UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Ingegneria Industriale

CORSO DI DOTTORATO DI RICERCA IN: Ingegneria Industriale
CURRICOLO: Ingegneria Chimica, dei Materiali e Meccanica
CICLO: XXIX

GRAIN PROCESS OPTIMIZATION BY DATA-DRIVEN MODELING

Coordinatore: Ch.mo Prof. Giovanni Meneghetti
Supervisore: Ch.mo Prof. Massimiliano Barolo

Dottorando : Filippo Dal Pastro

Foreword

The realization of the work included in this Dissertation involved the intellectual and financial support of many people and institutions, to whom the author is very grateful.
Most of the research activity that led to the results reported in this Dissertation has been carried out at CAPE-Lab, Computer-Aided Process Engineering Laboratory, at the Department of Industrial Engineering of the University of Padova (Italy). Part of the work was carried out at Bühler AG (Uzwil, Switzerland) during two stays (for an overall period of 8 months) under the supervision of Dr. Eliana Zamprogna. Furthermore, part of the experiments have been carried out at the industrial facility of Molino Quaglia (Vighizzolo d'Este, Padova). The author would like to thank Mr. Lucio Quaglia and Mr. Andrea Quaglia for providing access to their industrial facility.
Financial support to this study has been provided by Bühler AG (Uzwil, Switzerland).

All the material reported in this Dissertation is original, unless explicit references to studies carried out by other people are indicated. In the following, a list of the publications stemmed from this project is reported.

CONTRIBUTIONS IN INTERNATIONAL JOURNALS

Dal Pastro F., Facco P., Bezzo F., Zamprogna E., Barolo M. (2016). Data-driven modeling of milling and sieving operations in a wheat milling process. *Food Bioprod. Process.*, **99**, 99-108, doi: 10.1016/j.fbp.2016.04.007

Dal Pastro F., Facco P., Bezzo F., Zamprogna E., Barolo M. (2017). A model-based approach to the design and scale-up of wheat milling operations – proof of concept. *In preparation.*

CONTRIBUTIONS IN REFEREED CONFERENCE PROCEEDINGS

Dal Pastro F., Facco P., Bezzo F., Thomas H., Zamprogna E., Barolo M. (2015). Data-based multivariate modeling of a grain comminution process. In: Krist V. Gernaey Jakob K. Huusom and Rafiqul Gani, *Computer Aided Chemical Engineering*, 37, 2219-2224, Amsterdam:Elsevier, doi: 10.1016/B978-0-444-63576-1.50064-9

Dal Pastro F., Facco P., Bezzo F., Zamprogna E., Barolo M. (2016). Using PLS and NIR spectra to model the first-breakage step of a grain milling process. In: Zdravko Kravanja and Milos Bogataj, *Computer Aided Chemical Engineering*, 38, 1171-1176, Amsterdam: Elsevier, doi: 10.1016/B978-0-444-63428-3.50200-9

CONTRIBUTIONS IN UNREFEREED CONFERENCE PROCEEDINGS

Dal Pastro F., Facco P., Bezzo F., Zamprogna E., Barolo M. (2016). Verso la modellazione del processo di macinazione del grano. Presented at: GRICU Meeting 2016, September 12 - 14, 2016, Anacapri (NA), Italy

Abstract

Wheat milling is a complex process, in terms of plant layout, interactions between processing units and high variability of feed materials. Despite the rapidly increasing diffusion of analytical instruments to characterize important wheat and flour properties (e.g., near-infrared spectrometers and image analysis sensors), current wheat milling processes are still poorly instrumented. Therefore, process operation is mostly based on the operators' experience, and in most cases even the design of the process is based on experience only. On the other hand, severe quality requirements are usually set on the final product.
The development of science-driven (as opposed to experience-driven) tools to support process understanding, design and operation can be very beneficial for those industries that require wheat milling as a part of their manufacturing process. Intelligent manufacturing through digitalization is a trend in current manufacturing systems. This Dissertation moves along this trend by proposing an integrated, model-based approach to the optimization of wheat milling processes. In particular, two complementary approaches are investigated: development of data-driven models for single processing units, and simulation of entire milling processes.

The **models of single units** are used to tackle different issues that arise in the wheat milling operation: *i*) wheat classification, *ii*) process understanding, *iii*) prediction of product quality and *iv*) process design and scale-up.
Regarding **wheat classification**, a methodology based on the analysis of near-infrared (NIR) spectra in combination with latent variable modeling is developed. The objective is checking the ability of NIR spectra to discriminate different wheat varieties, an important aspect of the wheat milling operation, given the high intrinsic variability showed by natural products as wheat.
Wheat milling is a complex process which needs profound understanding to be operated in a satisfactory way. The analysis of the model parameters allows for a better comprehension of the breakage mechanism occurring in wheat milling, thus improving **process understanding**. The effect of the most influential process parameters and of the wheat properties onto the milled product characteristics is studied for different "passages" (i.e., combinations of processing units) of the process.
Expensive and time-consuming experiments can be partly avoided exploiting the predictive ability of data-driven models, which can be used to **predict the product quality** (i.e., the particle size distribution or the mass fraction of the milled or sieved material), thus supporting the operators in process operation. In this Dissertation, data-driven models are used to predict the product quality both for roller mills and for plansifters.

The problem of **process design** is related to the estimation of the process parameters to be used to obtain a product with a desired characteristic. Data-driven models provide a useful tool to this purpose, through the implementation of appropriate model inversion techniques. In this Dissertation, a latent variable model inversion methodology, that relies on an optimization framework and on data coming from different plant scales, is developed. The framework aims at suggesting the optimal process parameters to be used to obtain a product with desired properties, with the ultimate scope of accelerating the **scale-up** from the laboratory to the commercial plant.

The development of a platform for the **simulation of the whole milling process** is started in this Dissertation. Due to the complexity of the wheat milling operation, the process is still open to optimization, in terms of both layout and power consumption. The objective is developing a simulator of the whole milling process, that aims at modeling the pneumatic transport, pressure drops and power consumption. This tool can be used to get a better comprehension of the energy consumed throughout the mill and to optimize the process layout, minimizing the overall power consumption.

Riassunto

La macinazione del grano è un'operazione complessa, sia a causa delle complicate interazioni tra le unità, che risultano in una complessa struttura del processo stesso, che a causa dell'alta variabilità intrinseca di un prodotto naturale come il grano. Per queste ragioni, la macinazione è un processo difficile da esercire. Nonostante la grande diffusione di strumenti analitici per caratterizzare importanti qualità del grano e della farina (come ad esempio spettrometri al vicino infrarosso o analizzatori di immagini), l'attuale processo di macinazione di grano è tutt'ora poco equipaggiato con tali strumenti. Perciò il controllo di processo è principalmente basato sull'esperienza degli operatori e nella maggior parte dei casi persino la progettazione del processo è empirica. Tuttavia, al prodotto finale è richiesto un alto livello di qualità.
Le industrie che coinvolgono il processo di macinazione di grano possono beneficiare enormemente dello sviluppo di strumenti scientifici e sistematici (opposti a quelli empirici basati sull'esperienza) per assistere la comprensione, progettazione e l'operazione del processo. La produzione intelligente attraverso la digitalizzazione è la tendenza seguita dai moderni processi manifatturieri. Questa Tesi, assecondando la direzione indicata da questa tendenza, propone un approccio basato su modelli per ottimizzare il processo di macinazione di grano. A tal proposito, vengono esposti diversi approcci basati sulla modellazione matematica per assistere la macinazione di grano. In particolare, vengono considerati due approcci: modelli basati su dati delle singole unità di processo e simulazione dell'intero processo.

I **modelli delle singole unità di processo** sono stati utilizzati per risolvere diversi problemi che caratterizzano l'operazione di macinazione: *i*) classificazione delle diverse varietà di grano, *ii*) comprensione del processo, *iii*) predizione della qualità di prodotto e *iv*) progettazione di processo e scale-up.
La **classificazione delle diverse varietà di grano** è un aspetto cruciale del processo di macinazione di grano. A tal proposito è stata sviluppata una metodologia basata sull'analisi di spettri al vicino infrarosso combinata a modelli a variabili latenti, con l'obiettivo di testare le capacità discriminanti degli spettri.
L'effettivo meccanismo di rottura del grano è ancora relativamente poco conosciuto, soprattutto a causa della complessità del processo stesso: acquisire una migliore **comprensione del processo** può essere di grande aiuto nell'esercire l'operazione di macinazione. L'analisi dei parametri del modello permette da un lato di acquisire una migliore comprensione del meccanismo di rottura del grano e dall'altro di studiare e

quantificare l'effetto dei parametri di processo più influenti e delle proprietà del grano sulle caratteristiche del materiale macinato, per diversi passaggi del processo.
Le capacità predittive dei modelli basati su dati possono essere sfruttate per **predire la qualità del prodotto** (ad esempio, la granulometria o la frazione massiva del prodotto setacciato), in modo da evitare costosi, sia in termini di risorse che di tempo, esperimenti. In questa Tesi, modelli basati su dati sono utilizzati per predire la qualità di prodotto sia per mulini a rulli che per i setacci piani.
La **progettazione di processo** è la relativa al problema di stimare la migliore combinazione di parametri di processo che devono essere usati per ottenere un prodotto con la qualità desiderata. Una soluzione a questo problema è offerta dai modelli basati su dati, nel caso in cui vengano utilizzate appropriate tecniche di inversione di modello. In questa Tesi, è stata sviluppata una metodologia di inversione di modelli a variabili latenti che si basa su un problema di ottimizzazione che include dati provenienti da differenti scale del processo. L'obiettivo ultimo è suggerire i parametri di processo ottimali per ottenere un prodotto con una caratteristica desiderata, in modo da accelerare lo **scale-up** dalla scala di laboratorio a quella industriale.

In questa Tesi, si è cominciato a sviluppare una **simulazione dell'intero processo**. A causa della sua complessità, l'attuale processo di macinazione di grano è ancora aperto ad interventi di modernizzazione e ottimizzazione, sia riguardo la struttura del processo che i consumi di energia. L'obiettivo è sviluppare un simulatore dell'intero processo di macinazione, con lo scopo di modellare il trasporto pneumatico, le perdite di carico e i consumi di energia del processo. Questo strumento può essere utilizzato per comprendere meglio le fonti principali di consumo di energia nel mulino e per ottimizzare la struttura del processo minimizzando, ad esempio, il consumo di energia.

Table of contents

FOREWORD 3

ABSTRACT 5

RIASSUNTO 7

CHAPTER 1 – MOTIVATION AND STATE OF THE ART 13

1.1 Wheat: production, market and trade 13

1.2 Some features of the wheat milling process 14

1.3 The modeling contribution in the wheat milling field 15

1.3.1 Knowledge-driven models 17

1.3.2 Data-driven models 18

1.3.3 Analytical techniques: near-infrared spectroscopy 18

1.4 Objectives of the research 19

1.5 Dissertation roadmap 20

CHAPTER 2 – WHEAT AND THE WHEAT MILLING PROCESS 23

2.1 The structure of the wheat kernel and the objective of milling 23

2.2 General features of the milling process 26

2.2.1 Main unit operations: roller mills and plansifters 27

2.2.2 Cascade of roller mill and plansifter: a process passage 27

2.2.3 Overall milling process 28

CHAPTER 3 – MULTIVARIATE STATISTICAL MODELING 31

3.1 Latent variable models 31

3.1.1 Principal component analysis (PCA) 31

3.1.1.1 PCA data pre-treatment 34

3.1.1.2 Selection of the number of PCs 35

3.1.1.3 PCA diagnostics 36
3.1.2 Projection to latent structures (PLS) 37
3.1.2.1 PLS data pre-treatment, selection of number of LV and model diagnostics..... 39
3.2 Other latent variable model technique: joint-Y PLS 39
3.3 Latent variable model inversion 41
3.3.1 Optimization framework to solve latent variable model inversion 42
CHAPTER 4 – AVAILABLE DATASETS OF THE WHEAT MILLING PROCESS .. 45
4.1 General structure of the datasets 45
4.2 Small-scale dataset 46
4.2.1 B1 passage 46
4.2.2 B2 passage 47
4.2.3 R1 passage 48
4.3 Industrial-scale dataset 48
CHAPTER 5 – SMALL-SCALE MODELING 51
5.1 Introduction 51
5.2 First-breakage passage modeling 51
5.2.1 Roller mill model 52
5.2.1.1 Classification of wheat using NIR spectra 52
5.2.1.2 Process understanding 53
5.2.1.2 Estimation of the particle size distribution 54
5.3 Second-breakage passage modeling 55
5.3.1 Roller mill model 55
5.3.1.1 Process understanding 56
5.3.1.2 Estimation of the particle size distribution 57
5.3.2 Plansifter models 58
5.3.2.1 PSD plansifter model 58

5.3.2.2 Mass plansifter model 59
5.4 First reduction passage modeling 60
5.4.1 Roller mill model 60
5.4.1.2 Process understanding 61
5.4.1.3 Estimation of particle size distribution 62
5.4.2 Plansifter model 62
5.4.2.1 PSD plansifter model 62
5.4.2.2 Mass plansifter model 63
5.5 Concatenating the roller mill and plansifter models 63
5.6 A Simulink simulation of roller mill and plansifter models 65
5.7 Conclusions 68
CHAPTER 6 – A MODEL-BASED APPROACH TO THE DESIGN AND SCALE-UP OF WHEAT MILLING OPERATIONS – PROOF OF CONCEPT 71
6.1 Introduction 71
6.2 The modeling framework 74
6.2.1 The rationale behind the proposed approaches 75
6.2.2 Using PCA scores as a proxy of NIR spectra 76
6.2.3 Approach no.1: direct PLS modeling 78
6.2.4 Approach no.2: PLS model inversion 78
6.2.5 Approach no. 3: joint-Y PLS model inversion 79
6.3 Milling gap estimation results 80
6.3.1 Models diagnostics 81
6.3.2 Results using an extended industrial dataset 82
6.3.3 Results using a limited industrial dataset 83
6.4 Conclusions 84
CHAPTER 7 – SIMULATION OF THE PNEUMATIC TRANSPORT OF INDUSTRIAL WHEAT MILLING PLANTS 87

7.1 Introduction 87

7.2 Pneumatic transport of biomass particles: wheat 88

7.3 Pneumatic transport in an industrial wheat mill plant 89

7.4 Some features of the simulator 90

7.5 Results and discussion 91

7.5.1 Mass balance 91

7.5.2 Power and energy consumption 91

7.5.3 Pressure drops 91

7.5.4 Velocities 92

7.5.5 Simulation performances 92

7.6 Conclusions and future perspectives 93

CONCLUSIONS AND FUTURE PERSPECTIVES 95

APPENDIX A 101

REFERENCES 105

Chapter 1

Motivation and state of the art

This Chapter provides the background and the motivations of this Dissertation. First, a general overview about the wheat production, market and trade is provided. Additionally, some features of the wheat milling process are highlighted. Then, a review of the knowledge-driven and data-driven methodologies utilized for milling process modeling is presented. Furthermore, the state-of-the-art analytical techniques used in the wheat milling are presented. Finally, the objectives of the Dissertation and a roadmap to its reading are provided.

1.1 Wheat: production, market and trade

Wheat is a major global commodity. The Food and Agriculture Organization of United Nations (FAO) reported that the world wheat production in 2015 reached 734.1 million tons, a record compared to the previous years (FAO, 2016). Wheat and derived products are used in many sectors, such as food, animal feed, biofuels, cosmetics, and bio-plastics. Food is the main sector of use, exploiting over 68 % of the 2015 annual wheat production (FAO, 2016). World per capita wheat consumption were estimated around 67 kg per annum (60 kg in developing countries and 97 kg in developed countries). In food applications, wheat is mostly ground into flour and used, among others, for manufacturing bread, pasta and biscuits. In the United States, nearly 21 million tons of flour were produced in 2015, according to the National Agricultural Statistic Service (NASS) of the U.S. Department of Agriculture (NASS, 2016). Wheat used for animal feed was estimated around 19 % of the total production in 2015 and for industrial use, half of which for the production of starch, 3 % only. The different wheat utilizations are illustrated in Figure 1.1.

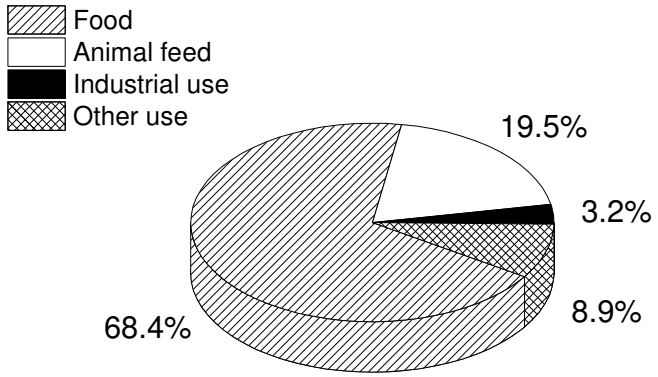


Figure 1.1. *Wheat utilization in 2015 (FAO, 2016).*

Wheat trade is one of the leading markets in the cereals field, with more than 154 million tons exchanged around the world in 2015 (FAO, 2016), for a total transfer of money of more than 29 billion of euros (considering an average wheat price of 193.5 €/ton, according to the United States Department of Agriculture; USDA, 2016). The wheat major exporters were the European Union, the Russian Federation, Canada and the United States; on the other side, the principal importers were: Egypt, Indonesia, Algeria and Brazil. Figure 1.2 provides details about wheat exports (Figure 1.2a) and imports (Figure 1.2b).

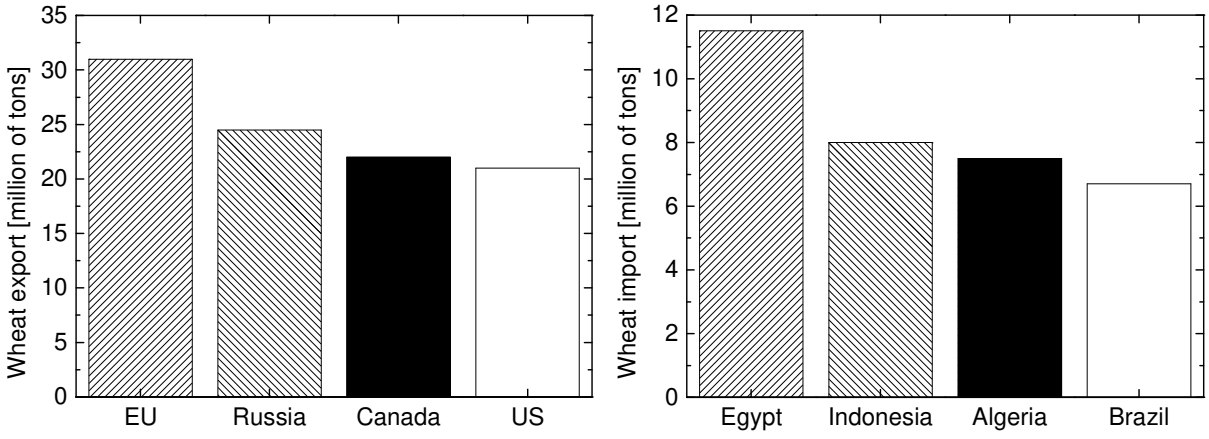


Figure 1.2. *World wheat trade: (a) major exporters and (b) major importers (FAO, 2016).*

1.2 Some features of the wheat milling process

In this section a short description of the wheat milling process is provided. More details will be given in Chapter 2. The objective of the wheat milling process is to break open the wheat kernel and recover as much pure endosperm as possible, of appropriate size, and at minimum cost (Campbell, 2007). In an industrial process, this is achieved by means of two main units

operation: a roller mill and a plansifter. A roller mill grinds the incoming material and a plansifter acts a rough separation between the wheat components (bran, semolina, germ and flour). The combination of a roller mill and a subsequent plansifter forms a so-called "passage" of the process. A single process passage does not guarantee the quality requirements in the produced flour; for this reason, the process uses a gradual reduction approach: repeated steps of milling and repeated steps of sieving are used to achieve purification of endosperm and its effective reduction into flour (up to seventeen different passages are used in a modern wheat milling process; Campbell, 2007). This approach ensures a high flour yield, but makes the process more complex and therefore more difficult to operate. Other issues affect a wheat milling process:

- complex layout of the process and complex interaction between the units operation;
- high intrinsic variability of the raw material (which depends on wheat type, season, climate, zone of cultivation, year of harvest, etc.);
- control mostly based on the operators' experience.

Despite the availability of analytical instruments (e.g., near-infrared spectrometers or image analysis sensors) to measure important characteristics of the wheat or flour properties, modern wheat milling processes are still poorly equipped from the measurement point of view. Therefore, the process is managed with an experience-driven approach, which relies on the operators' expertise to guarantee the required specification in flour quality. Wheat milling is a well-established process, but it works in an old fashioned way. The market demand for high quality product requires advanced tools to design and operate the process in a satisfactory way. Thus, what is needed now is to move from an experience-driven approach to a science-driven one.

1.3 The modeling contribution in the wheat milling field

A wheat milling process suffers of the above mentioned issues: complex process layout coupled to experience-driven operation.

In the last decade, other industries have been dealing with similar problems:

- the microelectronics and semiconductors industry. In the industrial practice, the process of semiconductors manufacturing were often operated empirically, with little knowledge of the underlying physics and chemistry. This was due to different reasons: complex series of semi-batch processes, low availability of important process measurements, poor process understanding and utilization of post-process measurements to assure the product quality (Edgar et al., 2000);
- the pharmaceutical industry. It was traditionally (and in many cases still is) based on experience, despite being perceived as innovative and as the spearhead of the industrial world, mostly for the relevance of the manufactured products. The pharmaceutical

industry lagged behind other industries with respect to the improvement of the manufacturing processes and of process understanding (Gernaey et al., 2012).

Initiatives were launched and big efforts were put to modernize the above mentioned industries moving from an experience-driven to a science-driven approach. In the semiconductors industry, many examples of such trend can be found in the literature. Considerable efforts were dedicated to develop systematic procedures to satisfy the product quality requirements (Guldi, 2004; Yaakobovitz et al., 2007). Statistical process control was used to monitor the variability of the process, to detect the abnormal conditions and to identify the cause for an anomaly: for example, Yue et al. (2000) investigated the suitability of optical emission spectroscopy for the fault detection and classification of plasma etchers. Regarding automatic process control, Facco et al. (2009) developed a multiresolution multivariate technique to identify the abnormalities on the surface of a photolithographed device. On the other side, in the pharmaceutical industries, the regulatory agencies such as the Food and Drug Administration (FDA), fostered the adoption of systematic and scientific approaches in the development of new processes and products and in the process management, by means of the Quality-by-Design and Process Analytical Technology initiatives (FDA, 2004). Supporting the pharmaceutical industry involved many aspects of the manufacturing process: for example, Tomba et al. (2013a) proposed a methodology to improve the process understanding of a continuous process for the manufacturing of paracetamol tablets; regarding process monitoring and control, Muteki et al. (2012) implemented a feedforward controller in a tablet manufacturing process, involving blending, dry granulation, milling and tableting. Burggraeve et al. (2011) developed a latent variable model procedure to monitor online a fluid bed granulation process. Many efforts were dedicated also to support the development of new products and the process scale-up: Tomba et al. (2013b) used latent variable model inversion to develop a new pharmaceutical high-shear wet granulation product; Liu et al. (2011) used latent variable model inversion to support the scale-up of a roller compaction process.

As in the previous examples, also in the wheat milling industry it is possible to move from an experience-driven approach to a science-driven one to design and operate the process. The latter approach will be based on mathematical modeling and simulation, which are key elements of process system engineering. Modeling and simulation can be powerful tools to support the wheat milling process, with the challenge to adapt well-established and mature methods (as in the example showed above) to the food industry, characterized by complex processes and natural, with an extremely high variability, products. Supporting the wheat milling operation by means of data-driven modeling can involve, among other aspects: improvement of process understanding, prediction of product quality, process design and scale-up.

In the following sections, a review of some relevant examples of knowledge-driven models, data-driven models and analytical techniques used in the wheat field, with particular attention to wheat milling and sieving, is provided.

1.3.1 Knowledge-driven models

Knowledge-driven (KD) models (also called mechanistic, first-principles or fundamental models) describe the underlying functional mechanisms of the system under investigation, relying on the fundamental knowledge of the system in terms of mass, energy and momentum balances, as well as constitutive equations. In other words, KD models are a mathematical representation of the available knowledge of a system (Bonvin et al., 2016). Usually, the time and efforts required to develop these type of models are very high (Troup and Georgakis, 2013). In some cases, such as in very complex processes for which the mechanistic knowledge is not easy to be described (such as biological processes involving natural products), KD models must rely on significant simplifying assumptions.

Wheat milling has been investigated with different modeling techniques. Several studies rely on the concept of breakage equation, namely a population balance with a breakage function that incorporates meaningful input and process parameters (Campbell and Webb, 2001; Campbell et al., 2001). Using this technique, Fang and Campbell (2003a) added the effect of roll disposition and Fang and Campbell (2003b) included the wheat kernel moisture content in the modeling framework. Campbell et al. (2007) studied the effect of kernel hardness and shape. Since a breakage function requires a large number of parameters, a normalized breakage function was developed (Mateos-Salvador et al., 2011; Campbell et al., 2012): the breakage function was simplified by normalising the output particle size distribution against the milling ratio raised to a power, thus reducing the number of parameters to four. Mateos-Salvador et al. (2013) extended the utilization of the normalized breakage function to the second breakage passage of the wheat milling process. Recently, Patwa et al. (2014) and Patwa et al. (2016) developed a discrete element method model for a wheat milling process. They used two different particle models, a sphere-shaped model with uniform bond strength and a kernel-shaped model with randomly distributed bond strength, but the results in predicting the average particle size of the milled material showed large deviations from the experimental results for both approaches.

Sieving is a simple operation, where in principle the achieved separation only depends on the material size. However, it is well know that non-idealities (such as particle aggregation or breakdown, or incomplete separation of material smaller than a given mesh size) can strongly affect the separation (Posner and Hibbs, 1997). Little has been published so far relating to the performance of plansifters. Posner and Hibbs (1997) concluded that plansifters are often a process bottleneck and that their efficiency is usually very low. Sultanbawa et al. (2001)

developed a graphical technique to evaluate the performances of a cascade of sifters. They assumed that the change of concentration of undersize particles remaining on a sieve as a function of sieving time can be described by a first-order rate equation with a screening rate constant that is inversely proportional to the particle size.

1.3.2 Data-driven models

Data-driven (DD) models (also called black-box models) do not require any prior knowledge of the physical mechanisms underlying a process, since the information useful to define mathematical relationships between its inputs and outputs is directly extracted by the analysis of the process data. Somehow, DD models are just a convenient representation of the available data.

Only very few examples can be found in the literature regarding data-driven models of wheat milling and sieving and no general modeling approach is available. Flores et al. (1991) developed an empirical model for a flour mill that relates the wheat commercial characteristics with flour ash and protein. The model allowed determining the impact of commercial parameters on the final product. Fang et al. (1998) developed a neural network model to predict particles characteristics as a function of the wheat properties, such as geometric mean diameter, specific surface area and break release. Yuan et al. (2003) used a logistic model to relate the commercial wheat properties with the output particle size distributions, but a high predictive accuracy was reached only for one unit operation. Li et al. (2002) described a two-dimensional computer simulation of the separation of crop seeds by screening. Their results showed that complete separation occurred only with a bed depth of around 30 mm or less. In deeper beds, the small particles could not segregate effectively and migrate to the screen apertures. As a result, a large number of undersize particles remained in the over tails product stream.

1.3.3 Analytical techniques: near-infrared spectroscopy

Near-infrared spectroscopy (NIRS) is nowadays a widespread analytical technique used in many fields, including food quality control (Huang et al., 2008; Miralbés, 2004), having the advantage of being fast, cheap, non-destructive and suitable for on-line implementation.

In association with chemometrics, NIRS has been used to predict many quality variables, such as protein (Manley et al., 2002; Delwiche et al., 1998; Kays et al., 2000; Wesley et al., 2001), moisture (Manley et al., 2002), dietary fibre content (Archibald and Kays, 2000), as well as wheat hardness (Manley et al., 2002). NIRS application has been extended to the estimation of functional and technological parameters, such as flour yield, percentage of damaged starch, water absorption, dough development, time extensibility and loaf volume (Delwiche et al., 1998). Furthermore, NIRS has been used to classify different wheat varieties (Delwiche and

Norris, 1993; Miralbés, 2008) and to discriminate between vitreous and non-vitreous durum wheat kernels (Dowell, 2000). NIRS has been applied also to characterize and classify wheat flours and doughs. Cocchi et al. (2005) discriminated different wheat flour varieties by means of NIR spectroscopy and Cocchi et al. (2006) exploited NIRS in order to quantify the degree of adulteration of durum wheat flour with common bread wheat flour. Arazuri et al. (2012) used NIRS to predict tenacity, extensibility and deformation energy of wheat doughs. Additionally, NIRS has been used to monitor the baking performance of wheat flours and doughs (Jirsa et al., 2008; Li Vigni et al., 2009; Li Vigni and Cocchi, 2013). In the food field, NIRS has also been used to classify vegetables oil (Sato, 1994), ginseng radix (Woo et al., 1999) and to classify soft, semi-hard and hard maize kernels (Williams et al., 2009).

1.4 Objectives of the research

The main objective of the research presented in this Dissertation is developing a science-driven platform, based on mathematical modeling and process simulation, to assist the wheat milling operation. Two approaches are used: data-driven models of single unit operations, and simulation of the entire process. The domain of application of the methodologies proposed in this Dissertation and the innovative contributions they provide are summarized in the following.

- **Data-driven models of single unit operations.** These models were used to tackle different issues that arise in the wheat milling operation.
    - **Wheat classification.** Wheat is a natural product, that shows a high intrinsic variability, deriving not only from the wheat type, but also from the geographical origin of the wheat, the climate, the harvesting zone and year of harvest. Describing this characteristics in a fast, cheap and non-destructive manner is crucial. A methodology based on the analysis of near-infrared (NIR) spectra in combination with latent variable modeling is developed. The purpose is checking the ability of NIR spectra to discriminate different wheat varieties.
    - **Process understanding.** Wheat milling is not a simple operation, since it involves a complex interaction between breakage mechanism and wheat properties. It is shown that the analysis of the model parameters allows for a better comprehension of the breakage mechanism occurring in wheat milling. The effect of the most influential process parameters and of the wheat properties onto the milled product characteristics is studied for different passages of the process.
    - **Product quality prediction.** The ability of predicting the output response (i.e., the particle size distribution or the mass fraction of the milled or sieved material) without carrying out expensive and time-consuming experiments can be of paramount importance in supporting the wheat milling operation. In this Dissertation, data-

driven models are used to predict the product quality both for roller mills and for plansifters.

- **Process design and scale-up.** The problem of process design is related to the estimation of the process parameters to be used to obtain a product with a desired characteristic. Data-driven models provide a useful tool to this purpose, through the implementation of appropriate model inversion techniques. In this Dissertation, a latent variable model inversion methodology, that relies on an optimization framework and on data coming from different plant scales, is developed. The framework aims at suggesting the optimal process parameters to be used to obtain a product with desired properties, with the ultimate scope of accelerating the scale-up from the laboratory to the commercial plant.

- **Simulation of an entire wheat milling process.** A wheat milling process is characterized by a complex layout, with complex interaction between the different units operation. Due to this complexity, the process is still opened to optimization, in terms of both layout and power consumption. The objective is developing a simulator of the whole milling process, that aims at modeling the pneumatic transport, pressure drops and power consumption. This tools can be used to get a better comprehension of the energy consumed throughout the mill and to optimize the process layout, minimizing the overall power consumption.

1.5 Dissertation roadmap

This Dissertation is organized as follows. Chapter 1 provides an overview of the modeling contribution in the wheat milling operation, along with the main objectives of the Dissertation. Chapter 2 provides a detailed discussion about the structure of the wheat grain, and describes the unit operations used within an industrial grain milling process. The overall process set-up is discussed as well. The description of the data-driven modeling techniques is provided in Chapter 3. Chapter 4 gives an overview of the available datasets of the wheat milling operation.

Figure 1.3 shows a sketch of the structure of the Dissertation, according to the two approaches used (models of single units operation and simulation of the overall process).

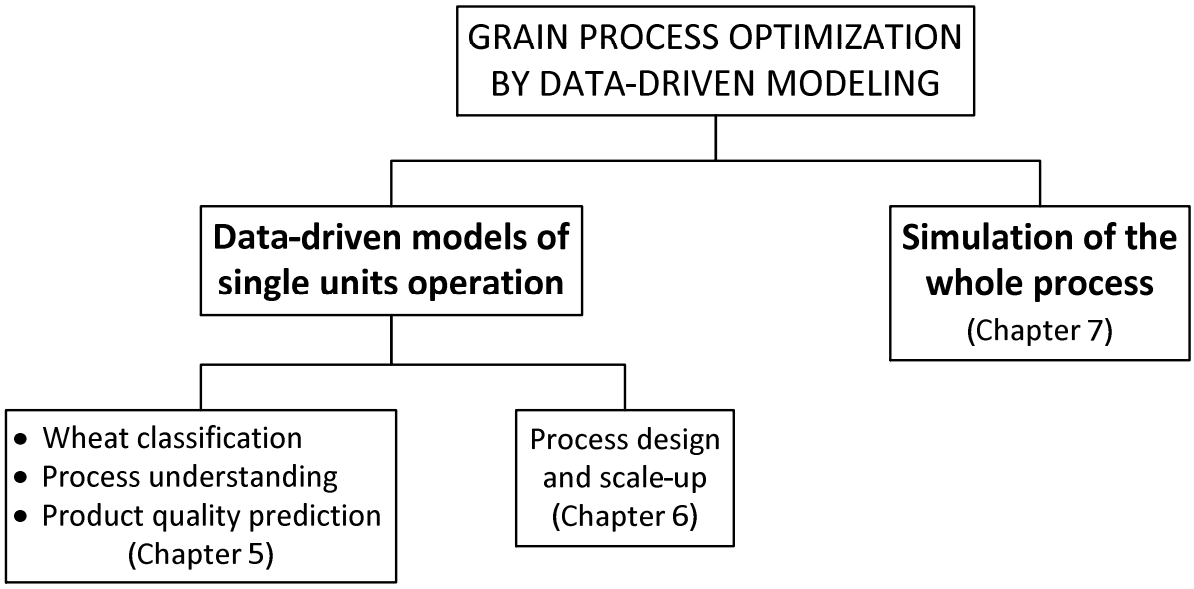


Figure 1.3. *Sketch of the structure of the Dissertation.*

Regarding the first approach, in Chapter 5 it is shown how the models developed can be used to support the milling operation in several ways: *i*) to classify different wheat varieties; to this purpose, a tool based on NIR spectra and latent variable modeling is developed, *ii*) to improve process understanding of the wheat breakage mechanism, analyzing the data-driven model parameters, and *iii*) to predict the product quality (namely, particle size distribution or mass fraction of the milled and sieved material).

Chapter 6 presents a data-driven model inversion framework to estimate the process parameters to be used to obtain a product with desired properties.

Chapter 7 addresses the problem of developing a simulation of the pneumatic transport of a wheat milling process, with particular attention to the power consumption and pressure drops throughout the mill.

Chapter 2

Wheat and the wheat milling process

The aim of the grain milling process is the separation and reduction of grain components into flour, bran and germ. In general as much white flour with constant baking quality parameters as possible should be produced from the milled grain. This Chapter provides details about the structure of the wheat grain, and describes the unit operations used within a grain milling process. Finally, the overall process set-up is shown and discussed.

2.1 The structure of the wheat kernel and the objective of milling

The first fundamental step for understanding the flour milling process is analyzing the structure of the wheat kernel.

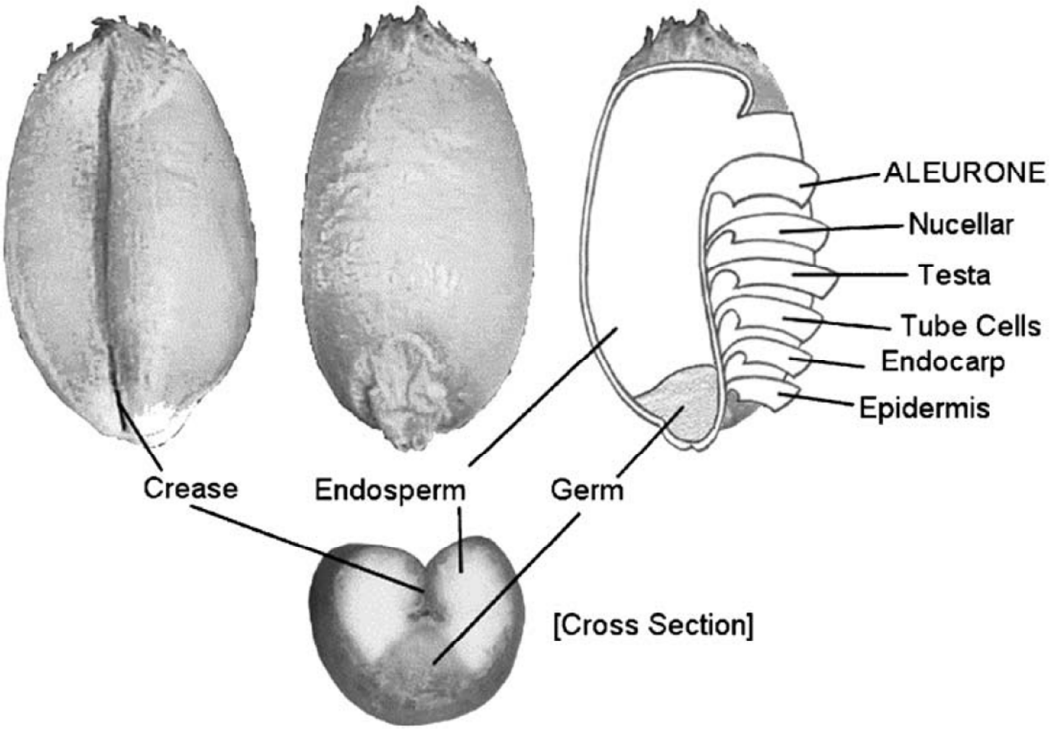


Figure 2.1. *Illustration of the wheat kernel structure, showing the germ, the endosperm and the layers forming the bran and the crease. Adapted from Campbell (2007).*

Figure 2.1 illustrates the structure of the wheat kernel. The wheat kernel consists of three major parts (Evers and Millar, 2002): the endosperm (81-84 % in mass of the whole kernel), the bran (14-16 %) and the germ (2.5-4 %).

The germ is the immature plant, consisting of an embryonic nucleus, which comprises a rudimentary sprout and root and the scutellum (a storage, digestive and absorbing organ).

The endosperm is the storage reserve for the nutritional needs of the germinating plant, until it emerges from the soil and it is able to begin photosynthesizing its food requirements. Therefore, the endosperm is rich in starch and protein (for energy and constructions needs, respectively).

Starch is produced by photosynthesis in green leaves. Shortly after, it is degraded into sucrose and, in the case of cereals, transported to seeds where it is deposited in the endosperm tissue as water insoluble granules that serves as the plant's main storage of carbohydrates (Delcour and Hoseney, 2010).

Starch granules are formed in amyloplasts. Wheat has two types of starch granules (Figure 2.2), corresponding to two different sizes of the granules: large lens-shaped granules (25-40 µm in the long dimension), and small spherical granules (5-10 µm). All of them are embedded in a protein matrix (Delcour and Hoseney, 2010). Starch granules cannot be damaged during milling, to avoid compromising flour properties and its baking quality.

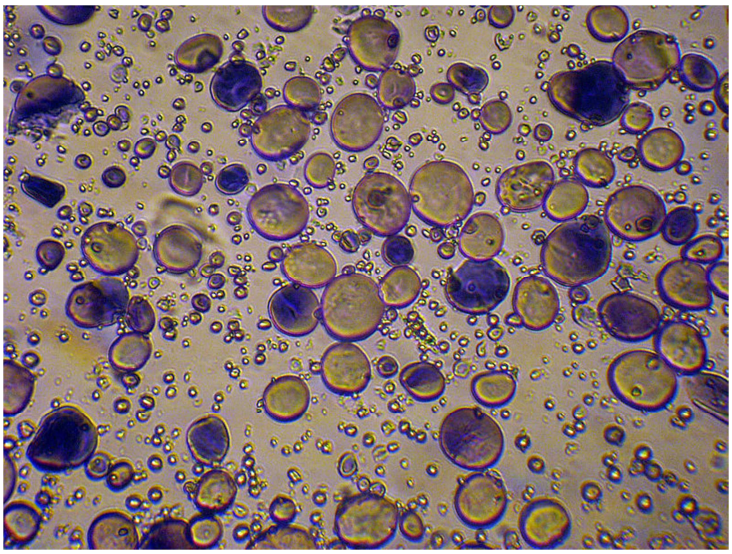


Figure 2.2. *Wheat starch granules, showing large lenticular and small spherical granules, photographed through a light microscope.*

The endosperm and the germ are also valuable for other kind of living organisms (humans, animals, insects and microorganisms). For this reason, the germ and the endosperm are covered with some protective layers, known collectively as bran, which furnish a physical barrier to intrusion. Even if physiological benefits of bran are widely recognized, it is nutritionally poor in comparison with endosperm. In addition, flour that comprises bran has a lower shelf-life and it produces a less appealing bread (Campbell, 2007).

The purpose of wheat milling is to break open the kernel and separate its structural components in such a way as to recover as much endosperm as possible (free from contamination with bran and germ), at minimum cost (Campbell, 2007). Pure white flour (high quality flour) can be obtained accepting a low extraction yield, but as endosperm extraction reaches its theoretical maximum (81-84 %), the flour becomes increasingly contaminated with bran (low quality flour). Therefore, the economic profitability of flour milling process is dictated by the compromise between yield and quality.

The main challenge millers have to deal with is to produce flour of consistent uniform quality, in the face of a constantly changing feedstock.

Recovering white pure flour is a challenge also due to the presence of the crease within the wheat kernel (Figure 2.1). The crease is a deep fold that covers the length of the kernel and extends the bran layers deep inside the kernel. The presence of the crease in the wheat kernel definitively requires to break open the kernel to separate bran from endosperm.

Another feature to be noticed in Figure 2.1 is the aleurone layer. It is a single layer of cells which surrounds almost completely the endosperm and the germ, separating them from the bran. The aleurone layer is a part of the endosperm which provides enzymes useful for seed germination, but it strongly adheres to the outer bran, so that it is removed as a part of the bran layers (Evers and Millar, 2002). This layer is rich in protein, vitamins and minerals. Hence, its inclusion in flour can increase both flour yield and nutritional qualities of the product. Accordingly, millers try to recover it as much as possible during the process, avoiding contamination of the flour due to bran.

One of the fundamental characteristics of wheat is the endosperm texture, namely the hardness of the wheat kernel. It affects significantly the milling process, influencing both milling yield and particle size of the flour. Hardness is mostly controlled by genetic factors but it can be affected by growing environment and other factors such as moisture, lipid content and agronomic practice (Turnbull and Rahman, 2002). Wheat are broadly divided into three major categories:

- soft wheat;
- semi-hard wheat;
- durum or hard wheat;

The former two categories arise from common wheat (*Triticum aestivum*) and durum wheat from *Triticum durum*. Given different hardness, wheat is used to manufacture different end products: soft wheat is usually used for cakes and biscuits, semi-hard wheat for bread and durum to produce pasta. This differentiation is caused by the different capability of various wheat both to produce CO_2 gas (leavening process) and to retain it inside the dough. The ability of retaining gas bubbles arises from the gluten proteins of wheat that, on hydration, form a viscoelastic strain-hardening network able to hold inflating gas bubbles (Gan et al., 1995).

It is important to say that the classification showed above is quite rough: wheat that nominally lay in the same category could considerably vary in properties due to different breeds, that introduce a large source of variation. Other sources of variation are the geographical origin of the wheat, the climate, the harvesting zone and year of harvest. Finally, being wheat a natural product, it decades during time and it is subject to pests (i.e. insects and molds), which affect also its storage properties. This high variability has a high impact on the overall milling process.

2.2 General features of the milling process

The main steps of the milling process (Figure 2.3) are here described:

- *step* 1, cleaning: the intake of the material is followed by a cleaning section, that usually comprises magnets (for metal undesired components removal), de-stoning machines and scourers (an intensive treatment to reduce bacteria count, scratching the outer part of the kernel);
- *step* 2, conditioning: conditioning is a fundamental operation of the milling process to adjust the wheat moisture content at a desired level. It has several aims: mellowing the endosperm (flour extraction can be increased and flour ash content reduced) and toughening the bran (at a higher moisture content the bran tends to break into larger particles, facilitating its separation from the smaller endosperm by sifting). Constant moisture level means both constant milling conditions and baking performance, increasing the profitability for the process. Conditioning is divided into two step in which water is added to the wheat grains and then the material is stored for a certain period, depending on the wheat characteristics. This step is followed by a second cleaning section, to contrast the bacteria growth due to the adding of water;
- *step* 3, milling;
- *step* 4, final product handling: a series of unit operations to control and assure the quality of the final product. This section usually includes control sifters, magnets, impact machines (to destroy insects' eggs) and sterilizers.

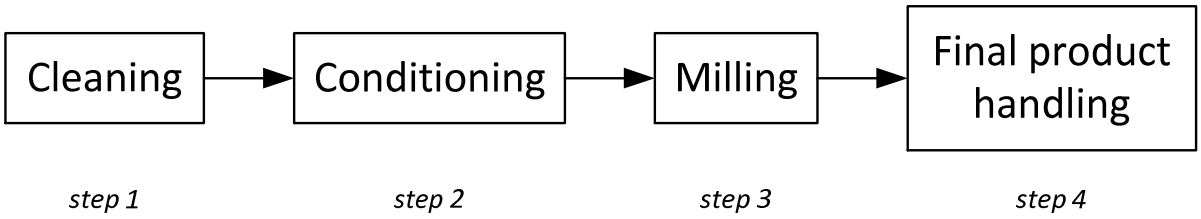


Figure 2.3. *General overview of the whole milling process.*

2.2.1 Main unit operations: roller mills and plansifters

The objective of the process, the recovery of pure endosperm of appropriate size, is achieved by means of two main unit operations: the roller mill and the plansifter. The roller mill grinds the incoming material whereas the plansifter separates the milled material into different fractions (of bran, semolina, and flour).

Roller mills give a better control over the process, compared to the behavior of the millstones, and allow processing a great variety of wheat of widely varying properties. The versatility of the roller mill can be explained in several ways (Campbell, 2007): the roll gap can be adjusted, rolls can operate at different relative speed to supply a better cutting action, different roll surfaces and fluting profiles can be used and successive unit operations can be arranged in a variety of configurations. This high versatility is also the weak point of the milling process, as small changes in roller mill operation can strongly modify the downstream processes and therefore final flour quality.

Plansifters are cascaded to each roller mill, so as to achieve a preliminary separation between bran and endosperm. A plansifter is made by a stack of sieves of decreasing sieve dimensions (i.e., mesh sizes) that usually vibrate to facilitate particle separation. The plansifter working principle is simple: the material to be sifted is fed to the upper section of the sieve stack, where it is graded depending on the particle size. The number of sieves and the mesh sizes are passage-dependent. The goal of sifting differs with the particular passage, but the flour must be removed from the process in each sifting step; if this is not well done, flour will undergo unnecessary further grinding and this might change its baking properties (Ulmer, 2011).

2.2.2 Cascade of roller mill and plansifter: a process passage

A process passage (Figure 2.4) is made by the sequence of a roller mill (which grinds the incoming material) and a plansifter (which separates the milled material into different fractions of bran, semolina, and flour). Passages represent key subsystems of the process: the overall process performance is determined by the performance of each passage, as well as by the interconnection between passages.

A process passage (one grinding and one separation step) is not enough to achieve an effective separation between endosperm, bran and germ and an effective reduction of endosperm into flour; therefore, the process uses the so called “gradual reduction” approach: repeated steps of milling (by roller mills) and repeated steps of separation (by plansifters) are used to fulfill the flour quality requirements. In a typical industrial wheat milling process (a simplified example is shown in Figure 2.5), up to seventeen different passages are used (Campbell, 2007).

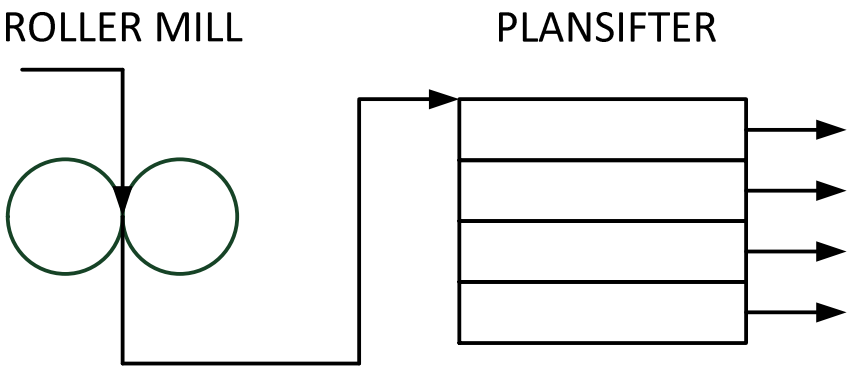


Figure 2.4. *Schematic representation of a process passage, formed by the cascade of a roller mill and a plansifter.*

2.2.3 Overall milling process

A simplified scheme of a wheat milling process is shown in Figure 2.5. Overall, the process can be divided in three sections: breakage, purifying and reduction.

The objective of the breakage section is to break-open the wheat kernel and to scrap the endosperm material from the bran particles, using fluted roller mills (Figure 2.6). In Figure 2.5, four different roller mills are indicated (B1, B2, B3 and B4), with finer material being processed as one moves from the upper mills to the lower ones. The fluting action is fundamental: it breaks open the wheat kernel such that bran tends to stay relatively intact and breaks into large particles, while the endosperm shatters into small particles, facilitating the separation of endosperm from bran by sifting. An important role is also played by the moisture content of wheat, a higher moisture content toughens the bran, that tends to break into larger particles, making the sifting separation easier.

The purpose of the purifying section is to improve the yield in the production of low ash flour (through processing units called purifiers). Semolina and middle fractions are obtained in each breakage passage. These fractions may contain pure endosperm particles, bran particles, heads (namely endosperm particles with adhering bran) and germ. The function of the purifier is to grade these products on the basis of their specific gravity into three different fractions: pure endosperm particles, heads and bran particles. The purifier grades the stock on the basis of the terminal velocity of the particles (that is a function of the specific gravity, the size and the shape of the product). The product is purified in several stages, corresponding to different screens (namely sieves); these screens form a carrier permeable to air for the product. In conjunction with screen motion, the air flowing in from below creates a fluidized bed which stratifies the different products (i.e., different particle sizes). Air velocity is varied so that bran particles are sucked along with air, whereas the heavier particles (heads and pure endosperm particles) tail over the screens (Ulmer, 2011).

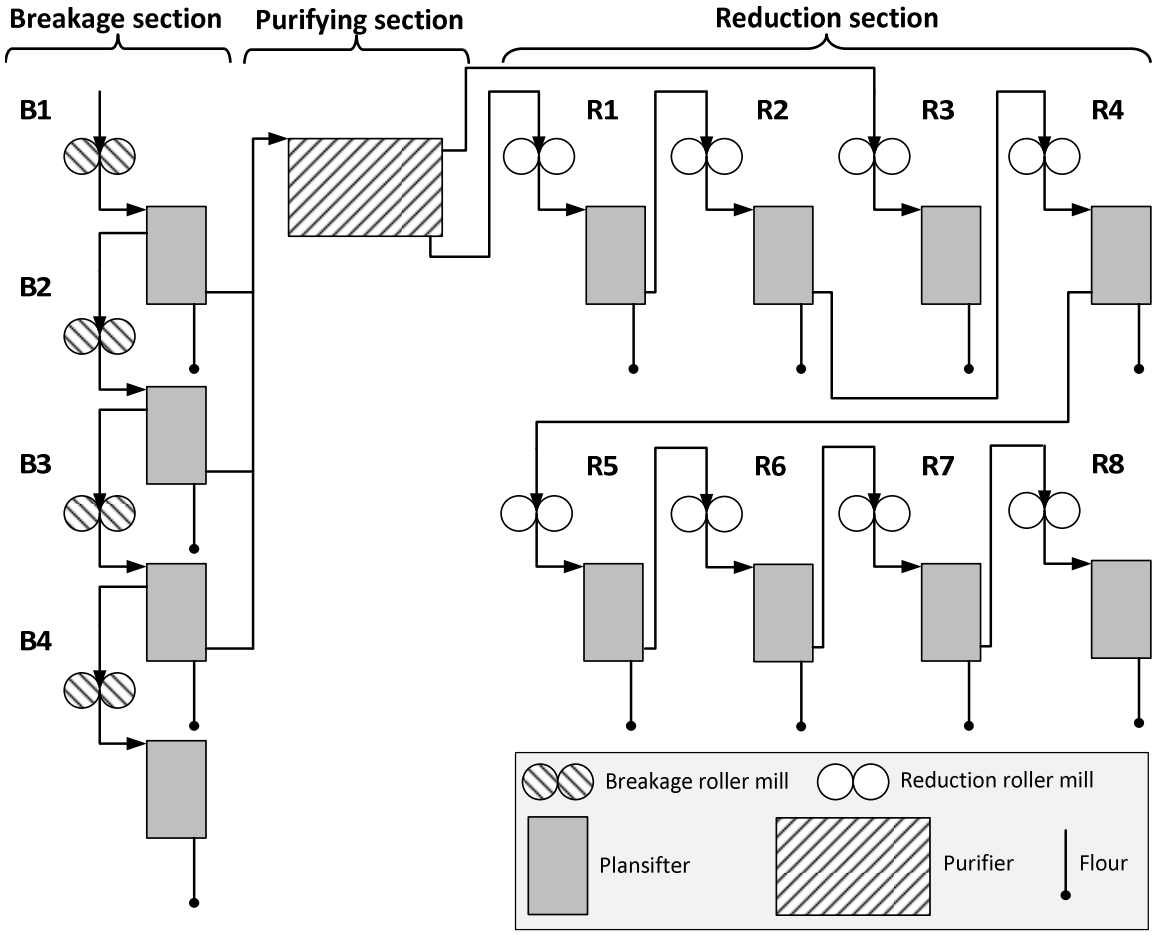


Figure 2.5. *Simplified scheme of a wheat milling process. The breakage, purifying and reduction sections are identified. Process passages are indicated by uppercase letters (where B stands for breakage and R stands for reduction) followed by numbers.*

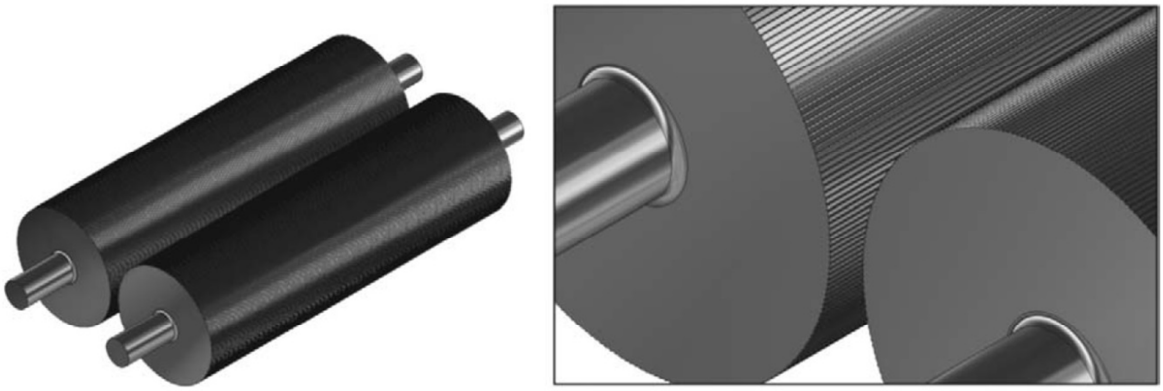


Figure 2.6. *Fluted rolls pair for the milling break system. Adapted from Campbell (2007).*

The reduction section aims at reducing endosperm into flour, using rolls with a smoothed surface (eight reduction roller mills are indicated in Figure 2.5). It is important to say that completely blank rolls will not grind semolina, but only press the material together producing

flakes and resulting in poor flour yield (Ulmer, 2011). Therefore a certain surface roughness of the roll is required, to increase the grinding action. Reduction rolls produce their action by their roll speed differential, namely the difference in velocity of the two rolls.

The quality of flour recovered from each passage (namely, from each roll pair) is different (Greffeuille et al., 2005). At least two different types of flour are usually produced (also depending on the country in which flour is manufactured): one white high-quality flour, and one flour of lower quality containing higher levels of minerals (i.e., higher bran contamination).

Chapter 3

Multivariate statistical modeling

This Chapter provides a general overview of the statistical techniques that have been used in this Dissertation. First of all, a background on latent variable models (LVMs) is presented. Principal component analysis (PCA) and projection to latent structures (PLS) are described from an algorithmic and practical point of view. Furthermore, more advanced techniques and the concept of LVM inversion are introduced.

3.1 Latent variable models

A LVM is a statistical model that relates a set of N observable (namely, measurable) variables to a set of A latent variables (LVs). LVs are found as linear combination of the observable variables, in order to summarize the information contained in the observable variables (Varmuza and Filzmoser, 2009). Therefore, a LVM is useful if A is smaller than N.
The general objectives of the different latent variable techniques are the following:

- dimension of the data space reduction: the observable variables space is transformed into a lower-dimensional LV space.
- data interpretation: the coordinates of the I samples on this LV space provide a compressed representation of the similarity between the observations, while the directions of the LV space provide a representation of the structure of covariance/correlation between variables.

In the following, theoretical and practical aspects of PCA and PLS are presented.

3.1.1 Principal component analysis (PCA)

Principal component analysis (PCA; Jackson, 1991) is a multivariate statistical technique used to summarize the information embedded in a dataset $\mathbf{X}$ of (correlate) variables, projecting the data onto a new space of latent variables that optimally describe the variability of data and the correlation among them. Each latent variable identifies a direction of maximum variability in the data and it is called principal component (PC).
The dataset $\mathbf{X}$ [$I \times N$] is represented as the sum of the $R_{\mathbf{X}}$ (the statistical rank of the $\mathbf{X}$ matrix) scores-loadings vector outer products:

$$\mathbf{X} = \sum_{a=1}^{R_{\mathbf{X}}} \mathbf{t}_a \mathbf{p}_a^{\mathrm{T}} \quad , \tag{3.1}$$

where $\mathbf{p}_a$ is the loading vector of the PC a and $\mathbf{t}_a$ is the score vector of the PC a.
The computation of the model scores and loadings can be performed by the optimization problem in Eq. (3.2). For the first PC, the optimization is represented by:

$$\begin{aligned} &\max_{\mathbf{p}}(\mathbf{p}_1^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{p}_1) \\ &\text{subject to } \mathbf{p}_1^{\mathrm{T}}\mathbf{p}_1 = 1 \quad , \end{aligned} \tag{3.2}$$

where $\mathbf{p}_1$ [N×1] is the loading vector, that represents the director cosines of the PC (namely, the direction of maximum variability in the data) with respect to the original variables. The original data are projected onto the PC direction, obtaining a vector $\mathbf{t}_1$ [I×1] of the coordinates into the PC space, called scores:

$$\mathbf{t}_1 = \mathbf{X}\mathbf{p}_1 \quad . \tag{3.3}$$

Consequently, the problem in (3.2) can be reformulated as in (3.4), maximizing the score vector length (Burnham et al., 1996):

$$\begin{aligned} &\max_{\mathbf{p}_1}(\mathbf{t}_1^{\mathrm{T}}\mathbf{t}_1) \\ &\text{subject to } \mathbf{t}_1 = \mathbf{X}\mathbf{p}_1 \\ &\quad \mathbf{p}_1^{\mathrm{T}}\mathbf{p}_1 = 1 \quad . \end{aligned} \tag{3.4}$$

It results that, geometrically, loadings are orthonormal and therefore PCs present the great advantage of being independent variables and no more correlated as the original ones. The solution of (3.4) is found by solving the eigenvalue problem (de la Fuente et al., 2010):

$$\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{p}_1 = \lambda_1\mathbf{p}_1 \quad , \tag{3.5}$$

where $\mathbf{p}_1$ is the eigenvector of the covariance matrix of $\mathbf{X}$ ($\mathbf{C} = \mathbf{X}^{\mathrm{T}}\mathbf{X}$), corresponding to the eigenvalue λ_1. λ_1 is a measure of the variance explained by the product $\mathbf{t}_1\mathbf{p}_1^{\mathrm{T}}$, namely it describes the amount of information embedded in the model by the first PC.
For the further PCs, the problem in (3. 5) can be solved iteratively on the deflated $\mathbf{X}_a$ matrix. The deflation process, for a=1,…,(A-1), being A the number of PCs, is defined as:

$$\mathbf{X}_{a+1} = \left(\mathbf{I}_I - \frac{\mathbf{t}_a\mathbf{t}_a^{\mathrm{T}}}{\mathbf{t}_a^{\mathrm{T}}\mathbf{t}_a}\right)\mathbf{X}_a \tag{3.6}$$

where $\mathbf{I}_I$ is the [I×I] identity matrix. Namely, at the a^{th} step the reconstructions of the dataset from the a^{th} estimated PC are subtracted to the dataset itself.
Given the equivalence with the eigenvalue problem in Eq. (3.5), PCs are ordered according to the variance of the original dataset they capture. When the variables of $\mathbf{X}$ are correlated (i.e., the matrix is not full rank), $\mathbf{X}$ can be represented with a number A of PCs, such that $A<N$. If two or

more original variables in **X** are correlated, they identify a common direction of variability, and this direction can be described by a unique PC when a PCA analysis is performed. A single PC will therefore capture the variability of all the variables which are correlated along the direction identified by the PC. One of the most important contributions of PCA is therefore that it allows to describe the original dataset with a lower number of variables, by projecting the data in **X** from the *N*-dimensional hyperspace of the original variables to the low-dimensional (i.e., *A*-dimensional) latent space of the PCs.

Therefore, the decomposition in (3.1) can be divided into two terms: the sum of the products of the scores and loadings on the first *A* PCs of the models and the sum of the products of the scores and loadings vectors on the last ($R_{\mathbf{X}}$-*A*) PCs:

$$\mathbf{X} = \sum_{a=1}^{A} \mathbf{t}_a \mathbf{p}_a^{\mathrm{T}} + \sum_{a=A+1}^{R_{\mathbf{X}}} \mathbf{t}_a \mathbf{p}_a^{\mathrm{T}} \quad . \tag{3.7}$$

On this basis, only the first *A* PCs are used to build the model and to effectively describe the variability in the dataset **X**. The first *A* scores vector and the first *A* loadings vectors can be collected in a score matrix **T** [*I*×*A*] and in a loading matrix **P** [*N*×*A*], respectively:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \quad , \tag{3.8}$$

while **E** [*I*×*N*] is the matrix generated by the ($R_{\mathbf{X}}$-*A*) PCs of the model. The matrix **E** contains the residuals when the **X** matrix is built using only the first *A* PCs ($\hat{\mathbf{X}}$):

$$\hat{\mathbf{X}} = \mathbf{T}\mathbf{P}^{\mathrm{T}} \quad , \tag{3.9}$$

$$\mathbf{E} = \mathbf{X} - \hat{\mathbf{X}} \quad . \tag{3.10}$$

Figure 3.1 shows the geometrical interpretation of PCA for a simplified example. A dataset **X** [6×2] collects 2 highly correlated variables (x_1 and x_2) measured for 6 times. As can be seen, data follow a defined (linear and uni-dimensional) trend in the bi-dimensional space of the original variables (x_1 and x_2). If a PCA is performed, this trend (namely, the direction of maximum variability) is identified by PC1. The loadings of the model (p_1 and p_2) represent the director cosines of PC1 with respect to x_1 and x_2, namely the cosines of the angles between the latent directions and the axes of the original variables space. The scores (i.e. t_1) represent the coordinates of the data samples in the new reference system represented by PC1. The lack of representativeness of the data by the PCA model is described by the residuals (i.e. e_1), represented by the perpendicular distance of the points from PC1. In Figure 3.1 the second PC (PC2) is also reported as a dashed line. It can be observed that PC2 is orthogonal to PC1, but it accounts for a very limited source of variability in the original data, meaning that a PC only is adequate to describe the entire content of information stored in **X**.

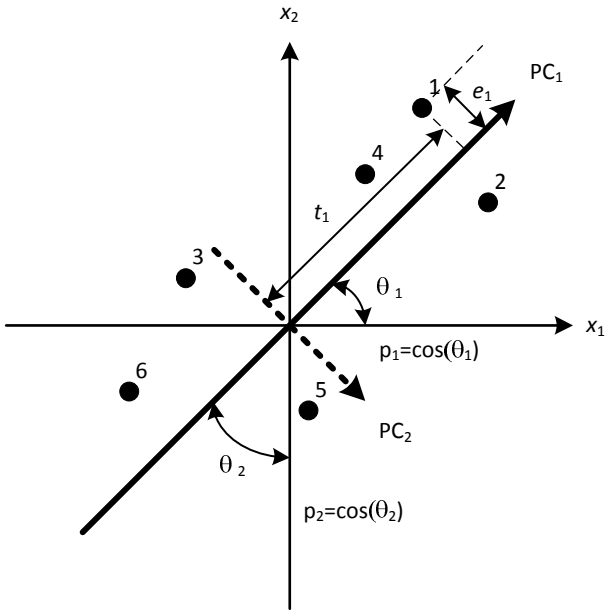


Figure 3.1. *Geometrical interpretation of PCA scores and loadings for a dataset* **X** *with 6 samples and 2 variables (x_1 and x_2).*

PCA scores and loadings are usually plotted and interpreted to gain information on how the observations relate to each other (through scores) and on the correlation among the original variables (through loadings). The parameters of a PCA model are usually calculated from the singular value decomposition (SVD; Meyer, 2000) of the **C** matrix or using the nonlinear iterative partial least squares algorithm (NIPALS; Wold, 1966; Geladi and Kowalski, 1986).

3.1.1.1 PCA data pre-treatment

It is usually convenient to pre-treat the data prior to the analysis to be performed, in order to better fulfill the assumptions of the method and improve the model performances. Pretreatments depend both on the characteristics of the available data and on the objectives of the analysis. They may include filtering, de-noising, transformations, advanced scaling and data compression (Eriksson et al., 2006).

It is important that variables are weighted in a similar way, in particular when dealing with datasets including many variables of different types and physical meaning. This can be achieved by auto-scaling, namely by mean-centering variables and scaling them to unit variance. Mean-centering consists in subtracting the mean value of the column itself to each column of the **X** matrix. This is fundamental to avoid the identification as significant directions of variability in the data due to differences between the variable mean values (Wise et al., 2006).

The scaling to unit variance is performed by dividing each column of the **X** matrix by its standard deviation, so that the total variance of the column is equal to one. This is essential to make the analysis independent of the measurement units. Furthermore, variables can undergo

further scaling or weighting operations to determine a different impact of each variable on the model (Kourti, 2003).

Other pre-treatment methods can be applied in addition to auto-scaling, in particular when dealing with spectral data: for example, Pareto scaling or standard normal variate (SNV). An extensive review of pre-treatment methods for spectral data can be found in van den Berg et al. (2006) and Rinnan et al. (2009).

3.1.1.2 Selection of the number of PCs

The determination of the dimension of the latent space in a PCA model (i.e., selecting the number of PCs to be used to build the model) is an important issue. Several methods have been proposed and the work of Valle et al. (1999) provides a comprehensive review and comparison of the most common ones.

To select the appropriate number of PCs, different issues must be considered, such as the number of samples present in the data, the total variance explained by the selected number of PCs, the values of the eigenvalues associated to the PCs (Johnson and Wichern, 2007).

In this Dissertation two methods have been applied:

- the scree test (Jackson, 1991);
- the eigenvalue-greater-than-one rule (Mardia et al., 1979).

The scree test is a graphical procedure, that is based on the analysis of an index indicating the variability of the original data described by each PC of the PCA model. This index can be either the explained variance R^2 or the eigenvalues. The chosen index is displayed in a simple line plot and the place where the smooth decrease of the index appears to level off to the right of the plot must be found (the so called "elbow"). No more than the number of factors to the left of this point should be retained, because additional PCs are supposed to describe the variability due to noise (i.e., the non-structural part of the data). The implementation of the method is relatively easy, but if the curve decreases smoothly it can be difficult to identify a "elbow" on it.

The eigenvalue-greater-than-one rule is a simple rule for which all the PCs whose corresponding eigenvalues are greater than one must be retained in the model. The idea behind this method is that, with auto-scaled data, the eigenvalue corresponding to a PC represents the number of original variables whose variability is captured by the PC. In this way, a PC describing less than one original variable is not included in the model. Although this method is very easy to implement, in some cases PCs are discarded even if their eigenvalue is very close to one and their contribution to explain the systematic variability is significant. In these cases, it may be reasonable to lower the threshold in order to include PCs whose eigenvalue may be slightly lower than one.

3.1.1.3 PCA diagnostics

There are several diagnostics that are used to evaluate the performances of a PCA model and they can be distinguished in model and data diagnostics (Eriksson et al., 2006).

Among model diagnostics, it is important to consider the amount of variability of the original data explained by the model, which is quantified by R^2:

$$R^2 = 1 - \frac{\sum_{i=1}^{I}\sum_{n=1}^{N}\left(x_{i,n} - \hat{x}_{i,n}\right)^2}{\sum_{i=1}^{I}\sum_{n=1}^{N} x_{i,n}^2} = 1 - \frac{ESS}{TSS} \quad , \tag{3.11}$$

where *ESS* and *TSS* are respectively the error sum of squares and the total sum of squares. $\hat{x}_{i,n}$ is the value of the i^{th} row and n^{th} column of $\mathbf{X}$ reconstructed with the PCA model. R^2 is calculated for each PC and a cumulative form (R^2_{CUM}) can be reported.

Data diagnostics are used to understand which samples are potential outliers or have a strong influence (i.e., leverage) in the model. To this purpose, Hotelling's T^2 and squared prediction error (SPE) are used.

The Hotelling's T^2 (Hotelling, 1933) measures the distance of the projections of the observations (namely, of the scores) of the dataset $\mathbf{X}$ from the origin of the PC space. The Mahalanobis distance is used to calculate it, to take into account the fact that each PC explains a different percentage of the data variance (Mardia et al., 1979):

$$T_i^2 = \mathbf{t}_i^{\mathrm{T}} \boldsymbol{\Lambda}^{-1} \mathbf{t}_i = \sum_{a=1}^{A} \frac{t_{a,i}^2}{\lambda_a} \quad , \tag{3.12}$$

where $\mathbf{t}_i$ [A×1] is the score vector of the i^{th} observation in the dataset $\mathbf{X}$, while $\boldsymbol{\Lambda}$ [A×A] is the eigenvalue diagonal matrix. The T^2 provides a check for observations adhering to multivariate normality (Eriksson et al., 2006) and it is used to assess the deviation of an observation from the average of the dataset. A sample with a large T^2, if it is well-represented by the model, provides useful information to expand the data space and ensures robustness to the model.

The SPE statistics describes how well an observation is represented by the model:

$$\mathrm{SPE}_i = \left(x_i - \hat{x}_i\right)^T \left(x_i - \hat{x}_i\right) = \mathbf{e}_i^{\mathrm{T}} \mathbf{e}_i \quad , \tag{3.13}$$

where $\mathbf{e}_i$ [N×1] is the residual vector of the reconstruction of the observation $\mathbf{x}_i$. SPE_i measures the orthogonal distance of the i^{th} observation from the latent space of the model, namely it accounts for the model mismatch to represent $\mathbf{x}_i$. Samples with a high SPE are not well-represented by the model and are characterized by a different correlative structure compared to the other samples.

3.1.2 Projection to latent structures (PLS)

Projection to latent structures (PLS; Wold et al., 1983; Höskuldsson, 1988) is a multivariate statistical technique that relates a dataset of regressors **X** (i.e. process parameters or raw material properties) with a dataset of response variables **Y** (i.e. product quality). In particular, PLS finds a (linear) transformation of the **X** data that maximizes the covariance of its latent variables (LVs) with the **Y** dataset variables. In other words, PLS finds the directions of maximum variability in **X** that are mostly correlated to the variability of **Y**. For the first LV it is represented by the following optimization problem (Burnham et al., 1996):

$$\begin{aligned} &\max_{\mathbf{w}_1}(\mathbf{w}_1^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{X}) \\ &\text{subject to } \mathbf{w}_1^{\mathrm{T}}\mathbf{w}_1 = 1 \quad , \end{aligned} \tag{3.14}$$

where $\mathbf{w}_1$ is the [N×1] weights vector for the first LV, representing the coefficients of the linear combination of the **X** variables determining the PLS scores $\mathbf{t}_1$:

$$\mathbf{t}_1 = \mathbf{X}\mathbf{w}_1 \quad . \tag{3.15}$$

The solution of the problem in (3.14) corresponds to the eigenvalue decomposition of the matrix $\mathbf{X}^{\mathrm{T}}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{X}$:

$$\mathbf{X}^{\mathrm{T}}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{X}\mathbf{w}_1 = \lambda_1\mathbf{w}_1 \quad . \tag{3.16}$$

To obtain the weights vector for the further LVs, the problem in (3.16) can be solved iteratively on the deflated $\mathbf{X}_a$ and $\mathbf{Y}_a$ matrices. The deflation process, for a=1,…,(A-1), being A the number of LVs, is defined as:

$$\mathbf{X}_{a+1} = \left(\mathbf{I}_I - \frac{\mathbf{t}_a\mathbf{t}_a^{\mathrm{T}}}{\mathbf{t}_a^{\mathrm{T}}\mathbf{t}_a}\right)\mathbf{X}_a \tag{3.17}$$

$$\mathbf{Y}_{a+1} = \left(\mathbf{I}_I - \frac{\mathbf{t}_a\mathbf{t}_a^{\mathrm{T}}}{\mathbf{t}_a^{\mathrm{T}}\mathbf{t}_a}\right)\mathbf{Y}_a \quad , \tag{3.18}$$

where $\mathbf{I}_I$ is the [I×I] identity matrix. Namely, at the a^{th} step the reconstructions of each dataset from the a^{th} estimated LV are subtracted to the datasets themselves. In particular, from the second terms of Eqs. (3.17)-(3.18) it results that:

$$\mathbf{p}_a^{\mathrm{T}} = \frac{\mathbf{t}_a^{\mathrm{T}}\mathbf{X}_a}{\mathbf{t}_a^{\mathrm{T}}\mathbf{t}_a} \tag{3.19}$$

$$\mathbf{q}_a^{\mathrm{T}} = \frac{\mathbf{t}_a^{\mathrm{T}}\mathbf{Y}_a}{\mathbf{t}_a^{\mathrm{T}}\mathbf{t}_a} \quad , \tag{3.20}$$

where $\mathbf{p}_a$ [N×1] and $\mathbf{q}_a$ [M×1] represent the loadings in the reconstruction of $\mathbf{X}_a$ and $\mathbf{Y}_a$, respectively. The datasets are then decomposed as follows:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E}_{\mathbf{X}} \tag{3.21}$$

$$\mathbf{Y} = \mathbf{T}\mathbf{Q}^{\mathrm{T}} + \mathbf{E}_{\mathbf{Y}} \tag{3.22}$$

$$\mathbf{T} = \mathbf{X}\mathbf{W}^{*} \quad . \tag{3.23}$$

$\mathbf{T}$ [I×A] is the score matrix, $\mathbf{P}$ and $\mathbf{Q}$ are the [N×A] and [M×A] loadings matrices, while $\mathbf{E}_{\mathbf{X}}$ [I×N] and $\mathbf{E}_{\mathbf{Y}}$ [I×M] are the residuals to account for the model mismatch. In Eq. (3.23) $\mathbf{W}^{*}$ [N×A] is the weight matrix calculated from the weight matrix $\mathbf{W}$:

$$\mathbf{W}^{*} = \mathbf{W}(\mathbf{P}^{\mathrm{T}}\mathbf{W})^{-1} \quad . \tag{3.24}$$

PLS offers a great advantage when dealing with two different datasets because it provides a model for the correlative structure of $\mathbf{X}$, a model for the correlative structure of $\mathbf{Y}$ and one for their mutual relation. Often in Eq. (3.22) the score matrix $\mathbf{T}$ is substituted by the score matrix $\mathbf{U}$ [I×A], with $\mathbf{U}$=$\mathbf{TB}$ (the so called inner relation; Geladi and Kowalski, 1986), that in this case is a linear relation between the scores of the regressors $\mathbf{X}$ and the scores of the response variables $\mathbf{Y}$. $\mathbf{B}$ is the matrix of the regression coefficients.

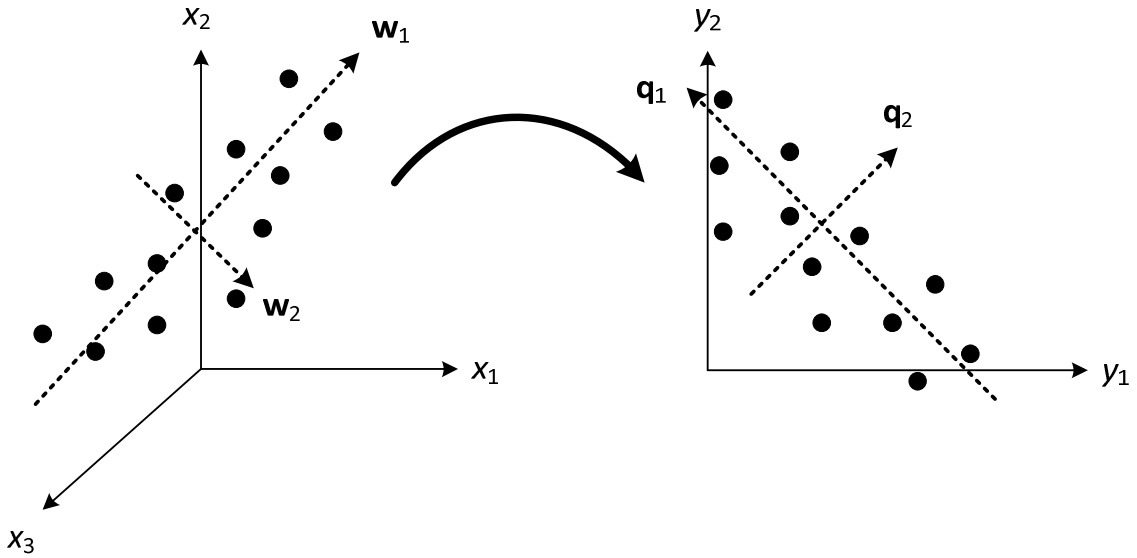


Figure 3.2. *Geometrical interpretation of PLS.*

Figure 3.2 shows the geometrical interpretation of a PLS model, in which a dataset $\mathbf{X}$ [12×3] and $\mathbf{Y}$ [12×2] are considered. As can be seen, data in $\mathbf{X}$ lay mainly on a plane, identified by two LVs. Latent directions are identified in $\mathbf{X}$ and $\mathbf{Y}$ to best approximate the directions of maximum variability of the original data and their correlative structure. The projection of the

original points on the latent directions represent the PLS scores, while loadings are the director cosines of the latent directions, as in the PCA case.
PLS model scores, weights and loadings can be interpreted to gain understanding on the similarity between different samples and on the correlation among variables within and between datasets. Several algorithms have been proposed in the literature to calculate the parameters of a PLS model, such as NIPALS (Wold et al., 1983) and SIMPLS (De Jong, 1993). The advantage in using these algorithms instead of solving the eigenvector decomposition in Eq. (3.16) or the optimization problem in Eq. (3.14) is that they are iterative, allowing to stop after calculating a given number of LVs. Furthermore they can easily handle datasets with missing data, providing robust models. An exhaustive discussion on PLS modeling techniques can be found in the works of Höskuldsson (1988), Burnham et al. (1996), Burnham et al. (1999a) and Burnham et al. (1999b).

3.1.2.1 PLS data pre-treatment, selection of number of LV and model diagnostics

Also in the case of PLS, appropriate data pre-treatment must be applied. The same consideration provided in Section 3.1.1.1 for PCA are valid. In particular, auto-scaling was applied in this Dissertation.
Different methods for selecting the appropriate number of LVs in a PLS model have been discussed and compared in literature (Li et al., 2002; Wiklund et al., 2007). In this Dissertation the scree test was used.
In a PLS model the diagnostics to evaluate its performances are the same of a PCA model (Eq. (3.11) - Eq.(3.13)), but in this case they can be calculated for both datasets $\mathbf{X}$ and $\mathbf{Y}$.

3.2 Other latent variable model technique: joint-Y PLS

Joint-Y PLS (JY-PLS; Garcìa-Muñoz et al., 2005) is a latent variable modeling technique which allows to relate two or more regressor datasets (i.e. $\mathbf{X}_A$ [$I \times N_A$] and $\mathbf{X}_B$ [$J \times N_B$]) through the joint space formed by their corresponding response variables datasets (i.e. $\mathbf{Y}_A$ [$I \times M$] and $\mathbf{Y}_B$ [$J \times M$]). The basic idea of JY-PLS is that, if the correlation structure of the response datasets $\mathbf{Y}_A$ and $\mathbf{Y}_B$ is similar, a common latent space can be identified between $\mathbf{Y}_A$ and $\mathbf{Y}_B$. Assuming that the regressor datasets are correlated with the corresponding response variables, this common latent space (or part of it) will be spanned by (part of) the LVs of the regressor datasets. Otherwise stated, there will exist a region in the latent space of the matrix $\mathbf{Y}_J$, obtained by joining the response variable datasets ($\mathbf{Y}_J = [\mathbf{Y}_A^T \, \mathbf{Y}_B^T]^T$), in which the LVs of the regressor datasets will be overlapped. This region can be exploited to relate the different datasets and to transfer information between them. For example, data from different plants or different plants scales can be used. To find this common region, the available datasets are decomposed on their latent structures in order to maximize at the same time the squared

covariance between $\mathbf{X}_A$ and $\mathbf{Y}_A$, and between $\mathbf{X}_B$ and $\mathbf{Y}_B$, together with the squared joint covariance between them (Garcìa-Muñoz et al., 2005):

$$\mathbf{Y}_J = \begin{bmatrix} \mathbf{Y}_A \\ \mathbf{Y}_B \end{bmatrix} = \begin{bmatrix} \mathbf{T}_A \\ \mathbf{T}_B \end{bmatrix} \mathbf{Q}_J^T + \mathbf{E}_{\mathbf{Y}_J} \tag{3.25}$$

$$\mathbf{X}_A = \mathbf{T}_A \mathbf{P}_A^T + \mathbf{E}_{\mathbf{X}_A} \tag{3.26}$$

$$\mathbf{X}_B = \mathbf{T}_B \mathbf{P}_B^T + \mathbf{E}_{\mathbf{X}_B} \tag{3.27}$$

$$\mathbf{T}_A = \mathbf{X}_A \mathbf{W}_A^* \tag{3.28}$$

$$\mathbf{T}_B = \mathbf{X}_B \mathbf{W}_B^* \quad , \tag{3.29}$$

where $\mathbf{Q}_J$ [$M \times A$] represents the joint loading matrix defining the common latent space of $\mathbf{Y}_J$ (A is the number of LVs used to build the model). The meaning of the other symbols is the same as in the PLS case. A modified version of the NIPALS algorithm has been proposed by Garcìa-Muñoz et al. (2005) to compute the parameters of the JY-PLS model. The JY-PLS method calculate two separate PLS models for datasets A ($\mathbf{X}_A$ and $\mathbf{Y}_A$) and datasets B ($\mathbf{X}_B$ and $\mathbf{Y}_B$) whose spaces are however rotated to be aligned with the common correlation structure of the variables in $\mathbf{Y}_J$.

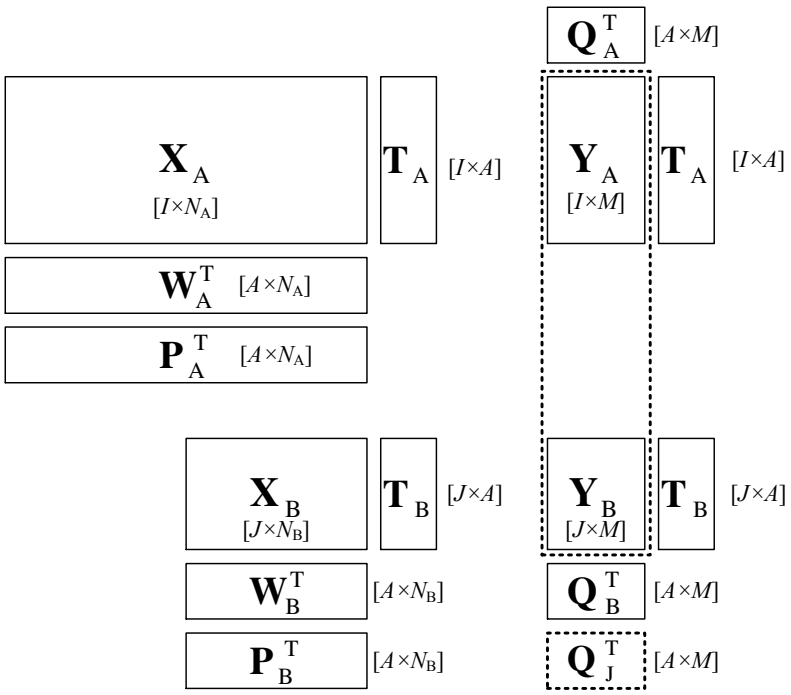


Figure 3.3. *Schematic of the data structure and parameters of a joint-Y PLS model. Adapted from Garcìa-Munoz et al. (2005).*

Figure 3.3 shows the typical structure of the datasets used in a joint-Y PLS model, with the representation of the model parameters for each dataset. As can be seen, the model does not

impose any restrictions on the number of columns in $\mathbf{X}_A$ and $\mathbf{X}_B$, which can be different between the datasets, or on the number of observation per regressor dataset. The only restriction is that the number of columns in the response variable matrices must be the same. Here the case with two regressors datasets was presented, but the technique can be easily extended to more datasets.

3.3 Latent variable model inversion

A LVM is usually used to predict a set of response variables $\hat{\mathbf{y}}^{NEW}$ [1×*M*] given a set of regressors $\mathbf{x}^{NEW}$ [1×*N*]. This is the direct use of a model and it is depicted in Figure 3.4a. However, if a LVM based on historical data is available and a set of desired response variables $\mathbf{y}^{DES}$ [1×*M*] is defined, the LVM can be exploited to estimate the set of regressors $\mathbf{x}^{NEW}$ [1×*N*] that, according to the model, provides the desired response variables (Figure 3.4b). This is the concept behind model inversion. Model inversion can be of paramount importance to assist product and process design and development. Assuming that $\mathbf{y}^{DES}$ is completely defined, the LVM inversion estimates the projections $\mathbf{t}^{DES}$ of $\mathbf{y}^{DES}$ onto the latent space of the model (Jaeckle and MacGregor, 1998):

$$\mathbf{t}^{DES} = (\mathbf{Q}^T\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{y}^{DES} \quad , \tag{3.30}$$

where $\mathbf{t}^{DES}$ can be used to reconstruct the set of regressors $\mathbf{x}^{NEW}$ corresponding to $\mathbf{y}^{DES}$ (this is the solution of direct inversion of the model). However, the solution of the model inversion can be not unique, depending on the dimensions of the latent spaces of $\mathbf{X}$ and $\mathbf{Y}$ (namely, on their statistical rank) and on the number *A* of LVs used to build the model.

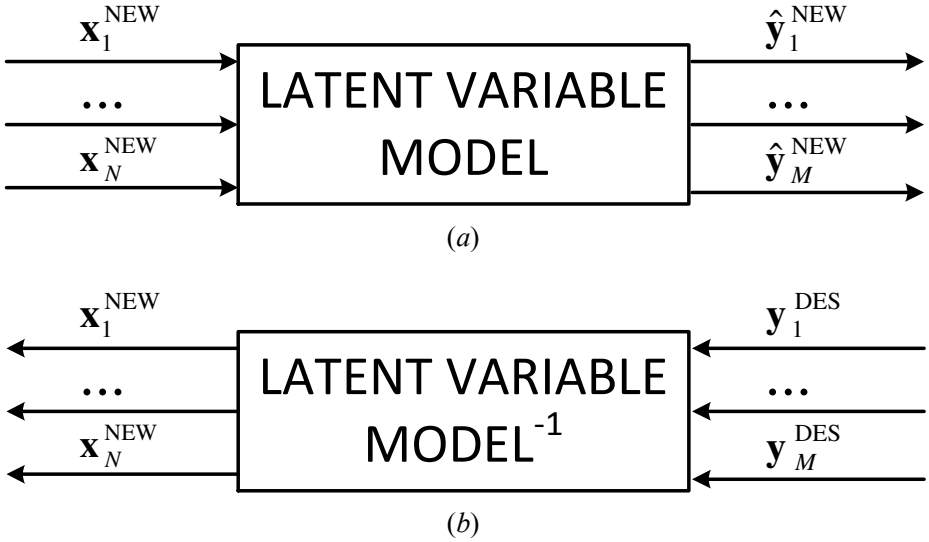


Figure 3.4. *Direct and inverse use of models: (a) direct use and (b) inverse use.*

Assuming that $R_{\mathbf{X}}$ is the statistical rank of matrix $\mathbf{X}$ and $R_{\mathbf{Y}}$ is the statistical rank of matrix $\mathbf{Y}$, usually A=max($R_{\mathbf{X}}$,$R_{\mathbf{Y}}$). Depending on the ranks of the dataset, three different cases may arise in model inversion:

1. $A = R_{\mathbf{Y}}$ (with $R_{\mathbf{Y}} \geq R_{\mathbf{X}}$): this is the most favorable case, when all the LVs of $\mathbf{X}$ have potentially an impact on the $\mathbf{Y}$ space (Burnham et al., 1999a). In this case model inversion corresponds to a projection from a higher dimensional space ($\mathbf{Y}$) onto a lower dimensional space ($\mathbf{X}$).
2. $A = R_{\mathbf{X}}$ ($R_{\mathbf{X}} \geq R_{\mathbf{Y}}$): this is the most common situation. There are some LVs in the $\mathbf{X}$ space that are statistically significant to explain the variability in the $\mathbf{X}$ dataset, but do not contribute in explaining the variability in the $\mathbf{Y}$ space. In other words, they account for a part of the $\mathbf{X}$ data that is not related (i.e., is orthogonal) to the $\mathbf{Y}$ space (Burnham et al., 1999a). In this case model inversion corresponds to a projection from a lower dimensional space ($\mathbf{Y}$) to a higher dimensional space ($\mathbf{X}$).
3. $A = R_{\mathbf{X}} = R_{\mathbf{Y}}$ but $\text{rank}([\mathbf{XY}]) > A$: in this case, even if the rank of the matrices is equal, $R_{\mathbf{XY}}$ is greater than A, meaning that there are $R_{\mathbf{XY}}$-A latent directions that do not overlap between the $\mathbf{X}$ and $\mathbf{Y}$ spaces. Therefore, the situation is similar to case 2.

In the first case, direct model inversion (Eq. (3.30)) can be applied and a unique solution exists. In the second and the third cases, the inversion problem is underdetermined and the set of solutions is infinite. The solution from Eq. (3.30) can be moved along the directions of the latent space not affecting $\mathbf{Y}$ (thus changing $\mathbf{t}^{\text{DES}}$), providing the same set of desired response variables $\mathbf{y}^{\text{DES}}$. These latent directions form a null space, which is a subspace of the model space representing the locus of the $\mathbf{X}$ projections with no influence on the response variables (Jaeckle and MacGregor, 1998). The direct inversion solution can therefore be moved along the null space, in order to find the solution that achieves the objectives and satisfies constraints the inversion problem may have (i.e., in product or process design). For this reason, optimization approaches have been proposed to solve LVM inversion (García-Muñoz et al., 2006; García-Muñoz et al., 2008). An exhaustive study on LVM inversion has been proposed by Tomba et al. (2012).

3.3.1 Optimization framework to solve latent variable model inversion

In this Dissertation LVM inversion is solved by means of an optimization framework. The optimization problem is formulated as follows, after data pre-tretament (Tomba et al., 2012):

$$\min_{\mathbf{x}^{\mathrm{NEW}}}\left\{\left(\mathbf{y}^{\mathrm{DES}}-\hat{\mathbf{y}}^{\mathrm{DES}}\right)^{\mathrm{T}}\left(\mathbf{y}^{\mathrm{DES}}-\hat{\mathbf{y}}^{\mathrm{DES}}\right)+G_1\left(\sum_{lv=1}^{\mathrm{LV}}\left(\frac{\mathbf{x}^{\mathrm{NEW}}\mathbf{w}_{lv}}{\mathbf{s}_{lv}}\right)^2\right)+G_2\mathrm{SPE}_{\mathbf{x}}\right\}$$

subject to

$$\begin{aligned}&\hat{\mathbf{y}}^{\mathrm{DES}}=\mathbf{x}^{\mathrm{NEW}}\mathbf{W}\mathbf{Q}^{\mathrm{T}}\\&\mathrm{SPE}_{\mathbf{x}}=\left(\mathbf{x}^{\mathrm{NEW}}-\mathbf{x}^{\mathrm{NEW}}\mathbf{W}\mathbf{P}^{\mathrm{T}}\right)^{\mathrm{T}}\left(\mathbf{x}^{\mathrm{NEW}}-\mathbf{x}^{\mathrm{NEW}}\mathbf{W}\mathbf{P}^{\mathrm{T}}\right)\\&\mathbf{x}^{\mathrm{NEW}}=\mathbf{c}^{eq}\quad,\end{aligned}\tag{3.31}$$

where $\mathbf{x}^{\mathrm{NEW}}$ is the estimated set of regressors to obtain $\mathbf{y}^{\mathrm{DES}}$, $\mathbf{W}$ are the weights of the LVM, $\mathbf{Q}$ are the $\mathbf{Y}$ loadings of the LVM, G_1 and G_2 are weighting parameters, LV is the number of latent variable used to build the model, $\mathbf{w}_{lv}$ is the lv^{th} column of $\mathbf{W}$, $\mathbf{s}_{lv}$ is the variance of the lv^{th} column of the score matrix $\mathbf{T}$ of the LVM, $\mathbf{P}$ are the $\mathbf{X}$ loadings in the LVM model, and $\mathbf{c}^{\mathrm{eq}}$ is an equality constraint on the estimated set of regressors.

Three different terms can be identified in the objective function:

- the first term is actually to obtain a set of regressors $\mathbf{x}^{\mathrm{NEW}}$ that provides, according to the model, the desired response variable $\mathbf{y}^{\mathrm{DES}}$;
- the second term represents a soft constraint on Hotelling's T^2 of the solution. Namely, it keeps the solution close to the average values of the historical dataset $\mathbf{X}$ used to build the model. G_1 is a weighting parameter and Tomba et al. (2012) suggests to use the reciprocate of the 95% Hotelling's T^2 limit of the LVM model;
- the third term identifies a soft constraint on the SPE of the estimated set of regressors. It forces the solution to lay close to the model space. G_2 is a parameter to weight the constraint and Tomba et al. (2012) suggests to use the reciprocate of the 95% SPE limit of the LVM model.

Chapter 4

Available datasets of the wheat milling process

In this Chapter the datasets of the wheat milling process are presented. A general overview of the structure of the available datasets and of the experimental campaigns for data collection is provided. Data have been collected for two different plant scales: small-scale and industrial-scale plant, with purposely designed experiments performed by the author.

4.1 General structure of the datasets

Data used in this Dissertation are related to the wheat milling process. Data were collected for different plant scales, different passages of the process and different unit operations.
In particular, two plant scales were studied:

- small-scale plant;
- industrial-scale plant.

The main unit operations of the process were analyzed:

- roller mills;
- plansifters.

In general, for the roller mills (superscript ^{r}), the experimental data were organized in an input matrix $\mathbf{X}^{r}$ [I×N] and in an output matrix $\mathbf{Y}^{r}$ [I×M], where I is the number of observations, N the number of input variables, and M the number of output responses. $\mathbf{X}^{r}$ collects the most relevant process parameters of the roller mill and the wheat properties of the feeding material; $\mathbf{Y}^{r}$ contains the particle size distribution (PSD) of the milled material. If the mass fractions are measured immediately after the roller mill, the matrix is defined as $\mathbf{Y}'^{r}$ [I×S], with S being the number of sieved fractions. The fractions of mass sieved in sieve s, with respect to the total mass sieved, is indicated by w_s. The data organization for the plansifters (superscript ^{p}) is similar, but the output matrix takes the form of a three-way array $\underline{\mathbf{Y}}^{p}$ [I×M×S], where S is the number of sieved fractions; $\underline{\mathbf{Y}}^{p}$ collects the PSDs of each sieved fraction. The mass fraction values after the plansifter are collected in a matrix $\mathbf{Y}'^{p}$ [I×S].
A detailed description of the dataset for each scale is provided in the following.

4.2 Small-scale dataset

Experiments were carried out for nine different passages: three breakage passages (B1, B2 and B3) and seven reduction passages (R1, R2, R3, R5, R6, R9 and R12). Two breakage passage (B1 and B2) and one reduction passage (R1) were considered as representative of the issues related to modeling the breakage and the reduction steps. Therefore, only results for these three passages will be reported in this Dissertation. Data were collected through an experimental campaign carried out in a small-scale roller mill. The main differences with an industrial-scale roller mill are the amount of material that can be processed and the length of the rolls. In the case of the first breakage passage the feeding material to the small-scale roller mill was raw wheat. For all the other passages, the feeding material was collected in an industrial-scale mill (e.g., the material used in the small-scale B2 roller mill was collected at the entrance of the industrial-scale B2 roller mill). The material processed by the small-scale mill was then sieved in a laboratory sifter. The sieve mesh sizes are reported in Table 4.1; the mesh size differences in the two passages reflect the fact the particle size range changes across the milling process.

Table 4.1. *Sieve mesh sizes for the breakage and reduction passages.*

B1 passage (after roller mill)		B passages		R passages	
Sieve number	**Mesh size [µm]**	**Sieve number**	**Mesh size [µm]**	**Sieve number**	**Mesh size [µm]**
1	1120	1	1120	1	300
2	0	2	700	2	250
		3	500	3	212
		4	250	4	180
		5	132	5	0
		6	0		

4.2.1 B1 passage

For the B1 passage, experiments were designed following a full factorial 2^3 design of experiments (DoE; Montgomery, 2008) applied to three parameters: feed mass flow (MF, kg/hr), milling gap between the rolls (MG, µm) and wheat kernel moisture content (MC, %). Some other process parameters can affect the milling operation, such as roll speed, roll differential and fluting type; all of them were set constant as in an industrial wheat milling process, in order to contain the scope of the investigation and focus on the most important process-specific variables. In order to take into account the variability of the raw material, the design was repeated for three different wheat varieties (semi-hard, hard and soft varieties were used). The ranges for MF, MG and MC were chosen to match typical industrial ranges and the actual values are not reported for confidentiality reasons. Each experiment was replicated twice, and average values are used. Each wheat sample feeding the roller mill was characterized using a NIR spectrum, measured with a NIR spectrometer (TANGO, Bruker

Optics, Germany). Each measurement was repeated three times, and averaged spectra are used.

The available dataset is organized as follows. Twenty-four different observations were organized in a $\mathbf{X}^{r}_{B1}$ matrix [24×953] collecting data on 953 variables: 2 process parameters (MF and MG), one wheat property (MC) and the absorbance corresponding to 950 wavenumbers that describe the NIR spectrum. The considered wavenumbers range from 11536 cm^{-1} to 3944 cm^{-1}, evenly spaced every 8 cm^{-1}.

The matrix $\mathbf{Y}^{r}_{B1}$ [24×35] is built, including 35 variables. Each row of $\mathbf{Y}^{r}_{B1}$ represents a PSD of a single observation. For a given observation *i*, the value taken by each $y_{i,m}$ variable of $\mathbf{Y}^{r}_{B1}$ refers to the volume fraction of particles whose Feret diameter (Merkus, 2009) is within the range of particle dimensions defined by bin *m* in the discretized representation of the PSD. Namely, $y_{i,1}$ corresponds to the volume fraction of the smallest particle diameter range (0-4 µm) in observation *i*, whereas $y_{i,35}$ represents the volume fraction of the biggest diameter range (3863-4798 µm) of the same observation. The details about the discretization of the PSD bins are reported in Appendix A. The PSDs were measured with the image analysis sensor QICPIC (Sympatec GmbH, Germany). The mass fractions after the roller mill are collected in a matrix $\mathbf{Y'}^{r}_{B1}$ [24×2]. Table 4.2 summarizes the variables used in the B1 model.

Table 4.2. *Variables used in the B1 passage modeling.*

$\mathbf{X}^{r}_{B1}$		$\mathbf{Y}^{r}_{B1}$		$\mathbf{Y'}^{r}_{B1}$	
no.	**name**	**no.**	**name**	**no.**	**name**
1	MF [kg/hr]	from 1 to 35	bins of the PSD	1	w_1
2	MC [%]			2	w_2
3	MG [µm]				
from 4 to 953	wavenumbers of the NIR spectrum [cm^{-1}]				

4.2.2 B2 passage

The experiments for the B2 passage were designed following a face-centered central composite DoE, the factors being MF, MC and MG. The intervals for each parameter were set to match typical industrial ranges. Each experiment, as well as each MC and PSD measurement, was repeated at least twice, and average values were then used for modeling. The PSDs were measured with the image analysis instrument QICPIC (Sympatec GmbH, Germany). The mass of each sieved fraction was also measured.

For the B2 passage, the roller mill data matrices are $\mathbf{X}^{r}_{B2}$ [15×3] and $\mathbf{Y}^{r}_{B2}$ [15×37], where the input variables are MF, MG, and MC, and the output variables are the PSDs at the mill exit. The B2 plansifter dataset has $\mathbf{Y}^{r}_{B2}$ as the input matrix, and $\underline{\mathbf{Y}}^{p}_{B2}$ [15×29×6] as the output matrix. The meaning of the 29 variables in $\underline{\mathbf{Y}}^{p}_{B2}$ is similar to the $\mathbf{Y}^{r}_{B2}$ variables, but the diameter ranges are different. The PSD bins for the roller mill and the plansifter are listed in

Appendix A. The mass fractions of each sieved fraction are collected in $\mathbf{Y'}^{p}_{B2}$ [15×6]. Variables used in the B2 passage modeling are listed in Table 4.3.

Table 4.3. *Variables used in the B2 passage modeling.*

$\mathbf{X}^{r}_{B2}$		$\mathbf{Y}^{r}_{B2}$		$\underline{\mathbf{Y}}^{p}_{B2}$		$\mathbf{Y'}^{p}_{B2}$	
no.	**name**	**no.**	**name**	**no.**	**name**	**no.**	**name**
1	MF [kg/hr]	from 1 to 37	bins of the PSD	from 1 to 29	bins of the PSD	1	w_1
2	MC [%]					2	w_2
3	MG [µm]					3	w_3
						4	w_4
						5	w_5
						6	w_6

4.2.3 R1 passage

The experiments for the R1 passage were designed with a full factorial design of experiments on the above mentioned three parameters, using 3 levels for MF and MC, and 9 levels for MG. The input variables for the R1 roller mill dataset $\mathbf{X}^{r}_{R1}$ [63×3] are the same as the for the B2 roller mill (MF, MC and MG). Only 63 observations were available, because not all the milled materials were measured. The outlet PSD bins in the $\mathbf{Y}^{r}_{R1}$ [63×24] matrix are reported in Appendix A. The R1 plansifter output matrix $\underline{\mathbf{Y}}^{p}_{R1}$ [6×24×5] includes only 6 observations, because not all the 63 materials milled in $\mathbf{Y}^{r}_{R1}$ were sieved and measured. $\mathbf{Y'}^{p}_{R1}$ [16×5] collects the mass fractions w_s for the R1 plansifter passage. Table 4.4 summarizes the variables used in R1 passage modeling.

Table 4.4. *Variables used in the R1 passage modeling.*

$\mathbf{X}^{r}_{R1}$		$\mathbf{Y}^{r}_{R1}$		$\underline{\mathbf{Y}}^{p}_{R1}$		$\mathbf{Y'}^{p}_{R1}$	
no.	**name**	**no.**	**name**	**no.**	**name**	**no.**	**name**
1	MF [kg/hr]	from 1 to 24	bins of the PSD	from 1 to 24	bins of the PSD	1	w_1
2	MC [%]					2	w_2
3	MG [µm]					3	w_3
						4	w_4
						5	w_5

4.3 Industrial-scale dataset

Experiments carried out in the industrial-scale plant were related to the B1 passage. Experiments were designed with a 2^1 DoE on the MG only, repeated for three wheat blends (a semi-hard, a hard and a soft blend were used).

The MG in an industrial-scale roller mill cannot be measured in µm, but it is specified with an index adjustable directly on the roller mill with a wheel. Additionally, a different MG can be fixed at each side of the rolls length, to counteract the not perfect alignment of the roll pair. This two MGs will be referred as MG_l and MG_r.

The industrial-scale dataset is structured as follows: the input matrix $\mathbf{X}^{r}_{B1,IS}$ [6×1052] collects the values of MG_l, MG_r and the 1050 wavelengths in nm describing the wheat type for six different experiments. The responses $\mathbf{Y}'^{r}_{B1,IS}$ [6×2] include the mass fractions of the milled material. Variables used in the industrial-scale modeling are listed in Table 4.5.

Table 4.5. *Variables used in the industrial-scale B1 passage modeling.*

$\mathbf{X}^{r}_{B1,IS}$		$\mathbf{Y}'^{r}_{B1,IS}$	
no.	**name**	**no.**	**name**
1	MG_l [-]	1	w_1
2	MG_r [-]	2	w_2
from 3 to 1052	wavelengths of NIR spectrum [nm]		

Six additional experiments, performed with a MG within the range defined by the first six experiments, were available. These experiments were used to validate the models. The structure of the dataset is showed in Figure 4.1 and the calibration and validation samples are highlighted.

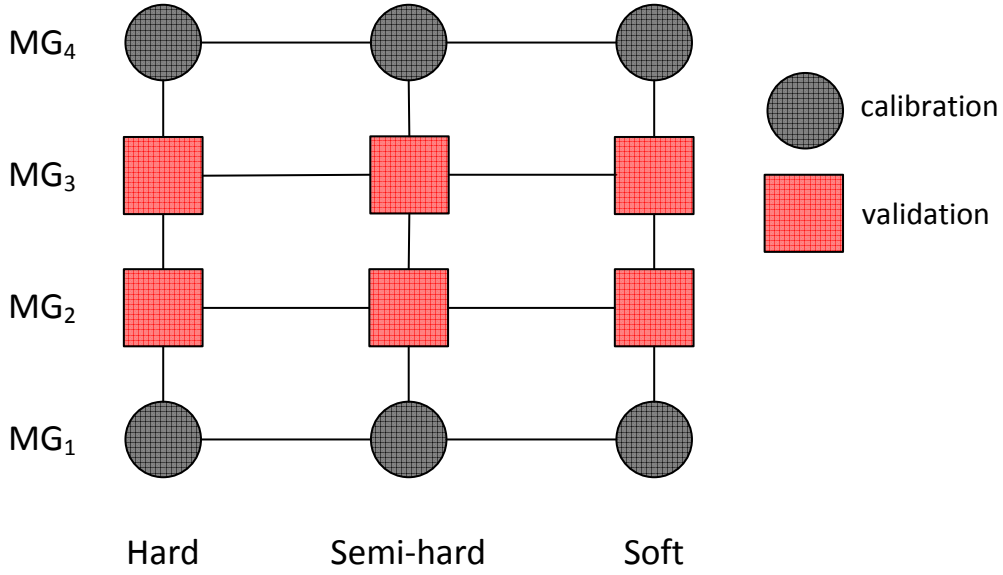


Figure 4.1. *Structure of the industrial-scale dataset. The calibration (circles) and validation (squares) samples are highlighted.*

Chapter 5

Small-scale modeling*

In this Chapter the models built on small-scale data are shown. In particular, the models of the first-breakage (B1), second-breakage (B2) and first reduction (R1) passages are presented. Then it is discussed how they can be used to support the milling and sieving operations.

5.1 Introduction

Models have been built for the two main unit operations of the process: roller mills and plansifters.

The roller mill models are used: *i*) to classify different wheat varieties, *ii*) to improve the understanding on the milling process, by studying and quantifying the correlation between input parameters (process parameters and wheat properties) and response variables (product PSD), and *iii*) to predict the product PSD from the values of the process parameters and wheat properties.

The plansifter models are utilized to predict the product quality and amount.

Additionally, the roller mill model and the plansifter model of different consecutive passages are linked together. In this case, the propagation of prediction error across subsequent models is also analyzed. This procedure presents two innovations: a first step to a data-based simulation of the entire milling process and the analysis of the error propagation into concatenated data-driven models.

Two breakage passages (B1 and B2) and one reduction passage (R1) were considered as representative of the issues related to modeling the breakage and the reduction steps. Therefore, only results for these three passages are reported in this Dissertation.

---

* Excerpts from this Chapter have been published in:

Dal Pastro F., Facco P., Bezzo F., Zamprogna E., Barolo M. (2016). Data-driven modeling of milling and sieving operations in a wheat milling process. *Food Bioprod. Process.*, **99**, 99-108, doi: 10.1016/j.fbp.2016.04.007.

Dal Pastro F., Facco P., Bezzo F., Thomas H., Zamprogna E., Barolo M. (2015). Data-based multivariate modeling of a grain comminution process. In: Krist V. Gernaey Jakob K. Huusom and Rafiqul Gani, *Computer Aided Chemical Engineering*, 37, 2219-2224, Amsterdam:Elsevier, doi: 10.1016/B978-0-444-63576-1.50064-9.

Dal Pastro F., Facco P., Bezzo F., Zamprogna E., Barolo M. (2016). Using PLS and NIR spectra to model the first-breakage step of a grain milling process. In: Zdravko Kravanja and Milos Bogataj, *Computer Aided Chemical Engineering*, 38, 1171-1176, Amsterdam: Elsevier, doi: 10.1016/B978-0-444-63428-3.50200-9.

5.2 First-breakage passage modeling

5.2.1 Roller mill model

A PLS model built using 3 LVs explains almost the entire variance (99.6 %) of the $\mathbf{X}_{B1}^{r}$ matrix (which collects the values of the feed mass flow, milling gap and wheat moisture content) together with a large fraction (83.1 %) of the $\mathbf{Y}_{B1}^{r}$ matrix variance (that contains the PSDs of the milled material). Given that the process parameters MF and MG, as well as the wheat property MC, were varied accordingly to a DoE strategy, they are uncorrelated. Three LVs are able to describe such a large amount of the $\mathbf{X}_{B1}^{r}$ matrix variability since there is a high correlation among the spectra at different wavenumbers, as well as between the parameters and the spectra.

The PLS model is used: *i*) to check if the NIR spectra are able to discriminate between different wheat varieties, *ii*) to improve process understanding of first-breakage by studying the effects of process parameters and wheat characteristics on the product PSD, and *iii*) to estimate the product PSD from the process parameters, moisture content and NIR spectra.

5.2.1.1 Classification of wheat using NIR spectra

First, the ability of NIR spectra to discriminate between the different wheat varieties was investigated. This can be done by checking the position of the observations on the score plot of the PLS model. If the NIR spectra have the ability to distinguish the different wheat varieties, the score representing each type of wheat should create a separated cluster in the latent variable space, with the variability within each cluster much smaller than the distance between clusters.

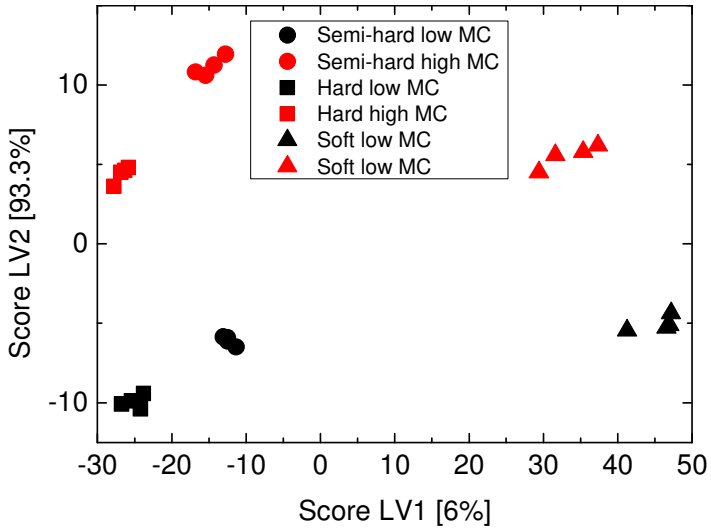


Figure 5.1. *Score plot of the first two LVs of the PLS model of first-breakage.*

After appropriate spectra pre-treatment (standard normal variate; Rinnan et al., 2009), the NIR spectra are able to discriminate not only different wheat varieties, but also different MCs. In

Figure 5.1, different symbols identify different wheat varieties, whereas different colors are associated to different MCs. It is interesting to note that the wheat varieties are positioned along the first LV. In particular, the semi-hard and hard wheat show negative values of LV1, whereas soft wheat has positive values of the same LV. The second latent variable explains the different moisture contents: samples characterized by high MC exhibit a positive value of LV2, whereas a negative LV2 value is exhibited for low MC. Given the same MC, the semi-hard wheat shows higher values of the second LV (namely, it is shifted in the second latent variable respect to the other two varieties), meaning that the effect of MC is not linear and depends on the wheat variety. The first LV explains a very low percentage of the $\mathbf{X}_{B1}^{r}$ matrix variability (only 6 %), whereas the second LV explains more than 93 % of the $\mathbf{X}_{B1}^{r}$ variance. Since the first LV explains the difference in variety and the second LV the difference in MC, MC represents a more significant source of variability in the $\mathbf{X}_{B1}^{r}$ matrix than the wheat variety. It is important to note that this PLS model is not supervised; the discrimination of the wheat varieties was not imposed in the model, that was created with the purpose of studying the information of the input parameters that affects the response and of predicting the product quality, but the discriminant ability is directly "hidden" in the NIR spectra.

5.2.1.2 Process understanding

A combined reading of the loading plots **P** and **Q** of the PLS model allows one to understand the effects of the model inputs (MF, MG, MC and NIR spectra) on the PSD of the product leaving the first-breakage roller mill.

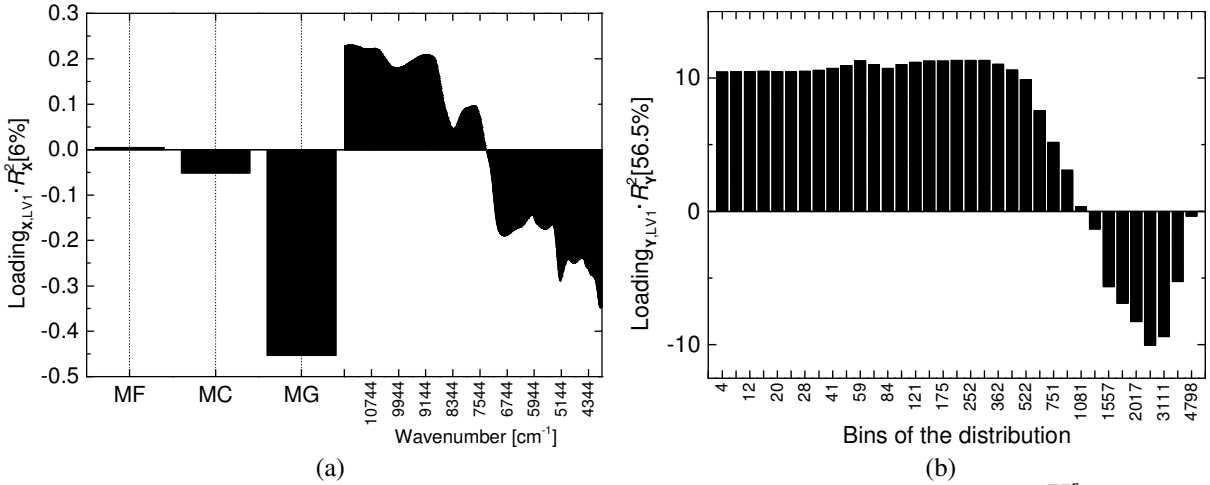


Figure 5.2. *Loading plots of the first LV of the PLS model: (a) loading plot of the* $\mathbf{X}_{B1}^{r}$ *matrix and (b) loading plot of the* $\mathbf{Y}_{B1}^{r}$ *matrix.*

As an illustrative example, in Figure 5.2 the loading **P** (Figure 5.2a) and **Q** (Figure 5.2b) of the first LV of the PLS model are shown. Each loading is weighted by the explained variance of the relevant LV (in this case 6 % and 56.5 % for the $\mathbf{X}_{B1}^{r}$ and $\mathbf{Y}_{B1}^{r}$ matrices, respectively). In Figure 5.2a, the dominant variable is the MG. Compared with the loading plot of the $\mathbf{Y}_{B1}^{r}$

matrix in Figure 5.2b, the effects of the MG on the output PSD from the roller mill can be assessed. When the MG increases, the amount of "big" particles (in particular, those bigger than 1297 µm) increases, whereas the amount of "small" particles decreases. It is interesting to note that, even if the MG explains only a small fraction of explained variance of the $\mathbf{X}_{B1}^{r}$ matrix (6 %), it explains the majority of the variability of the PSDs in the $\mathbf{Y}_{B1}^{r}$ matrix (56.5 %), meaning that the MG is the variable that mostly affects the output PSDs.

The MG is not the only important variable in the first LV. Two significant ranges of wavenumbers of the NIR spectra can be identified as important, too: the first one is around 5000 cm^{-1}, whereas the other one is located around 4000 cm^{-1}. The first wavenumber is related to the OH bonds of water (Burns and Ciurczak, 2008): when the sample MC increases, the amount of particles bigger than 1297 µm increases as well. It is known that particles bigger than 1297 µm are mainly made of bran, and the bran breakage behavior is highly dependent on the MC (namely, the industrial practice suggests that a wetter bran becomes less fragile and it breaks in bigger particles). Note that also the MC shows a positive correlation with the wavenumber range around 5000 cm^{-1}, but this correlation is hidden by the stronger one showed between the wavenumbers. The second wavenumber range (the one around 4000 cm^{-1}) can be associated to different components, such as cellulose, wheat gluten and wheat starch (Burns and Ciurczak, 2008). Interestingly, this range is related to cellulose, the principal component of bran. This means that when the sample contains a higher amount of cellulose (namely, a higher amount of bran), a larger fraction of particles that are greater than 1297 µm are produced (and these particles are made of bran). In summary, an increase of MG, MC and/or cellulose amount in the feed is associated to the production of particles bigger than 1297 µm, and these particles can be identified as bran.

5.2.1.2 Estimation of the particle size distribution

The objective of this subsection is to estimate the product quality (PSD), once the process parameters (MF and MG), the wheat moisture content (MC) and the wheat NIR spectrum are known. To estimate the PSD, a jackknife modeling methodology was used (Shao and Tu, 2012). Namely, one of the observations was removed from the dataset and was used to validate the model calibrated on the remaining observations; the procedure was then iterated, by re-including that observation in the dataset and removing a different observation, until all observations were removed once. Jackknifing is not the only validation technique, other possibilities include, for example, K-fold cross validation (Triba et al., 2015) or bootstrapping (Duchesne and MacGregor, 2001).

PSD predictions errors are reported in terms of root-mean squared error (RMSE). The root-mean squared error for observation i ($RMSE_i$) is defined as:

$$\mathrm{RMSE}_i = \sqrt{\frac{\sum_{m=1}^{M}(y_{i,m} - \hat{y}_{i,m})^2}{M}} \quad , \tag{5.1}$$

where $y_{i,m}$ is the measured value of the response variable for bin m, and $\hat{y}_{i,m}$ is the corresponding predicted value. The RMSE_i values are averaged across all the I observations to give RMSE.

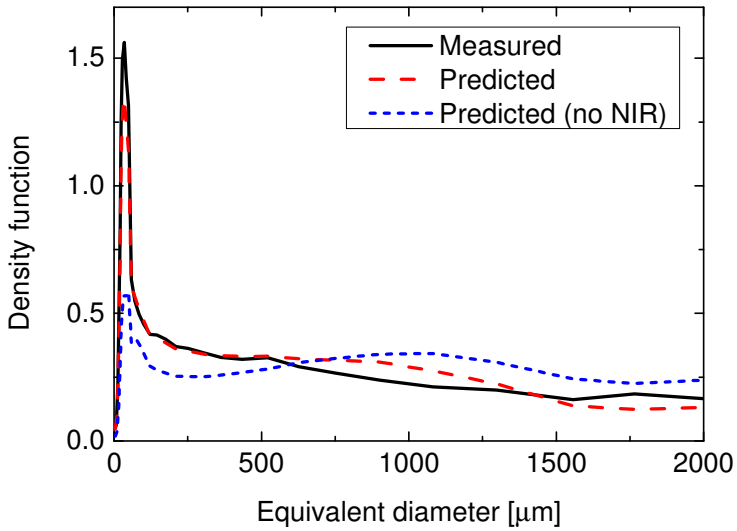


Figure 5.3. *Representative example of prediction of output PSD from the first-breakage roller mill.*

Figure 5.3 shows a representative example of PSD estimation from the first-breakage roller mill, in which the measured PSD (solid line) is compared to the predicted one (dashed line). The prediction is good, with a RMSE of 0.075. The predictions are quite accurate for all observations, and the $\mathrm{RMSE}^{\mathrm{r}}_{\mathrm{B1}}$ (averaged on all the jackknife iterations) was 0.081. Interestingly, if the NIR spectra are removed from the input variables, the estimation performance of the model is severely degraded, with an average RMSE of 0.21. In the example of Figure 5.3, the PSD estimated without the NIR spectrum as input variable is shown as a short-dashed line, and it exhibits a RMSE significantly higher than the one predicted using the NIR spectra (0.34 respect to 0.075). This demonstrates the importance of including the information on wheat variety in an efficient modeling framework.

5.3 Second-breakage passage modeling

5.3.1 Roller mill model

The roller mill model is calibrated on $\mathbf{X}^{\mathrm{r}}_{\mathrm{B2}}$ and $\mathbf{Y}^{\mathrm{r}}_{\mathrm{B2}}$ and the number of LVs to be retained in the model is determined by the scree test in such a way as to explain a sufficiently large fraction of the variance of the product quality. Three LVs were retained in the model,

explaining 88% of the variance of the response matrix $\mathbf{Y}^{r}_{B2}$ (since the experiments were designed through DoE, in order to explain the entire three-parameter input space three LVs are needed, therefore 100 % of $\mathbf{X}^{r}_{B2}$ variance is explained).

5.3.1.1 Process understanding

The effect of the process parameters and wheat properties on the breakage behavior (i.e., on the PSD) of wheat is analyzed. This can be achieved through a combined reading of the **P** and **Q** loadings of the roller mill PLS model.

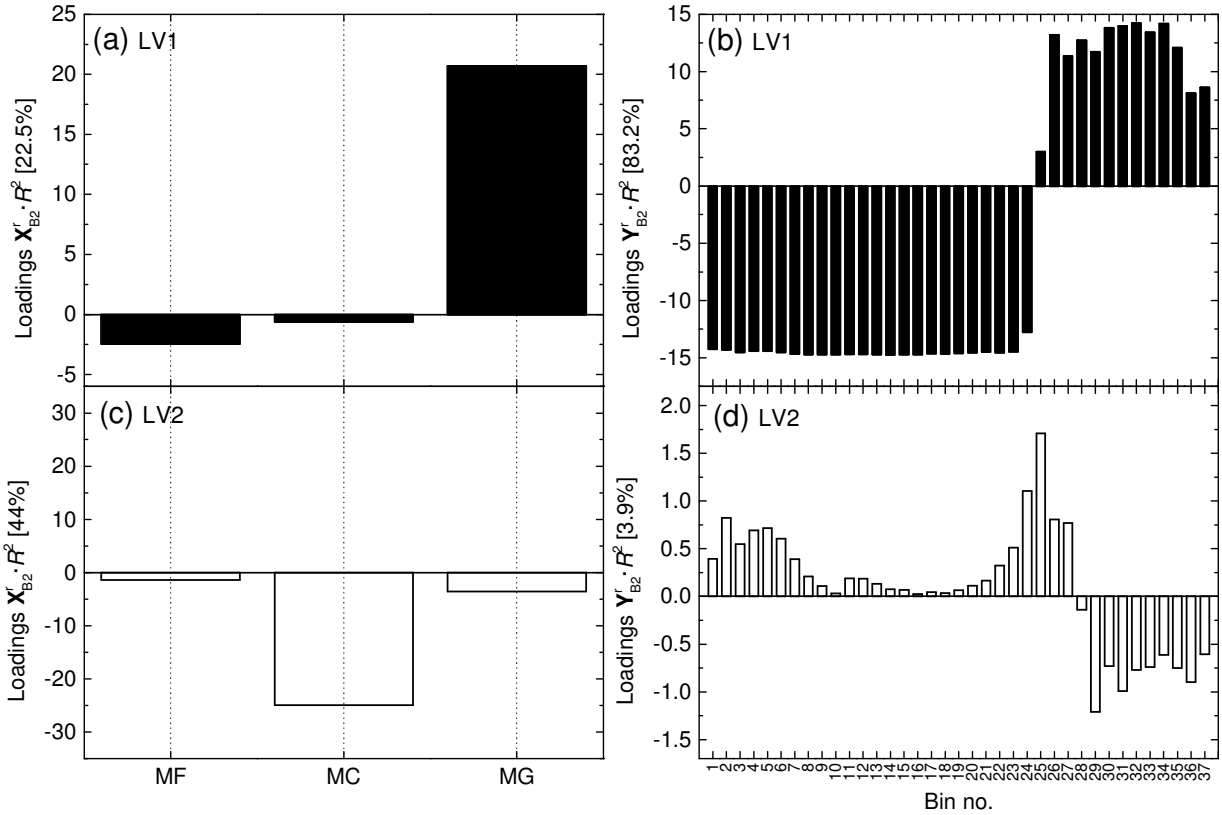


Figure 5.4. *Loading plots of the B2 roller mill model: (a) loadings for the first LV of matrix* $\mathbf{X}^{r}_{B2}$*, (b) loadings for the first LV of matrix* $\mathbf{Y}^{r}_{B2}$*, (c) loadings for the second LV of matrix* $\mathbf{X}^{r}_{B2}$ *and (d) loadings for the second LV of matrix* $\mathbf{Y}^{r}_{B2}$*.*

In Figure 5.4 the plots of the **P** loadings (Figure 5.4a and 5.4c) and **Q** loadings (Figure 5.4b and 5.4d) of the B2 passage roller mill model are shown; note that each loading is weighted for the explained variance of the relevant LV.

The first latent variable (Figure 5.4a) is dominated by the MG, whose effect can be understood by comparing Figure 5.4a with Figure 5.4b (loading plot of first LV of matrix $\mathbf{Y}^{r}_{B2}$). The MG shows a strong correlation with all bins of the PSD; whether the correlation is direct or inverse depends on the bin number (i.e., on the particle diameter). Namely, when the MG increases, the amount of particles smaller than 751 µm (bin no. 25) decreases (bars with negative values in Figure 5.4b), whereas the amount of particles bigger than that increases

(positive bars). This is in agreement with the studies of Campbell et al. (2001), even though they refer to the first breakage. Therefore, for the processed feed, this is the threshold particle diameter for which a change in particle amount can be observed in response to a change on the milling gap. Note that, since the first LV explains about 83 % of the output PSD variability, nearly all the PSD variability is due to the variability of the MG, meaning that the MG is the process parameter that most affects the milling operation in the second breakage. This result is consistent with the findings of Mateos-Salvador et al. (2013).

The second LV (Figure 5.4c) is dominated by the MC. MC is positively correlated with particles that are bigger than 1081 µm (bin no. 28; negative bars in Figure 5.4d). This implies that when MC increases, also the amount of these particles increases. This behavior is explained by the fact that particles bigger than 1081 µm can be identified as bran, and bran breakage behavior is highly dependent on MC: a higher MC has the effect of hardening the bran, with the consequence that it breaks in larger particles (Campbell, 2007).

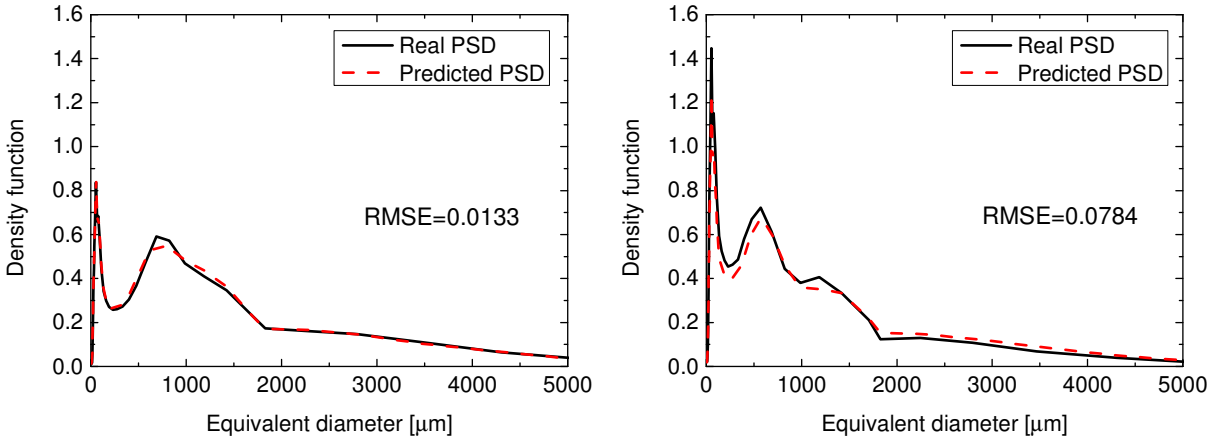
A threshold diameter range can be observed around 751–1081 µm (bins no. 25–28), where the PSD shows an inversion of correlation with MG and MC (namely, the relevant loadings change sign). The proposed physical explanation of this evidence is the difference in the shape of the milled particles, as recognized in common industrial practice. Particles bigger than 1081 µm (bin no. 28) are identified as bran and are characterized as flat large particles (two-dimensional particles); smaller particles are made of endosperm and have a more spherical shape (three-dimensional particles). Therefore, the difference in shape can explain why the PSD bins behave differently when this process parameter is varied.

The loading plots for the third LV are not reported because, even though this LV is explaining a large amount of data variability for matrix $\mathbf{X}_{B2}^{r}$ (33.5 %, almost entirely due to the variability of MF), it explains a very low percentage of variance for $\mathbf{Y}_{B2}^{r}$ (only 1.5 %), meaning that the output PSD is hardly correlated to MF.

5.3.1.2 Estimation of the particle size distribution

The predictive performance of the roller mill models is now evaluated. The idea is to predict the product quality (i.e., the PSD at the mill exit) once the process parameters (MF and MG) and wheat properties (MC) are known for a given feed material. Also in this case a jackknife approach was used.

The best and the worst PSD prediction performance obtained by the B2 roller mill model are shown in Figure 5.5a and Figure 5.5b, respectively. The average performance index is $\mathrm{RMSE}_{B2}^{r} = 0.0338$.



(a) (b)

Figure 5.5. *Prediction of the output PSD from the B2 roller mill: (a) best prediction and (b) worst prediction.*

5.3.2 Plansifter models

The objective of plansifter modeling is to predict the PSDs and the amount of all the sieved fractions, once the PSD of the material entering a plansifter is known.

Two distinct models are developed: one ("PSD plansifter model") has the PSDs of the sieved fractions as predicted outputs, the other one ("mass plansifter model") has the mass fractions of the sieved fractions as predicted outputs. The predictive performance of the models is analyzed using a jackknife approach.

5.3.2.1 PSD plansifter model

The PSD plansifter model relates the PSD of particles exiting from the roller mill (collected in $\mathbf{Y}_{B2}^{r}$ for the B2 passage) to the PSD of each sieved fraction (collected in $\underline{\mathbf{Y}}_{B2}^{p}$). Different models for different fractions were developed. Namely, for the B2 passage 6 different sieving models were built, each one relating the PSD of the material milled in the roller mill with the output PSD of a single sieved fraction (i.e., to a vertical "slice" of matrix $\underline{\mathbf{Y}}_{B2}^{p}$). The number of LVs to be retained in the models was chosen using the scree test. For all models, the optimal number of LVs was found to be 3. In the B2 passage models, the explained variance of the response variables ranged from a minimum of 46 % to a maximum of 72 %.

Figure 5.6 shows an example of prediction of the PSDs of the sieved fractions, where the predicted PSD (red dashed line) is compared to the real one (black solid line). For all sieves, the predictions are quite accurate, as can be seen from the RMSE indicated in Table 5.1.

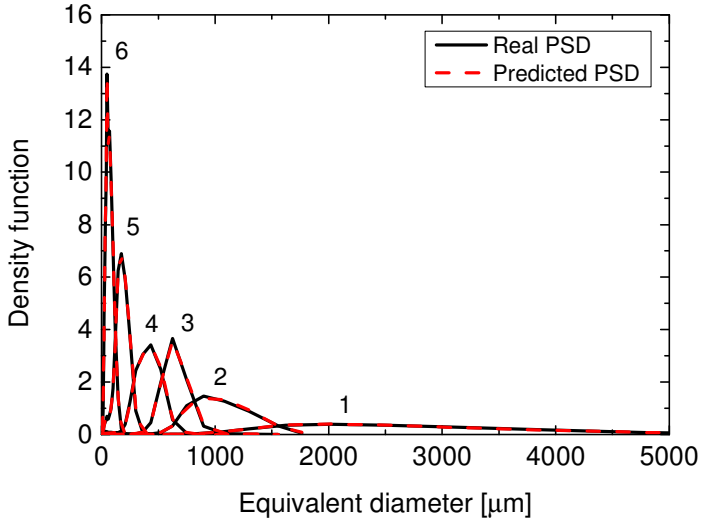


Figure 5.6. *Representative examples of prediction of sieved fraction PSDs for the B2 passage.*

Notice that the prediction error tends to increase moving from the first (coarser) fraction to the last (finer) one. This is due to the fact that, once the sieved fraction gets finer, the PSD becomes more peaked, making the PSD itself more difficult to predict (this is confirmed by the results averaged on the iterations of a jackknife approach, reported in Table 5.2).

Table 5.1. *RMSEs of the example reported in Figure 5.6.*

Sieve number	$RMSE^{p}_{B2}$
1	0.0048
2	0.0182
3	0.0228
4	0.0235
5	0.0332
6	0.1013

Table 5.2. *Averaged RMSEs on the jackknife approach of the B2 PSD plansifter models.*

Sieve number	$RMSE^{p}_{B2}$
1	0.0080
2	0.0320
3	0.0303
4	0.0301
5	0.0541
6	0.1413

5.3.2.2 Mass plansifter model

The mass plansifter model relates the PSD exiting from the roller mill with the mass fractions of the sieved fractions (which are collected in $\mathbf{Y}'^{p}_{B2}$ for the B2 passage). Namely, the objective is predicting how the material distributes through a sieve stack, depending on the PSD of the

material feeding the stack. Three LVs were able to describe 99 % of the $\mathbf{Y}^{'p}_{B2}$ variability. Prediction errors are reported in terms of mean absolute error (MAE) for a specific fraction, that is defined as:

$$\mathrm{MAE}_s = \frac{\sum_{i=1}^{I} \left| w_{i,s} - \hat{w}_{i,s} \right|}{I} \quad , \tag{5.2}$$

where $w_{i,s}$ is the measured mass fraction of fraction *s*, and $\hat{w}_{i,s}$ is its predicted value. The mass fractions can be predicted with good accuracy: the biggest MAE values were found to be 0.006 (on an average fraction of 0.137) for the B2 passage. On average, the errors on the estimated mass fractions never exceeded 6 %, which is conveniently smaller than a 10 % error that is typically acceptable in the industrial practice.

5.4 First reduction passage modeling

5.4.1 Roller mill model

The roller mill model for the R1 passage is calibrated on $\mathbf{X}^{r}_{R1}$ and $\mathbf{Y}^{r}_{R1}$. The number of LVs to be retained in the model is also in this case determined by the scree test. In the R1 roller mill model 3 LVs were used, which explain 72 % of the variability of $\mathbf{Y}^{r}_{R1}$.

5.4.1.2 Process understanding

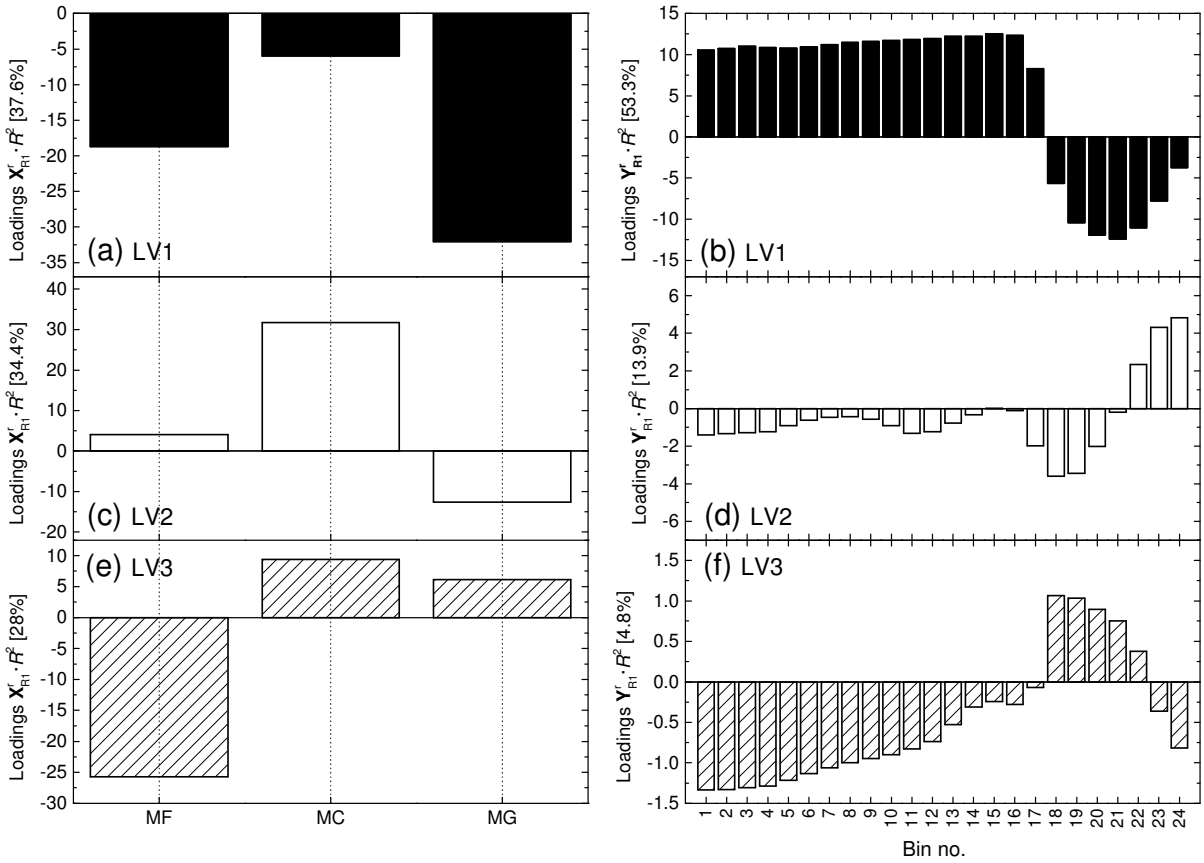


Figure 5.7. *Loading plots of R1 roller mill model: (a) loadings for the first LV of matrix* $\mathbf{X}^r_{R1}$*, (b) loadings for the first LV of matrix* $\mathbf{Y}^r_{R1}$*, (c) loadings for the second LV of matrix* $\mathbf{X}^r_{R1}$*, (d) loadings for the second LV of matrix* $\mathbf{Y}^r_{R1}$*, (e) loadings for the third LV of matrix* $\mathbf{X}^r_{R1}$ *and (f) loadings for the third LV of matrix* $\mathbf{Y}^r_{R1}$*.*

The **P** loadings plots and **Q** loadings plots for the roller mill model in the R1 passage are reported in Figure 5.7. Also in this case each loading is weighted for the explained variance of the relevant LV.

The first LV is dominated by the MG (Figure 5.7a). Again, the MG is the most important process parameter (it explains, with a minor contribution of MF on the first LV, more than half of the output PSD variability). When the MG is decreased, the amount of "big" particles (namely, those with a diameter larger than 210 µm, bin no. 18) decreases as well, whereas the number of "small" particles increases (Figure 5.7b).

The second LV is dominated by the MC (Figure 5.7c). In Figure 5.7d it can be seen that MC has quite a strong positive correlation with particles whose diameters lay in the range 326–626 µm (bins no. 22–24). When the MC increases, the amount of these particles increases as well. Physically, this behavior can be explained considering that a higher MC facilitates the agglomeration of endosperm material, with the result that more "big" particles are formed (Matz, 1991; Domian and Poszytek, 2005).

The third LV is dominated by MF (Figure 5.7e). MF is positively correlated with "small" particles (i.e., with a diameter between 0 and 59 µm; bins 1–11 in Figure 5.7f): when MF increases, the amount of those particles increases. This behavior finds a physical explanation in the fact that a higher MF involves a higher compression force between the particles themselves and between the rolls and the particles, and therefore a larger number of "small" particles is produced during milling.

5.4.1.3 Estimation of particle size distribution

The best and the worst predictions performance are shown in Figure 5.8a and Figure 5.8b (respectively) for the R1 passage: again, the predictions are quite accurate and the average RMSE is $\mathrm{RMSE}^{r}_{R1} = 0.2308$.

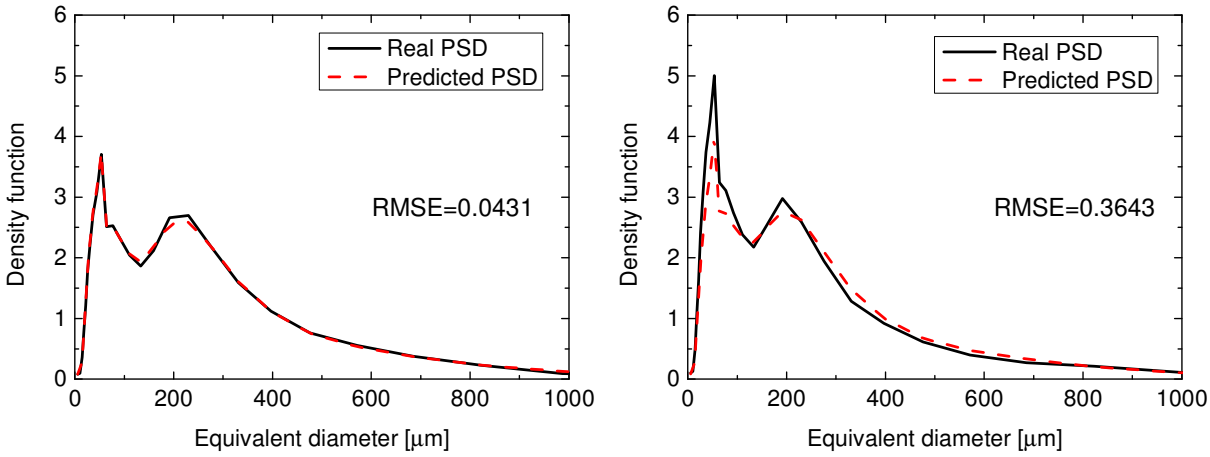


Figure 5.8. *Prediction of the output PSD from the R1 roller mill: (a) best prediction and (b) worst prediction.*

5.4.2 Plansifter model

5.4.2.1 PSD plansifter model

The R1 passage PSD plansifter model is built on $\mathbf{Y}^{r}_{R1}$ (that collect the PSDs exiting from the roller mill) and $\underline{\mathbf{Y}}^{p}_{R1}$ (that contains each PSD of the sieved fractions). For the R1 passage, 5 different sieving models were developed. 3 LVs were selected to build all the PSD plansifter models, which were able to explain from 66 % to 95 % of the response variables variability. Figure 5.9 shows an example of prediction of the PSDs of the sieved fractions, where the predicted PSD (red dashed line) is compared to the real one (black solid line). The predictions are quite accurate, as can be seen from the RMSE indicated in Table 5.3. Results averaged on the iterations of the jackknife approach are shown in Table 5.4.

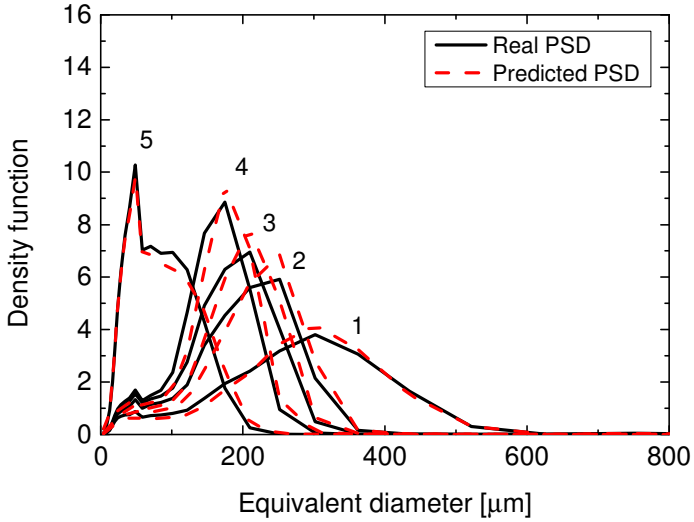


Figure 5.9. *Representative examples of prediction of sieved fraction PSDs.*

Table 5.3. *RMSEs of the examples reported in Figure 5.9.*

Sieve number	$RMSE^{p}_{R1}$
1	0.1576
2	0.3939
3	0.4906
4	0.5989
5	0.2997

Table 5.4. *Averaged RMSEs on the jackknife approach of the R1 PSD plansifter model.*

Sieve number	$RMSE^{p}_{R1}$
1	0.1395
2	0.2918
3	0.3510
4	0.3932
5	0.2492

5.4.2.2 Mass plansifter model

The mass plansifter model relates the PSD exiting from the roller mill with the mass fractions of the sieved fractions (which are collected in $\mathbf{Y}'^{p}_{R1}$). Three LVs were able to describe 93 % of $\mathbf{Y}'^{p}_{R1}$ variability. The mass fractions can be predicted with good accuracy: the biggest MAE values were 0.019 (on an average fraction of 0.373) for the R1 passage.

5.5 Concatenating the roller mill and plansifter models

As a first step toward the simulation of an entire industrial milling process, we now consider the concatenation of a roller mill model with a PSD plansifter model, both models belonging

to the same passage (B2 or R1). Specifically, we assess how the roller mill prediction error impacts on the prediction accuracy for the sieved fractions.

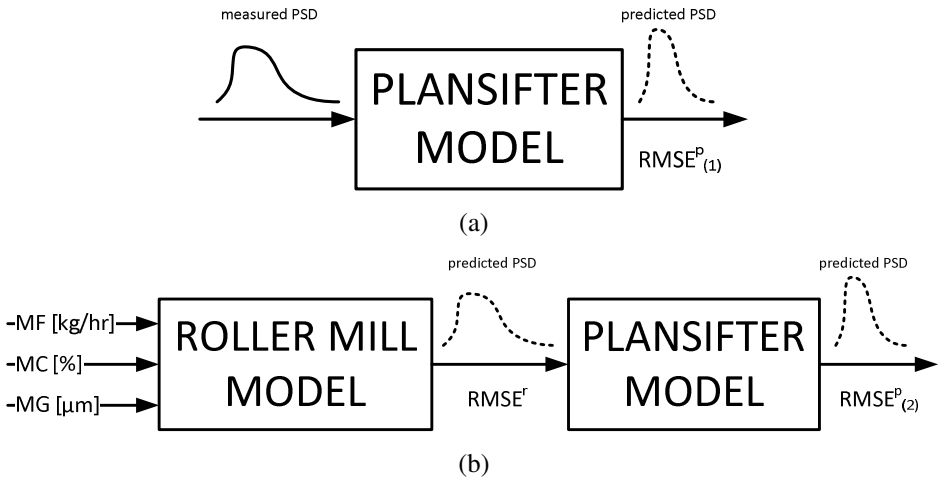


Figure 5.10. *Rationale followed in combining the models for a sequence of milling and sieving operations: (a) Scenario 1 and (b) Scenario 2.*

As illustrated in Figure 5.10, two different scenarios are investigated:

- Scenario 1 (subscript $_{(1)}$; Figure 5.10a): a plansifter model for each sieve is built using the measured (inlet and outlet) PSD data to calibrate it. To test the model performance, the PSD of each sieved fraction is predicted from the PSD measured at the roller mill outlet.
- Scenario 2 (subscript $_{(2)}$; Figure 5.10b): the plansifter model is the same as in Scenario 1, but to validate it, the PSD predicted by the roller mill model is used as the plansifter model input.

For both scenarios, the plansifter RMSE is calculated ($RMSE^{p}_{(1)}$ and $RMSE^{p}_{(2)}$, respectively). Notice that, while $RMSE^{p}_{(1)}$ accounts for the plansifter model prediction errors only, $RMSE^{p}_{(2)}$ accounts for the errors of both the plansifter model and the roller mill model. Hence, Scenario 2 allows one to evaluate how $RMSE^{r}$ impacts on the plansifter model performance, an issue that should be investigated whenever the unit operation models are concatenated to obtain the model for an entire wheat milling process.

For the B2 passage, Table 5.5 reports the prediction results averaged across all the jackknife iterations for both scenarios. Whereas the prediction accuracy achieved in Scenario 2 is generally worse than that in Scenario 1, the RMSE values for all sieved fractions are comparable. Therefore, we conclude that the sensitivity of the plansifter model to the performance of the roller mill model is limited, making the plansifter model robust to input changes. This in turn means that, as long as the performance of the roller mill model is adequate ($RMSE^{r}_{B2} = 0.0338$ in this case), the design of the plansifter model can be done independently of that of the roller mill model, even if the two models are eventually concatenated.

Table 5.5. *RMSEs for error propagation analysis on the B2 passage.*

$RMSE^{r}_{B2}$	**Sieve number**	$RMSE^{p}_{(1)}$	$RMSE^{p}_{(2)}$
0.0338	1	0.0080	0.0108
	2	0.0320	0.0363
	3	0.0303	0.0330
	4	0.0301	0.0362
	5	0.0541	0.0591
	6	0.1413	0.1243

Results for the R1 passage are reported in Table 5.6. The difference between $RMSE^{p}_{(2)}$ and $RMSE^{p}_{(1)}$ is larger than in the B2 passage, due to a greater $RMSE^{r}_{R1}$ (0.2308). While the loss of performance is marginal for the first three (coarser) sieved fractions, it is more significant for the finer fractions. However, even though a greater error is observed for the finer fractions, all the RMSEs are of the same order of magnitude. Therefore, also in this case, it would be possible to decouple the modeling of the roller mill and that of the plansifter.

Table 5.6. *RMSEs for error propagation analysis of the R1 passage.*

$RMSE^{r}_{R1}$	**Sieve number**	$RMSE^{p}_{(1)}$	$RMSE^{p}_{(2)}$
0.2308	1	0.1395	0.1911
	2	0.2918	0.3279
	3	0.3510	0.4639
	4	0.3932	0.6814
	5	0.2494	0.6100

5.6 A Simulink simulation of roller mill and plansifter models

All the developed models, both for roller mills and plansifters, were implemented in a Simulink® (MathWorks Inc., Massachusetts, U.S.A.) library. The objective is having an easy and intuitive way to use the models to predict the output PSD from either the roller mills and plansifters.

The whole library with all the available models and all the utility blocks are shown in Figure 5.11. Three type of blocks can be identified:

- utility blocks: these blocks are used to give inputs to the simulation and to manage the results. They comprise:
  - `Wheat Properties` block: it is used to define the moisture content of the wheat kernel;
  - `Ambient Conditions` block: useful to give as input the values of the ambient humidity and ambient temperature;
  - `Process Parameters` block: gives to the simulation the values of the milling gap and of the feed mass flow;

- `Save variable to Workspace` block: this block saves the desired response variables to the Matlab® (MathWorks Inc., Massachusetts, U.S.A.) workspace with the specified name;
- `Plot` block: it plots the response variable specified in y-axis with the x-axis defined in x-axis;
- `Read PSD from .txt file` block: this block is useful to read a PSD from an external text file (whose name is specified in the block interface);

- roller mill blocks: they contain the data-driven roller mill models (for example, the model shown in Section 5.3.1). Nine different roller mill models, for nine different passages, are available;
- plansifter blocks: they represent the data-driven plansifter models, as the one showed in Section 5.3.2. Seven passages are available for simulation.

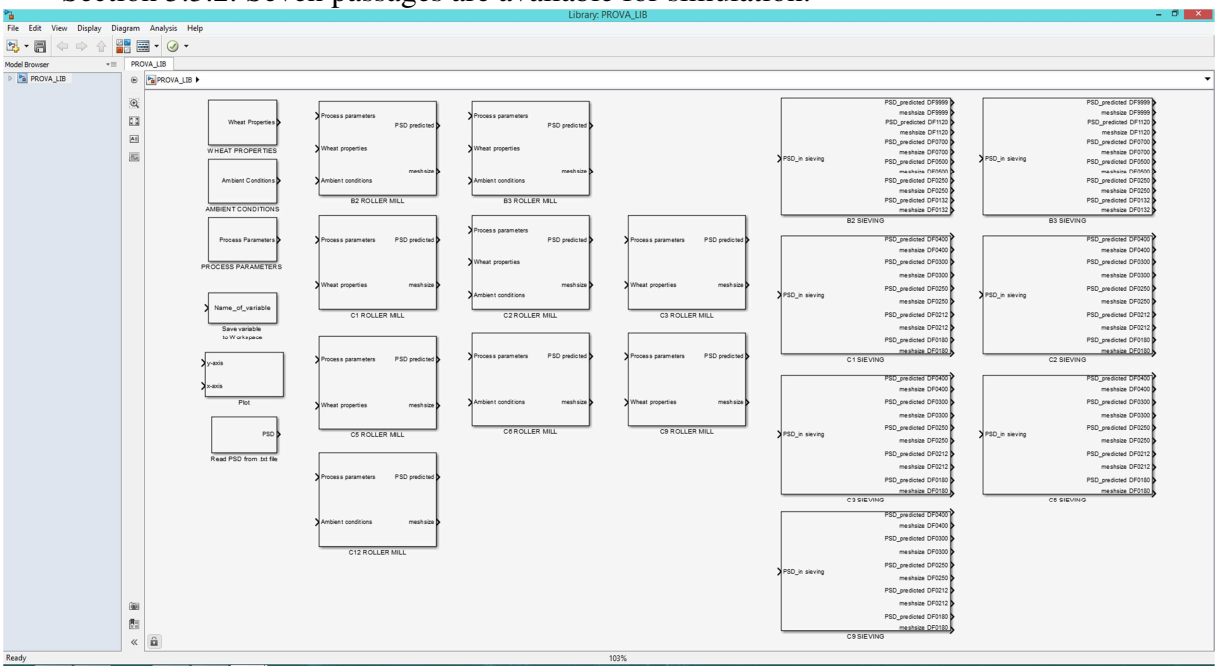


Figure 5.11. *Simulink library of roller mills and plansifter models.*

An example of how the library can be used to simulate a passage of the process is shown in Figure 5.12. The example reported here concerns the second-breakage passage. The blocks `Process Parameters`, `Wheat Properties` and `Ambient Conditions` are used to give the values of the input parameters (milling gap and mass flow in the `Process Parameters` block, wheat moisture content in the `Wheat Properties` block and ambient humidity and ambient temperature in the `Ambient Conditions` block). The interface to insert the input parameters is shown in Figure 5.13, with the example of the `Process Parameters` block, in which the values of the milling gap and of the mass flow can be inserted. The PSD is predicted in the `B2 ROLLER MILL` block and it is both plotted (using the `Plot` block) and saved in the workspace (with the name `PSD_roller_mill`, using the `Save variable to Workspace` block). The

predicted PSD is then used as input to the `B2 SIEVING` block and the PSD of the coarsest fraction is then saved to the Matlab® workspace, with the name `PSD_sieving_DF9999` (DF9999 refers to the mesh size of the sieve).

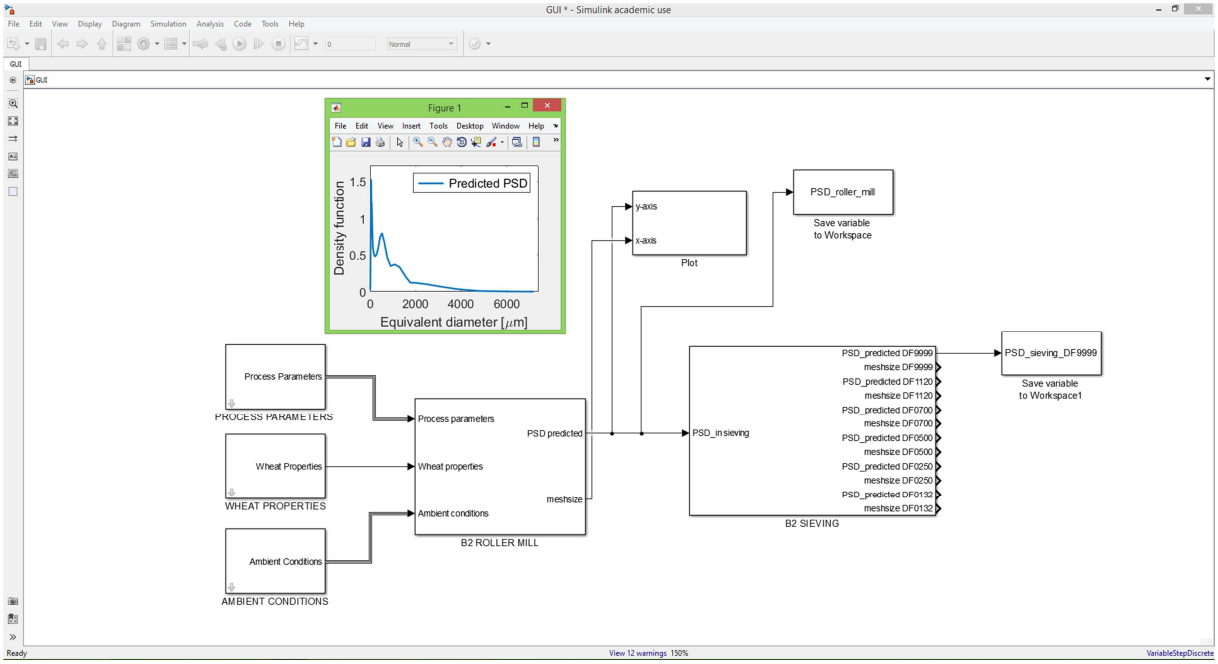


Figure 5.12. *Example of simulation of the second-breakage passage of the process. The PSD from the roller mill is both plotted and saved in the workspace, whereas the PSD of the coarsest sieved fraction is saved to the workspace only.*

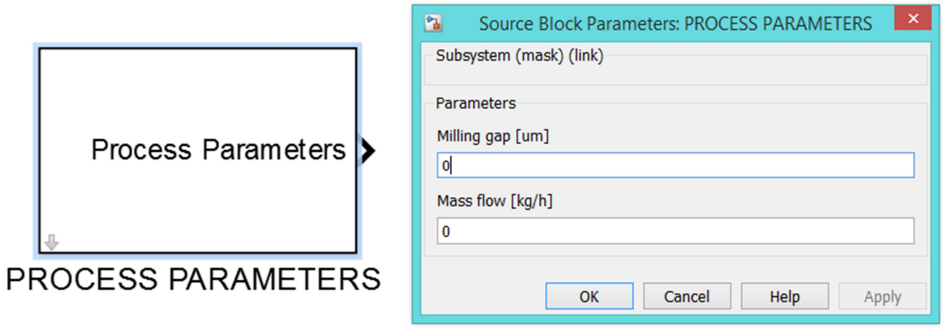


Figure 5.13. *Interface to insert the values of the process parameters in the `Process Parameters` block.*

The unit operations models can be used jointly to simulate a passage, as in the example showed in Figure 5.12, but also as standalone blocks to simulate single units (i.e. the roller mill or the plansifter only). An example of the simulation of the second-breakage plansifter is showed in Figure 5.14. The PSD to be used as input to the `B2 SIEVING` block is read from an external text file, using the `Read PSD from .txt file`. The `B2 SIEVING` block predicts all the PSDs of the sieved fractions, of which the PSD of the finest one is both plotted and saved into the workspace with the name `PSD_sieving_DF0132` (also in this case DF0132 is a reference to the sieve mesh size).

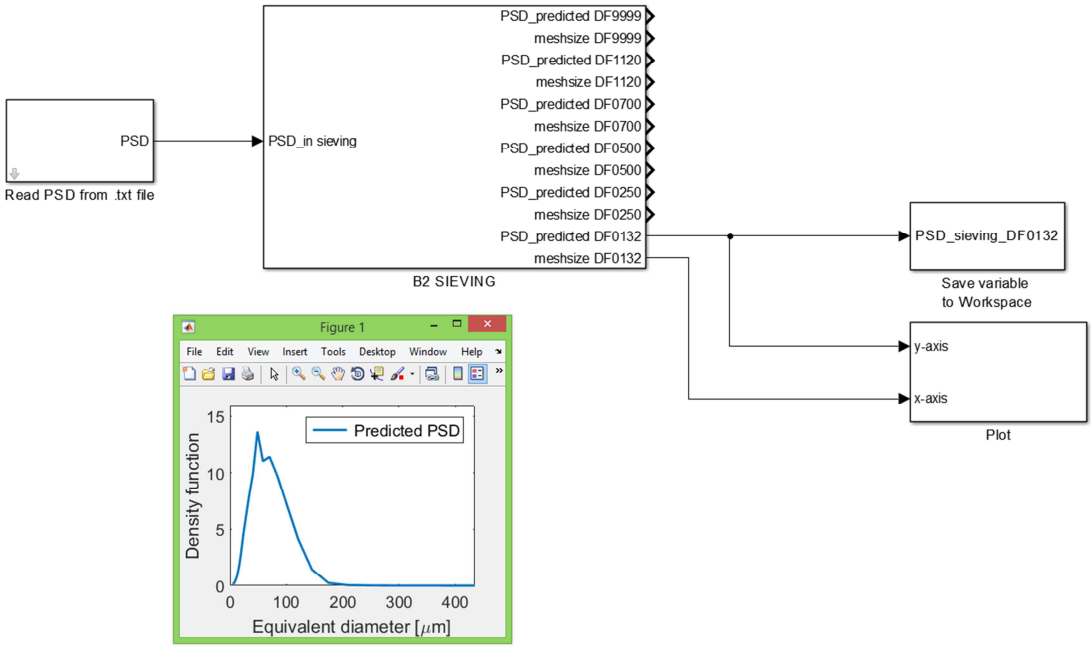


Figure 5.14. *Example of simulation of the second-breakage plansifter.*

The simulations are pretty fast: on average, only 30 seconds are required to simulate a passage of the process (as the example of Figure 5.12) and just 5 seconds to simulate a single unit operation (such as in Figure 5.14), both from cold start (computer specifications: 8 GB, 2.4 GHz, 64 bit, Intel i-7, Windows 8.1). The Simulink® simulation here presented can be of great support of the wheat milling operation: for example, if an operator wants to get an idea about the PSD after a specific roller mill, costly and time-expensive experiments can be avoided and the simulator can estimate the PSD. Therefore, the advantages of the simulation are: easy to handle models, no costs for additional experiments and fast estimation of the PSD.

5.7 Conclusions

In this Chapter it has been demonstrated how it is possible to use data-driven models of single unit operations to assist the wheat milling operation. The examples of the first-breakage, second-breakage and first-reduction passages have been shown and discussed. The roller mill models have been used:

- to classify wheat varieties in a cheap, fast and non-destructive way, using NIRS: NIR spectra have the great potential to discriminate not only different wheat varieties, but also different moisture contents of the wheat kernel; three different varieties (a soft, semi-hard and a hard one) and two moisture contents were discriminated with very good results. It has been shown how semi-hard and hard varieties are more similar respect to the soft one;

additionally, the effect of moisture content is not linear and it depends on the wheat variety;

- to improve the process understanding of the milling operation: loading analysis allowed to study and quantify the effect of process parameters (milling gap and mass flow), wheat properties (moisture content of the wheat) and wheat characteristics (described by the NIR spectra) on the product quality (PSD of the milled material). This analysis improved the general comprehension of the breakage mechanism of wheat; in particular, the milling gap was found to be the most influential process parameter on the PSD of the milled material. Also a significant impact of the wheat moisture content was noticed.
- to predict the product quality: the PSD of the milled wheat is predicted with good accuracy given the values of the milling gap, mass flow, wheat moisture content and NIR spectra describing the wheat type. This kind of prediction can be used to avoid expensive and time-consuming experiments.

The plansifter models have been used to predict both the PSDs and the amount of each sieved fractions. Prediction errors never exceed 10 %, which is an acceptable error in the industrial practice.

Furthermore, the roller mill models and plansifter models were then connected, in order to assess how the prediction errors in the roller mill model affect the subsequent sieving model. The plansifter models were shown to be quite robust to the input accuracy, thus suggesting that, even in a data-driven approach like the one considered in this Dissertation, the development of the roller mill models can be decoupled from that of the plansifter models.

On a long-term perspective, the developed models would act as the core of an intelligent grain processing system that provides a science-driven (rather than experience-driven) platform to assist several operations, such as optimal selection of processing conditions, early detection of faults in real-time processing, real-time product quality prediction, and operator training.

Chapter 6

A model-based approach to the design and scale-up of wheat milling operations- proof of concept

In this Chapter, the problem of determining the operating conditions of industrial-scale wheat milling processes is addressed. Namely, a proof-of-concept study is presented where the problem is estimating the milling gap to be used in an industrial-scale roller mill in order to obtain a product with a desired mass distribution from a specific wheat variety. The framework is proposed with the intent of minimizing the number of experimental trials that need to be carried out in the industrial-scale plant to find the appropriate operating conditions for the industrial plant under the required quality targets.

6.1 Introduction

In the operation of an industrial wheat milling plant, two different problems often need to be faced. When the plant is to be installed and operated for the first time, very limited knowledge about its operation is available, often deriving from a preliminary experimental campaign carried out in a smaller (e.g., laboratory or pilot) scale equipment. The question therefore arises about whether the small-scale plant data can be used to find the conditions at which the industrial-scale plant should be operated to obtain a given product from an assigned wheat variety. In this Chapter, we will refer to this problem as to a process scale-up problem.

On the other side, if the industrial-scale plant has long been in operation, it may be required to determine the operating conditions that are needed to manufacture a product with a quality profile that is different from those already obtained, or to manufacture a given product from a wheat variety that is different from those used in previous production campaigns. This second problem, which is in essence similar to the scale-up one, will be called a process design problem.

In this Chapter we consider different alternatives to address the process scale-up or process design problem. Given the proof-of-concept of the study, the analysis is not extended to an entire milling plant, but is limited to a roller mill. Namely, the objective is estimating the

milling gap to be used in an industrial-scale roller mill to obtain a product with a desired mass distribution from an assigned wheat variety.

From a practical standpoint, the scale-up problem, as well as the process design one, can be tackled by carrying out an extensive experimental campaign on the industrial-scale plant and estimating the new operating conditions based on the experiments and on the miller's experience. However, besides being time consuming, this trial-and-error approach may be very expensive due to the cost of the industrial experiments and to the potential off-specification of the resulting flour product. To just quote some figures, the typical cost of a single experimental run in an industrial roller mill is on the order of 700 € (Molino Quaglia, 2016), whereas it is about four times smaller (i.e., ~174 €/ run) in a pilot-scale mill (Bühler AG, 2016; Figure 6.1). Therefore, developing a process scale-up methodology that allows one to reduce the number of experiments that need to be carried out at the industrial level would be greatly beneficial.

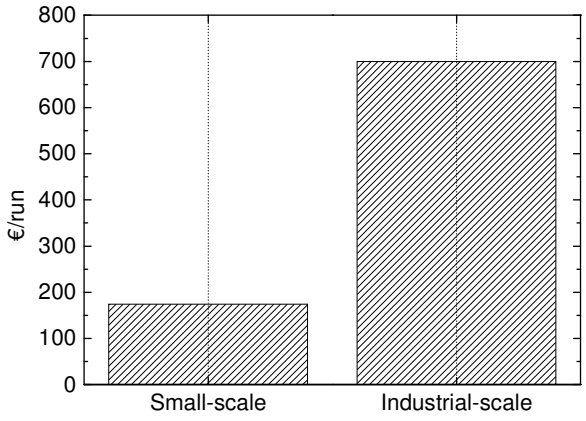


Figure 6.1. *Overall costs per experimental run in a small-scale and in an industrial-scale roller mill.*

The basic idea we explore in this Chapter is to move most of the experimentation from the industrial scale to the small scale. Particularly, we assess the possibility to reduce the coverage of the industrial-scale experimental campaign by creating an extended dataset at the small-scale level that can be used to complement the industrial-scale one. Under this perspective, the small-scale dataset would act as a sort of "universal" dataset, which is collected once and for all *by the plant manufacturer*, and that can then be used to support the scale-up exercise on *any* industrial-scale plant produced by that manufacturer. Notice that the advantage of this approach would be twofold: on the one side, the cost of experimental trials at the industrial site (the customer) would be significantly reduced; on the other side, the commissioning of the industrial plant would be significantly accelerated, even in the case of limited or no experience of the customer on a particular plant or product. To the best of the

author's knowledge, so far no attempt has been done to investigate the scale-up problem for a wheat roller mill under this perspective.

The use of historical data to assist product and process design was first reported by Moteki and Arai (1986), who used PCA to analyze historical data from a low-density polyethylene process and to infer process conditions for new grades of products. Other authors proposed the use of expert system tools (such as fuzzy logic or artificial neural networks) to model the relations between input variables and product properties and use them to estimate the inputs corresponding to new product characteristics (Borosy, 1999; Sebzalli and Wang, 2001).

Jaeckle and MacGregor (1998 and 2000a) developed a methodology based on the inversion of latent variable models to estimate the required process conditions for a plant to yield a desired set of properties in the final product. Tomba et al. (2012) proposed a general methodology to design processes and products using a constrained optimization framework to invert a latent-variable model (see Chapter 3 for a short review on this framework).

Jaeckle and MacGregor (2000b) pioneered the use of process historical database to transfer products and processes between different plants (Jaeckle and MacGregor, 1998; Jaeckle and MacGregor, 2000a). They related the process variable datasets available in different process scales/sites through the common products manufactured in different scales/sites, and used latent-variable regression model inversion to estimate the process conditions for a target plant to manufacture a new product of assigned properties. This methodology was then used in several studies. Zhou and Titchener-Hooker (1999) determined operational windows in a bioprocess by plotting pairs of process variables. Chen and Wang (2000) used the same approach with historical data from a fluid catalytic cracking unit, but with multivariate operational regions defined by the scores of a PCA model. Lakshminarayanan et al. (2000) used the Jaeckle and MacGregor (2000b) methodology introducing non-linearity by means of a neural network.

The methodology was further refined by Garcìa-Muñoz et al. (2005), which developed the joint-Y PLS (JY-PLS) modeling approach. JY-PLS is a latent variable regression modeling technique that allows relating datasets of regressors (e.g., raw materials properties and processing conditions) from different sources through the latent space generated by the response variables, which account for the product quality. The basic assumption of JY-PLS is that, if similar products are manufactured in different plants (or equipment) exploiting the same chemical and physical process, the product properties should share a common correlation structure, represented by their latent variables. Assuming that the regressors are correlated with the product quality within each plant (within-plant correlation), the latent spaces of the regressor datasets will span a common region, represented by the latent space of the product properties (or a subset of it). This latent space can therefore be used to relate regressor datasets from different sources (between-plant correlation). Details about the methodology have been provided in Chapter 3. Liu et al. (2011) used JY-PLS to support the

scale-up of a roller compaction process in a pharmaceutical product. Tomba et al. (2014) used the JY-PLS model approach to transfer a nanoparticle product between mixers of different geometry. An application of latent-variable modeling to support process scale-up has been reported by Muteki et al. (2011), who combined PLS modeling and inversion to estimate the end point for a high-shear wet granulation process in the target plant using data archived from past batches and drug lot properties.

Three different scale-up approaches, all based on latent-variable modeling, are investigated in this Chapter. The approaches differ for the coverage of the available industrial-scale dataset with respect to the operating conditions that would be needed to obtain a product with assigned quality.

6.2 The modeling framework

This section describes the methodology that is proposed to support the process design in an industrial-scale wheat roller mill. The main objective is estimating the milling gap to be used in an industrial-scale roller mill to manufacture a product with a desired mass distribution over a sieve from a specific wheat variety, while at the same time minimizing the experiments that need to be carried out in the industrial plant. To this purpose, three different modeling approaches are investigated: direct PLS modeling, PLS model inversion, and JY-PLS model inversion. The first two approaches use industrial-scale data only, whereas in the latter one the industrial-scale dataset is complemented with a dataset coming from a small-scale roller mill, which allows one to assess whether data integration can accelerate the scale-up exercise.

Data from two plants (small-scale and industrial-scale) are used, and the available datasets are described in Chapter 4. Namely, $\mathbf{X}^{r}_{B1}$ [24×953] and $\mathbf{Y}'^{r}_{B1}$ [24×2] collect the input and output data from a small-scale roller mill, whereas $\mathbf{X}^{r}_{B1,IS}$ [6×1052] and $\mathbf{Y}'^{r}_{B1,IS}$ [6×2] include industrial-scale roller mill data. The structure of the available datasets is summarized in Table 6.1 and Table 6.2 for the small-scale and industrial-scale plants, respectively. The small-scale dataset (Table 6.1) collects the values of feed mass flow MF, moisture content MC, milling gap MG and NIR spectrum (950 wavenumbers) describing the wheat variety for 24 observations. The response variable is the corresponding mass distribution on a 1120 µm sieve (described by w_1 and w_2, where w_1 is the mass fraction above the sieve and w_2 is the mass fraction below the sieve). The industrial-scale dataset (Table 6.2) is composed of 6 different observations, which are used for model calibration: the input variables are the milling gaps (MG_l and MG_r) and the wheat variety NIR spectrum; the output variables are the mass fraction above (w_1) and below (w_2) a 1120 µm sieve. Two milling gaps are used, since they can be set independently, as described in Chapter 4. Note that the measured variables and the dataset structure for the two plants are different.

The three different modeling approaches are investigated in two different situations: *i*) the full industrial-dataset of Table 6.2 is available (namely, all 6 observations are available to calibrate the models) and, *ii*) only a limited industrial dataset is available (namely, only some of the six observations are used to calibrate the models). The latter case is considered to mimic the situation where the desired product significantly differs, in terms of wheat variety and mass distribution, from those explored by the available industrial dataset.

Table 6.1. *Structure of the small-scale dataset.*

$\mathbf{X}^{r}_{B1}$ [24×953]		$\mathbf{Y}^{r}_{B1}$ [24×2]	
no.	**Name**	**no.**	**name**
1	MF [kg/hr]	1	w_1
2	MC [%]	2	w_2
3	MG [µm]		
from 4 to 953	wavenumbers of the NIR spectrum [cm^{-1}]		

Table 6.2. *Structure of the industrial-scale dataset.*

$\mathbf{X}^{r}_{B1,IS}$ [6×1052]		$\mathbf{Y}^{r}_{B1,IS}$ [6×2]	
no.	**name**	**no.**	**name**
1	MG_l [-]	1	w_1
2	MG_r [-]	2	w_2
from 3 to 1052	wavelengths of NIR spectrum [nm]		

6.2.1 The rationale behind the proposed approaches

Regardless of the modeling approach used, the objective is estimating the milling gap that should be used in an industrial-scale roller mill to obtain an assigned product distribution from a specific wheat variety, using data available from previous experimental campaigns.

Approach no.1 is a PLS model, as sketched in Figure 6.2. The model relates the wheat type and the desired product mass distribution (used as input variables) to the milling gaps (output variables), thus allowing one to directly estimate the milling gap that is required to obtain the desired product. Clearly, only industrial-scale data can be exploited to build the PLS model.

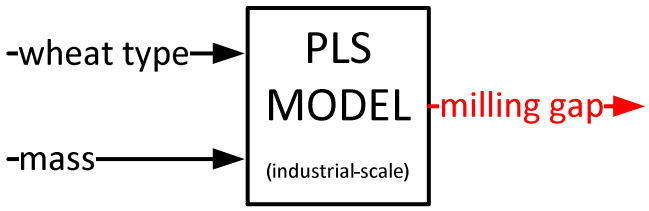


Figure 6.2. *Approach no.1 to solve the roller-mill scale-up problem by PLS modeling.*

Also Approach no.2 relies on PLS modeling. However, in this case the PLS model is built by relating the milling gap and the wheat type to the mass distribution (Figure 6.3a); the model is

then inverted to estimate the milling gap to be used to obtain the desired product, for a given wheat variety (Figure 6.3b). Also this approach can use industrial-scale data only to build the PLS model.

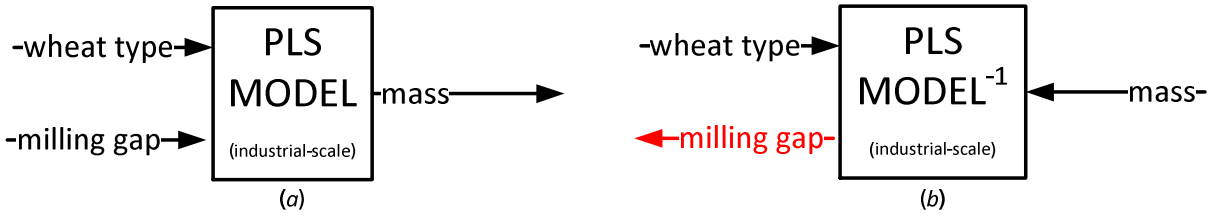


Figure 6.3. *Approach no.2 to solve the roller-mill scale-up problem: (a) PLS model calibration and (b) model inversion to estimate the milling gap.*

Approach no.3 is based on a JY-PLS model, which can exploit data from two (or more) different plants simultaneously. According to this approach, a JY-PLS model is first built to relate the wheat type and milling gap to the product mass distribution in the small-scale and large-scale plants simultaneously (Figure 6.4a); the model is then inverted as shown in Figure 6.4b to estimate the milling gap in the industrial-scale plant given the desired mass distribution in the same plant.

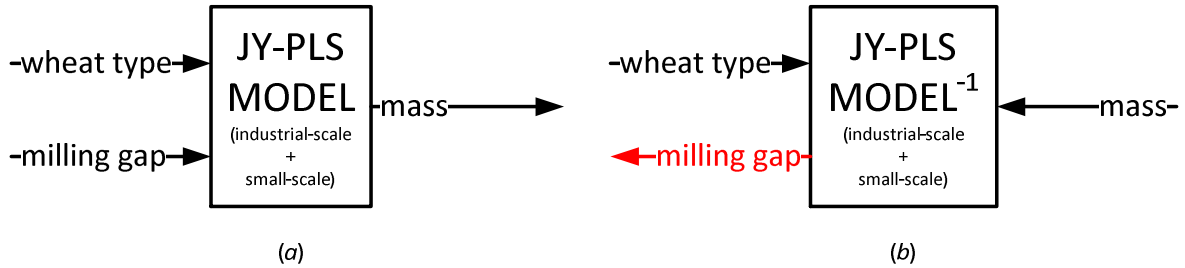


Figure 6.4. *Approach no.3 to solve the roller-mill scale-up problem: (a) joint-Y PLS model calibration, and (b) model inversion to estimate the milling gap.*

This approach allows one to assess whether data coming from an experimental campaign carried out in a small-scale plant can effectively complement an industrial-scale dataset when the industrial-scale plant milling gap is to be determined. Note that the small-scale and the industrial-scale datasets will generally have a different structure. The JY-PLS modeling approach can easily handle different dataset structures, provided that the responses of all datasets lie in the same space.

6.2.2 Using PCA scores as a proxy of NIR spectra

The input matrix $\mathbf{X}$ to a model can be split into two submatrices: submatrix $\mathbf{X}^{\text{process}}$ including the process operating parameters, and submatrix $\mathbf{X}^{\text{NIR}}$ containing the feed material NIR spectra; hence, $\mathbf{X} = [\mathbf{X}^{\text{process}}, \mathbf{X}^{\text{NIR}}]$. The $\mathbf{X}$ matrices for the small-scale and industrial-scale plants are denoted with $\mathbf{X}^{\text{r}}_{\text{B1}} = [\mathbf{X}^{\text{r,process}}_{\text{B1}}, \mathbf{X}^{\text{r,NIR}}_{\text{B1}}]$ and $\mathbf{X}^{\text{r}}_{\text{B1,IS}} = [\mathbf{X}^{\text{r,process}}_{\text{B1,IS}}, \mathbf{X}^{\text{r,NIR}}_{\text{B1,IS}}]$, respectively.

The ability of NIR spectra to discriminate different wheat varieties was already proven in Chapter 5 for the small-scale roller mill dataset. This ability was checked also for the industrial-scale plant. Figure 6.5 shows the score plot of the PCA model built on the NIR spectra (after standard normal variate pre-treatment; Rinnan et al., 2009) of the wheat used in the industrial-scale experimental campaign. The pattern is similar to that observed in the small-scale plant (compare with Figure 5.1), but with some differences. The hard and semi-hard varieties are closer in the score plot, due to the fact that blends of "pure" wheat varieties were used in the industrial-scale plant, and many pure varieties used to generate the hard and semi-hard blends were in common. Additionally, the variability observed in the second component is not due to the moisture content as in Figure 5.1 (the moisture content was constant in fact), but it is due to a variability of the wheat blend.

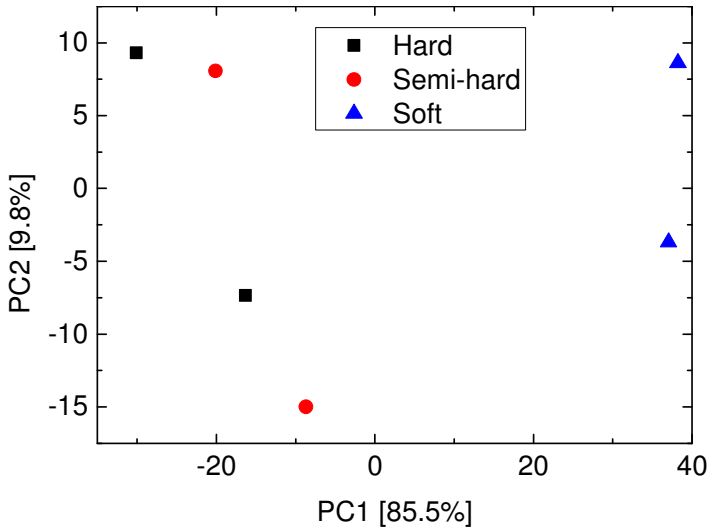


Figure 6.5. *Score plot of the PCA model built on NIR spectra collected in the industrial-scale plant.*

Two PCs are a very good representation of the NIR spectra variability, since they describe 95.3 % of the NIR spectra variance. Therefore, to reduce the number of inputs that need to be processed, $\mathbf{X}^{\text{NIR}}$ (whose column dimension is equal to the number of wavelengths) is compressed by PCA modeling, using two PCs at each plant scale. The corresponding score matrices will be denoted as $\mathbf{T}_{\text{B1}}^{\text{wheat}}$ and $\mathbf{T}_{\text{B1,IS}}^{\text{wheat}}$ for the small-scale and industrial scale plants, respectively. The PCA scores are used as a proxy of the whole spectra to describe the wheat variety. Therefore, for an assigned new product, the input matrices will be indicated by $\mathbf{X}_{\text{B1}}^{\text{r}} = [\mathbf{X}_{\text{B1}}^{\text{r,process}}, \mathbf{T}_{\text{B1}}^{\text{wheat}}]$ for the small-scale plant, and $\mathbf{X}_{\text{B1,IS}}^{\text{r}} = [\mathbf{X}_{\text{B1,IS}}^{\text{r,process}}, \mathbf{T}_{\text{B1,IS}}^{\text{wheat}}]$ for the industrial-scale plant. Note that the scores are also used to assign the wheat type in the constrained optimization problem in the framework of Eq. (3.31) (to be discussed later).

6.2.3 Approach no.1: direct PLS modeling

As sketched in Figure 6.6, the input variables of the PLS model are the mass distribution over a sieve (collected in $\mathbf{Y}^{'r}_{B1,IS}$) and the wheat type (described by $\mathbf{T}^{wheat}_{B1,IS}$) for the industrial scale-plant. The output variables are the milling gaps, collected in $\mathbf{X}^{r,process}_{B1,IS}$. Therefore, according to Approach no.1 the milling gap is estimated directly for the given wheat type and desired mass distribution of the product.

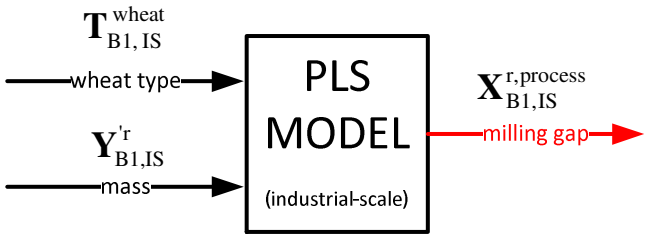


Figure 6.6. *Approach no.1: detailed view of model inputs and outputs.*

The model is thus formulated as:

$$[\mathbf{Y}^{'r}_{B1,IS}, \mathbf{T}^{wheat}_{B1,IS}] = \mathbf{TP}^{T} + \mathbf{E}_{\mathbf{X}} \tag{6.1}$$

$$\mathbf{X}^{r,process}_{B1,IS} = \mathbf{TQ}^{T} + \mathbf{E}_{\mathbf{Y}} \,. \tag{6.2}$$

6.2.4 Approach no.2: PLS model inversion

First, a PLS model is built for the industrial-scale plant by relating $\mathbf{X}^{'r}_{B1,IS}$ (i.e., the milling gap and wheat type) to $\mathbf{Y}^{'r}_{B1,IS}$ (the mass distribution over a 1120 µm), as shown in Figure 6.7a. It may be noticed that the way the model is built reflects how the roller mill works (namely, milling gaps and wheat type are the input variables, and the mass distribution is the response). Once the model is calibrated, it is then inverted to estimate the milling gap to be used to obtain a desired product, for a given wheat type (Figure 6.7b). To this purpose, a constrained optimization problem is formulated and solved as in (6.5).

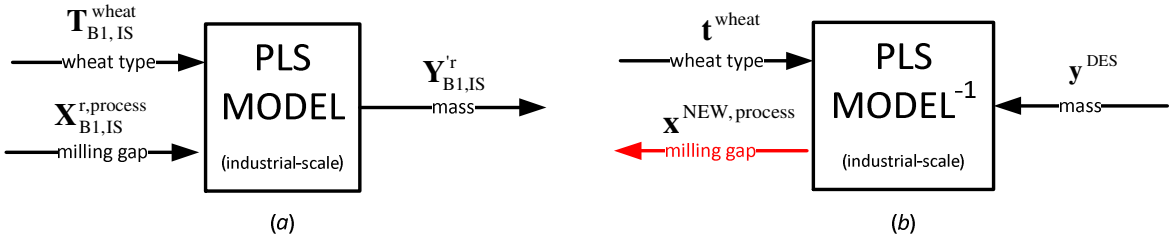


Figure 6.7. *Approach no.2: detailed view of model inputs and outputs.*

The model is formulated as follows:

$$\mathbf{X}_{\mathrm{B1,IS}}^{\mathrm{'r}} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E}_{\mathbf{X}} \tag{6.3}$$

$$\mathbf{Y}_{\mathrm{B1,IS}}^{\mathrm{'r}} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E}_{\mathbf{Y}} \tag{6.4}$$

To accomplish the model inversion, the optimization framework (3.31) is rewritten as:

$$\min_{\mathbf{x}^{\mathrm{NEW}}} \left\{ \left(\mathbf{y}^{\mathrm{DES}} - \hat{\mathbf{y}}^{\mathrm{DES}}\right)^{\mathrm{T}} \left(\mathbf{y}^{\mathrm{DES}} - \hat{\mathbf{y}}^{\mathrm{DES}}\right) + G_1 \left(\sum_{lv=1}^{LV} \left(\frac{\mathbf{x}^{\mathrm{NEW}} \mathbf{w}_{lv}}{\mathbf{s}_{lv}} \right)^2 \right) + G_2 \mathrm{SPE}_{\mathbf{X}} \right\} \tag{6.5}$$

subject to

$$\hat{\mathbf{y}}^{\mathrm{DES}} = \mathbf{x}^{\mathrm{NEW}} \mathbf{W}\mathbf{Q}^{\mathrm{T}} \tag{6.6}$$

$$\mathrm{SPE}_{\mathbf{X}} = \left(\mathbf{x}^{\mathrm{NEW}} - \mathbf{x}^{\mathrm{NEW}} \mathbf{W}\mathbf{P}^{\mathrm{T}}\right)^{\mathrm{T}} \left(\mathbf{x}^{\mathrm{NEW}} - \mathbf{x}^{\mathrm{NEW}} \mathbf{W}\mathbf{P}^{\mathrm{T}}\right) \tag{6.7}$$

$$\mathbf{t}^{\mathrm{new,wheat}} = \mathbf{t}^{\mathrm{wheat}} . \tag{6.8}$$

The different terms of the objective function were described in detail in Chapter 3. The set of input variables is $\mathbf{x}^{\mathrm{NEW}} = [\mathbf{x}^{\mathrm{NEW,process}}, \mathbf{t}^{\mathrm{new,wheat}}]$, where $\mathbf{x}^{\mathrm{NEW,process}}$ is the set of estimated process parameters that should be used to obtain the desired product, and $\mathbf{t}^{\mathrm{new,wheat}}$ are the scores describing the wheat variety for the new product to be obtained. The latter are assigned in Eq. (6.5) using the equality constraints $\mathbf{t}^{\mathrm{new,wheat}} = \mathbf{t}^{\mathrm{wheat}}$, where $\mathbf{t}^{\mathrm{wheat}}$ are the scores obtained by projecting the NIR spectrum onto the PCA model built on the NIR spectra of the calibration samples.

6.2.5 Approach no. 3: joint-Y PLS model inversion

The third approach is based on a JY-PLS model, which exploits data from different plant scales: a small-scale roller mill and an industrial-scale roller mill. The small-scale roller mill is described by $\mathbf{X}_{\mathrm{B1}}^{\mathrm{'r}}$ and $\mathbf{Y}_{\mathrm{B1}}^{\mathrm{'r}}$, whereas the industrial-scale plant uses $\mathbf{X}_{\mathrm{B1,IS}}^{\mathrm{'r}}$ as inputs and $\mathbf{Y}_{\mathrm{B1,IS}}^{\mathrm{'r}}$ as responses. Details about the matrices are reported in Chapter 4. Note that the industrial-scale and small-scale data cannot be simply merged into an augmented matrix to be used as input of a standard PLS model, since the measured input variables are different in the two plants.

The model is calibrated as shown in Figure 6.8a, and then inverted (as in Figure 6.8b) to estimate the milling gap, which is calculated according to framework (3.31), in which constraints on the PCA scores are used to assign the wheat variety.

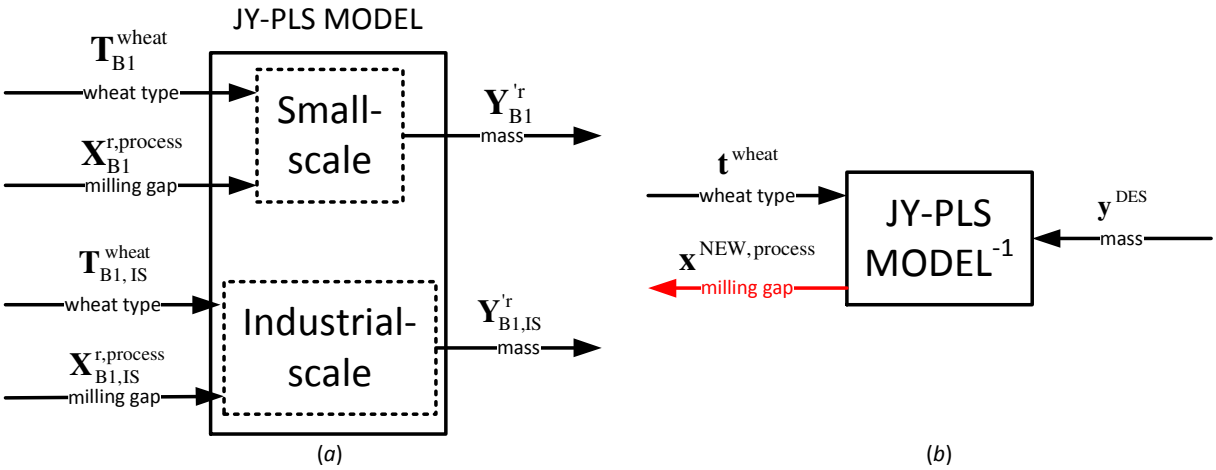


Figure 6.8. *Approach no.3: detailed view of model inputs and outputs.*

The model is formulated as:

$$\mathbf{Y}_{\mathrm{J}} = \begin{bmatrix} \mathbf{Y}_{\mathrm{B1}}^{\mathrm{'r}} \\ \mathbf{Y}_{\mathrm{B1,IS}}^{\mathrm{'r}} \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{\mathrm{A}} \\ \mathbf{T}_{\mathrm{B}} \end{bmatrix} \mathbf{Q}_{\mathrm{J}}^{\mathrm{T}} + \mathbf{E}_{\mathbf{Y}_{\mathrm{J}}} \tag{6.9}$$

$$\mathbf{X}_{\mathrm{B1}}^{\mathrm{'r}} = \mathbf{T}_{\mathrm{A}} \mathbf{P}_{\mathrm{A}}^{\mathrm{T}} + \mathbf{E}_{\mathbf{X}_{\mathrm{A}}} \tag{6.10}$$

$$\mathbf{X}_{\mathrm{B1,IS}}^{\mathrm{'r}} = \mathbf{T}_{\mathrm{B}} \mathbf{P}_{\mathrm{B}}^{\mathrm{T}} + \mathbf{E}_{\mathbf{X}_{\mathrm{B}}} \,. \tag{6.11}$$

The optimization framework (3.31) is formulated as:

$$\min_{\mathbf{x}^{\mathrm{NEW}}} \left\{ \left(\mathbf{y}^{\mathrm{DES}} - \hat{\mathbf{y}}^{\mathrm{DES}} \right)^{\mathrm{T}} \left(\mathbf{y}^{\mathrm{DES}} - \hat{\mathbf{y}}^{\mathrm{DES}} \right) + G_1 \left(\sum_{lv=1}^{LV} \left(\frac{\mathbf{x}^{\mathrm{NEW}} \mathbf{w}_{\mathrm{B},lv}}{\mathbf{s}_{lv}} \right)^2 \right) + G_2 \mathrm{SPE}_{\mathbf{X}} \right\} \tag{6.12}$$

subject to

$$\hat{\mathbf{y}}^{\mathrm{DES}} = \mathbf{x}^{\mathrm{NEW}} \mathbf{W}_{\mathrm{B}} \mathbf{Q}_{\mathrm{j}}^{\mathrm{T}} \tag{6.13}$$

$$\mathrm{SPE}_{\mathbf{X}} = \left(\mathbf{x}^{\mathrm{NEW}} - \mathbf{x}^{\mathrm{NEW}} \mathbf{W}_{\mathrm{B}} \mathbf{P}_{\mathrm{B}}^{\mathrm{T}} \right)^{\mathrm{T}} \left(\mathbf{x}^{\mathrm{NEW}} - \mathbf{x}^{\mathrm{NEW}} \mathbf{W}_{\mathrm{B}} \mathbf{P}_{\mathrm{B}}^{\mathrm{T}} \right) \tag{6.14}$$

$$\mathbf{t}^{\mathrm{new,wheat}} = \mathbf{t}^{\mathrm{wheat}} \,. \tag{6.15}$$

6.3 Milling gap estimation results

In this section, the performances of the three approaches in the process design problem are presented and critically compared. The objective is estimating the milling gap to be used in the industrial-scale roller mill to manufacture a product with a given mass distribution over a sieve from an assigned wheat variety. It should be recalled that *two* milling gaps need to be estimated, because in the industrial-scale plant two independent knobs are used to set the

milling gaps on either side of the rollers. Two different situations are analyzed: an extended dataset (namely, the full dataset presented in Chapter 4) is available for the industrial-scale plant, or only a limited dataset is available.

6.3.1 Models diagnostics

In Approach no.1, the PLS model is built using 2 LVs, which are able to describe 75.6 % of the input data variability and 99.3 % of the output variables variability.
In Approach no.2, the PLS model is also built using 2 LVs, which explain 76.6 % of the variability of $\mathbf{X}^{'r}_{B1,IS}$ and 99.4 % of the $\mathbf{Y}^{'r}_{B1,IS}$ variability.
The joint-Y PLS model (Approach no.3) is built using 2 LVs and the variability described by each LV is reported in Table 6.3. Two LVs are able to explain the entire variability of $\mathbf{X}^{'r}_{B1,IS}$, 56 % of the $\mathbf{X}^{'r}_{B1}$ variability, and almost the entire variability of $\mathbf{Y}^{'r}_{B1,IS}$ and of $\mathbf{Y}^{'r}_{B1}$. Interestingly, one LV alone is able to describe almost the entire variance of both response matrices, whereas the second LV explains a very significant amount of the $\mathbf{X}^{'r}_{B1}$ (small-scale) variability as well as a significant amount of the $\mathbf{X}^{'r}_{B1,IS}$ (industrial-scale) variability.

Table 6.3. *Explained variance for each LV of the joint-Y PLS model.*

	$R^2_{\mathbf{X}^{'r}_{B1,IS}}$ [%]	$R^2_{\mathbf{X}^{'r}_{B1}}$ [%]	$R^2_{\mathbf{Y}^{'r}_{B1,IS}}$ [%]	$R^2_{\mathbf{Y}^{'r}_{B1}}$ [%]
LV1	90.1	20	99.9	98.2
LV2	9.9	36	0	0

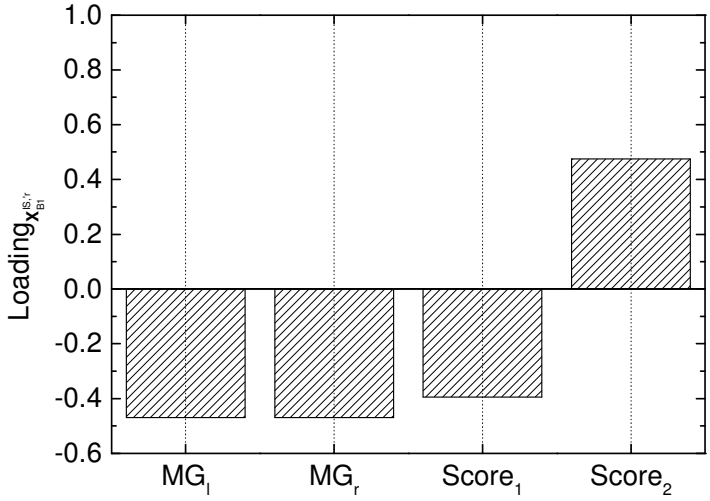


Figure 6.9. *Loading plot of the first LV of* $\mathbf{X}^{'r}_{B1,IS}$ *in the joint-Y PLS model.*

Figure 6.9 shows the loading plot of the first LV of $\mathbf{X}^{'r}_{B1,IS}$ in the joint-Y PLS model. As expected, the dominating variables are the two milling gaps, but the wheat type contribution (through the relevant two scores) is not negligible.

6.3.2 Results using an extended industrial dataset

For each approach, the models are calibrated as described in Section 6.2. Subsequently, they are validated on six additional experiments, as shown in Figure 4.1 of Chapter 4. The milling gap prediction results for the three approaches are shown in Table 6.4 and Table 6.5. The objective was obtaining a product with mass distribution $\mathbf{y}^{DES}$ with the wheat type specified in "wheat variety". Table 6.4 lists the real milling gap used to obtain that product and the milling gap estimations from the three modeling approaches. In Table 6.4 only one milling gap is reported, even though two are estimated, for conciseness. On the other side, Table 6.5 reports the relative errors for the estimations shown in Table 6.4.

Table 6.4. *Validatory milling gap estimations for the three approaches to obtain a* $\mathbf{y}^{DES}$ *mass distribution with the specified wheat variety, when a full industrial dataset is available. For each validation point, the approach providing the best performance is indicated in boldface.*

$\mathbf{y}^{DES}$	Wheat variety	MG_{real}	$MG_{Approach\_1}$	$MG_{Approach\_2}$	$MG_{Approach\_3}$
[79.4,20.6]	Hard	6.25	**6.16**	**6.16**	6.08
[74,26]	Hard	7.17	6.89	**6.91**	6.85
[80.5,19.5]	Semi-hard	6.25	**5.95**	**5.95**	5.88
[74.5,25.5]	Semi-hard	7.17	6.82	**6.83**	6.77
[80,20]	Soft	6.25	**6.23**	**6.23**	6.09
[72.4,27.6]	Soft	7.17	6.84	6.83	**6.85**

Table 6.5 *Relative milling gap estimation errors for the validatory examples of Table 6.4.*

$\mathbf{y}^{DES}$	Wheat variety	$RE_{Approach\ 1}$ [%]	$RE_{Approach\ 2}$ [%]	$RE_{Approach\ 3}$ [%]
[79.4,20.6]	Hard	1.4	1.3	2.6
[74,26]	Hard	3.8	3.6	4.4
[80.5,19.5]	Semi-hard	6.7	4.7	5.8
[74.5,25.5]	Semi-hard	4.8	4.6	5.5
[80,20]	Soft	0.2	0.2	2.5
[72.4,27.6]	Soft	4.5	4.7	4.5

All approaches perform very good milling gap estimations, with relative estimation errors well below 7 %. Overall, slightly better estimations are provided by Approach no.2 (PLS model inversion); in fact, the average relative error is 3.2 %, whereas it is 3.6 % for Approach no.1 and 4.2 % for Approach no.3. We conclude that, in this case, including the small-scale data in the modeling framework does not provide a major advantage. Nevertheless, the

industrial-scale experimental effort can be somewhat reduced, and the milling gap estimations are still acceptable.

6.3.3 Results using a limited industrial dataset

The models were calibrated under the assumption that only three observations are available from the industrial-scale plant (namely, two belonging to the hard wheat variety and one to the semi-hard variety; circles in Figure 6.10). The models were then validated on two observations of the soft variety (squares in Figure 6.10), namely those whose sieve mass distribution is outside the range explored in the calibration set. Therefore, the industrial-scale calibration samples do not explore the entire set of desired mass distribution and wheat variety; conversely, this set is covered by the small-scale experimental dataset (dashed line in Figure 6.10).

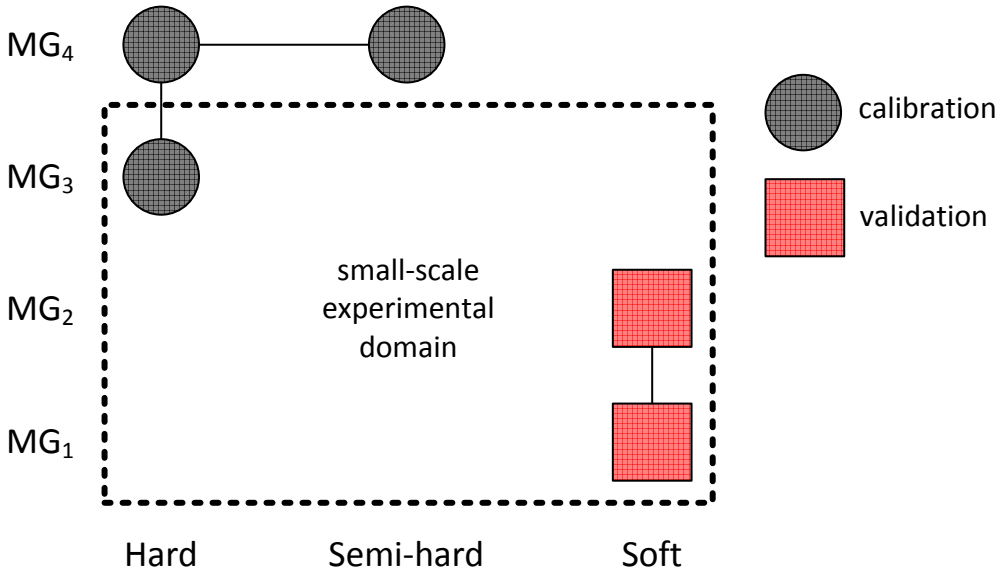


Figure 6.10. *Sketch of the situation where only a limited dataset is available from the industrial-scale plant. The available industrial-scale data (circles) cover only two wheat varieties and two "large" milling gaps. The small-scale data (dashed square) cover a wide range of milling gaps and wheat varieties. The products to be obtained (small squares) lie within a subspace not covered by the industrial-scale dataset.*

Results for the three approaches are shown in Table 6.6 and Table 6.7. Table 6.6 shows the milling gap estimations for the three approaches. In Table 6.7 the corresponding relative errors are reported. Note that, due to the limited domain explored by the industrial-scale experimentation, Approach no.2 (PLS model inversion) failed to provide a solution.

Table 6.6. *Validatory milling gap estimations for the three approaches to obtain a* $\mathbf{y}^{DES}$ *mass distribution with the specified wheat variety, when a limited industrial dataset is available. For each validation point, the approach providing the best performance is indicated in boldface.*

$\mathbf{y}^{DES}$	Wheat variety	MG_{real}	$MG_{Approach\ 1}$	$MG_{Approach\ 2}$	$MG_{Approach\ 3}$
[62.7,37.3]	Soft	8.00	6.71	constraints not satisfied	**7.83**
[72.4,27.6]	Soft	7.17	6.01	constraints not satisfied	**6.65**

Table 6.7. *Relative milling gap estimation errors for the validatory examples shown in Table 6.6.*

$\mathbf{y}^{DES}$	Wheat variety	$RE_{Approach\ 1}$ [%]	$RE_{Approach\ 2}$ [%]	$RE_{Approach\ 3}$ [%]
[62.7,37.3]	Soft	16.1	unfeasible	2.1
[72.4,27.6]	Soft	15.8	unfeasible	7.2

It can be observed that the JY-PLS model inversion approach (Approach no.3) largely outperforms the other approaches in estimating the milling gap that is required to obtain the desired product. Notably, Approach no.2 (PLS model inversion) does not even provide a result, as PLS model inversion turns out to be unfeasible. Despite the fact that the industrial-scale data cover a very narrow experimental domain only, complementing them with data from a small-scale plant enables one to precisely estimate the milling gap required to obtain a product that lies in a region that was not explored by the industrial experimentation.

6.4 Conclusions

In this Chapter a latent-variable methodology to solve the problem of process design and scale-up of an industrial-scale roller mill model was presented. The objective was to calculate the milling gap to be used to obtain a milled product with a desired mass distribution over a sieve from a given wheat variety. To this purpose, three different modeling approaches were critically compared: direct PLS modeling, PLS model inversion, and joint-Y PLS model inversion. The approaches were tested using two different industrial datasets: in the first one, all wheat varieties were used for calibrating the models and the desired product quality was inside the range defined by the calibration samples. The second calibration dataset was characterized by a reduced number of industrial-scale observations (three) and by two wheat varieties only (hard and semi-hard).

It has been shown how a joint-Y PLS modeling approach (which uses data coming from an industrial-scale plant together with data from a small-scale plant) allows better milling gap estimations when the calibration dataset is limited and the solution is extrapolated. The relative milling gap estimation errors were from 2 to 7 times smaller than those obtained

using a standard PLS modeling approach. This approach demonstrates that in particular situations including small-scale data within the industrial-scale modeling framework can improve the model performance. Although a supplementary small-scale experimental campaign is needed, fewer industrial-scale experiments are required, and this is particularly useful when a new machine is installed (namely, few industrial-scale experiments are available). Additionally, the small-scale dataset may be interpreted as a universal dataset, that is measured only once and can be transferred to all industrial-scale plants, and needs to be integrated with only a few industrial-scale experiments. The plant manufacturer could think of directly integrating the small-scale dataset in the machine to be sold to the customer, that needs only to integrate the small-scale dataset with few experiments. The approach we propose can lower the barriers to the development and use of modeling for wheat milling process optimization.

On the other side, the use of more conventional modeling approaches (PLS modeling and PLS model inversion) is suggested when the calibration dataset is more complete (namely, all wheat varieties and milling gaps are represented within the industrial dataset, so that the solution is not extrapolated). In this case, the PLS model inversion allows the best milling gap estimations, with an average relative estimation error of 3.2 %.

Chapter 7

Simulation of the pneumatic transport of industrial wheat mill plants

In this Chapter a simulator of the pneumatic transport of a reference industrial mill plant is shortly presented. Due to confidentiality reasons, most of the technical details cannot be reported. Nevertheless, a general overview of the issues related to model development is provided, together with an excerpt of the major results of the simulation exercise.

7.1 Introduction

The transportation of fluids is not a modern concept. Its usage is tracked back to antiquity: the Romans used lead pipes for water supply and sewerage disposal (Marcus *et al.*, 1990), while the Chinese conveyed natural gas through bamboo canes (Marcus *et al.*, 1990). The first record of pipeline transportation of solids in air is more recent and dated to 1866 (Marcus *et al.*, 1990). The first large scale application of pneumatic conveying was the vacuum transportation of grain in the late 19^{th} century (Marcus *et al.*, 1990).

Pneumatic conveying offers the following advantages (Molerus, 1996):

- dust free transportation of a variety of products;
- flexibility in routing;
- distribution and pick-up to many different areas of the plant;
- low maintenance and low manpower costs;
- multiple product use of pipeline;
- security;
- ease of automation and control.

On the other hand, the disadvantages include:

- high power consumption;
- wear and abrasion of equipment;
- particle abrasion in case of incorrect design;
- limited distance of transportation;

When solid materials are transported in conventional pneumatic systems, three different flow modes can be observed (Pan, 1999):

1. smooth transition from dilute-phase to fluidized dense-phase;

2. dilute-phase, unstable-zone and slug-flow;
3. dilute-phase only.

The flow mode for a solid material is largely determined by the material properties, which involve particle-air interaction (i.e. permeability, air retention and de-aeration). The particle-air interaction characteristics usually are a function of basic particle properties, such as particle size, size distribution, density, shape and hardness.

7.2 Pneumatic transport of biomass particles: wheat

Handling and conveying biomass particles (e.g. wheat) are challenging tasks due to unusual physical properties of those particles. Unusual characteristics commonly include a combination of relatively large mean sizes, wide size distributions, extreme shapes (i.e. flakes and chips), pliability, flexibility, compressibility and general heterogeneity. Biomass particles are highly anisotropic and they may also contain significant quantities of moisture, affecting inter-particle forces and density distribution (Cui and Grace, 2006).

Light and free flowing granular products (such as wheat) are characterized by a dilute-phase, unstable-zone and slug-flow flow mode (Pan, 1999). Figure 7.1 describes the behavior of this flow mode.

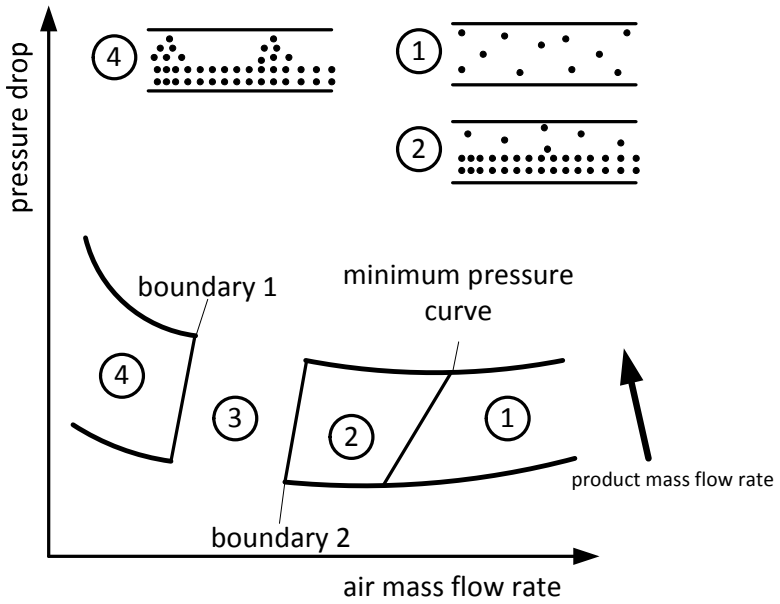


Figure 7.1. *General form of pneumatic conveying characteristics for dilute-phase, unstable-zone and slug-flow. Adapted from Pan (1999).*

In dilute-phase (zone 1 in Figure 7.1), the particles are distributed evenly over the entire cross section of the pipe. While decreasing the air mass flow rate, a point of minimum pressure is reached (minimum pressure curve in Figure 7.1). The critical air velocity corresponding to this point is called saltation velocity. It is observed that, for an effective dilute-phase, the system must be operated at a velocity slightly greater than the saltation velocity (Marcus *et*

al., 1990). With a lower air mass flow rate, a layer of particles is being conveyed at the bottom of the pipe (zone 2 in Figure 7.1). Some particles become stationary at the bottom of the pipeline and most of the material is transported in small clusters and dunes. If the air mass flow is reduced again, the air velocity is not enough to pick up *all* solid particles and some of them accumulate on the bottom of the pipe to form long slugs. These slugs are forced through the pipeline and produce high fluctuations in pressure and vibration (unstable-zone, zone 3 in Figure 7.1). As the air mass flow is decreased further, particles are conveyed gently in the form of slugs (zone 4 in Figure 7.1).

Güner (2007) observed that wheat conveying requires the highest pressure drop and power requirement, respect to other agricultural seeds such as barley, sunflower seeds and lentils.

7.3 Pneumatic transport in an industrial wheat mill plant

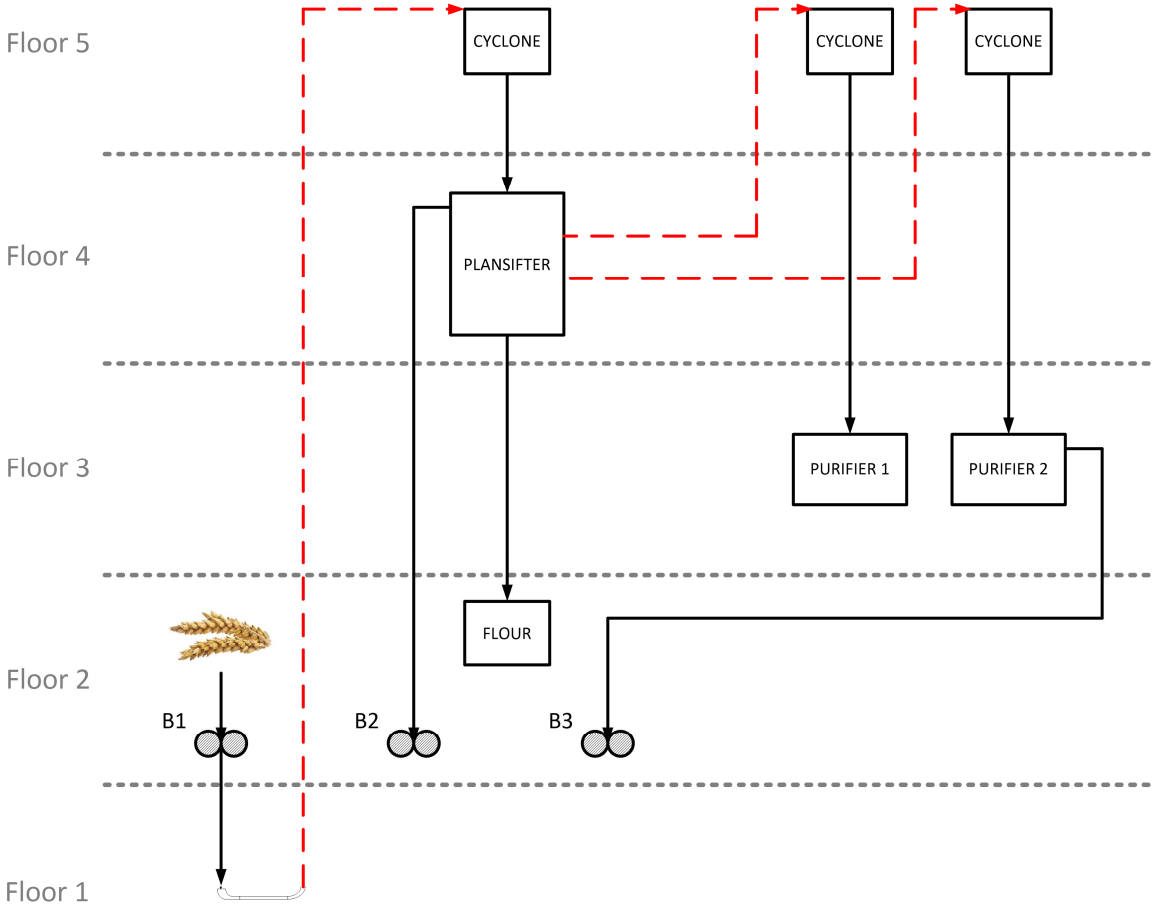


Figure 7.2. *Example of the layout of an industrial wheat mill.*

A typical pneumatic conveying system for an industrial wheat mill plant comprises four distinct zones (Marcus *et al.*, 1990), namely the prime mover, the feeding zone, the conveying zone and the gas-solid separation zone. In an industrial wheat mill they can be identified with:

1. prime mover: fans are used to provide the necessary energy to the conveying gas;

2. feeding zone: the solid material directly enters in the first roller mill;
3. conveying zone: it consists of all the piping system;
4. gas-solid separation zone: the solids are separated from the gas stream in which they have been conveyed using cyclones.

The machines are positioned on different floors of a building, meaning that the material has to be lifted to move from one processing unit to another (in a typical industrial mill, depending on the layout, the material is lifted up to 15 times). This causes very high power consumption in the milling operation.

In the example reported in Figure 7.2, the machines are located on five different floors. Wheat is fed to the system directly through the first roller mill (labelled B1 in Figure 7.2), which is located on the second floor of the building. After the material is milled, it falls for one floor and then is lifted up to the fifth floor where it enters into a cyclone. After the air is removed, the milled wheat falls into a plansifter positioned at the third floor: at this step, depending on the size of the material, it is sent to different units. In the example of Figure 7.2, the coarsest fraction falls in the second roller mill (named B2) at the second floor, the middle fractions are lifted to the cyclones at the fifth floor, and the smallest fraction is collected as flour.

A pneumatic conveying system provides on one side great advantages in the transportation of solid materials, but from the other side it shows a rather high power consumption. Therefore, efforts to reduce the power consumption are of great interest. Simulation and modeling can be powerful tools in supporting the development and the operation of a pneumatic transportation system, and this is the objective of the simulator presented below.

7.4 Some features of the simulator

The main objective of the simulator is to model the pneumatic transport in a reference industrial wheat mill, calculating the pressure drops and the power consumption throughout the flowsheet. To this purpose, all the processing units and the piping system must be modeled. In the simulator, different types of models are used, such as first-principles models (e.g. the cyclone model, specifically modified to adapt it to the different geometry of the mill's cyclones; Chen and Shi, 2007) or empirical models (e.g. the one used to calculate the roller mill power consumption).

The following assumptions are made:

- only dilute-phase flow mode occurs in the system. This is a meaningful assumption if the velocity of the system is greater than the saltation velocity, as observed by Marcus *et al.* (1990);
- the system is at steady-state;
- the ambient temperature is constant.

7.5 Results and discussion

Once the models and the flowsheet are implemented in the simulator, the simulation can be performed. In this subsection the key results are reported.

7.5.1 Mass balance

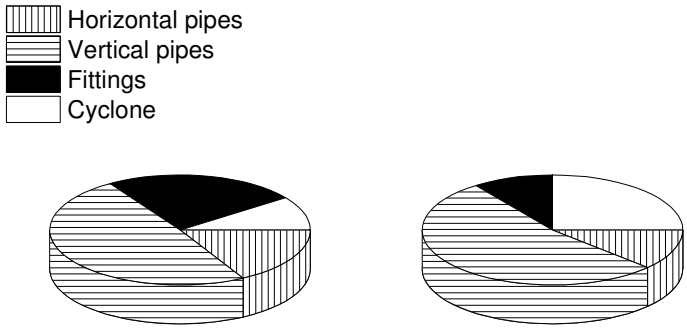
Closure of the solid phase mass balance is important to ensure that the simulation of how the wheat kernel is transformed into flour reflects the true flowrates of the different mill streams. The air phase mass balance is important to ensure that the correct gas flow is considered throughout the mill, since the flow affects the calculation of the main quantities of interest, such as pressure drops and fan power consumption.
Closure of the solid and air phase mass balances was verified.

7.5.2 Power and energy consumption

All the machines in a mill are operated by motors that consume energy. The reduction of energy required to run the mill is a key economic driver from the design and the operation point of view. Therefore, correctly calculating the power and energy consumed by the motors of all the machines in the mill is very important.
The simulation results indicate that the roller mills consume most of the power (more than half of the total energy consumed), whereas fans consume about a quarter of the overall power.

7.5.3 Pressure drops

The pressure and pressure losses in the pneumatic conveying system throughout the mill were calculated. In order to get a better understanding of the phenomena occurring in a passage, it is interesting to examine the different contribution to the total pressure drop. These contributions include (for each passage): horizontal pipes, vertical pipes, fittings and cyclone. Figure 7.3 summarizes the results for two passages of the process (namely, the first breakage and the first reduction one). As can be seen, for both passages vertical pipes account for the majority of the pressure drops, which is not surprising. It is interesting to note that the first breakage and the first reduction passage exhibit an opposite behavior regarding cyclones and fittings. Namely, for the first breakage passage the fittings contribute more to the total pressure drop respect to the cyclones; on the other hand, for the first reduction passage the opposite is true. This fact suggests that the simulator can be used to get a better understanding of the phenomena occurring in the mill, and that it can be used as a tool to increase the efficiency of the process design.



(a) (b)

Figure 7.3. *Pie chart of the pressure drop contributions: (a) First breakage passage and (b) First reduction passage.*

7.5.4 Velocities

The pressure drop in the pneumatic lines increases significantly if the air velocity falls below the saltation velocity. Additionally, the assumption of a dilute-phase flow is not true anymore. Therefore, it is important to verify the minimum velocities in all lift lines. The air velocity has its lowest magnitude at the end of the line; for this reason, the cyclone inlet velocity was considered to monitor the velocities throughout the mill. It was verified that the velocities in all pneumatic lifts are greater than the saltation velocity.

7.5.5 Simulation performances

In Table 7.1 some background metrics related to the performance of the simulator are reported.

Table 7.1. *Some metrics of the simulation.*

Simulation time (warm start) [s]	109
Simulation time (cold start) [s]	412
Number of variables	93232
Number of units	2115
Computer specifications	16 GB, 3.6 GHz, 64 bit, Intel i-7, Windows 7 SP1

Although the high number of variables and units implemented in the flowsheet (93232 and 2115, respectively), the simulation does not require much time to be completed (about 7 minutes in cold start).

7.6 Conclusions and future perspectives

In this Chapter, a computer simulator of the pneumatic transportation system of a reference mill was presented. The objective of the simulator was to simulate the pneumatic conveying of the milling process, calculating the pressure drops and power consumption throughout the mill. Preliminary simulations represent a good basis for further development of the simulator. There does not seem to be any anomalies in the results indicating modeling defects or predictions inconsistent with the true mill behavior, although a systematic validation and further refinement of the models is still required.

At the moment, the simulator can be used to study:

- the impact of changes of mass flow distribution on pneumatics and machines;
- the impact of changes of pneumatic set up, geometries and configuration on energy consumption and pneumatic transport;
- the effect of changes of fan operating conditions on material transport;
- changes of mill layout and impact on pneumatic transport system and energy consumption.

Therefore, the simulator represents a powerful tool that can be used to improve the design and support the operation of a an industrial mill.

Future development of the simulator may include the implementation of an optimization framework of the mill design and layout (changing the position of the unit operations inside the mill building) to minimize the energy consumption. Also a transient simulation of the plant can be implemented, in order to analyze the dynamic phases of the process, such as the start-up and the raw material changes.

Conclusions and future perspectives

Although wheat milling is an old and well-established operation, industrial wheat milling processes are still difficult to design and operate because of several issues:

- complex plant layout;
- complex interaction between the unit operations;
- disturbances in the feeding material, which derive from the strong variability that wheat is subject to as a natural product;
- possible lack of operators' experience and high personnel turnover;
- changing of environmental conditions;
- changing of quality requirements for final product.

Most industrial wheat milling processes are still poorly instrumented and are operated empirically. As a result, the final product quality strongly depends on the operators' expertise and may be subject to variability. Additionally, the energy requirements of a wheat milling plant strongly depends on the plant layout, which is rarely subject to optimization.

The main objective of this Dissertation has been developing a platform to assist wheat milling processes with the intent to move from an experience-driven approach to process understanding, design and operation, to a science-driven one centered on mathematical modeling and process simulation. Table 1 summarizes the main achievements of the Dissertation, with indication about the origin of the data used in the Dissertation, as well as a reference to related papers that have been published in journal or conference proceedings.

Table 1. *Summary of the main achievements of this Dissertation, with indication of the data origin and relevant references.*

Chapter	Main achievement	Data origin	References
Chapter 5	Development of a general data-driven modeling approach for roller mills and plansifters. Models have been used to assist the wheat milling operation in several ways.	Small-scale plant	Dal Pastro F., Facco P., Bezzo F., Zamprogna E., Barolo M. (2016). Data-driven modeling of milling and sieving operations in a wheat milling process. *Food Bioprod. Process.*, **99**, 99-108, ISSN: 0960-3085, doi: 10.1016/j.fbp.2016.04.007 Dal Pastro F., Facco P., Bezzo F., Thomas H., Zamprogna E., Barolo M. (2015). Data-based multivariate modeling of a grain comminution process. In: Krist V. Gernaey Jakob K. Huusom and Rafiqul Gani, Computer Aided Chemical Engineering 37, 2219-2224, Amsterdam:Elsevier, ISBN: 9780444634290, doi: 10.1016/B978-0-444-63576-1.50064-9 Dal Pastro F., Facco P., Bezzo F., Zamprogna E., Barolo M. (2016). Using PLS and NIR spectra to model the first-breakage step of a grain milling process. In: Zdravko Kravanja and Milos Bogataj, Computer Aided Chemical Engineering 38, 1171-1176, Amsterdam: Elsevier, ISBN: 9780444638731, doi: 10.1016/B978-0-444-63428-3.50200-9
Chapter 6	Development of a model inversion framework to tackle the problem of process scale-up in an industrial roller mill. The methodology exploits data from different plant scales.	Industrial-scale and small-scale plant	In preparation.
Chapter 7	Development of a simulator of the wheat milling process that aims at modeling the pneumatic transport, pressure drops and power consumption in a reference mill.	Industrial-scale plant, small-scale plant and literature	

Namely, a general data-driven modeling methodology has been proposed in this Dissertation to mathematically describe roller mills and plansifters used in wheat milling processes. The models have been used:

- to **classify different wheat varieties**. A methodology based on the analysis of wheat near-infrared (NIR) spectra and latent variable modeling was proposed. In particular, the analysis of the model scores allowed one to classify not only the wheat variety, but also different moisture contents within the same variety. The method was tested on three wheat varieties and two moisture contents of the wheat kernel. A general model built on all three varieties and all moisture contents achieved a perfect wheat classification.
- to **improve process understanding**. The loading analysis allowed one to quantify the effect of process parameters and wheat properties on the particle size distribution (PSD) of the milled material, improving the comprehension of the breakage mechanism occurring during wheat milling. Different passages of the process were analyzed and, in general, the milling gap was found to be the most influential parameter affecting the PSD of the milled material. As a representative example, in the first-breakage passage, also a significant effect of moisture content and cellulose amount was observed. Additionally, threshold diameters for which a change in particle amount can be observed, in response to a change on process parameters and wheat properties, were identified.
- to **predict the product quality**. The predictive abilities of the models were used to predict the PSD of the milled material from roller mills and both the PSD and amount of each sieved fraction from plansifters. The prediction accuracy was very good for both units, with root mean squared errors in predicting the PSD that do not exceeded 0.36 (for roller mills) and 0.59 (for plansifters). The amount of each sieved fraction were predicted with errors lower than 10 %. All errors are acceptable in the industrial practice. Additionally, an interface based on Simulink® was developed, to simulate single units as well as a passage of the process.

The models of roller mills and plansifters were connected (thus simulating a “passage” of the process) in order to assess how the prediction errors in the roller mill model affect the subsequent sieving model. The plansifter models were shown to be quite robust to the input accuracy, thus suggesting that, even in a data-driven approach like the one considered in this Dissertation, the development of the roller mill models can be decoupled from that of the plansifter models.

The **problems of process scale-up and process design** for industrial-scale roller mills have been tackled by proposing a data-driven model inversion framework and exploiting data from different plant scales. The objective was to calculate the milling gap to be used to obtain a milled product with a desired mass distribution of a given wheat variety, while at the same time minimizing the experiments that need to be carried out in the industrial plant. It was

shown how a joint-Y PLS modeling approach (which uses data coming from an industrial-scale plant together with data from a small-scale plant) allows better milling gap estimations when the available industrial-scale calibration dataset is limited. This approach demonstrated that including small-scale data within the industrial-scale modeling framework can improve the model performance, and that a reduced operating space coverage of the industrial-scale experimental campaign may be complemented by a dataset deriving from small-scale experimentation. Under this perspective, the small-scale dataset would represent a "universal" dataset, which is collected once and for all by the equipment manufacturer/vendor and which can then be used to support the scale-up exercise on *any* industrial-scale plant.

The **simulation of an entire wheat mill process** was also addressed (for the first time ever, to the best of our knowledge) with the objective of modeling the pneumatic conveying of the milling process, calculating the pressure drops and power consumption throughout a reference mill. Preliminary simulations represent a good basis for further development of the simulator, although a systematic validation and further refinement of the models are still required. At the moment, the simulator can be used to study:

- the impact of changes of mass flow distribution on pneumatic transport system and machines;
- the impact of changes of geometries and configuration on energy consumption and pneumatic transport;
- the effect of changes of fan operating conditions on material transport;
- changes of mill layout and impact on the pneumatic transport system and energy consumption.

Therefore, the simulator represents a powerful tool that can be used to improve the design and support the operation of a an industrial mill.

In summary, this Dissertation has shown how latent variable models can be considered as an advanced tool whose potential can be exploited to move from the experience-driven approach used nowadays to a science-driven one to support the milling operation. On a long-term perspective, the models developed in this project would act as the core of an intelligent grain processing system that provides a science-driven (rather than experience-driven) platform to assist industrial wheat milling operations. Ideally, the models could directly be embedded by the vendor within the process equipment and used by the customers to predict the PSD of the milled material and to get suggestions on which process parameters should be used to obtain a desired product.

This Dissertation has demonstrated how useful information can be gained from analytical instruments (such as NIR spectrometers), which are rarely (if ever) installed in current

industrial wheat milling equipment. It can be envisaged that future industrial mills will take advantage of being equipped with NIR spectrometers to assist plant operation.
Additionally, given that data storage is not an issue, process operation data could be stored for industrial milling processes, following what is routinely done for the chemical and process industry. Having access to historical datasets can represent a great advantage in setting the most appropriate process operating parameters when a change in the feed material or a new product quality profile are requested, thus significantly reducing the need of expensive industrial-scale experiments. Given the fact that small-scale operational data can effectively complement industrial-scale ones in process scale-up/design exercises (at least under the conditions discussed in Chapter 6), one may also envisage that future industrial equipment for wheat milling may be marketed with a set of accompanying small-scale operational data to be used by any customer to complement the data at the industrial scale, at least as long as not enough industrial-scale data are available to satisfactorily address the process scale-up problem.

On the modeling side, other units (such as purifiers) need to be included in the simulator. Finally, even if the product PSD is an important characteristic defining the product quality profile, other quality properties, such as ash content and starch damage, should be considered, since they have an important impact in the final baking quality of flour.

Appendix A

Table A.1. *Distribution of the particle diameters across the PSD bins for the B1 passage. All diameters are reported in μm.*

bin no.	Y^{r}_{B1}
1	0-4
2	4-8
3	8-12
4	12-16
5	16-20
6	20-24
7	24-28
8	28-34
9	34-41
10	41-49
11	49-59
12	59-70
13	70-84
14	84-101
15	101-121
16	121-146
17	146-175
18	175-210
19	210-252
20	252-302
21	302-362
22	362-435
23	435-522
24	522-626
25	626-751
26	751-901
27	901-1081
28	1081-1297
29	1297-1557
30	1557-1765
31	1765-2017
32	2017-2505
33	2505-3111
34	3111-3863
35	3863-4798

Table A.2. *Distribution of the particle diameters across the PSD bins for the B2 passage. All diameters are reported in μm.*

bin no.	$\mathbf{Y}^{r}_{B2}$	$\underline{\mathbf{Y}}^{p}_{B2}$ (sieve 1)	$\underline{\mathbf{Y}}^{p}_{B2}$ (sieve 2)	$\underline{\mathbf{Y}}^{p}_{B2}$ (sieve 3)	$\underline{\mathbf{Y}}^{p}_{B2}$ (sieve 4)	$\underline{\mathbf{Y}}^{p}_{B2}$ (sieve 5)	$\underline{\mathbf{Y}}^{p}_{B2}$ (sieve 6)
1	0-4	63-78	0-4	0-4	8-12	4-8	0-4
2	4-8	78-357	4-8	4-12	12-16	8-12	4-8
3	8-12	357-443	8-12	12-16	16-20	12-16	8-12
4	8-16	443-550	12-16	16-20	20-24	16-20	12-16
5	16-20	550-683	16-20	20-24	24-28	20-24	16-20
6	20-24	683-848	20-24	24-28	28-34	24-28	20-24
7	24-28	848-1053	24-28	28-34	34-41	28-34	24-28
8	28-34	1053-1308	28-34	34-41	41-49	34-41	28-34
9	34-41	1308-1624	34-41	41-49	49-59	41-49	34-41
10	41-49	1624-2017	41-49	49-59	59-70	49-59	41-49
11	49-59	2017-2505	49-59	59-70	70-84	59-70	49-59
12	59-70	2505-3111	59-70	70-101	84-101	70-84	59-70
13	70-84	3111-3863	70-101	101-121	101-121	84-101	70-84
14	84-101	3863-4798	101-121	121-175	121-146	101-121	84-101
15	101-121	4798-5958	121-146	175-210	146-175	121-146	101-121
16	121-146	5958-7399	146-175	210-252	175-210	146-175	121-146
17	146-175	7399-9189	175-210	252-302	210-252	175-210	146-175
18	175-210		210-252	302-362	252-302	210-252	175-210
19	210-252		252-302	362-435	302-362	252-302	210-252
20	252-302		302-362	435-522	362-435	302-362	252-302
21	302-362		362-435	522-626	435-522	362-435	302-362
22	362-435		435-522	626-751	522-626	435-522	362-435
23	435-522		522-626	751-901	626-751	522-626	
24	522-626		626-751	901-1081	751-901	626-751	
25	626-751		751-901	1081-1297	901-1081	751-901	
26	751-901		901-1081	1297-1557	1081-1297		
27	901-1081		1081-1297		1297-1557		
28	1081-1297		1297-1557				
29	1297-1557		1557-1765				
30	1557-1765						
31	1765-2017						
32	2017-2505						
33	2505-3111						
34	3111-3863						
35	3863-4798						
36	4798-5958						
37	5958-7399						

Table A.3. *Distribution of the particle diameters across the PSD bins for the R1 passage. All diameters are reported in μm.*

bin no.	$\mathbf{Y}_{R1}^{r}$ and $\underline{\mathbf{Y}}_{R1}^{p}$ (sieve 1)	$\underline{\mathbf{Y}}_{R1}^{p}$ (sieve 2) and (sieve 3)	$\underline{\mathbf{Y}}_{R1}^{p}$ (sieve 4)	$\underline{\mathbf{Y}}_{R1}^{p}$ (sieve 5)
1	0-4	0-4	0-4	0-4
2	4-8	4-8	4-8	4-8
3	8-12	8-12	8-12	8-12
4	12-16	12-16	12-16	12-16
5	16-20	16-20	16-20	16-20
6	20-24	20-24	20-24	20-24
7	24-28	24-28	24-28	24-28
8	28-34	28-34	28-34	28-34
9	34-41	34-41	34-41	34-41
10	41-49	41-49	41-49	41-49
11	49-59	49-59	49-59	49-59
12	59-70	59-70	59-70	59-70
13	70-84	70-84	70-84	70-84
14	84-101	84-101	84-101	84-101
15	101-121	101-121	101-121	101-121
16	121-146	121-146	121-146	121-146
17	146-175	146-175	146-175	146-175
18	175-210	175-210	175-210	175-210
19	210-252	210-252	210-252	210-252
20	252-302	252-302	252-302	252-302
21	302-362	302-362	302-362	
22	362-435	362-435	362-626	
23	435-522	435-522		
24	522-626			

References

Arazuri, S., Arana, J.I., Arias, N., Arregui, L.M., Gonzalez-Torralba, J., Jaren, C., 2012. Rheological parameters determination using Near Infrared technology in whole wheat grain. *J. Food Eng.* **111**, 115–121.

Archibald, D.D., Kays, S.E., 2000. Determination of total dietary fiber of intact cereal food products by near-infrared reflectance. *J. Agric. Food Chem.* **48**, 4477–4486.

Bonvin, D., Georgakis, C., Pantelides, C.C., Barolo, M., Rodrigues, D., Schneider, R., Dochain, D., Grover, M.A., 2016. Linking Models and Experiments. *Ind. Eng. Chem. Res.* **55**, 6891–6903.

Borosy, A.P., 1999. Quantitative composition–property modelling of rubber mixtures by utilising artificial neural networks. Chemom. Intell. Lab. Syst. 47, 227–238.

Bühler AG, 2016. Personal communication. http://www.buhlergroup.com.

Burggraeve, A., Van den Kerkhof, T., Hellings, M., Remon, J.P., Vervaet, C., De Beer, T., 2011. Batch statistical process control of a fluid bed granulation process using in-line spatial filter velocimetry and product temperature measurements. *Eur. J. Pharm. Sci.* **42**, 584–592.

Burnham, A.J., MacGregor, J.F., Viveros, R., 1999a. Latent variable multivariate regression modeling. *Chemom. Intell. Lab. Syst.* **48**, 167–180.

Burnham, A.J., MacGregor, J.F., Viveros, R., 1999b. A statistical framework for multivariate latent variable regression methods based on maximum likelihood. *J. Chemom.* **13**, 49–65.

Burnham, A.J., Viveros, R., MacGregor, J.F., 1996. Frameworks for latent variable multivariate regression. *J. Chemom.* **10**, 31–45.

Burns, D.A., Ciurczak, E.W., 2008. Handbook of Near-Infrared analysis, Third edition. ed. CRC Press, United States.

Campbell, G.M., 2007. Roller milling of wheat, in: Handbook of Powder Technology. Elsevier Ltd, Amsterdam (Netherlands).

Campbell, G.M., Bunn, P.J., Webb, C., Hook, S.C.W., 2001. On predicting roller milling performance Part II. The breakage function. *Powder Technol.* **115**, 243–255. doi:10.1016/S0032-5910(00)00349-1

Campbell, G.M., Fang, C., Muhamad, I.I., 2007. On predicting roller milling performance VI: effect of kernel hardness and shape on the particle size distribution from First Break milling of wheat. *Food Bioprod. Process.* **85**, 7–23.

Campbell, G.M., Sharp, C., Wall, K., Mateos-Salvador, F., Gubatz, S., Huttly, A., Shewry, P., 2012. Modelling wheat breakage during roller milling using the Double Normalised

Kumaraswamy Breakage Function: effects of kernel shape and hardness. *J. Cereal Sci.* **55**, 415–425.

Campbell, G.M., Webb, C., 2001. On predicting roller milling performance Part I: the breakage equation. *Powder Technol.* **115**, 234–242. doi:10.1016/S0032-5910(00)00348-X

Chen, F.Z., Wang, X.Z., 2000. Discovery of operational spaces from process data for production of multiple grades of products. *Ind. Eng. Chem. Res.* **39**, 2378–2383.

Chen, J., Shi, M., 2007. A universal model to calculate cyclone pressure drop. *Powder Technol.* **171**, 184–191.

Cocchi, M., Corbellini, M., Foca, G., Lucisano, M., Pagani, M.A., Tassi, L., Ulrici, A., 2005. Classification of bread wheat flours in different quality categories by a wavelet-based feature selection/classification algorithm on NIR spectra. *Anal. Chim. Acta* **544**, 100–107.

Cocchi, M., Durante, C., Foca, G., Marchetti, A., Tassi, L., Ulrici, A., 2006. Durum wheat adulteration detection by NIR spectroscopy multivariate calibration. *Talanta* **68**, 1505–1511.

Cui, H., Grace, J.R., 2006. Pneumatic conveying of biomass particles: a review. *China Particuology* **4**, 183–188.

De Jong, S., 1993. SIMPLS: an alternative approach to partial least squares regression. *Chemom. Intell. Lab. Syst.* **18**, 251–263.

de la Fuente, R.L.-N., García-Muñoz, S., Biegler, L.T., 2010. An efficient nonlinear programming strategy for PCA models with incomplete data sets. *J. Chemom.* **24**, 301–311.

Delcour, J.A., Hoseney, R.C., 2010. Principles of cereal science and technology, 3rd ed. AACC International, Minnesota (U.S.A.).

Delwiche, S.R., Graybosch, R.A., Peterson, C.J., 1998. Predicting protein composition, biochemical properties, and dough-handling properties of hard red winter wheat flour by near-infrared reflectance. *Cereal Chem.* **75**, 412–416.

Delwiche, S.R., Norris, K.H., 1993. Classification of hard red wheat by near-infrared diffuse reflectance spectroscopy. *Cereal Chem.* **70**, 29–29.

Domian, E., Poszytek, K., 2005. Wheat flour flowability as affected by water activity, storage time and consolidation. *Int. Agrophysics* **19**, 119–124.

Dowell, F.E., 2000. Differentiating vitreous and nonvitreous durum wheat kernels by using near-infrared spectroscopy. *Cereal Chem.* **77**, 155–158.

Duchesne, C., MacGregor, J.F., 2001. Jackknife and bootstrap methods in the identification of dynamic models. *J. Process Control* **11**, 553–564.

Edgar, T.F., Butler, S.W., Campbell, W.J., Pfeiffer, C., Bode, C., Hwang, S.B., Balakrishnan, K.S., Hahn, J., 2000. Automatic control in microelectronics manufacturing: Practices, challenges, and possibilities. *Automatica* **36**, 1567–1603.

Eriksson, L., Kettaneh-Wold, N., Trygg, J., Wikström, C., Wold, S., 2006. Multi-and megavariate data analysis: Part I: basic principles and applications.

Evers, T., Millar, S., 2002. Cereal grain structure and development: some implications for quality. *J. Cereal Sci.* **36**, 261–284.

Facco, P., Bezzo, F., Barolo, M., Mukherjee, R., Romagnoli, J.A., 2009. Monitoring roughness and edge shape on semiconductors through multiresolution and multivariate image analysis. *AIChE J.* **55**, 1147–1160.

Fang, C., Campbell, G.M., 2003a. On predicting roller milling performance IV: effect of roll disposition on the particle size distribution from first break milling of wheat. *J. Cereal Sci.* **37**, 21–29.

Fang, C., Campbell, G.M., 2003b. On predicting roller milling performance V: effect of moisture content on the particle size distribution from first break milling of wheat. *J. Cereal Sci.* **37**, 31–41.

FDA, 2004. Guidance for industry. PAT – A framework for innovative pharmaceutical development, manufacturing and quality assurance.

Food and Agriculture Organization (FAO), 2016. Food Outlook: Biannual Report on Global Food Markets. FAO Trade and Markets Division.

Gan, Z., Ellis, P.R., Schofield, J.D., 1995. Gas cell stabilisation and gas retention in wheat bread dough. *J. Cereal Sci.* **21**, 215–230.

García-Muñoz, S., Kourti, T., MacGregor, J.F., Apruzzese, F., Champagne, M., 2006. Optimization of batch operating policies. Part I. Handling multiple solutions. *Ind. Eng. Chem. Res.* **45**, 7856–7866.

Garcìa-Muñoz, S., MacGregor, J.F., Kourti, T., 2005. Product transfer between sites using Joint-Y PLS. *Chemom. Intell. Lab. Syst.* **79**, 101–114.

García-Muñoz, S., MacGregor, J.F., Neogi, D., Latshaw, B.E., Mehta, S., 2008. Optimization of batch operating policies. Part II. Incorporating process constraints and industrial applications. *Ind. Eng. Chem. Res.* **47**, 4202–4208.

Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: a tutorial. *Anal. Chim. Acta* **185**, 1–17. doi:10.1016/0003-2670(86)80028-9

Gernaey, K.V., Cervera-Padrell, A.E., Woodley, J.M., 2012. A perspective on PSE in pharmaceutical process development and innovation. *Comput. Chem. Eng.* **42**, 15–29.

Greffeuille, V., Abecassis, J., L'Helgouac'h, C.B., Lullien-Pellerin, V., 2005. Differences in the aleurone layer fate between hard and soft common wheats at grain milling. *Cereal Chem.* **82**, 138–143.

Guldi, R.L., 2004. In-line defect reduction from a historical perspective and its implications for future integrated circuit manufacturing. *IEEE Trans. Semicond. Manuf.* **17**, 629–640.

Güner, M., 2007. Pneumatic conveying characteristics of some agricultural seeds. *J. Food Eng.* **80**, 904–913.

Höskuldsson, A., 1988. PLS regression methods. *J. Chemom.* **2**, 211–228.

Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **24**, 417.

Huang, H., Yu, H., Xu, H., Ying, Y., 2008. Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review. *J. Food Eng.* **87**, 303–313.

Jackson, J.E., 1991. A user's guide to principal analysis. John Wiley & Sons Inc, New York (U.S.A.).

Jaeckle, C.M., MacGregor, J.F., 2000a. Industrial applications of product design through the inversion of latent variable models. *Chemom. Intell. Lab. Syst.* **50**, 199–210.

Jaeckle, C.M., MacGregor, J.F., 2000b. Product transfer between plants using historical process data. *AIChE J.* **46**, 1989–1997.

Jaeckle, C.M., MacGregor, J.F., 1998. Product design through multivariate statistical analysis of process data. *AIChE J.* **44**, 1105–1118.

Jirsa, O., Hrušková, M., Švec, I., 2008. Near-infrared prediction of milling and baking parameters of wheat varieties. *J. Food Eng.* **87**, 21–25.

Johnson, R.A., Wichern, D.W., 2007. Applied multivariate statistical analysis, 6th ed. Pearson Education Inc., Upper Saddle River, NJ (U.S.A.).

Kays, S.E., Barton, I.I., Franklin, E., Windham, W.R., 2000. Predicting protein content by near infrared reflectance spectroscopy in diverse cereal food products. *J. Infrared Spectrosc.* **8**, 35–43.

Kourti, T., 2003. Multivariate dynamic data modeling for analysis and statistical process control of batch processes, start-ups and grade transitions. *J. Chemom.* **17**, 93–109.

Lakshminarayanan, S., Fujii, H., Grosman, B., Dassau, E., Lewin, D.R., 2000. New product design via analysis of historical databases. *Comput. Chem. Eng.* **24**, 671–676.

Li, B., Morris, J., Martin, E.B., 2002. Model selection for partial least squares regression. *Chemom. Intell. Lab. Syst.* **64**, 79–89.

Li Vigni, M., Cocchi, M., 2013. Near infrared spectroscopy and multivariate analysis to evaluate wheat flour doughs leavening and bread properties. *Anal. Chim. Acta* **764**, 17–23.

Li Vigni, M., Durante, C., Foca, G., Marchetti, A., Ulrici, A., Cocchi, M., 2009. Near infrared spectroscopy and multivariate analysis methods for monitoring flour performance in an industrial bread-making process. *Anal. Chim. Acta* **642**, 69–76.

Liu, Z., Bruwer, M.-J., MacGregor, J.F., Rathore, S.S., Reed, D.E., Champagne, M.J., 2011. Scale-up of a pharmaceutical roller compaction process using a joint-Y partial least squares model. *Ind. Eng. Chem. Res.* **50**, 10696–10706.

Manley, M., Van Zyl, L., Osborne, B.G., 2002. Using Fourier transform near infrared spectroscopy in determining kernel hardness, protein and moisture content of whole wheat flour. *J. Infrared Spectrosc.* **10**, 71–76.

Marcus, R.D., Leung, L.S., Klinzing, G.E., Rizk, F., 1990. Pneumatic conveying of solids. Chapman and Hall, London, U. K.

Mardia, K.V., Kent, J.T., Bibby, J.M., 1979. Multivariate analysis. Academic Press Limited, London, U. K.

Mateos-Salvador, F., Sadhukhan, J., Campbell, G.M., 2013. Extending the Normalised Kumaraswamy Breakage Function for roller milling of wheat flour stocks to Second Break. *Powder Technol.* **237**, 107–116.

Mateos-Salvador, F., Sadhukhan, J., Campbell, G.M., 2011. The normalised Kumaraswamy breakage function: a simple model for wheat roller milling. *Powder Technol.* **208**, 144–157.

Matz, S.A., 1991. Milling of wheat and rye, in: Chemistry and Technology of Cereals as Food and Feed. Springer Science & Business Media.

Merkus, H.G., 2009. Particle size measurements: fundamentals, practice, quality. Springer Science & Business Media.

Meyer, C.D., 2000. Matrix analysis and applied linear algebra. SIAM, Philadelphia, PA (U.S.A).

Miralbés, C., 2008. Discrimination of European wheat varieties using near infrared reflectance spectroscopy. *Food Chem.* **106**, 386–389.

Miralbés, C., 2004. Quality control in the milling industry using near infrared transmittance spectroscopy. *Food Chem.* **88**, 621–628.

Molerus, O., 1996. Overview: pneumatic transport of solids. Powder Technol. 88, 309–321.

Molino Quaglia, 2016. Personal communication. http://www.molinoquaglia.org/HOME/.

Montgomery, D.C., 2008. Design and analysis of experiments. John Wiley & Sons.

Moteki, Y., Arai, Y., 1986. Operation planning and quality design of a polymer process. *IFAC DYCORD* 159–165.

Muteki, K., Swaminathan, V., Sekulic, S.S., Reid, G.L., 2012. Feed-Forward Process Control Strategy for Pharmaceutical Tablet Manufacture Using Latent Variable Modeling and Optimization Technologies. *IFAC Proc.* **45**, 51–56.

Muteki, K., Yamamoto, K., Reid, G.L., Krishnan, M., 2011. De-risking scale-up of a high shear wet granulation process using latent variable modeling and near-infrared spectroscopy. *J. Pharm. Innov.* **6**, 142–156.

NASS, 2016. Flour milling products report.

Pan, R., 1999. Material properties and flow modes in pneumatic conveying. *Powder Technol.* **104**, 157–163.

Patwa, A., Ambrose, R.K., Casada, M.E., 2016. Discrete element method as an approach to model the wheat milling process. *Powder Technol.* **302**, 350–356.

Patwa, A., Ambrose, R.P., Dogan, H., Casada, M.E., 2014. Wheat mill stream properties for discrete element method modeling. *Trans. ASABE* **57**, 891–899.

Posner, E.S., Hibbs, A.H., 1997. Wheat flour milling, 2nd ed. American Association of Cereal Chemists Inc., Minnesota (U.S.A.).

Rinnan, A.A., van den Berg, F., Engelsen, S.B., 2009. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends Anal. Chem.* **28**, 1201–1222.

Sato, T., 1994. Application of principal-component analysis on near-infrared spectroscopic data of vegetable oils for their classification. *J. Am. Oil Chem. Soc.* **71**, 293–298.

Sebzalli, Y.M., Wang, X.Z., 2001. Knowledge discovery from process operational data using PCA and fuzzy clustering. *Eng. Appl. Artif. Intell.* **14**, 607–616.

Shao, J., Tu, D., 2012. The jackknife and bootstrap. Springer Science & Business Media.

Sultanbawa, F.M., Owens, W.G., Pandiella, S.S., 2001. A new approach to the prediction of particle separation by sieving in flour milling. *Food Bioprod. Process.* **79**, 211–218.

Tomba, E., Barolo, M., García-Muñoz, S., 2012. General framework for latent variable model inversion for the design and manufacturing of new products. *Ind. Eng. Chem. Res.* **51**, 12886–12900.

Tomba, E., De Martin, M., Facco, P., Robertson, J., Zomer, S., Bezzo, F., Barolo, M., 2013a. General procedure to aid the development of continuous pharmaceutical processes using multivariate statistical modeling–An industrial case study. *Int. J. Pharm.* **444**, 25–39.

Tomba, E., Facco, P., Bezzo, F., García-Muñoz, S., 2013b. Exploiting historical databases to design the target quality profile for a new product. *Ind. Eng. Chem. Res.* **52**, 8260–8271.

Tomba, E., Meneghetti, N., Facco, P., Zelenková, T., Barresi, A.A., Marchisio, D.L., Bezzo, F., Barolo, M., 2014. Transfer of a nanoparticle product between different mixers using latent variable model inversion. *AIChE J.* **60**, 123–135.

Triba, M.N., Moyec, L.L., Amathieu, R., Goosens, C., Bouchemal, N., Nahon, P., Rutledge, D.N., Savarin, P., 2015. PLS/OPLS models in metabolomics: the impact of permutation of datset rows on the K-fold cross-validation quality parameters. *Mol. Biosyst.* **11**, 13–19.

Troup, G.M., Georgakis, C., 2013. Process systems engineering tools in the pharmaceutical industry. *Comput. Chem. Eng.* **51**, 157–171.

Turnbull, K.-M., Rahman, S., 2002. Endosperm texture in wheat. *J. Cereal Sci.* **36**, 327–337.

Ulmer, K., 2011. Technology and equipment - Grain milling. Bühler AG, Uzwil (Switzerland).

USDA, 2016. Grain: world markets and trade.

Valle, S., Li, W., Qin, S.J., 1999. Selection of the number of principal components: the variance of the reconstruction error criterion with a comparison to other methods. *Ind. Eng. Chem. Res.* **38**, 4389–4401.

van den Berg, R.A., Hoefsloot, H.C., Westerhuis, J.A., Smilde, A.K., van der Werf, M.J., 2006. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* **7**, 142–157.

Varmuza, K., Filzmoser, P., 2009. Introduction to multivariate statistical analysis in chemometrics. CRC Press, Boca Raton (FL, U.S.A).

Wesley, I.J., Larroque, O., Osborne, B.G., Azudin, N., Allen, H., Skerritt, J.H., 2001. Measurement of gliadin and glutenin content of flour by NIR spectroscopy. *J. Cereal Sci.* **34**, 125–133.

Wiklund, S., Nilsson, D., Eriksson, L., Sjöström, M., Wold, S., Faber, K., 2007. A randomization test for PLS component selection. *J. Chemom.* **21**, 427–439.

Williams, P., Geladi, P., Fox, G., Manley, M., 2009. Maize kernel hardness classification by near infrared (NIR) hyperspectral imaging and multivariate data analysis. *Anal. Chim. Acta* **653**, 121–130.

Wise, B.M., Gallagher, N.B., Bro, R., Shaver, J.M., Windig, W., Koch, R.S., 2006. PLS_Toolbox Version 4.0 for use with MATLAB™. Eigenvector Research Inc., Wenatchee, WA (U.S.A.).

Wold, H., 1966. Estimation of principal components and related models by iterative least squares, in: Multivariate Analysis. Academic Press Limited, New York (U.S.A.).

Wold, S., Martens, H., Wold, H., 1983. The multivariate calibration problem in chemistry solved by the PLS method, in: Matrix Pencils. Springer, pp. 286–293.

Woo, Y.-A., Kim, H.-J., Chung, H., 1999. Classification of cultivation area of ginseng radix with NIR and Raman spectroscopy. *Analyst* 1**24**, 1223–1226.

Yaakobovitz, B., Cohen, Y., Tsur, Y., 2007. Line edge roughness detection using deep UV light scatterometry. *Microelectron. Eng.* **84**, 619–625.

Yue, H.H., Qin, S.J., Markle, R.J., Nauert, C., Gatto, M., 2000. Fault detection of plasma etchers using optical emission spectra. *IEEE Trans. Semicond. Manuf.* **13**, 374–385.

Zhou, Y.H., Titchener-Hooker, N.J., 1999. Visualizing integrated bioprocess designs through "windows of operation." *Biotechnol. Bioeng.* **65**, 550–557.