12.213	2.377	20.246	23.027	27.432	27.000	199.143	2.002

Appendice 2

Figure e codici di calcolo

Nell'Appendice 2 vengono riportate le Tabelle dove è riportata la lista delle Figure presenti nei Capitoli della Tesi ed i relativi *file* grafici da cui sono selezionate, reperibili nella cartella `tesiDeMartin\Figure\`. Inoltre vengono riportate le Tabelle dove è riportato l'elenco dei codici di calcolo contenuti nella cartella `tesiDeMartin\Codici\`.

In Tabella A2.1 sono elencate le Figure del Capitolo 2, presenti nella cartella `tesiDeMartin\Figure\Capitolo2`.

Tabella A2.1 *Figure del Capitolo 2.*

Figura	File
Figura 2.5	M4.opj
Figura 2.7	water.opj

In Tabella A2.2 sono elencate le Figure del Capitolo 3, presenti nella cartella `tesiDeMartin\Figure\Capitolo3`.

Tabella A2.2 *Figure del Capitolo 3.*

Figura	File
Figura 3.1	BnoNaN.opj
Figura 3.2a	pcaZ.opj
Figura 3.2b	
Figura 3.4a	pcaX1.opj
Figura 3.4b	
Figure 3.5a	pcaX1medi.opj
Figura 3.5b	
Figure 3.6a	pcaY1.opj
Figura 3.6b	
Figure 3.7a	pcaY2.opj
Figura 3.7b	
Figura 3.8a	pcaY2C3.opj
Figura 3.8b	
Figura 3.9a	
Figura 3.9b	

Figura 3.10a	pcaY21.opj
Figura 3.10b	
Figura 3.12a	pcaB.opj
Figura 3.12b	
Figura 3.13a	
Figura 3.13b	
Figura 3.15	pcaB345.opj
Figura 3.17a	pcaZX1X2Y1.opj
Figura 3.17b	
Figura 3.18a	
Figura 3.18b	
Figura 3.20a	pcaC.opj
Figura 3.20b	
Figura 3.22	plsZwY1.opj
Figura 3.24a	mbplsZX1Y1.opj
Figura 3.24b	
Figura 3.24c	
Figura 3.27a	pcaX1noZw.opj
Figura 3.27b	
Figura 3.28a	pcaX1new.opj
Figura 3.28b	
Figura 3.29	
Figura 3.30a	pcaY2new.opj
Figura 3.30b	
Figura 3.31	parX1dapcanewX1.opj

In Tabella A2.3 sono elencate le Figure del Capitolo 4, presenti nella cartella `tesiDeMartin\Figure\Capitolo4`.

Tabella A2.3 *Figure del Capitolo 4.*

Figura	File
Figura 4.1a	predTSpar.opj
Figura 4.1b	
Figura 4.2	
Figura 4.3	predTSnew.opj
Figura 4.4a	
Figura 4.4b	
Figura 4.5	

In Tabella A2.4 sono elencate le Figure del Capitolo 5, presenti nella cartella `tesiDeMartin\Figure\Capitolo5`.

Tabella A2.4 *Figure del Capitolo 5.*

Figura	File
Figura 5.1a Figura 5.1b Figura 5.2 Figura 5.3	pcaX12f.opj
Figura 5.4	lotto5.opj
Figura 5.5a Figura 5.5b Figura 5.6 Figura 5.7a Figura 5.7b Figura 5.8a Figura 5.8b Figura 5.9a Figura 5.9b	pcaX15f.opj
Figura 10a Figura 10b Figura 11	plsX1Y12f.opj
Figura 12a Figura 12b Figura 13a Figura 13b Figura 14a Figura 14b Figura 15a Figura 15b	plsX1Y15f.opj

In Tabella A2.5 sono elencati i codici di calcolo usati in questa tesi, presenti nella cartella `tesiDeMartin\Codici`.

Tabella A2.5 *Codici di calcolo*

Codici di calcolo	Dati	Risultati
pcasc.m	ZX1X2Y1Y2.mat	blocchisc.mat
mbplsZX1Y1.m	ZX1X2Y1Y2.mat	mbplsr.mat
mbplsbb.m	ZX1X2Y1Y2.mat	mbplsbb.mat
pcasec.m	X1Y1Y2new.mat	blocchinew.mat
predTSpar.m	Y2par.mat	TSpar.mat
predTSsec.m	Y2new.mat	TSsec.mat
plsY1par.m	ZX1X2Y1Y2.mat	resplsY1par.mat
plsY1sec.m	X1Y1Y2new.mat	resplsY1sec.mat
pcafault.m	processo.mat	respcafault.mat
plsfault.m	processo.mat	plsfault.mat

Riferimenti bibliografici

- Appelgren, C. e C. Eskilson (1990). A novel method for the granulation and coating of pharmacologically active substances. *Drug Dev. Ind. Pharm.*, **16**, 2345-2351.
- Bergman, R., M. Johansson, T. Lundstedt, E. Seifert e J. Aberg (1998). Optimization of a granulation and tableting process by sequential design and multivariate analysis. *Chemom. Intell. Lab. Syst.*, **44**, 271–286.
- Chatterjee, S. (2012). FDA Perspective on Continuous Manufacturing. *Presentato al IFPAC Annual Meeting*. Baltimore (U.S.A).
- Choi, S. W. e I.-B. Lee (2005). Multiblock PLS-based localized process diagnosis. *J. Process Control*, **15**, 295-306.
- Chong, I.-G. e C.-H. Jun (2005). Performance of some variable selection methods when multicollinearity is present. *Chemom. Intell. Lab. Syst.*, **78**, 103-112.
- Conlin, A., E. Martin e A. Morris (2000). Confidence limits for contribution plots. *J. Chemom.*, **14**, 725–736.
- Facco , P., F. Doplicher, F. Bezzo e M. Barolo, (2009). Moving average PLS soft sensor for online product quality estimation in an industrial batch polymerization process. *J. Process Control* , **19**, 520–529.
- Facco, P. (2005). Monitoraggio di un processo semicontinuo di polimerizzazione con metodi statistici multivariati. *Tesi di Laurea in Ingegneria Chimica*. DIPIC, Università di Padova (Italy).
- Faure, A., P. York e R. Rowe, (2001). Process control and scale-up of pharmaceutical wet granulation processes: a review. *Eur. J. Pharm. Biopharm.*, **52**, 269 - 277.
- FDA. (2004). *Final report on pharmaceutical cGMPs for the 21st century – A risk based approach*. Rockville, (MD), USA.: Department of Health and Human Services, U.S. Food and Drug Administration.
- FDA. (2006). Guidance for industry: Q8 pharmaceutical development. *Center for Drug Evaluation and Research, U.S. Food and Drug Administration*. Rockville (MD), USA.
- Geladi, P. e B. Kowalski (1986). Partial least-squares regression: a tutorial. *Anal. Chim. Acta*, **185**,1-17.
- Gernaey, K. V., Albert E. Cervera-Padrell, John. M.Woodley, (2012). A perspective on PSE in pharmaceutical process development and innovation. *Computers Chem. Eng.*, doi:10.1016/j.compchemeng.2012.02.022

- Hayes, G. (1987). *Food Engineering Data Handbook*. Longman Scientific and Technical, Harlow, UK.
- Jackson, J. (1991). *A user's guide to principal components*. J. Wiley e Sons Inc New York (U.S.A.).
- Jiménez-González, C., P. Poechlauer, Q.B. Broxterman, Bing-Shiou Yang, David am Ende, J. Baird, C. Bertsch, R.E. Hannah, P. Dell'Orco, H. Noorman, S. Yee. R. Reintjens, A. Wells, V. Massonneau e J. Manley. (2011). Key green engineering research areas for sustainable manufacturing: a perspective from pharmaceutical and fine chemicals manufacturers. *Org. Process Res. Dev.*, **15**, 900-911.
- Kano, M., S. Hasebe e I. Hashimoto (2004). Evolution of multivariate statistical process control: application of independent component analysis and external analysis. *Computers Chem. Eng.*, **28**, 1157-1166.
- Kourti, T. (2003). Abnormal situation detection, three-way data and projection methods; robust data archiving and modeling for industrial applications. *Annual Reviews Control*, **27**,131-139.
- Kourti, T. (2003). Multivariate dynamica data modeling for analysis and statistical process control of batch process, start-ups and grade transitions. *J. Chemom.*, **17**, 93-109.
- Kourti, T. e J. MacGregor (1994). Process analysis, monitoring and diagnosis, using multivariate projection methods. *Chemom. Intell. Lab. Syst.*, **28**, 3-21.
- Kristensen, H. e T. Schaefer (1987). A review on pharmaceutical wet-granulation. *Drug. Dev. Ind. Pharm.*, **13**, 803-872.
- Li, B., J. Morris e E. B. Martin (2002). Model selection for partial least squares regression. *Chemom. Intell. Lab. Syst.*, **64**, 79-89.
- Liu, Z., M.-j. Brower, J. MacGregor, S.S. Rathore, D.E. Reed e M. J. Champagne (2011). Modeling and optimization of a tablet manufacturing Line. *J. Pharm. Innov.*, **6**, 170-180.
- MacGregor, J., C. Jaeckle, C. Kiparissides e M. Koutoudi (1994). Process monitoring and diagnosis by multiblock PLS methods. *AIChE J.*, **40**, 826-838.
- Nomikos, P. e J. MacGregor (1995). Multivariate SPC charts for monitoring batch process. *Technometrics*, **37**,41-59.
- Parikh, D. M. (1997). *Handbook of pharmaceutical granulation technology*. Dekker, New York, (U.S.A).
- Perry, R. e D. Green, (1997). Size reduction and size enlargement. In *Perry's Chemical Engineers' Handbook (7th Edition)*. McGraw-Hill, New York (U.S.A.).
- Schulze, D. (2007). *Powders and bulk solids: behavior, characterization, storage and flow*. Springer, New York (U.S.A).
- Soh, J., F. Wang, N. Boersen, R. Pinal, G. Peck e M. Carvajal (2008). Utility of multivariate analysis in modeling the effects of raw material properties and operating parameters

- on granule and ribbon properties prepared in roller compaction. *Drug Dev. Ind. Pharm.*, **34**,1022–1035.
- Westerhuis, J., T. Kourti e J. MacGregor (1998). Analysis of multiblock and hieratical pca and pls models. *J. Chemom.*, **12**, 301-321.
- Wise, B. e N. Ricker, (1991). Recent advances in multivariate statistical process control: improving robustness and sensitivity. *IFAC Symp. ADCHEM'91*, Toulouse (France), 125–130.
- Wise, B. e N. Gallagher, (1996). The process chemometrics approach to process monitoring and fault detection. *J. Process Control*, **6**, 329-348.
- Wise, B., N. Gallagher, R. Bro, J. Shaver, W. Windig e R. Koch (2006). Chemometrics Tutorial for PLS_Toolbox and Solo. Eigenvector Research, Wenatchee, WA (U.S.A).
- Woodcock, J. (23-27 October 2011). 2011 AAPS Annual Meeting and Exposition Webcasts. *AAPS Annual meeting*. Washington, D.C. (U.S.A).
- Yacoub, F., J. Lautens, e L. Lucisano (2011). Application of quality by design principle to legacy drug products. *J. Pharm. Innov.*, **6**, 61-68.
- Yihong , Q., C. Yisheng, L. Lirong e G. Geoff , (2009). *Developing solid oral dosage forms: pharmaceutical theory and practice*. Academic Press, New York, (U.S.A.).
- Yu, L. (2008). Pharmaceutical quality by design: product and process development, understanding, and control. *Pharma. Research*, **25**, 781-791.