INTEGRATED SIMULATION AND OPTIMIZATION OF CONTINUOUS PHARMACEUTICAL MANUFACTURING

by

FANI BOUKOUVALA

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Chemical and Biochemical Engineering

written under the direction of

Marianthi G. Ierapetritou, PhD

Fernando J. Muzzio, PhD

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

January, 2013

ABSTRACT OF THE DISSERTATION

Integrated Simulation and Optimization of Continuous Pharmaceutical Manufacturing

By FANI BOUKOUVALA

Dissertation Director:

Marianthi G. Ierapetritou, PhD

Fernando J. Muzzio, PhD

The goal of this work focuses in the development of reduced-order modeling tools for the design and optimization of manufacturing of pharmaceutical oral solid dosage form products. This work- as part of a collective effort of the Engineering Research Center for Structured Organic Particulate Systems (ERC-SOPS)- aims to combine all the research findings of the Center in terms of material properties and process understanding in order to design efficient continuous integrated manufacturing processes. Different surrogate-based methodologies are assessed for reduced-order model building based on expensive experimental data or computer simulation data. In addition, the effects of uncertainty of the available data, material properties and model parameters on model performance are addressed systematically. Finally, unit operation models are combined into a flowsheet simulation modeling framework which can be used for process design through the

development of a simulation-based optimization strategy for noisy expensive flowsheets with complex constraints.

Acknowledgement and Dedication

Looking back to the past four years of my journey as a PhD Candidate, I realize that I have been so lucky to have received abundant help and support from my mentors, family and friends. I couldn't have done this without all of you.

Firstly, I would like to express my deepest gratitude to Professor Marianthi G. Ierapetritou, who has not only been my academic mentor, but also my close friend. Under her guidance, I have grown both academically and personally, while her support, passion for research, patience and trust have made this journey very productive, rewarding and enjoyable! I would also like to thank Professor Fernando Muzzio, for his guidance and support, and mostly for his initiative for the creation of the Engineering Research Center for Structured Organic Particulate Systems. Without this Center, none of this exciting multidisciplinary work would ever be possible. However, I feel that my journey in academia would have never started if I had not come across my first mentor, Professor Ioannis Palivos, who insisted in me pursuing a PhD. I could not ask for better role models and could not be prouder of my academic family.

I would also like to sincerely thank my committee members, Dr. Rohit Ramachandran and Dr. Salvador Garcia Munoz for their constructive and valuable feedback and detailed comments. I had the privilege of working closely with all of my committee members and I realize now how valuable their help has been towards the completion of my work.

Moreover, I would like to acknowledge my family, who has been the greatest source of support and encouragement in everything that I have chosen to do in my life, and I cannot even express the gratitude and appreciation I feel for all of them. My sister

Natassa and her family, Aris and little Zeta, have in a sense adopted me since the first day I moved to the United States. I am so blessed to have a sister who has been my best friend and mother at the same time since the first day I was born. Undoubtedly, the most important source of inspiration have always been my parents, Manolis and Keti, who no matter how far, have always believed in me, given me strength and supported me. I cannot thank you enough and I cannot wait to pay you back for everything you have done for me. At this point, I would especially like to thank my father, to whom I would like to dedicate this work. My father has always been my ultimate role model, because of his values, his sense of logic, his objectivity and his persistence on my education. Thank you for making me the person I am and I will always try my best to make you proud.

Last but not least, I would like to thank my close friends and particularly Pantelis and Popi who have been there for me every step of the way. They are both so special to me in many different ways, but mostly for being happy with my happiness and successes as well as comforting and supporting during bad times and difficulties. According to a Greek saying: "When you share happiness it doubles, when you share sadness it is halved" and it is a great privilege to be surrounded by friends to whom this applies perfectly. I am blessed with many good friends who have helped me in many ways that they don't even know, and most importantly still support me even though I sometimes have neglected them, Mary, Foteini, Vasilis, Marilena, Eleni, Heather and Rob. Finally, I must also acknowledge the support from all of my colleagues at the Laboratory for Optimization and Systems Analysis and ERC-SOPS, Yijie, Mehmet, Zukui, Nikisha, Jinjun, Nihar, Amalia, Ravendra, Zhaojia, Amanda, Hong,Alice, Aditya, Juan, Bill, Sara, Sarang, Eric and Pallavi, who have all been a pleasure to work and travel with.

Table of Contents

List of tables

List of illustrations

Chapter 1

1. Introduction

1.1. Integrated process design and optimization in pharmaceutical manufacturing

Chemical engineers are often involved in product and process development, where they are asked to design experiments and formulate theories to explain mechanisms of processing operations. Alternatively, they can be involved in process and product design, where they try to identify compositions to provide desired properties, create flowsheets and select operating conditions to produce desired products with high yield and low costs, and/or design configured industrial consumer products. Design problems are usually open ended and may have many solutions that are attractive and near optimal, and in fact design has been characterized as the most creative of all engineering activities, with many opportunities to invent novel products and processes. All of the above tasks have been made possible through the development of algorithmic methods which are used to solve, analyze and optimize process simulations.

Pharmaceutical products and manufacturing is a sector which recently is involving more and more chemical engineers. There are many special considerations that are needed for designing pharmaceutical products, which involves a number of stages which form the typical Development Cycle (Discovery, Preclinical Development, Clinical Trials and Approval). This work focuses on this final stages of manufacturing, specifically the processing of raw powder materials in order for their transformation to the final pharmaceutical product (i.e. tablet, capsule). Historically, the pharmaceutical

industry has been very innovative and successful in the field of new drug discovery and development. However, this has drawn the focus away from the development of efficient manufacturing methods and process understanding (Gernaey & Gani, 2010; Hamad, Bowman, Smith, Sheng, & Morris, 2010; J. Huang, et al., 2009; Klatt & Marquardt, 2009). The lack of knowledge of how critical material attributes and process parameters affect end-point product attributes often hinders the implementation process control, resulting to pharmaceutical manufacturing processes which generate products with relatively high variability (McKenzie, Kiang, Tom, Rubin, & Futran, 2006). One additional challenge for the establishment of efficient, controlled, and automated manufacturing methods is the considerable variability in new raw material properties. In addition, the majority of pharmaceutical products (~80%) are in a solid based form of tablets or capsules, composed from bulk powder materials, which are far more complex and challenging to handle than liquid or gas phase materials. Even though significant progress has been made recently in particle technology research, there is a gap between fundamental science and applied engineering due to the need for integration of multiscale knowledge (Ng, 2002).

All of the above reasons have been the source of consensus and legacy based heuristic production strategies, conducted overwhelmingly in batch mode; with product quality being traditionally verified offline through acceptance sampling. This approach has lead to additional sources of variability, which are the effects of the analytical method, and the human factor, since it is common for operators to regulate the process based on their individual knowledge and experience. Moreover, due to the insofar typical batch manufacturing procedures, which involve the processing of material one stage at a

time, integrated process design, has not been a priority of the industry. Even in cases where one specific process stage has been studied in detail and potentially optimized, its operation would typically be considered independently of the next and previous operating stages. As a result, each process operation would be optimized individually. However, if one considers all of the process stages as a whole integrated procedure, the global operating conditions may be drastically different due to process interactions and conflicting objectives.

Recently, the industry as well as the Food and Drug Administration (FDA) have recognized the need for modernizing pharmaceutical manufacturing and have launched an initiative for enhancing process understanding through Quality by Design (QbD) and Process Analytical Technology (PAT) tools (Garcia, Cook, & Nosal, 2008; Lionberger, Lee, Lee, Raw, & Yu, 2008; Nosal & Schultz, 2008; L. Yu, 2008). The major goals of these efforts include the development of scientific mechanistic understanding of a wide range of processes; harmonization of processes and equipment; development of technologies to perform online measurements of critical material properties during processing; performance of real-time control and optimization; minimization of the need for empirical experimentation and finally, exploration of process flexibility or design space (Lepore & Spavins, 2008). To achieve these goals, the industry needs the modeling tools and databases for measuring, controlling and predicting quality and performance.

1.2. Transition towards continuous pharmaceutical manufacturing

During the last five years, one of the main approaches for modernizing pharmaceutical manufacturing, transition of production from batch to continuous mode, is becoming increasingly more appealing to the industry and regulatory authorities (Betz,

Junker-Bürgin, & Leuenberger, 2003; Gonnissen, Gonçalves, De Geest, Remon, & Vervaet, 2008; Leuenberger, 2001; Plumb, 2005). The advantages of this change have been proven very beneficial in many aspects when applied in other fields, such as petrochemicals and specialty chemicals (Gorsek & Glavic, 1997). Firstly, continuous manufacturing could potentially minimize the need for scale-up studies and decrease the time-to-market significantly (Leuenberger, 2001). At the present time, processes developed in small-scale equipment used for initial clinical studies must be scaled-up and subsequently validated experimentally and further optimized, since their operation is always potentially different at the larger scale. This leads to another advantage of continuous integrated manufacturing, which is the minimization of the plant footprint, since the entire continuous process typically fits inside a much smaller space. A well controlled continuous process involves the handling of small aliquots of material throughout the unit operations, increasing the ability to monitor a significant fraction of the process streams, which is impossible in a large-scale batch process. In addition, continuous operations can produce higher throughputs under better control, which implies the optimal use of the invested capital (space, raw materials and equipment), as well as the reduction of waste (Plumb, 2005). Also, in a continuous setting, the human factor can be further decreased, compared to a batch setting, through automation of the operation of integrated systems and thus labor costs can be reduced. A detailed economical analysis and comparison of batch versus continuous operating mode for the production of pharmaceuticals has been performed by Schaber et al. (Schaber, et al., 2011), demonstrating the possibilities and advantages of the latter. However, continuous systems also possess certain disadvantages, such as the need to dedicate specific

equipment and facilities to a certain integrated process and the need for advanced Process Analytical Technology (PAT) systems which are costly and not yet developed for granular systems.

However, a switch from the already established batch to continuous operation involves many challenges, and could lead to failure if not performed after careful design and planning. Firstly, pharmaceutical substances are highly sensitive to environmental conditions, such as humidity and temperature and a possible larger residence time than required can cause significant material degradation. This can cause dangerous product contaminations and should be avoided. Subsequently, in a continuous production, the process does not reach steady state from the beginning, and this may cause off-specification product to be produced during a particular time interval. However, because regulatory authorities require detailed and time-consuming documentation for the establishment of a manufacturing strategy, and because in many cases companies currently have unused capacity, it is still debatable by pharmaceutical companies whether the modification of an already established batch manufacturing procedure to a continuous one is worth the risk and the up-front expense.

All of the above risks and challenges, however, can be reduced through detailed and accurate process knowledge and the exploitation of advanced computer-aided simulation tools to ensure effective performance and optimal integrated design of both existing batch or future developed continuous processes.

1.3. *Process systems engineering tools and their connection to Quality by Design in pharmaceutical Engineering*

As it is stated by Oksanen et al (Oksanen & García Muñoz, 2010), “from a chemical engineer's perspective, QbD, Design Space and Process Analytical Technology (PAT) can be understood as the application of Process Systems Engineering (PSE) to the development and manufacture of pharmaceutical products”. In fact, this is one of the primary goals of this research, to serve as an initial building block towards the implementation of PSE tools for designing pharmaceutical manufacturing.

PSE has been introduced as a distinct community during the last 50 years and has sprung out from computer-aided engineering practices. Stephanopoulos et al. (Stephanopoulos & Reklaitis, 2011), recently published a comprehensive review of the history and applications of this community from the 1800’s to today. The authors describe the stages that chemical and process engineering have been through during these decades and interestingly, the period of the 1920s to 1950s, is categorized as the ‘‘Waiting Period’’. The reason for this distinction is due to the fact that during this time engineers studied the individual components of a process (unit operations) and not the overall integrated process- from raw materials to final product. It was a period during which robust models were developed and validated for all the basic unit operations of the chemical engineering field, which later allowed researchers to start combining these models and develop integrated systems science techniques. Specifically, from the 1960’s to 2000 there was a huge leap in the publications and developments on integrated design, optimization and control which were all greatly facilitated through the advances in computational capabilities. Engineers started relying on computational tools for designing their processes, minimizing the need to physically experiment with their processes or rely on heuristics and empirical knowledge. Even though the most common objective of

design and optimization is profit, more recent developments include additional aspects to integrated process design, such as systems to improve environmental impact, simultaneous process and product design, process intensification combining multiple operations into fewer components and multi-plant integration.

All the above are mentioned in order to make what is being realized during the past few years even more obvious: the pharmaceutical industry can benefit tremendously by all the tools developed in the PSE field. The pharmaceutical industry can be considered to be coming out of its own "Waiting Period" during the past few years, so it is now the time for engineers to start thinking about integrated design and optimization. Certainly, implementation of such tools for pharmaceutical processes and products is a very challenging task, due to the complexity and variability of powder materials and the specifications that a product must meet in terms of physical and biological properties. However, if practitioners in this industry become aware of the capabilities of the years of research that PSE community has to offer, as well as success stories and applications from other industries (i.e. petrochemical, polymers and consumer products), then perhaps the effort will be put in order to gather the necessary information for the implementation of PSE tools.

Table 1 is a simple summarized set of connections between goals of the FDA's QbD initiative and PSE tools. Process understanding is equivalent to the development of predictive models, preferably first-principle based when available. The developed models can then be integrated into a flowsheet simulator in order to build a computer version of an actual process and be able to schedule operating procedures and identify integration bottlenecks. In addition, the very popular design space in QbD guidances can be

correlated with the flexibility of a process. Finally, steady-state and dynamic optimization strategies can facilitate process and product design and control respectively, while risk analysis can be facilitated by sensitivity analysis tools.

Table 1.Connection between QbD goals and PSE tools

Quality by Design initiative Goal	**Process Systems Engineering Tool**
Process understanding	*Predictive models* (mechanistic, data-based, hybrid)
Process integration and simulation	*Flowsheet modeling*
Design Space	*Flexibility Analysis*
Process and product design	*Steady-state optimization* (Flowsheet synthesis)
Process controllability	*Dynamic optimization* (model predictive control)
Risk analysis	*Sensitivity Analysis*

1.4. Motivation: Integration of process knowledge and experimental or simulation data for the design of continuous pharmaceutical processes

The main purpose of this work is the integration of various types of process models along with experimental and computational data into a flowsheet simulation framework, in order to describe and analyze pharmaceutical manufacturing processes. The development of the integrated models enables the next two stages of Design Space (DS) characterization and process optimization of integrated pharmaceutical processes. However, the computational complexity and stochastic nature of the aforementioned models have brought up the need for development of special algorithmic procedures for the identification of the DS and optimization, which are also the focus of this work.

Due to the aforementioned current efforts to enhance process understanding in the pharmaceutical manufacturing field, large amounts of data for the characterization of processes or materials have been collected. In certain cases, these data have provided

insight about the governing underlying mechanisms and have lead to the development of first-principle models, while in other cases data have been used to build empirical response surface models, or multivariate latent variable models in cases of high dimensionality. In fact, even in the well-established liquid-based flowsheet models, certain physical phenomena which are insufficiently understood, such as chemical reaction kinetics or heat and mass transfer effects are represented through empirical correlations which are incorporated into first-principle models to create hybrid or semi-empirical modules (Kahrs & Marquardt, 2007).

Moreover, the complex nature of powder based-processes has also lead to the development and use of high-fidelity discrete type simulation models that aim to capture the complex behavior of interacting particles in order to provide insight into detailed critical material attributes such as particle size distributions. This introduces another challenge towards the final goal of flowsheet model building, which is the incorporation of multiscale computational data, derived from very expensive computer models. In this work, it is shown that often, handling of such computational data is very similar to that of experimental data.

This work aims to summarize all of the identified challenges that originate from the incorporation of experimental and computational data into a dynamic flowsheet model library for pharmaceutical processes, in order to develop a generic model library which may be used for a wide range of raw materials and processes, integrated to form complete production lines.

The current work aims to tackle only a small fraction of the aspects of Table 1 and act potentially as an initial “success story”, towards the implementation of certain of the

aforementioned tools to models and data from actual pharmaceutical processes. It is not one of the goals of this work to develop first-principle models for specific unit operations, but rather incorporate such models either from the literature or from the ERC-SOPS databases-when available- in order to develop an integrated flowsheet simulation and initiate the work of process design using this integrated model. When these models are not available, data-based approaches may be employed. Overall, the models necessary for design and optimization strategies must be computationally efficient, thus a strong focus on reduced-order modeling techniques is evident throughout this work.

1.4. Role of reduced-order models

In many aspects of this work, it was realized that the use of reduced-order or data-based models can be of great importance in modeling of complex pharmaceutical processes. In the literature, the term "reduced-order model" is used to describe any type of model which is of lower dimensionality and/or computational cost from the original full-scale model, such as surrogate-based or data-driven models, hybrid or grey-box models and models of reduced dimensionality based on multivariate analysis methods such as Principal Component Analysis (PCA) and Partial Least Squares (PLS). The first category of surrogate models refers to data-based response surface type methods with unknown parameters fitted based on a set of experimental or simulated data (Boukouvala, Muzzio, & Ierapetritou, 2010; Boukouvala, Muzzio, & Ierapetritou, 2011; Donald R. Jones, 2001; J. P. C. Kleijnen & Beers, 2004). Hybrid or grey-box models are typically referred to as low-order because they are composed of a combination of first-principle but also empirical correlations which are used as fast approximations of certain complex and demanding correlations (L. Chen, Hontoir, Huang, Zhang, & Morris, 2004; Kahrs & Marquardt, 2007; Psichogios & Ungar, 1992; Romijn, Özkan, Weiland, Ludlage, &

Marquardt, 2008). Finally, model reduction based on PCA methodologies is an efficient tool used to create inexpensive approximations of large scale distributed systems, such as those resulting from discretization of partial differential equation systems (Cizmas, Palacios, O'Brien, & Syamlal, 2003; Deane, Kevrekidis, Karniadakis, & Orszag, 1991; Frouzakis, Kevrekidis, Lee, Boulouchos, & Alonso, 2000; Lucia, Beran, & Silva, 2004; My-Ha, Lim, Khoo, & Willcox, 2007), or identify structures and reduce dimensionality of large datasets. Model reduction is becoming very essential for allowing compressed or simplified versions of computationally expensive models in dynamic simulation, real-time control and optimization. However, model reduction techniques come at certain costs, such as the limited validity within the range of the available data. A more detailed description of the methods, advantages and disadvantages of ROMS is provided in chapter 3.

Even in the optimization literature, where rigorous optimization techniques require large number of function evaluations and calculation of accurate derivatives, optimization based on surrogate and faster models is becoming more and more popular. The reason for this is that derivative information is not available if the process model is very computationally expensive (i.e. each function evaluation requires weeks) or when the process knowledge is not yet available in order for any type of mechanistic model to be developed. This has given rise to the development of many techniques for what is referred to as black-box, or surrogate-based or derivative-free optimization (Auger, Hansen, Perez Zerpa, Ros, & Schoenauer, 2009; Barron, Duvall, & Barron, 2004; Booker, et al., 1999; Henao & Maravelias, 2011; Queipo, et al., 2005) . Response surface optimization has been applied to many fields including automotive engineering and

design, structural optimization, geology, and food engineering, to name a few. Although surrogate-based optimization can usually guarantee local optimality, global methods have also been studied (Donald R. Jones, 2001).

Throughout this work, reduced modeling aspects are used within flowsheet simulations and design spaces- in order to form hybrid models or add correlations for which no first-principle model exists- as well as in the developed optimization framework.

1.5. Outline of dissertation

This dissertation is broken up into five main chapters. Chapter 2 describes the process of flowsheet model building for a variety of different manufacturing scenarios, namely direct compaction, dry granulation and wet granulation. Chapter 3 focuses on the challenges of integration of various types of models, as well as the role of reduced order data-based models as unit operation models in the developed flowsheets. The integration of a reduced-order model for capturing the effects of material properties is also discussed in this chapter. Chapter 4 describes the proposed methodology for the mapping and analysis of process design space, for both individual unit operations and integrated flowsheet models. The methodologies described in this chapter is parallelized with the PSE concepts of process feasibility and flexibility. Next, chapter 5, focuses on the optimization of integrated flowsheets, which are stochastic and of high computational cost. Specifically, already existing methodologies for the optimization of such models are discussed, and a new approach for the optimization of such models subject to constraints is developed, tested and applied to a direct compaction case study. The final chapter,

summarizes this work and discusses critical issues that have risen through this work, and are proposed to be solved in the future.

Chapter 2

2. Integrated simulation and analysis of continuous pharmaceutical manufacturing

2.1. Dynamic flowsheet modeling for particulate processes

Flowsheet modeling is one of the most influential achievements of computer-aided process systems engineering, which has enabled the design, analysis and optimization of robust processes in the chemical industry. A robust and detailed flowsheet simulation is an approximate representation of the actual plant operation, which also helps in the establishment of successful control strategies that will regulate the process when given a desired set-point change or when a problem occurs during the operation of the integrated process (R. Ramachandran, Arjunan, Chaudhury, & Ierapetritou). Through accurate simulations, one can predict the time interval of the transitional stage during which product has not yet reached desired state, and the control actions to take when the system deviates from the desired state (feedback control) or when perturbations are detected as they enter into the system (feed-forward control). Accurate modeling of the residence time distribution of material in the process also allows discarding a small percentage of faulty product when necessary, which is more profitable than production of failing batches.

A developed flowsheet simulation can also facilitate the identification of possible process integration bottlenecks, conflicting design and control objectives, simulation of the effect of recycle streams as well as process start-up and shut-down. For these and other reasons, flowsheet synthesis is an extremely important first step in a wide range of industries, during which the optimal process configuration is decided upon according to

the desired design objectives (Biegler, Grossmann, & Westerberg, 1997). This procedure enables the investigation of design alternatives through the formulation of superstructure networks and the solution of mixed integer optimization problems which have operating conditions and design parameters as decision variables (Biegler & Grossmann, 2004; Biegler T L, Grossmann E I, & Westerberg W.A, 1997; Henao & Maravelias, 2011). In the literature and in industrial practice, flowsheet simulations have helped identify global optimal operating conditions and design configurations that lead to robust, flexible and economically profitable processes.

Research in flowsheet building for fluid-based processes common to the chemical industry (i.e. petrochemicals) has become a mature activity, resulting in a variety of state of the art software packages (e.g. ASPEN, gPROMS, CHEMCAD etc.) that contain all the needed capabilities. Using the developed software is easy since a user can simply 'drag-and-drop' the necessary unit operations from established model libraries and connect them appropriately to simulate a specific integrated process. On the contrary, flowsheet models and software for solid based processes are only in a primitive stage for a variety of reasons (Gruhn, Werther, & Schmidt, 2004; Ng, 2002; Ng & Fung, 2003; Werther, Reimers, & Gruhn, 2008, 2009).

Firstly, the lack of knowledge of the critical material properties, design, and process variables, as well as the lack of unit operation process models have inhibited the development of model-based flowsheet simulators. Due to the high complexity of the raw materials and the lack of standardized procedures, another obstacle is the inexistence of comprehensive material property databases for a wide range of used excipients and APIs. In fact, material characterization of pharmaceutical powders is a non standardized

procedure which differs amongst the different companies. Universal material property libraries are vital to model libraries, if one wants to produce a wide ranging multipurpose flowsheet model, independent of specific product applications. Lastly, solid-based flowsheet modeling has been held back due to the lack of software with capabilities for handling dynamic changes of distributed parameters (i.e. particle size distributions) (Werther, Toebermann, Rosenkranz, & Gruhn, 2000). In this work, gPROMS is used as a platform for building a flowsheet model for various production schemes for pharmaceutical tablets. gPROMS is an equation-based (Oh & Pantelides, 1996; Winkel, Zullo, Verheijen, & Pantelides, 1995) and dynamic software, which is widely accepted in a variety of fluid based product industries. The aspect of flowsheet model building which is mostly addressed through this dissertation is integration of different model types and material property databases.

2.1.1. Model types

The model library built in order to simulate tablet production scenarios: direct compaction, dry granulation and wet granulation include: feeders, mixers, hoppers, roller compactors, mill, tablet press and feed frame and wet granulator. This is a collaborative effort to combine all the know-how of the ERC-SOPS center into this flowsheet simulation.

Feeders

In the majority of powder handling industries (i.e. pharmaceutical, food, ceramics, catalysts), feeding of powder materials into a continuous processing line is performed through gravimetric feeders. The main purpose of the unit operation of feeding is the supply of raw materials to the preceding unit operation at desired and consistent flowrates

in order to match the desired mixture composition and total material throughput. Pharmaceutical manufacturing also employs LIW feeders for semi-continuous operations such as jet milling and roller compaction. The output feedrate of a gravimetric feeder is calculated based on the loss in the weight of the material contained in the feeder over time. This mode of operation allows the accurate monitoring of the exact amount of material provided to the next unit operation, since any accumulation of powder in the hopper will be captured by the change in the total weight. Currently available feeders are operated in closed-loop. Even though significant work has been performed in the area of feeder design development from an equipment vendor perspective, powder feeding has not been studied extensively from a modeling viewpoint. In order to capture the dynamics of the process, a model is required for the accurate calculation of the output flowrate and the description of the feedrate response in the case of possible feedrate set-point changes. Thus, an extensive experimental design was performed in order to collect dynamic data of the output flowrate after different magnitudes of step changes under different operating conditions. Analysis of the collected data suggests that the closed-loop feeder dynamics can be described by a first order delay differential equation. The parameters of this model consist of the process gain parameter (k), the time constant (τ), and the time delay factor (θ) (Equation 1). The optimum parameter values were obtained through minimization of the least-squares error of the observed versus predicted flowrate values. Throughout the process of feeding, the material is assumed to retain its original particle size distribution and bulk density.

$$F_{out}(s) = \frac{Ke^{-\theta s}}{\tau s + 1} \qquad (1)$$

Mixers

Several modeling approaches exist in the literature for powder mixing processes. The current modeling approaches can be categorized into Monte-Carlo methods (Mizonov, Berthiaux, Marikh, Ponomarev, & Barantzeva, 2004), continuum and constitutive models (Muzzio, Sudah, Chester, Kowalski, & Beeckman, 2002), data-driven statistical models (Boukouvala, et al., 2010; M. G. Ierapetritou, Boukouvala, & Muzzio, 2010; Muzzio, Portillo, & Ierapetritou, 2009; Wu, Heilweil, Hussain, & Khan, 2007), compartment models (M. G. Ierapetritou, Portillo, & Muzzio, 2006; Portillo, Muzzio, & Ierapetritou, 2008), RTD modeling approaches (Y. Gao, M. Ierapetritou, & F. Muzzio, 2011a; M. Ierapetritou, Gao, & Muzzio, 2011; M. Ierapetritou, Gao, Vanarase, & Muzzio, 2011), hybrid-models (M. G. Ierapetritou, Portillo, & Muzzio, 2007; Wassgren, Freireich, Li, & Litster, 2011) and discrete element method (DEM) based models (Bertrand, Leclaire, & Levecque, 2005; Glasser, Remy, Canty, & Khinast, 2010; Glasser, Remy, & Khinast, 2009, 2010; Hassanpour, et al., 2011; Muzzio, Dubey, Sarkar, Ierapetritou, & Wassgren, 2011; Sarkar & Wassgren, 2009; Wassgren & Sarkar, 2010). The broad portfolio of mixing models showcase the potential of using different models for a specific use. For instance, statistical/RTD models could be used for online control and dynamic optimization whilst constitutive and hybrid models could be used for design and simulation. Hybrid models especially have the potential to incorporate multi-scale information from the particle level to the unit-operation level. In this work, an alternative hybrid methodology based on population balance modeling is employed. A multi-dimensional population balance model is incorporated into the flowsheet model library, based on the work of the research group of Dr. Ramachandran (Fani Boukouvala, et al.,

2011a; Poon, et al., 2009; Rohit Ramachandran, et al., 2012a; Rohit Ramachandran & Barton, 2010; Rohit Ramachandran & Chaudhury, 2012b; Rohit Ramachandran, Immanuel, Stepanek, Litster, & Doyle III, 2009). Specifically, a multidimensional population balance model has been developed for blending processes that accounts for *2* solid components, two external coordinates (axial and transverse directions in the blender) and one internal coordinate (size distribution due to segregation). The main equation of the PBM is shown below:

$$\frac{\partial F(n,z_1,z_2,r,t)}{\partial t}+\frac{\partial}{\partial z}\left[F(n,z_1,z_2,r,t)\frac{dz_1}{dt}\right]+\frac{\partial}{\partial z}\left[F(n,z_1,z_2,r,t)\frac{dz_2}{dt}\right]+\frac{\partial}{\partial r}\left[F(n,z_1,z_2,r,t)\frac{dr}{dt}\right]$$
$$=\Re_{formation}(n,z_1,z_2,r,t)-\Re_{depletion}(n,z_1,z_2,r,t)$$

(2)

In Equation 2, z_1 is the spatial coordinate in the axial direction, z_2 is the spatial coordinate in the radial direction, *r* is the internal coordinate that depicts particle size and $n=2$ to indicate presence of two components (Active Pharmaceutical Ingredient and excipient). Hence $\frac{dz_1}{dt}$ and $\frac{dz_2}{dt}$ represent the axial and radial velocity respectively. More details on the development of the PBM can be found in Boukouvala et al. (Fani Boukouvala, et al., 2011b).

Roller Compaction

Roller compaction is the process of compacting a blended mixture of powders into a thin ribbon between two rotating rolls. Compaction of powders into ribbons aims to produce compacts which are subsequently milled to give powder mixtures with improved

mechanical and uniformity characteristics (am Ende, et al., 2007; Cunningham, Winstead, & Zavaliangos, 2010; Hein, Picker-Freyer, & Langridge, 2008; Soh, et al., 2008). Roller compaction has been studied in literature and several modeling approaches have been proposed, ranging from simpler data-based (Turkoglu, Aydin, Murray, & Sakr, 1999) to more complex FEM and/or DEM simulations (Cunningham, et al., 2010; Dec, Zavaliangos, & Cunningham, 2003; Zinchuk, Mullarney, & Hancock, 2004), that link input material properties and process variables to properties of the produced ribbon. An interesting review of the most significant modeling attempts of roller compaction is found in (Dec, et al., 2003). The first significant effort was a contribution of Johanson *et al.* in the 1960's which can predict the steady-state ribbon density and thickness of a produced ribbon given various processing conditions, inlet powder material properties and design aspects. Even though this model is based on a series of simplistic assumptions, it has been verified against experimental data and has been used often in the roller compaction literature. More recently, Hsu *et al.* (Hsu, Reklaitis, & Venkatasubramanian, 2010) extended the capabilities of this model to capture the dynamics of the process and it is used in this flowsheet.

Milling

A three-dimensional population balance model is implemented to describe the dynamics of the milling process. Unlike the mixing process, the milling process is assumed to be homogenous with respect to spatial position (i.e., uniform velocities within the process geometry) but heterogeneous with respect to its internal coordinates which in this case are particle size, bulk density and API composition. Similar to Verkoeijen et al., (Verkoeijen, Pouw, Meesters, & Scarlett, 2002) a volume-based PBM is described in this

study and is characterized by the internal coordinates API volume, excipient volume, and gas volume. This PB model is also developed by the same group lead by Dr. Ramachandran and it is solved via a finite volume method, previously developed by Immanuel and Doyle III (Immanuel & Doyle Iii, 2005) for multi-dimensional systems, in combination with a backward differential formula (BDF) implicit integrator (in-built in gPROMS).

Hoppers

Hoppers are unit operations which are a supplementary component of processes such as tablet presses and roller compactors and aim to collect powder from an upper opening and feed material to the actual process from the bottom orifice. Modeling the behavior of coarse (< 500 μm) particulate mixtures in handling systems such as conical hoppers has been studied extensively in the literature (S. B. Savage, 1965; Stuart B. Savage & Sayed, 1981; Weir, 2004). From the 1920's it was realized that the outflow of powder materials from steep-walled hoppers is independent of the height of the contained material, while it is a function of the outlet diameter and the mean particle diameter (Nedderman, Tüzün, Savage, & Houlsby, 1982). This conclusion was derived from the analysis of the stress, velocity and density fields of the material that is flowing through the system.

Conical shaped hoppers have been also studied experimentally, where a number of problems have been observed depending mainly on the properties of the processed material and the geometry design of the process. Typical examples include arching or bridging- which prevents material from flowing out of the hopper, and ratholing- which limits the capacity of the hopper due to a formation of a stagnant layer of material around

the walls of the hopper. In order to overcome these problems, modifications to the geometry of the hopper may be performed (i.e. change of slope or shape), addition of inserts such as inverted cones and finally, addition of flow enhancing material such as glidants (i.e. Silicon Dioxide or Magnesium Stearate) can be added to the mixture. Above all, the properties of the processed raw materials (i.e. cohesiveness, friction angle, flowability) can suggest the optimal design in order to attain the operation of the hopper in the 'mass flow' regime, where all of the material inside the hopper is moving towards the exit, where it is discharged with a relatively constant flowrate. The output flow rate from a hopper in tableting or capsule filling is sometimes controlled through the presence of a feeding screw (system equivalent to gravimetric feeders), or through the feedframe in the tablet press. It is important, however, to make sure that the hopper is carefully designed (height, angle and outlet diameter) such that mass flow is achieved and stagnation of material in "dead" zones is avoided.

For the materials processed in this study, experimental studies will be employed to determine conditions that achieve 'mass flow' independently from the geometry design of the conical hopper, due to the concentration of MgSt present in the mixture (Faqih, Alexander, Muzzio, & Tomassone, 2007). A mass balance on the hopper system will be of the following form (Equation 3):

$$\frac{d\dot{m}}{dt} = F_{in} - F_{out} \tag{3}$$

where $\dot{m}$ is the mass holdup inside the hopper. Assuming constant bulk density throughout the hopper the height of the material inside the hopper can be correlated to the mass holdup through (Equation 4):

$$\dot{m} = H \cdot A(H) \cdot \rho \tag{4}$$

where H is the height, A is the area of the hopper and ρ is the bulk density. The area of the conical hopper is not constant and it is assumed to be a linear function of the height. In addition, since it is assumed that all the material entering the hopper flows out at a constant flowrate, it is safe to assume that the material is no further mixed and according to its mean residence time all properties of the blend will propagate at the output of the hopper accordingly. This is a very simple model implementation of the hopper, aiming to capture the residence time of the material under mass-flow and no mixing assumptions. As pointed out in the conclusions, there are many opportunities to improve this process step by incorporating a reduced order DEM model coupled with material property correlations, such as flowability and flow index.

Tablet compaction and feed frames

Finally, the next process of an integrated tablet production line is tablet compaction, during which small portions of blended powders are compressed into a tablet of desired size and shape between two punches. The most common type of tablet press equipment used in the pharmaceutical industry are rotary tablet presses which have been studied both experimentally and computationally (Hein, et al., 2008; Jain, 1999; Michaut, et al., 2010; Sinka, Cunningham, & Zavaliangos, 2003). Initially, the material, which has usually passed through a hopper, enters the "feed frame" which is a small chamber with rotating blades that fill the dies of the tablet press. The powder in the dies is then compressed. Significant work has been performed in order to model the effect of process parameters of the feed frame on powder properties and die filling (Mendez, Muzzio, & Velazquez, 2010).

The models available in literature can be categorized between two extremes: purely empirical (Akande, Rubinstein, & Ford, 1997; Gonnissen, et al., 2008; Haware, Tho, & Bauer-Brandl, 2009; Michaut, et al., 2010; Seitz & Flessland, 1965) and high fidelity FEM or DEM simulation models. Since the aim of this work is to generate a computationally tractable flowsheet model that can be used for optimization purposes and aims to capture the dynamics of the process, expensive models in the form of FEM or DEM simulations should be avoided or introduced as reduced order models based on the actual expensive simulation. In this work, the tablet press is represented as a simple empirical model, based on the popular Heckel equation (Jain, 1999; Seitz & Flessland, 1965). According to Heckel analysis, the compression force of the powder can be linked to the porosity of the produced tablets, which is a very significant product quality attribute (Equation 5).

$$\ln\frac{1}{\varepsilon} = kP + A \tag{5}$$

where ε represents the porosity of a tablet, P is the compaction force, while k and A are empirical parameters which should be calculated through experimentation for the material which is being processed. The porosity of a tablet is highly correlated to the dissolution and bioavailability of a tablet. The compression pressure has been experimentally observed to follow a first order linear response given a set-point change, and this is introduced into the model in order to capture the dynamics of the process. In order to account for the residence time of the powder in the feed frame, a response surface model is used, which correlates the die disc speed (x_1) and the feed frame speed (x_2) to the average residence time of the powder in the feedframe section (Equation 6).

The model parameters are fitted based on the experimental data obtained in (Mendez, et al., 2010) for the processed formulation (Equation 6) through minimization of the least-squares error between measured and predicted values. The investigated ranges of variables x_1 and x_2 are [29-57] and [24-72] rpm respectively, but they are normalized so that they belong in the same range ([0-1]).

$$\text{Average residence time} = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 \quad (6)$$

It is important to account for the time necessary for powder that is processed upstream to propagate to the tablet press and thus affect properties of out coming tablets. However, evidence from the same experimental study shows that a fraction of the material processed through the feed frame is further mixed according to their residence time, which might affect their flow properties and RSD. In this work, however, it is assumed that no further mixing of the material is performed prior to die filling. It is realized that this is a simple model, which disregards potentially significant information such as the effects of particle size distribution and powder bulk density on the properties of the produced tablets. However, the online prediction of porosity is a significant result, which can be used to evaluate the quality of the produced tablets online.

Wet granulation

Granulation is a size enlargement process which aids in obtaining a product of more uniform particle size distribution and also helps in optimizing the bioavailability of the drug product. The process of granulation is still not very well understood and is carried out with large recycle ratios, thus making the operation extremely inefficient. Many approaches have been into existence which can be used to model such processes

(Salman, Hounslow, & Jonathan, 2007), but population balance modeling is very suitable for describing granulation because of its discrete nature and the ability to depict particulate processes. Although agglomeration of fine particles is most prominent, the final outcome of granulation is a combined effect of the various sub processes like aggregation, breakage, growth and nucleation, acting together. Apart from the particle size, binder content and particle porosity are also very important attributes that affect granulation, so lumping these parameters would not be advisable. The three-dimensional population balance equation for granulation can be written as:

$$\frac{\partial}{\partial t}F(s,l,g,t)+\frac{\partial}{\partial s}\left[F(s,l,g,t)\frac{ds}{dt}\right]+\frac{\partial}{\partial l}\left[F(s,l,g,t)\frac{dl}{dt}\right]+\frac{\partial}{\partial g}\left[F(s,l,g,t)\frac{dg}{dt}\right]$$
$$=\Re_{aggregation}+\Re_{breakage}+\Re_{nucleation} \quad (7)$$

where $F(s,l,g,t)$ represents the population density function and s,l,g are the internal coordinates namely solid, liquid and gas volumes respectively. The partial derivative terms with respect to the internal coordinates represent the various growth terms that indicate an overall increase or decrease in the particle size but there is no change in the number of particles associated with it. The partial derivative with respect to the solid volume denotes layering, which shows the growth due to fines getting deposited on the bigger particles, the partial derivative with respect to the liquid volume expresses the drying or rewetting of particles due to the addition or evaporation of liquid and the partial derivative with respect to the gas volume represents consolidation, which is a negative growth term and is associated with the compaction of granules while gas escapes out of it. The right hand side of equation 7 represent the source terms, comprising of aggregation, breakage and nucleation. Aggregation and breakage are accompanied by a

change in the number of particles. It is important to represent all of the underlying mechanisms in the population balance equation.

Drying

The next processing step after wet granulation is drying of the formed granules. Drying of granulated particles using fluidized beds is a common and important step in the pharmaceutical industry. Compared with other drying methods, fluidized bed drying provides more efficient air–solids contact and, hence, faster drying rate because better mixing and uniformity can be achieved through the process of fluidization. In this work, a simple mathematical model is developed to consider the change of particle size distribution (PSD) during drying after granulation.

Specifically, the evaporation rates of water in all periods of drying can be determined by (Burgschweiger and Tsotsas, 2002):

$$m_{evap} = A_p H_s \rho_g \left(X_s - X_e \right) L \tag{8}$$

where m_{evap} represents the evaporation rate of liquid of each particle; ρ_R represents density of the hot air used for drying; H_s denotes the mass transfer coefficient; L is the length of the dryer; X_s and X_e stand for the saturated moisture content of drying medium on the particle surface and the equilibrium moisture content of the particle respectively. Specifically, H_s can be expressed based on the diffusion coefficient (Radford, 1997).

$$H_s = 2D / \left(10^{-6} d_p \right) \tag{9}$$

$$D = 2.6 \times 10^{-5} \left(\frac{T_p}{273.15} \right)^{1.5} \tag{10}$$

where *D* represents the diffusion coefficient of water; d_p and T_p denote respectively the diameter and temperature of the particle. Finally, X_s is dependent on the saturation pressure of water in the pre-warming and constant-rate period. The relationship between saturation moisture content and saturation pressure is (Wang et al., 2007):

$$X_s = 0.622 \frac{p_{sat}}{\left(p_{sat} - p_{oper} \right)} \tag{11}$$

$$p_{sat} = 101325 e^{13.869 - \frac{5173}{T_p}} \tag{12}$$

where p_{sat} and p_{oper} are the saturate pressure of water and operating pressure of the fluidized bed dryer. Here, it should be noted that it is assumed that the particle size distribution of the granules will only change in the pre-warming and constant-rate periods.

Finally, the parameters used within the current drying model are based on literature and empirical knowledge about the operation of drying. Specifically, the temperature of the particles is assumed to be 348 K; the temperature for the drying air is set to 393 K and therefore its corresponding density is equal to 0.898 kg/s based on ideal gas law. Finally, the operating pressure (p_{oper}) and equilibrium moisture content (X_e) are set at 1 atm and 0.05 kg/kg respectively.

Dissolution and disintegration

Tablet dissolution and disintegration is a key quality attribute of the end product, which is, in most cases, correlated to bioavailability of the API. Therefore the integration of a dissolution component as the final unit of the dynamic flowsheet simulation is valuable since there are several physical tablet and powder properties determined upstream, which affect the drug release mechanism. These properties are tablet porosity and composition, as well as the particle size distribution of the powder after milling. Through this integration, the factors which affect the final tablet dissolution time will be determined in order to be able to identify and propose strategies for mitigating upstream disturbances before they reach the final product.

In most low-dosage drug formulations, the set of excipients added form the bulk of the tablet and control the dissolution mechanism of the entire tablet. On the other hand, interactions between the different components as well as the granule arrangement and homogeneity of the final tablet structure have been found as critical towards the final dissolution time. Dissolution has been studied both experimentally as well as from a modeling perspective in literature, leading to a variety of models ranging from empirical (Costa and Lobo, 2001) , to even two-dimensional or three-dimensional distributed rigorous models (Borgquist et al., 2006; Kimber et al., 2011; Muschert et al., 2009), which allow the investigation of effects such as tablet homogeneity and granule structure to the final release profile.

The model which is implemented is this work is based on the work of Kimber et al. (2011) and it is a first-principle distributed model in which drug dissolution is

captured by Fick's second Law equations (Equations 13-14).

$$\frac{\partial}{\partial t}C_i = -\nabla\left(-D_i \nabla C_i\right) + s_i \tag{13}$$

where

$$s_i\left(x\right) = k_i\left(c_i^{sat} - c_i\left(x\right)\right) \tag{14}$$

where (s_i) is the source term which is non equal to zero when the presence the API component in location (*x*) is true. Otherwise, the source term is equal to zero. The term (C_i) represents the dissolved concentration of each component i, while (D_i) is the diffusion coefficient of each component which is related to the solid fraction of the component as well as its mean particle size based on the following equations:

$$D_i = D_{i0}\left(1 - \sum \varphi_j\right)^{\alpha_i} \tag{15}$$

and

$$D_{i0} = \frac{k_i T}{6 d_{50i} \pi \mu} \tag{16}$$

The first equation ensures that the diffusivity is lower when location x is purely filled with solid component, and as it is filled with liquid (diffuses) the value of diffusivity will reach the bulk diffusivity of the component(D_{i0}). The bulk diffusivity is a function of the temperature (T), viscosity (μ) mean particle size of the granule (d_{50}) and k which is Boltzmans' constant. This equation holds due to the fact that diffusion is the

result of thermal fluctuations in the suspension and is often referred to as Brownian motion (Crowe, 2006).

Once each unit operation model is developed and validated individually, the formation of the aforementioned integrated system is achieved through the creation of a flowsheet model with the desired configuration. The major challenges involved in the integration of unit operation modules of different types and complexity is the identification of the necessary interconnecting variables which should be included in the process streams which connect process i to process i+1. This information should include the necessary properties which characterize the powder stream as well as the required input data for each downstream process i+1 which depends upon the output of process i. In certain cases, certain properties remain unaffected during one processing step, but should be passed on through the process streams since they are considered as inputs to a further downstream process. Once again, it should be noted at this point that even though this procedure may seem trivial for a fluid/gas based flowsheet, it is a great challenge in solids handling where there is still a lot of uncertainty in terms of the properties which characterize the powder material and the models which characterize each process are of different fidelity and detail. In fact, the authors realize that there are still gaps in the integration of specific unit operation steps, due to the non-existence of a critical parameter that is a result of process i in process model i+1. These missing links are described in the following paragraphs and will be the focus of future work, either though the incorporation of empirical correlations or significant model modifications to account for additional effects.

2.2. Integration of models for Direct Compaction

Tablet production through direct compaction is the simplest method involving the least number of processing steps of the raw powders to actual pharmaceutical tablets. The system has as many feeders as the number of components in the pharmaceutical blend, which feed the material into the mixer where it is blended. Subsequently, the blended material is sent to a hopper from which it is fed to the tablet press through the feed frame. This type of processing is used when the raw materials are easy to handle and there are no problems in terms of segregation, cohesiveness and flowability, which are determined through a series of material property characterization tests prior to production.

Through the development of the model library for the unit operations described in the previous section, a dynamic flowsheet model for this case study can be simply built by dragging-and-dropping into a flowsheet and setting material property, design and model parameter values for a specific case study. Here, the production of Acetaminophen tablets is simulated, where the formulation consists of: 3% Acetaminophen (API), 96% Avicel (Excipient) and 1% MgSt (lubricant). This formulation is chosen for enabling comparison of computational results to experimental data as well as the use of empirical correlations from data, which have been collected within the ERC-SOPS for this specific formulation. Lubrication is a very important aspect of tablet manufacturing since the presence and concentration of lubricants can improve flowability and reduce powder adherence to metal walls (Wang, Wen, & Desai, 2010). However, an increased concentration of MgSt can lead to tablets with reduced hardness and lower dissolution rate. In fact it has been shown that the number of revolutions that a lubricated blend undergoes inside the mixer is correlated to the compaction force of a tablet (Kushner &

Moore, 2010). Thus, it is critical to know how much MgSt reaches the tablet press and its history in terms of the mixer residence time at the point it reaches the tablet press.

For any desired formulation, the simulation requires information about the total flowrate or throughput, and the composition of the blend, as well as the particle size distribution and bulk density of each raw powder material. Through this information, the flowsheet can calculate the necessary rotation rates of each of the screw feeders in order to supply the system with the desired powder blend at steady state. In addition, due to the experimentally observed variability of the processed powder flowrates, noise is added to the dynamic feeder outlet flowrates by drawing values from a normal distribution with a standard deviation of 0.05 every 2 seconds of simulation time. This feature can aid in identifying whether this added variability can be handled by the developed models, and how much is filtered at the output stream of the process. However, this noise is completely filtered out by the mixing model, and has no effect to further downstream properties and it is realized that for a detailed analysis of the effects of noise, a stochastic approach should be followed. Due to the assumption of mass flow through the hopper, no mixing of the material is assumed. Thus, properties of the material such as concentration and Relative Standard Deviation are passed through the unit operation with a time delay equal to the mean residence time of the material in the hopper. This assumption of a plug-flow reactor behavior is one of the aspects of this current flowsheet which need further investigation in the future, since the effects of dispersion are not taken into account. . The mean residence time can be calculated approximately equal to the hopper volume divided by the volumetric powder flowrate. Similarly, the same assumptions of no mixing and average residence time delayed properties are incorporated for the feed frame

operation. Finally, the final tablet porosity is controlled solely by the process parameter of compaction force, but the average composition of materials in the tablets is approximated based on the powder material properties exiting the feed frame. Figure 1 is a more detailed representation of the system containing all the variables, parameters and design aspects which should be defined in each process and are passed from one process to the next.

The first assessment of the flowsheet simulation is performed through the verification of the overall mass balances for each unit operation as well as for the overall system, such that all mass that enters from the feeders exits in the form of tablets. A further assumption for the uniformity of the average weight of tablets is necessary for this calculation. However, this type of dynamic simulation can be very useful in identifying how possible perturbations during the operation of a process can affect the output of the same process or further processes downstream. For example, a typical perturbation which will be inevitably encountered during continuous direct compaction is feeder refilling. In fact, strategies and equipment for feeder hopper refilling is a major concern in the powder processing industry (Engisch & Muzzio, 2010), since it is important to balance between flowrate fluctuation, frequency of refilling and equipment cost. Through the developed flowsheet model, however, it is possible to simulate any operating sequence in order to study its effects on final and intermediate product properties. Figure 2 represents the overshoot in the API flowrate which is caused when the hopper is refilled with material at t=1000sec.

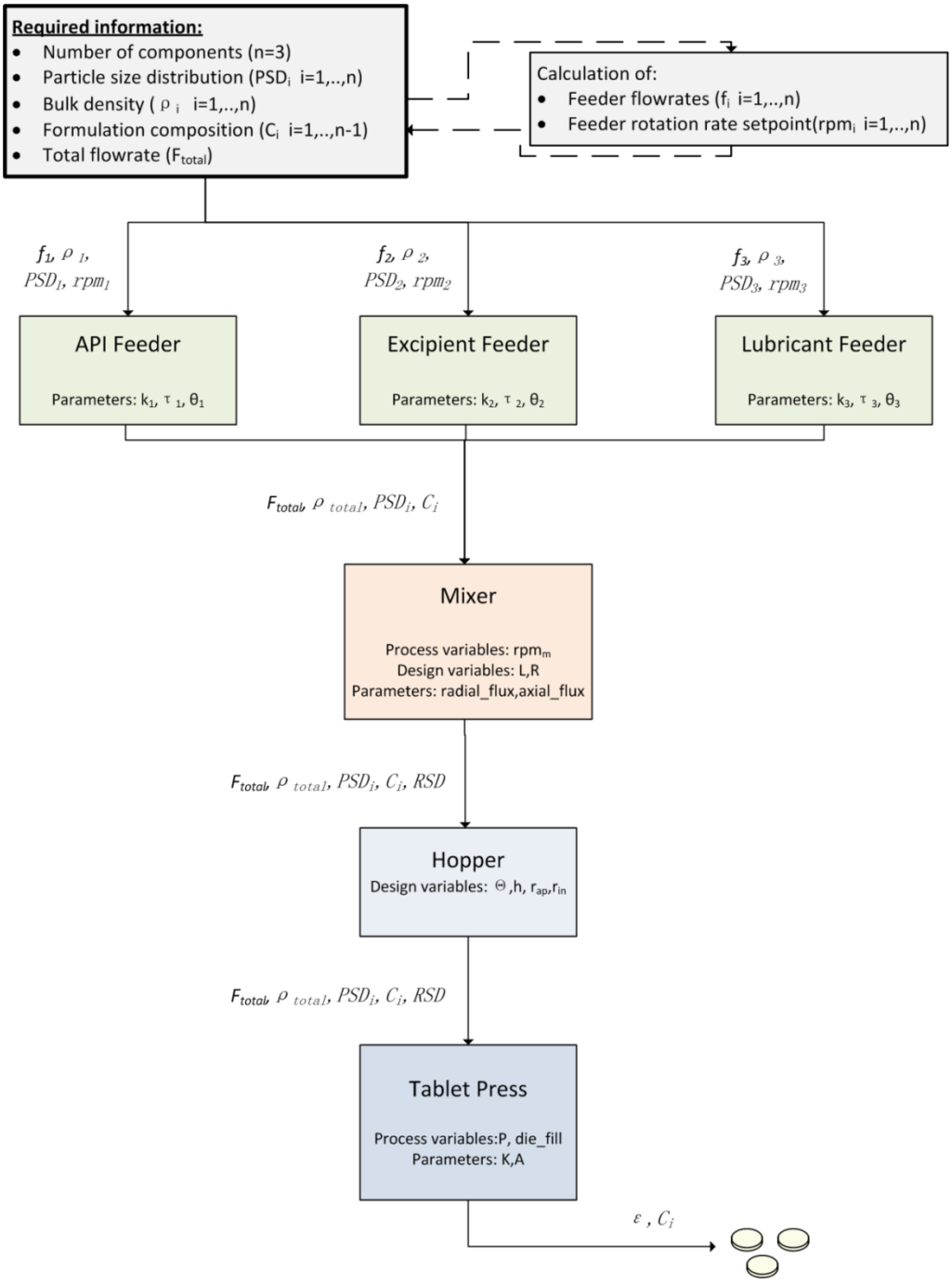


Figure 1. Detailed representation of flowsheet model for direct compaction

The downstream effects of this fluctuation can be seen in Figure 3. Specifically, in Figure 3 the API concentration down to the tabletting stage can be predicted. Thus, based on this

integrated dynamic model, the imposed perturbation causes the API concentration to increase from the desired value of 0.03 to 0.034 within a small time window.

In addition, due to the knowledge of the residence times of the material in the hopper and feed frame, it is possible to approximate which fraction of the produced tablets to discard, if the predicted API concentration is not within the acceptable bounds (CQA).

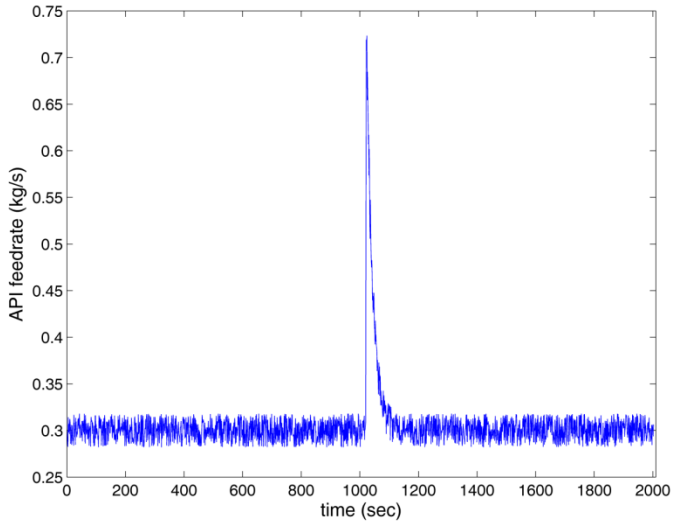


Figure 2. Feedrate of API during simulation of refilling

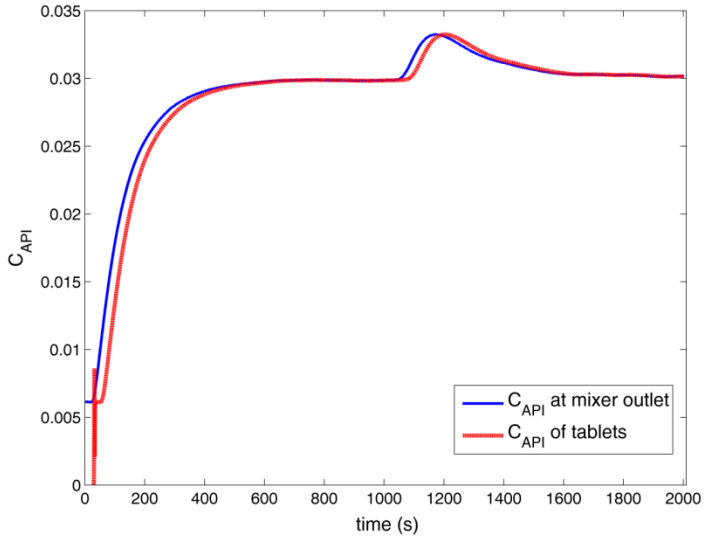


Figure 3. API concentration throughout the different units during refilling

Another advantage of flowsheet simulation is the simulation of different process operation scenarios including recycle dynamics. Thus for this specific case study the effect of adding a second mixing tank in the process through which a fraction of the mixer outlet will be re- processed and sent to the mixer inlet is assessed and it shows promising filtering abilities of feedrate fluctuations. However, this increases the cost of the system significantly and can cause the system response to be slow in response to online control actions.

2.3. Integration of models for Dry Granulation

The second case study is a more complex processing line, which pre compresses the powder blends into thin ribbons which are then milled and then finally compressed into tablets. This production scenario is selected when powder problems such as segregation or aggregation due to cohesion are observed, or in cases where the raw materials are highly sensitive to water, or also in cases where high API content causes the uncompressed powder to be poorly flowable. Due to the particle deformation mechanisms, compression of powders into ribbons, which are then milled to form particles in the 50-400 micron range, produces powders with better flow abilities and more consistent API composition. Following the same approach as in the previous section, the necessary processes are linked to build a flowsheet model, using the model types described earlier (Figure 4). Through this case study, the aim is not only to show the capabilities of the resulting model, but also the problems and gaps of this integration, due to the limits of the available models.

In this case study, the initial stages of feeding and mixing are identical to direct compaction, referring to feeding and mixing. The available model for this process can predict ribbon output properties as a function of process and design variables , but does not consider the composition of ribbons and effects of RSD. Thus, this information enters the process, but cannot be communicated to further downstream processes. A simple assumption of the average ribbon composition (API, excipient and lubricant) to be equal to the composition of the material entering the rolls is used, however, further investigation is necessary for the validity of this assumption. Despite the fact that the integration of these two models is missing certain significant correlations, it can still capture some interesting effects. For example, the integration can successfully capture the effect of perturbations in the mixture composition to the average particle size of the milled particles. In fact, during the same schedule for API feeder refilling as in the first case study, as more API enters the milling process, a small increase in the average particle size is observed (Figure 5). The result of Figure 5, which leads to the conclusion that a small fraction of the milled particles has a higher particle size after a feeder refill is contradictory to what would be expected in reality, due to the smaller particle size of the API particles. However, a possible perturbation in the concentration of the blend, also affects the density of the produced ribbons, which is the main explanation of this result. This result requires both further experimental validation and verifies the need for a more sophisticated roller compaction model, which will be discussed in the future work section. It is very important to be able to accurately predict the average particle size of milled particles, since this will play a very important role in tablet properties, such as hardness and dissolution.

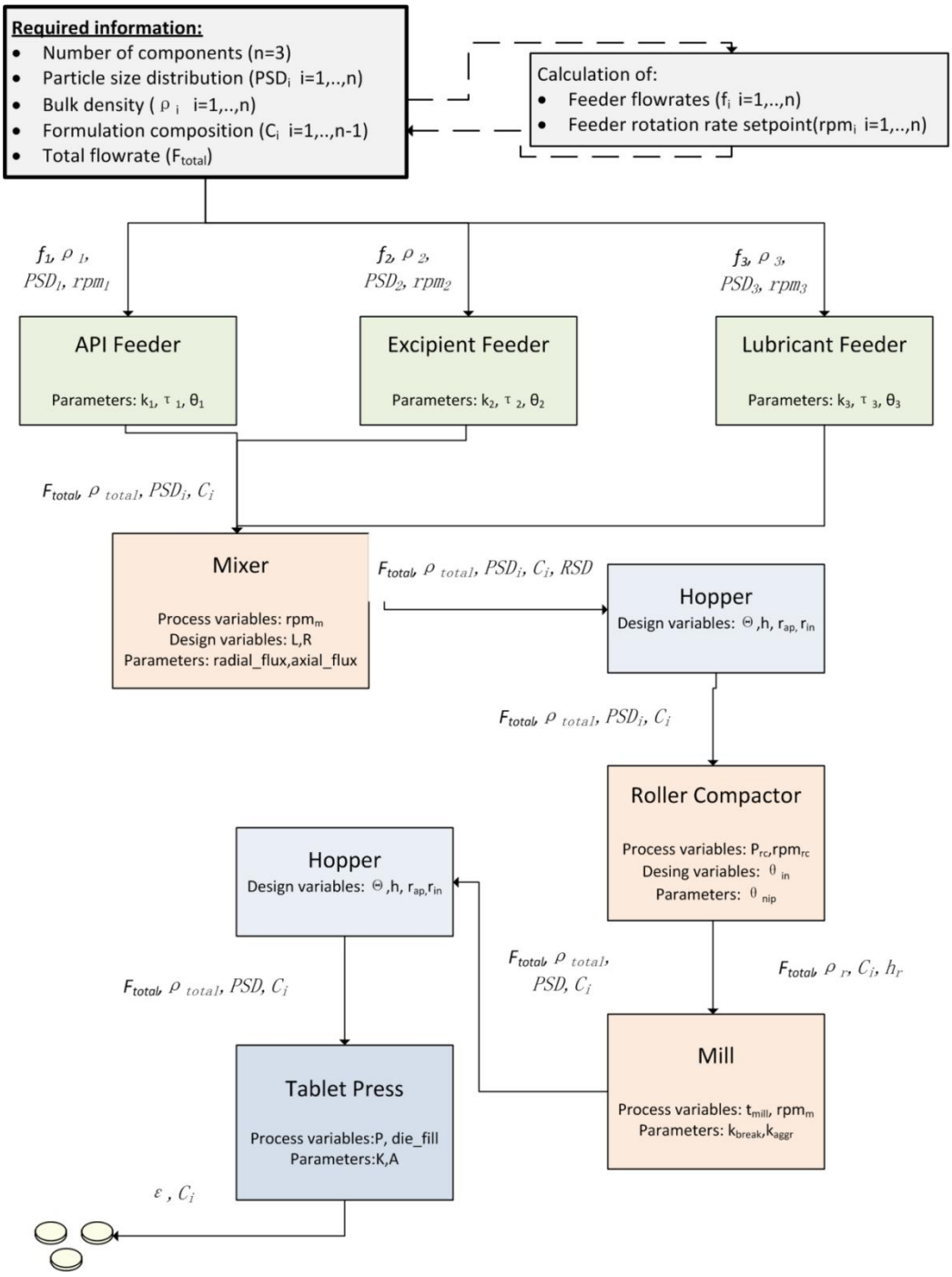


Figure 4. Detailed description of flowsheet model for dry granulation

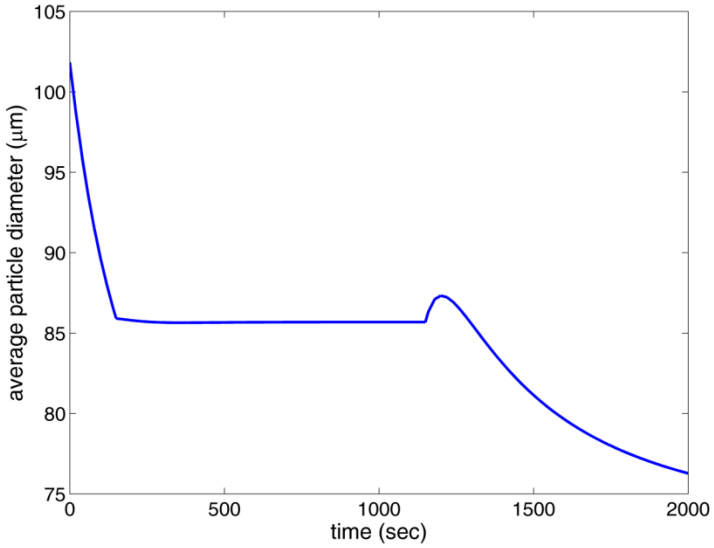


Figure 5. Effect of refilling in average particle size of milling output stream

2.4. Integration of models for Wet Granulation

The integrated system consists of two feeders (API and excipient) that feed raw material into a blender where the components are mixed due to both convective and diffusive forces. The mixture of API and excipient is then continuously transported into a granulator whereby through the addition of liquid binder, the particles are transformed into larger granules. Next, drying of particles is necessary for the decrease of the moisture content which may have detrimental effects in compaction and in the final tablet properties. The dry granules are then passed through a mill in order to achieve the desired particle size distribution and finally, the milled granules enter the hopper which feeds material to the feed frame and the tablet press for the production of tablets at a desired production rate. At the final stage, a dissolution component is included which provides

the characteristic dissolution rate in real time based on the properties of the produced tablets. The integrated design is shown in detail in Figure 6.

Specifically, particle size distribution (PSD), bulk density and mass flowrate are tracked throughout the process across units up to the tablet compaction step. The API concentration which is initially defined in the mixer and then is further affected through granulation, drying and milling, is also a critical property which affects the average concentration of the drug present in the tablets. Moisture content and porosity are introduced after granulation and drying and are critical to downstream processes such as milling, tablet compaction and dissolution. The outlet stream of the tablet press should be expressed in terms of tablets instead of powder, where additional parameters such as tablet hardness, porosity, density and production rate are included. Finally, the critical inputs to the dissolution are the PSD of the powder which reaches the tablet press, tablet porosity and the API concentration, which are all properties defined in upstream processes. In terms of the aforementioned missing links, the most significant deficiencies of the current flowsheet model are the ignorance of the effect of tablet hardness on the disintegration and dissolution mechanism and the effect of moisture content of the powder after drying in all downstream processes and dissolution. Even though it has been experimentally observed that the above quality attributes affect the quality of the final product, the simplified models which are used for the hopper, tablet press and dissolution cannot explain or predict the effects caused the additional variability introduced by these inputs.

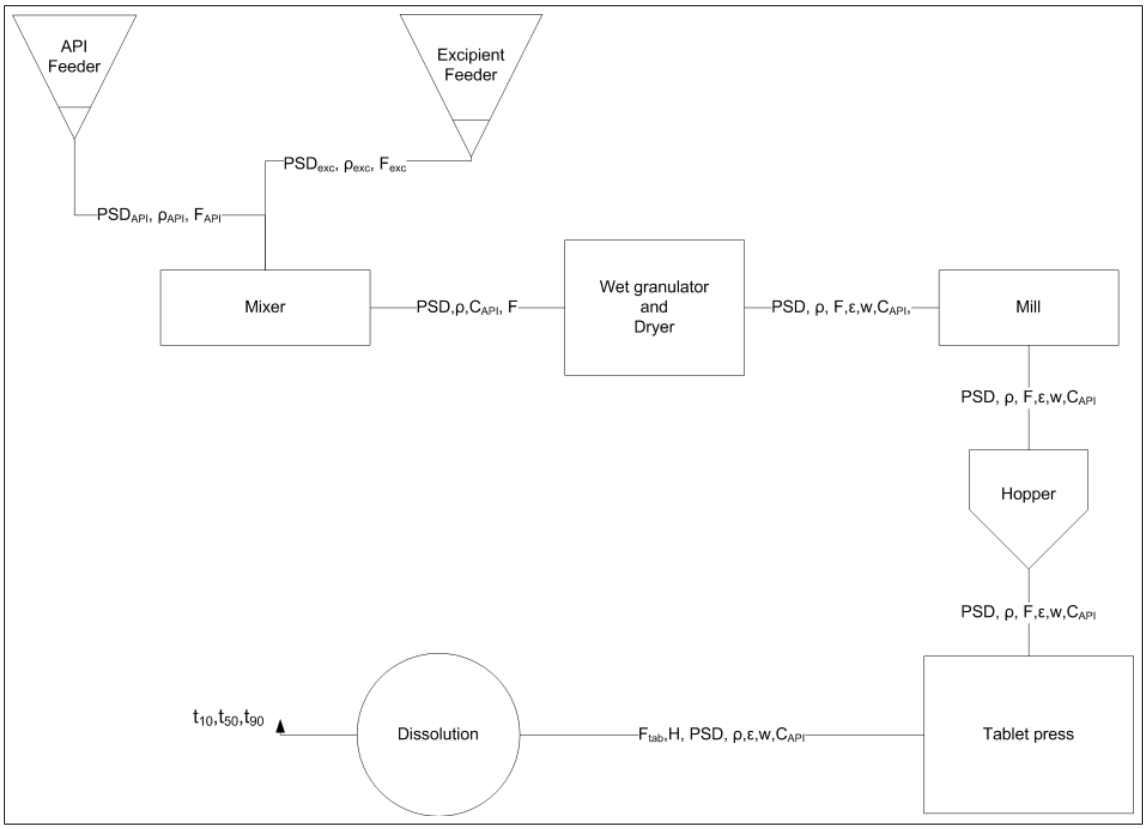


Figure 6: Schematic of the integrated process design for continuous wet granulation

The simulation requires the input of the desired total throughput and mixture composition in order to calculate the required operating conditions of the two feeders which will introduce the correct ratio of material to the entire system. In addition, material properties of each individual component such as particle size distribution parameters and bulk density are necessary input parameters of the simulation. If more than two materials are introduced into the system (i.e. lubricant), then this can be achieved by the addition of more feeders into the flowsheet which is possible in the current setting. However, the prediction of the behavior of multiple materials through the multidimensional population balance models is an aspect that still requires further validation, thus, a two component mixture is described in this case study. Specifically, a

mixture of 30% Acetaminophen and 70% Avicel at a total throughput of 80 kg/h is simulated, producing tablets at a rate of 800000 tablets/h.

Initially, it is important to verify the overall performance of the flowsheet through validating that the mass balances are satisfied throughout the process and the residence times of the powder material inside each process are correctly captured.

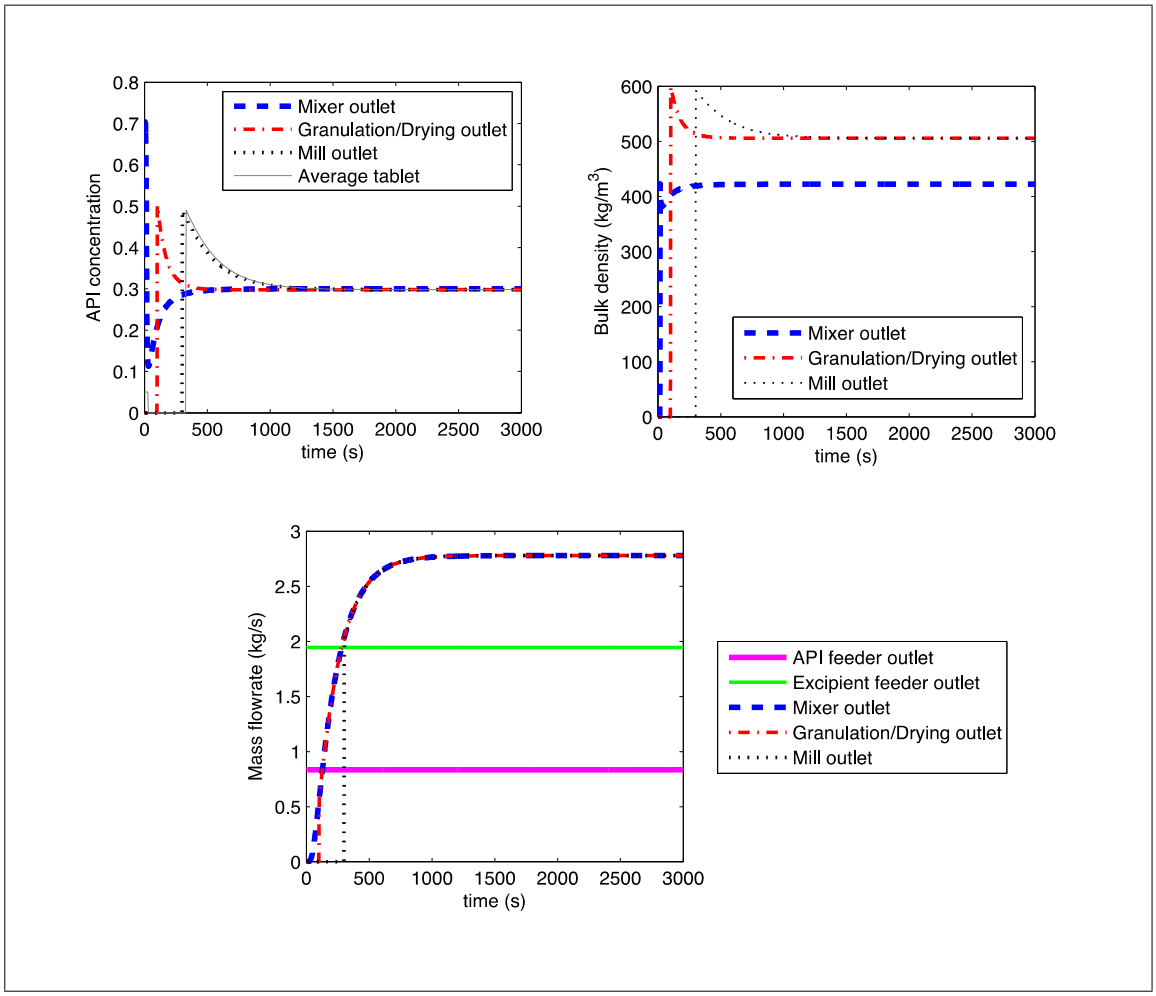


Figure 7: (a) API concentration of process streams, (b) bulk density of process streams and (c)mass flowrate of material, during dynamic simulation

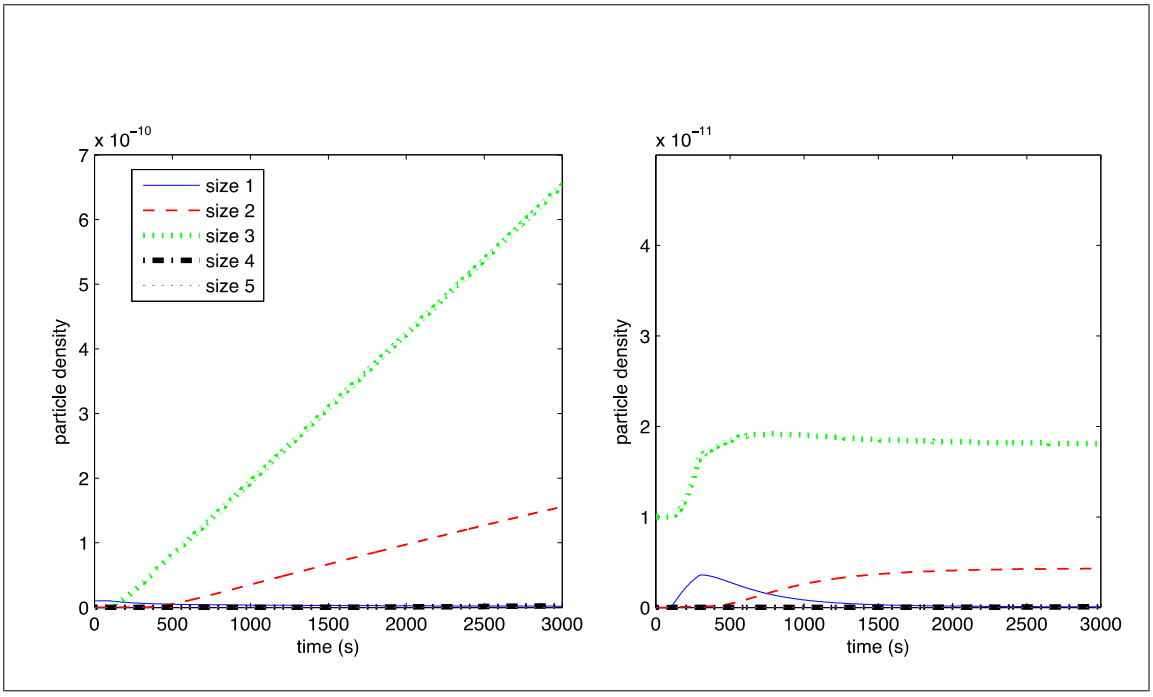


Figure 8: (a)PSD evolution of granules during wet granulation, (b) PSD evolution of particles during milling

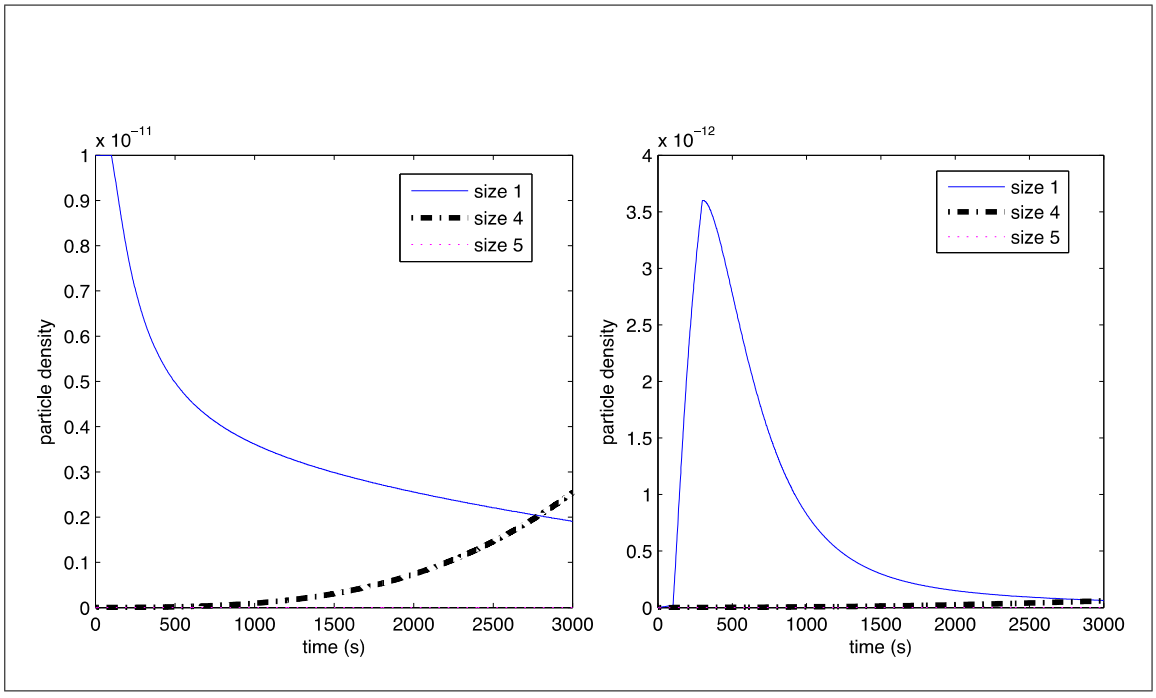


Figure 9: (a)PSD evolution of granules during wet granulation, (b) PSD evolution of particles during milling, only of sizes 1,4 and 5

In Figure 7 the dynamic profiles of the mass flowrate, the API concentration and the bulk density of the processed powder stream are shown at different locations along the process line. It can be observed that the process reaches steady-state at approximately 1100 seconds, where all of the powder streams have obtained constant characteristics. The mean residence times of the granulator and the mill are set to 100 seconds and 200 seconds respectively. In addition, the results verify the assumption of zero loss of mass, since all of the incoming material is equal to the material which reaches the tablet press. Moreover, the results validate the approach which was used to integrate the processes of granulation and milling, since not only the conservation of mass, but also that of the desired concentration of API is achieved. Finally, the model can capture the expected change in the bulk density of the material as it is transformed from dry particles to dried wet granulated agglomerates. Specifically, the bulk density of the material has increased from the point of the blender to the granulator, which can be explained due to the transformation of the particle size of the material. In fact, experimental studies have shown that there are many process parameters, such as wetting conditions and mean residence time, as well as raw material properties, such as particle size composition, which affect the final bulk density of the granulated material. Thus, by choosing the optimal operating conditions, a material with high bulk density can be produced, which is easier to store and transport. In (Gluba, Obraniak, & Gawot-Mlynarczyk, 2004) it was shown that droplet diameter, mean particle size of raw materials and saturation of the granulated bed were the most critical towards the final bulk density of silica flour. In addition, it is certain that the drying process also affects the bulk density of the material significantly, since once the liquid binder is removed, the

particles may retain their granules or break into smaller particles. One of the limitations of this work, however, is the inability to decouple the effects of granulation and drying, since they are modeled as one unit operation. Based on the results it can also be observed that the process of milling does not affect the bulk density of the material significantly in this case, under the specific operating conditions. This fact implies that the configured milling process does not significantly change this material attribute, however, this does not mean that the particle size distribution of the material has remained unchanged, and this will be shown through Figures 8 and 9.

Figure 8 is a representation of the evolution of the frequency of different grades of particle sizes during wet granulation and milling. The size ranges that are most common during granulation are of size 2 and 3, which correspond to the particle sizes ranges shown in Table 2. In fact, based on the used parameter values, the population of size 3 particles is increasing at a constant rate and does not reach steady-state even after 3000 seconds. The remaining finer and larger sizes (1,4 and 5) have negligible concentrations. On the contrary, after the process of milling, the mean particle size has decreased in value, which is reflected by the significant increase in the population of the smallest size particles relative to the total amount of particles (size 1 and 2). There exist a certain amount of size 3 particles, which can be explained by the continuous birth in the previous process step, but their population reaches a steady-state at 2000 seconds, due to their simultaneous depletion and the fraction of this grade of material within the entire population of particles is smaller. The largest grades of particles (size 4 and 5)are depleted at a much faster rate. In Figure 9, only the groups of 1,4 and 5 are included in order to distinguish their behavior better, since their frequency ranges are much lower

compared to groups 2 and 3. During granulation, smaller size particles are depleted and larger size particles are formed. On the contrary, during milling coarse particles are not formed and the finest size particles increase initially, but later are depleted. The above results are encouraging since the desired particle size should lie in the middle size groups after milling, in order to avoid agglomerates but also fines which can cause flowability issues in the hoppers.

Table 2.Grid size bounds for Particle Size Distribution

Bounds	minimum radius	maximum radius
Size1	0.0	5.05×10^{-5}
Size2	5.05×10^{-5}	6.08×10^{-5}
Size3	6.08×10^{-5}	7.10×10^{-5}
Size4	7.10×10^{-5}	8.12×10^{-5}
Size5	8.12×10^{-5}	9.14×10^{-5}

In a continuous operation of a wet granulation line, the refilling of the materials in the feeder hoppers is inevitable and it has been shown that this procedure introduces significant perturbations in the system in the form of a feedrate pulse of the material that is being refilled (Engisch et al., 2010). Experimentally, this pulse has been shown to propagate downstream and therefore efficient control strategies or design modifications (i.e. recycling) are necessary to mitigate this effect. The objective is to correct the perturbation in the composition of the powder stream before it reaches the final step of compaction and it significantly affects the composition and dissolution of the tablets. Thus, it is important to investigate whether the current flowsheet model can capture the

propagation of effects caused by a pulse in the feedrate of the API, since this is the first step in the identification of the optimal strategies which could be used to filter this perturbation further.

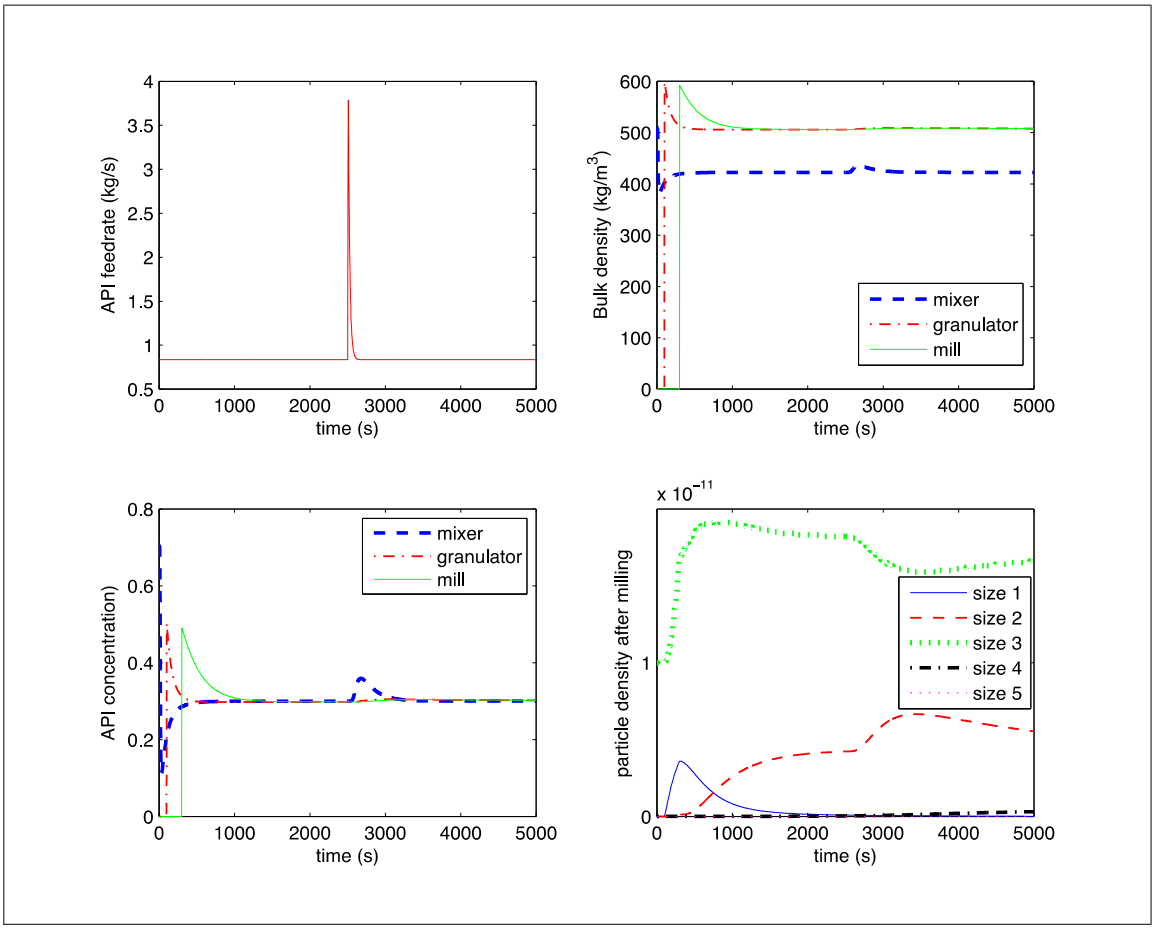


Figure 10: (a)pulse of API material during API feeder refilling, (b)effect of feeder refilling on API concentration in downstream process units

The simulated pulse of the active ingredient material is shown in Figure 10a, and it is of a relatively large magnitude in order to better observe the effects of the perturbation downstream. However, the magnitude of this pulse is realistic, since experimental studies have verified that the size of the overshoot is highly dependent on the fill level of the

gravimetric feeder hopper. According to the model results, the perturbation at 2500 seconds, affects the bulk density of the powder blend exiting the mixer, but this effect is mitigated further downstream during granulation and milling. Similarly, even if mixing process has not filtered out the API flowrate pulse completely, the effects are further filtered out during granulation and milling. In other words, it is not probable to get significant effects on macroscopic properties of granulated materials after such a short time perturbation. This result is very encouraging for the continuous simulated design, however, it needs further experimental validation. Finally, in Figure 10c, the effects of the performed perturbation on microscopic properties, such as particle density are shown to be clearer and more significant. Specifically, after the API refilling perturbation, the larger size particles (size 3) decrease, while the smaller grade of material (size 2) increase in population. This effect suggests a shift in the mean particle size of the milled particles towards smaller sizes.

The performance of each unit operation is controlled by a set of critical operating conditions which will affect the properties of the outlet powder stream and thus affect the operation of further downstream processes and the final product properties. Therefore, it is important that the developed flowsheet model can capture the effects of modifications in certain key operating conditions, based on experimentally observed results. If this is validated, then the flowsheet model can be further used for the identification of optimal control strategies of the integrated granulation line. As an example, the most important in-process parameter which has been studied here is the liquid binder addition rate during granulation.

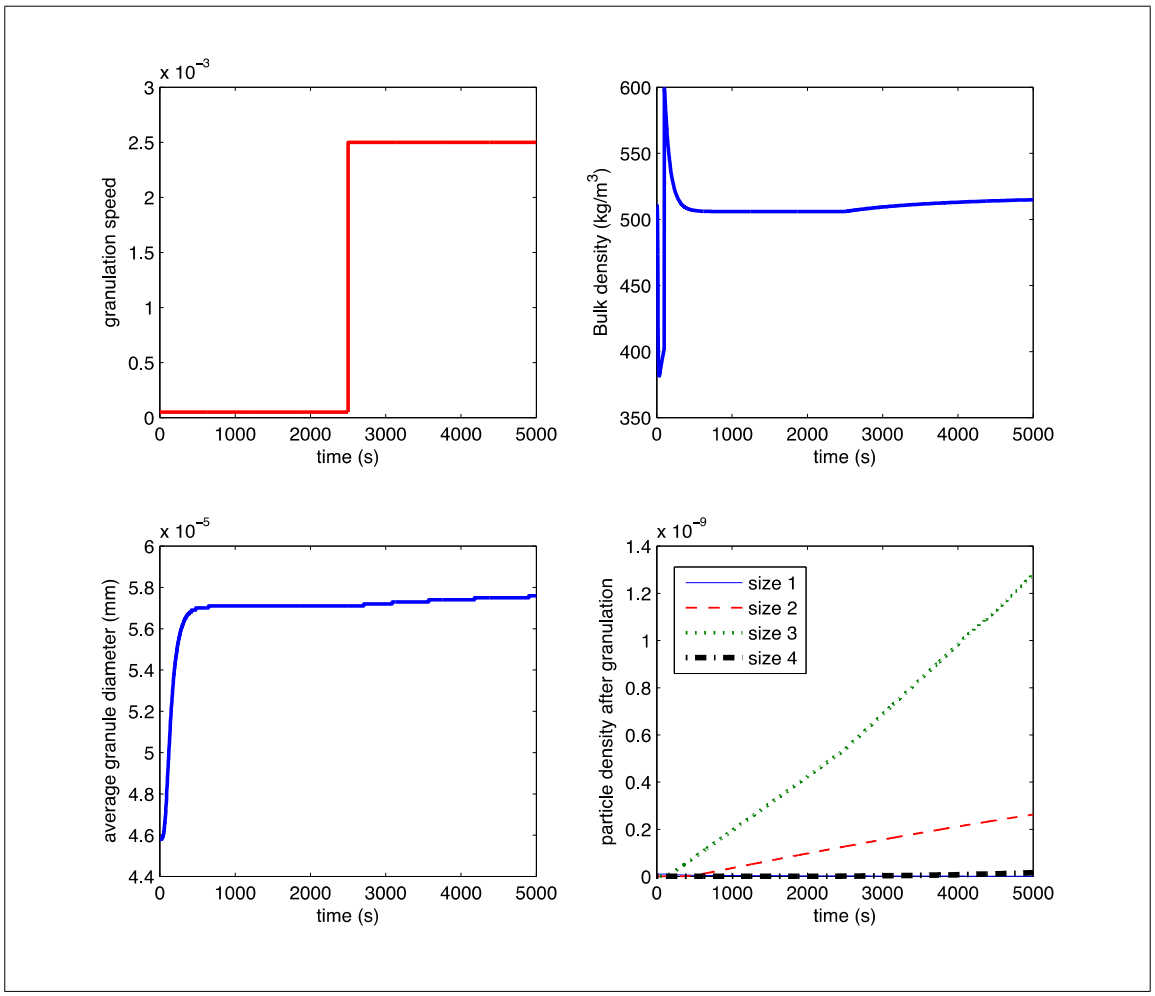


Figure 11: (a)Step change in granulation binder rate addition, (b) effect of binder rate step change on granule bulk density, (c) effect of binder rate step change on PSD of granules

According to (Gluba, et al., 2004) experimental studies have shown that the moisture content of particles, which is dependent on the rate at which the liquid binder is added in the process is very critical towards the properties of the final granules. In another study that was performed using a batch wet granulation process (A.B. Yu, Standish, & Lu, 1995) it was observed that the agglomerate density increased with moisture content, up to the point where it reached a maximum and then the effect was inverted. In Figure 11, it is observed that the simulation results agree with experimental

findings in the sense that an increase in the binder addition rate causes an increase in the bulk density of the produced material. Since this is a continuous operation, where granules are simultaneously removed from the process it is expected that this effect will not have a maximum, but rather reach a new steady-state. In terms of the PSD of the granules, an increase in the binder rate modifies the slope of the birth of larger size agglomerates, such that they are produced at a higher rate. This result shows that the developed granulation model accurately captures the expected favoring of agglomerate formation, when the amount of liquid binder is larger.

The properties of the powder stream reaching the tablet press as well as the properties of the produced tablets at each time interval, highly affect the disintegration and dissolution of each of the produced tablets. In order to quantify the performance of dissolution through a single metric the ratio $\frac{t_{10}}{t_{90}}$ is introduced which corresponds to the ratio of the time needed for the dissolution of 10% of the API over the time needed for the dissolution of 90% of the API (Figure 12). This metric should lie between strict and specific upper and lower bounds based on the specifications of the produced tablet, which are the most critical especially in controlled release products. Any tablets which are produced which violate these bounds should be discarded, hence it is very valuable for a simulator to be able to predict the exact time interval within which a fraction of the continuous production should be diverted. In Figure 13, it can be seen that the dissolution profile metric of the tablets produced while feeder refilling and after a step change in the granulation operating conditions is affected. In the first case, the feeder refilling has affected the tablet properties after a significant time interval and is shown to have a wide time range effect. In other words the effect of the instantaneous pulse has been delayed

and dispersed over time, at the point it has reached the tablets. The slight increase of the $\frac{t_{10}}{t_{90}}$ index signifies that the API has been released faster, which can be explained by the slight decrease in the mean particle size of the powder which reaches the tablet press. On the other hand, after the step change in the binder amount, and due to the slight increase in the mean particle size of the granules, a slight decrease in the dissolution metric is observed. Even though the magnitude of the effects is not large, it is a matter of how strict the product quality specifications are in order to decide whether the produced tablets are Out of Specitications (OOS).

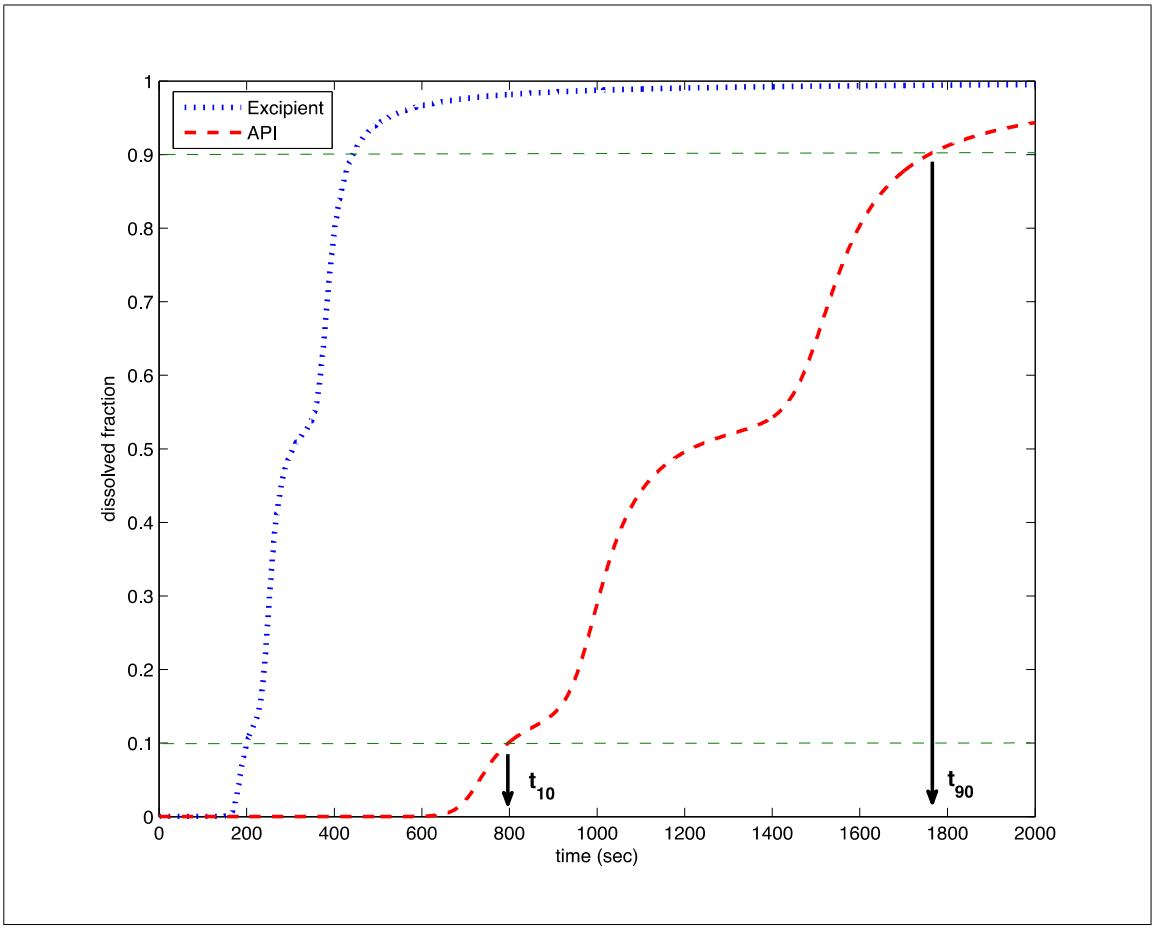


Figure 12: Dissolution profile of API and excipient for steady-state operation

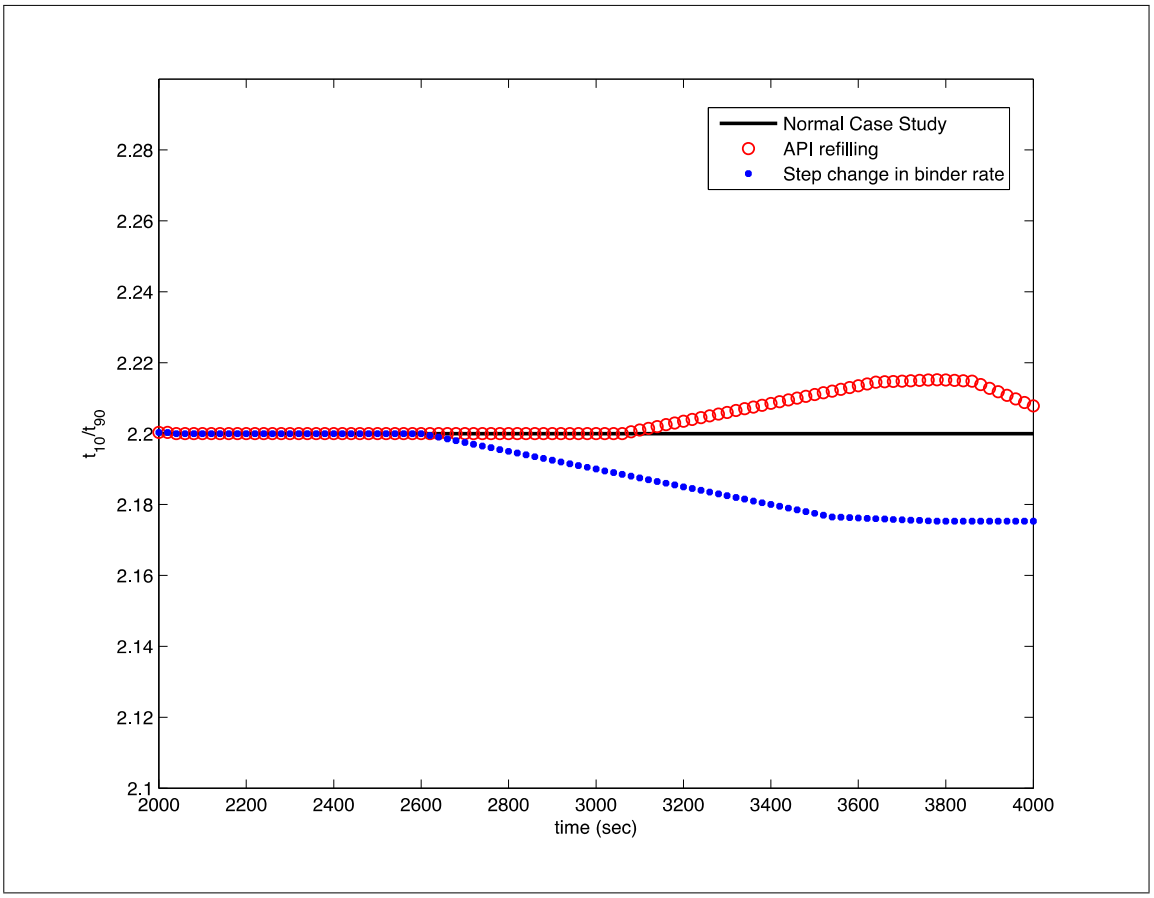


Figure 13: $\frac{d_{10}}{d_{90}}$ for normal steady-state operation, during feeder refilling and after step change in binder addition rate.

In conclusion, this chapter is focused on the description of the types of models which are available for modeling powder processes and their implementation and integration to form continuous tablet manufacturing scenarios. However, due to the lack of first-principle knowledge and process understanding, certain unit operations are described by very simple models, for which the connectivity with upstream and downstream processes is not very realistic. There are certain cases, where a unit operation model does not capture all of the effects the process has on the properties of the powder stream. In other cases, the inputs of one process model does not even include

the entire set of outputs of a previous process, which leads to the incorporation of the specific variable as a delayed response, which remains unchanged through a process, however, the powder stream which enters the process will exit the process after a specific residence time.

Chapter 3

3. Model improvement through the integration of reduced-order models

3.1. Response Surface Techniques

Response surface methods refer to a set of different basis functions with a number of parameters, which are tuned based on experimental or computational data, in order to explain the correlation between a set of inputs to a set of outputs. Response surfaces are used in many cases where the a first-principle model is not available, or when a model of a process is in the form of a proprietary non-access code, or finally in cases where the process model is in fact known but very expensive to evaluate. In the last case, it is common to employ response surface techniques, and develop a faster approximation model of the expensive one, based on a set of designed samples. Based on the final goal of response surface building, the required data-base might have different forms. If the purpose is to accurately represent the underlying process over the entire range of the input space, then carefully designed experiments which must sufficiently fill the whole experimental region must be performed, based on which the response surface is built. If, however, the final goal is the identification of a single optimal value, then local response surface modeling techniques have been developed since it is realized that the accuracy of the response surface must be good only in regions which are likely to contain an optimum. This minimizes sampling to a smaller set of regions of the entire investigated space and subsequently minimizes the required sampling overall. In chapter 4, it is shown that response surface techniques may be used to approximate the Design Space of a

process, so sampling in this case is targeted towards regions where the probability of the existence of a boundary is high. However, the challenge in these cases is relocated towards defining the search criteria in order to identify regions in which the probability of optimality or boundary existence is high. This challenge is even greater when the process is stochastic (presence of noise) and there are multiple process constraints which affect the overall feasible region.

There are many advantages and disadvantages to response surface techniques. Response surfaces are not predictive outside the range of the data which was used to build the model, or in other words, they are not good extrapolators. This should be taken into account whenever a DoE is designed, or whenever such a model is used for a prediction. In addition, the model parameters usually have no physical meaning, while sometimes they can provide information about the variable, linear, quadratic or interaction term which they are multiplied by. However, if there is no knowledge at all about the true underlying physics, one may build a very good model by simply adding more terms and parameters. This however, would lead to over fitting the data, which often captures the noise of the measurements and not the true correlations and trends. In most cases, even if a first-principle model is not available, empirical knowledge is available for a process, thus the predictions and conclusions for a reduced-order model can be compared to such knowledge in order to test its validity.

On the other hand, response surface methods are very fast to implement, build and use for multiple purposes, such as local prediction, optimization and control. In many fields of industry, an engineer cannot wait for the time it is required to develop a first-principle model, according to the scientific knowledge that is available at the time, thus

turning to these models is definitely better than simply empirical approaches. If the samples are collected based on a good design, the experiments are performed meticulously, and the appropriate basis function is chosen such that it does not overfit the data, then the developed model can be a very powerful tool.

There exist multiple types of response surfaces, such as Response Surface Methods (quadratic polynomials), Kriging, Neural Networks, High Dimensional Model Representations, Radial Basis Functions and many more, which are used in the literature for multiple purposes and applications. Most of these methods are available through commercial packages and anyone can simply enter the input- output data and get the model parameters, predictive ability and statistics of the model. The performance of different techniques has been compared for different datasets in the past by us for both steady-state and dynamic data (Boukouvala, et al., 2010; F. Boukouvala, et al., 2011), however, the final choice is user-dependent and should be made based on the type of data (noisy or deterministic) and the expected performance of the true input-output correlation. In this dissertation, only the two most common and useful to us types of functions are described in detail, Response Surface Methods and Kriging.

In the pharmaceutical industry, based on insofar experience, as soon as QbD initiatives started to be implemented, there was a very high interest in just using such models to improve processes. This was done to such an extent, that today the meaning of response surface has lost its value. It is no where advised to employ response surface models, when there is a chance that a first-principle model can be developed. However, as a complementary tool to the first-principle models, reduced-order models can enhance the understanding, capture more missing correlations and speed-up the design. This,

however, should not hinder the drive for obtaining more knowledge about the true mechanisms and physics that govern a specific phenomenon or process. What was also found, is that often in powder based processes, the only way to explain the behavior of particles is through detailed particle simulation software, which lead to first-principle models which are highly computationally expensive. These models are impressive and can integrate multiscale information, from the particle level to bulk powder properties. However, these models could never be used for any purpose which would require multiple function calls, such as optimization and control. It is then when again one can turn to response surfaces.

When building a direct input-output response surface mapping to data, it is commonly assumed that the set of inputs are independent to each other, as well as the set of outputs. This is usually the case when the data comes from a designed experiment, after the performance of screening studies and analysis of variance, in order to identify the most important variables which explain the variability in a process output. However, in cases of historical data, or high dimensional data which are not based on a designed experiment, there is a high probability that the data is correlated and thus latent variable model techniques are more appropriate in order to identify not only input-output correlations, but also the structure and true dimensionality of the data. These approaches are not classified under the Response Surface techniques, even though they share some common characteristics due to their data-based nature- and will be discussed later in the chapter.

Sampling methods for physical and computational experiments

Distinctions and similarities can be found between data which comes from a physical experiment and data that comes from a computer simulation. Physical data are always subject to experimental error (human error, systematic error, random error), which cause the same exact experiment to lead to slightly different results when performed multiple times. It is this noise that response surfaces should filter, if they are built with the right number of parameters. Replication of experiments is also desired, in order to approximate the true level of noise for a specific measurement. Computer simulations may also be noisy, due to numerical instabilities and/or rounding-off errors of computer calculations, however, in most cases this numerical noise is systematic and the computer simulation will give the same output when it is run for the same exact input conditions. In this case, replication would be completely unnecessary since it would not add any new information to the problem and add computational cost. There are however cases, where a simulation can be made stochastic, by allowing several input conditions to take values from their known distributions each time a call is made to the simulation. This is an approach which is followed for the optimization of flowsheets, since it is desired to take the effect of the variability of powder materials. Overall, the amount of 'smoothing-out' ability of a response surface should be related to the level of noise present in the data. If it is desired to completely interpolate the measurements coming from a computer simulation, then the response surface should be interpolating.

The designing of the experimental plan based on which the samples from a process model or actual process will be collected, is another important step in response surface building. Throughout this work, whenever a response surface model is built, it is

assumed that screening techniques have been already performed in order to identify the important variables for a process output. Once the dimensionality of the problem has been identified, the user has multiple options for designing an experiment which should have good space-filling properties, uniform spread throughout the investigated space, good spanning properties, poisedness, all of which are properties of DoEs used in different literatures (statistics, engineering, optimization)(Conn, Scheinberg, & Vicente, 2009). In physical experimentation, full- factorial designs are often used which require the sampling of every possible combination of every design variable value. Due to the curse of dimensionality, full-factorial designs are not used in cases of very high dimensional problems, which lead to the design of fractional factorial designs and many more which have different optimality criteria (Forrester, Sobester, & Keane, 2008).

Another common option for designing an experiment is Latin Hypercube Sampling (LHS) , which aims to span a space as well as possible by maximizing the sum of differences between each chosen sample, given a maximum number of samples. These designs are used more often in computer simulations and can further reduce the sampling cost when compared to a factorial approach. The objective of LHS is to generate points whose projections on all of the variable axes are uniform. Compared to factorial designs, LHS designs do not usually generate points at vertices of the experimental region. LHS are random in nature, meaning that based on the objective used, there are multiple combinations of points which satisfy the criterion, thus a different set of points is generated every time a LHS algorithm is called.

The choice of the design should always be subject to the purpose of the collected samples and the type of response surface used to fit to the data. In other words, if a

quadratic model is fit to the data, multiple level design points are necessary, increasing computational cost. Latin Hypercube samples are usually chosen as initial designs for a multidimensional problem due to their good spanning abilities, in order to get a first cheap approximation of a true function. However, they are usually followed by further sampling in regions of interest since the initial model approximation cannot be trusted.

Response Surface Methodology

Response surface methodology (RSM) was first introduced by Box and Wilson in 1951 and is a tool that has been widely employed for the optimization of noisy processes. There are three basic steps to the RSM algorithm: 1) specification of a sampling set within the investigated region, accomplished with design of experiment tools (DoE), 2) construction of a model through optimization of the model parameters based on the experimental data , 3) model optimization in order to determine the location at which process improvement is maximized. The basis function chosen for a response surface may be as simple as linear, with or without interaction terms or it may have quadratic and even cubic terms, depending on the nature of the output surface. Theoretically, the type of basis function may be of any functional form and it is the decision of the user which should be based on the available knowledge about the process. This decision greatly affects the DoE which should be chosen. If one can afford to run a multiple-level DoE, then out of all the regression coefficients for all the candidate basis functions, only those terms that are found to be statistically significant using F-tests for lack-of-fit and prediction error sum of squares are retained. A general second-order response surface model with interactions has the following form (Equation 17):

$$y = \beta_0 + \sum_j \beta_j x_j + \sum_{i<j} \beta_{ij} x_i x_j + \sum_j \beta_{jj} x_j^2 + \varepsilon \tag{17}$$

where x_j are input variables, β_0 β_j, β_{ij}, and β_{jj} are model coefficients, and y represents the response, or predicted output behavior.

RSM models are used very frequently for pharmaceutical process design and product design, when a first-principle model is not available and a DoE data set is performed to collect the necessary information. Even in optimization, quadratic models have been used often as approximations and for calculation of derivatives in the vicinity of a local optimum, due to the fact that the behavior of a model is quadratic within this region (Taylor expansion) (E. Davis & Ierapetritou, 2008). However in many cases it has been observed that quadratic response surfaces fail to capture the true response of a function, especially in cases where the underlying correlation is non-smooth and non-convex within the region of interest(Donald R. Jones, Schonlau, & Welch, 1998). In addition, since the number of sampling points needed to construct reliable response surfaces can increase exponentially as the dimensions of **x** increase, response surfaces have been used mainly for low-dimensional problems.

Kriging

Kriging was first developed as an inverse distance weighting method to describe the spatial distribution of mineral deposits- specifically it was designed to interpolate random fields and it belongs in the linear least-squares estimation category of methods (Cressie, 1993).The introduction of kriging to the optimization literature by Sacks (REF) and the use of the model's ability to identify regions which have been poorly sampled towards global optimization are the main reasons which have made it so compelling to various other literatures and mostly optimization. In addition to this, the interpolating abilities of kriging with a low number of parameters, has made kriging very popular to a

specific range of problems which depend on deterministic computer data. The concept of kriging is similar to many other regression techniques, and it is that the value of any new unsampled point is a weighted function of the values or nearby samples, where the weights are dependent on the relative spatial location between sampled points. Having this in mind, the main difference of kriging, to an ordinary least squares regression, is that the error between all the already existing samples is modeled explicitly via an assumed functional form (i.e. exponential, linear, cubic etc). The assumption in kriging, also referred to as Gaussian process modeling, is that an output is given as a distribution with a mean equal to the kriging prediction and standard deviation equal to the kriging error (Forrester, Sóbester, & Keane, 2008). The kriging variance provides information about regions where subsequent sampling is required, and that is the key component of kriging that has been exploited in optimization. The popularity of kriging has made it somehow a benchmark in surrogate-based optimization and reduced-order modeling, and this is one of the reasons why it has been used in this work in various applications. However, as it will be seen in the following sections, the choice of the most appropriate method for metamodeling is a user-defined task and it is highly dependent on the application.

In order to build a kriging model, a number of sampling points must be collected (n) based on the chosen experimental design procedure. The sample data forms the input matrix $\mathbf{X} = \left[\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, .., \mathbf{x}^{(n)}\right]^T$, where each $\mathbf{x}^{(i)}$ vector has m elements, equal to the number of inputs. The observed responses are denoted as $\mathbf{y} = \left[y^{(1)}, y^{(2)}, .., y^{(n)}\right]^T$. The case of one output for building a kriging model will be described, since if there are multiple independent outputs, one kriging surface is built for each one. The observed

responses can be considered as random variables (Equation 18), correlated with each other based on Equation 19:

$$\mathbf{Y} = \begin{bmatrix} Y(\mathbf{x}^{(1)}) \\ \vdots \\ Y(\mathbf{x}^{(n)}) \end{bmatrix} \tag{18}$$

$$cor\left[Y(\mathbf{x}^{(i)}), Y(\mathbf{x}^{(j)})\right] = \exp\left(-\sum_{l=1}^{m} \theta_l \left| x_l^{(i)} - x_l^{(j)} \right|^{p_l} \right) \tag{19}$$

The covariance matrix of the n measurements is equal to:

$$Cov = \sigma^2 \begin{pmatrix} cor\left[Y(\mathbf{x}^{(1)}), Y(\mathbf{x}^{(1)})\right] & \cdots & cor\left[Y(\mathbf{x}^{(1)}), Y(\mathbf{x}^{(n)})\right] \\ \vdots & \ddots & \vdots \\ cor\left[Y(\mathbf{x}^{(n)}), Y(\mathbf{x}^{(1)})\right] & \cdots & cor\left[Y(\mathbf{x}^{(n)}), Y(\mathbf{x}^{(n)})\right] \end{pmatrix} = \sigma^2 \boldsymbol{\Psi} \tag{20}$$

where σ^2 represents the variance of $\mathbf{Y}$ in equation (Equation18). In kriging, the correlation of $\mathbf{Y}$ is modeled using the basis function of Equation 19 assuming that the true function which is modeled behaves in a smooth and continuous way. The parameters of kriging θ_l and p_l are fit to capture how the distance between two points $\left| x_l^{(i)} - x_l^{(j)} \right|$ affects the change in $\left| Y(\mathbf{x}^{(i)}) - Y(\mathbf{x}^{(j)}) \right|$. This form of basis function has the desired characteristic that when the distance between two points is almost zero, their correlation is close to 1 (Equation 21), since their values should be very similar. On the other hand, their correlation is close to zero when their distance increases. Parameter p captures how smooth this change is happening while parameter θ captures how fast this change is happening. For example, a value of p=2 produces a smooth change in the basis function as the distance increases (resembling a quadratic response), while further reducing p leads to a steeper initial decrease in correlation as the distance increases. Parameter θ captures the extent of a sample point's influence with distance. A very low value of θ

signifies similar values of correlation across all points. Parameters θ and can be different for every dimension of **x,** meaning that if a specific input parameter is not that significant, it will have a lower value of **θ**. For this reason the absolute value of parameter θ, can be used as a measure of how important the input variable is.

$$\exp(-\theta\left|x^{(i)}-x^{(j)}\right|^{p}) \rightarrow \begin{cases} 1, \text{ when } x^{(i)}-x^{(j)} \rightarrow 0 \\ 0, \text{ when } x^{(i)}-x^{(j)} \rightarrow \infty \end{cases} \quad (21)$$

Modeling of both parameters p and θ, in all dimensions increases the number of parameters in the model, and complicates the optimization problem of parameter estimation based on the sampled data. In general, it is assumed that p is constant throughout the entire space, but θ may vary with dimension.

The next question is how do one estimates the parameters of kriging, and this can be done via two methods: maximum likelihood estimation, or variogram model fitting(E. Davis & Ierapetritou, 2007, 2008, 2009). The former aims to identify θ and p in order to maximize the likelihood of y which is the objective function of kriging, while the latter uses the variogram plot which is **y** versus the distance $\left|x^{(i)}-x^{(j)}\right|$, based on which various forms of the basis function are fitted through least-squares error minimization. In this work, both approaches have been used to estimate kriging parameters and found that MLE is more reliable, however it is necessary to ensure the optimization problem is solved efficiently. The cost of the computation is due to the matrix algebra which is necessary for computing the likelihood, which requires matrix inversions.

Assuming the errors between **Y** (Equation 18) are independently randomly distributed and after certain amount of matrix algebra, the optimum $\hat{\mu}$ and $\hat{\sigma}^2$ which maximize the maximum likelihood function of **y** given the test samples **X** are:

$$\hat{\mu} = \frac{\mathbf{1}^{\mathrm{T}} \boldsymbol{\Psi}^{-1} \mathbf{y}}{\mathbf{1}^{\mathrm{T}} \boldsymbol{\Psi}^{-1} \mathbf{1}}$$
$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{1}\mu)^{\mathrm{T}} \boldsymbol{\Psi}^{-1} (\mathbf{y} - \mathbf{1}\mu)}{n} \tag{22}$$

where Ψ is a function of parameters **θ** and **p**.

In order to predict a new point using an already existing kriging model, it is now required to maximize the likelihood of the sample data and the prediction, assuming the correlation parameters are sufficient to describe the behavior of the new point. In order to solve this problem, the vector of correlations between all sampled points and the new point is formed as:

$$\boldsymbol{\psi} = \begin{bmatrix} cor\left[Y(\mathbf{x}^{(1)}), Y(\mathbf{x}^{\mathbf{new}})\right] \\ \vdots \\ cor\left[Y(\mathbf{x}^{(n)}), Y(\mathbf{x}^{\mathbf{new}})\right] \end{bmatrix} \tag{23}$$

Thus it is required to maximize the likelihood of the augmented matrix

$$\boldsymbol{\Psi}_{aug} = \begin{bmatrix} \boldsymbol{\Psi} & \boldsymbol{\psi} \\ \boldsymbol{\psi}^{\mathrm{T}} & 1 \end{bmatrix} \tag{24}$$

which leads to the final prediction of $\hat{y}(\mathbf{x}^{(new)}) = \hat{\mu} + \boldsymbol{\psi}^{\mathrm{T}} \boldsymbol{\Psi}^{-1} (\mathbf{y} - \mathbf{1}\hat{\mu})$. Matrix **Ψ** is in general positive-definite and symmetric, however, it can become close to singular if multiple points are collected which are clustered in a region. In the developed algorithms, it is avoided by two ways, firstly not allowing a new sample to be used in the database for kriging if it is closer than a tolerance distance to any of the already existing points, and secondly- in cases of noisy and/or stochastic data for which replication is performed- by allowing kriging to be non-interpolating which requires a modification to the **Ψ** matrix.

Multiple versions of Kriging algorithm have been used throughout the years of this research, based on work of Davis et al. (E. Davis & Ierapetritou, 2007) and the

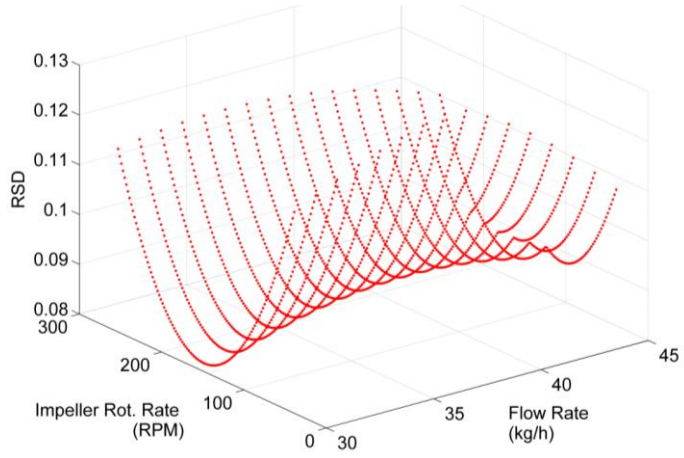
DACE toolbox (Lophaven, Nielsen, & Sondergaard, 2008) which are modified to solve the parameter estimation optimization problem using the TOMLAB LGO toolbox which was found to be very reliable. One final modification made to kriging DACE toolbox is its ability to be a regressor instead of a pure interpolator, by the addition of a constant w to the diagonal of matrix $\boldsymbol{\Psi}$. This is also found in the kriging literature as the so called nugget effect (Cressie, 1993) which causes the correlation of two points which are very close to each other to be $1+w$ instead of 1. The magnitude of w should reflect the amount of noise in the data, since it controls the smoothing ability of kriging. In addition, based on recent work of (Yin, Ng, & Ng, 2011) the nugget effect parameter can also be a function of $\mathbf{x}$, in cases where the level of noise present in the data is higher or lower in different parts of the input space. One can employ this concept to search for optima, since in areas where the uncertainty is higher, the sampling may be denser to get a more reliable prediction of the response.

Response surface techniques are used in many parts of this dissertation, each time for a different purpose. Firstly, in many cases one is provided with experimental data from designed experiments for the characterization of unit operations such as continuous mixing and feeding. A comparison of the performance of each method for the prediction an output for which no existing first-principle correlation exists is performed in (Boukouvala, et al., 2010).

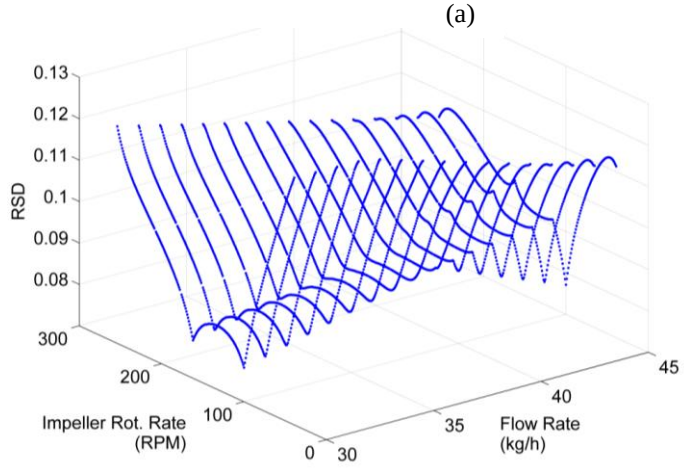
As an example, a set of experimental data is used to correlate the impeller rotation rate and total powder throughput of a continuous mixer to the Relative Standard Deviation (RSD) of the outlet mixture API concentration using an interpolating kriging

model and a Response Surface model (Figure 14 a and b). It is obvious that the performance of the RSM model for the specific experimental data set is better, since the interpolating nature of kriging which requires the model to pass through the exact experimental noisy data reveals a behavior that cannot be explained in reality. The performance of the blender is observed to be best at mid range rotation rates, due to the balance of the effects of radial mixing and axial mixing within the blender. Thus RSM is due to its smoothing ability captures the true behavior of the process.

.



(a)



(b)

Figure 14. (a) RSM prediction, (b) Kriging prediction of RSD in the blender output as a function of Impeller Rotation Rate and Flow Rate.

Response surface modeling and specifically Kriging is also used in this work for developing an optimization strategy for computationally expensive simulations, of multiple dimensions and unknown complexity (Chapter 5).

3.2. Latent Variable Models

As mentioned in the introduction, a different set of modeling techniques, those who correlate latent variable spaces are extremely important when the set of inputs and outputs to be correlated are high dimensional and possibly not independent. Specifically, Principal Component Analysis (PCA) and Partial Least Squares (PLS) modeling are also employed in various parts of this work(S. Wold, Martens, & Wold, 1983; Svante Wold, Sjöström, & Eriksson, 2001, 2002). In the case where a large set of input variables is present, PCA is used to decompose the data into a new coordinate set (scores) and loadings, which correspond to the weights of each of the original variables to the new defined reduced space. PCA has been used in various disciplines for model reduction including random variables, image processing, signal analysis, structural analysis, data compression, process identification and control (Aquino, 2007; Christofides, 2001; Cizmas, et al., 2003; Gay & Ray, 1995; Liang, et al., 2002; My-Ha, et al., 2007; Tabib & Joshi, 2008; A. Varshney & Armaou, 2008; Amit Varshney, Pitchaiah, & Armaou, 2009). In cases of dynamic data, PCA has also been named Proper Orthogonal Decomposition (POD), however, the theory behind these two methods is identical. In the case where output data is also present and is required to be correlated to input data, PLS is used to

correlate a reduced latent space of X, to a reduced latent space of Y and build a correlation between the two. In other words such methods are used to identify correlations and structures of the original data.

There are two main cases where it was found that the use of LMV methods extremely useful, namely in the integration of high dimensional Discrete Element Method data within flowsheet simulations and in the incorporation of material property databases in within flowsheet simulations. These two concepts are described in the next sections.

3.2.1. Reduced-order DEM data

In this section the application of a model reduction scheme for expensive distributed variable data from DEM simulations using a PCA methodology is described. The application of reduced-order models (ROMs) for the flow of fluids in complex geometries has been studied extensively in literature, where ROMs have been used to identify hidden underlying flow patterns and to capture the dynamics of expensive distributed simulations for control purposes (Anttonen, King, & Beran, 2003; Berkooz, Holmes, & Lumley, 1993). Specifically, in dynamic ROMs the weights of the decomposition vectors have a temporal component and dynamic adaptive sampling techniques are developed in order to collect samples from spatial locations which contribute to the model as the process evolves (Briesen & Marquardt, 2000; Galbally, Fidkowski, Willcox, & Ghattas, 2010; Amit Varshney, et al., 2009). Another interesting approach for model reduction of dynamic multiscale systems is based on the decomposition of different time-scale dynamics, in order to identify variables which are involved in slow processes and can be transformed into simpler algebraic pseudo steady-

state equations (Baldea & Daoutidis, 2006; Contou-Carrere & Daoutidis, 2008; Sotiropoulos, Contou-Carrere, Daoutidis, & Kaznessis, 2008).

The integration of POD methods with surrogate response surface building has been discussed in (Romijn, et al., 2008), where the computational cost can be further reduced through the use of fast metamodels in replacement of complex non-linear equation systems. In addition, this coupling- which is also employed in the current work- allows the decomposition of mappings between input-output and input-state space, which is a desirable aspect in optimization and flowsheet simulation. When a different correlation between input to output streams is produced, it is possible to develop a ROM which has the necessary inputs and outputs which will allow it's integration with upstream and downstream units in a flowsheet simulation. In addition, process optimization can be performed based on objective function which involves only the output as a function of the input parameters, which reduces the complexity of the optimization problem. Based on the identified optimum input conditions, the optimal distributed state-space can be reconstructed. Recent studies tested the applicability of such a model as part of an ASPEN flowsheet simulation, which was shown to be as accurate as the full CFD simulation at steady-state (Y. Lang, Zitney, & Biegler, 2011). The data snapshot PCA approach involves three spaces, according to the methodology developed in (Y.-d. Lang, et al., 2009): (i) input variable space U, (ii) output variable space Y and (iii) the state variable space X. The interconnection of these three spaces is shown in Figure 16.

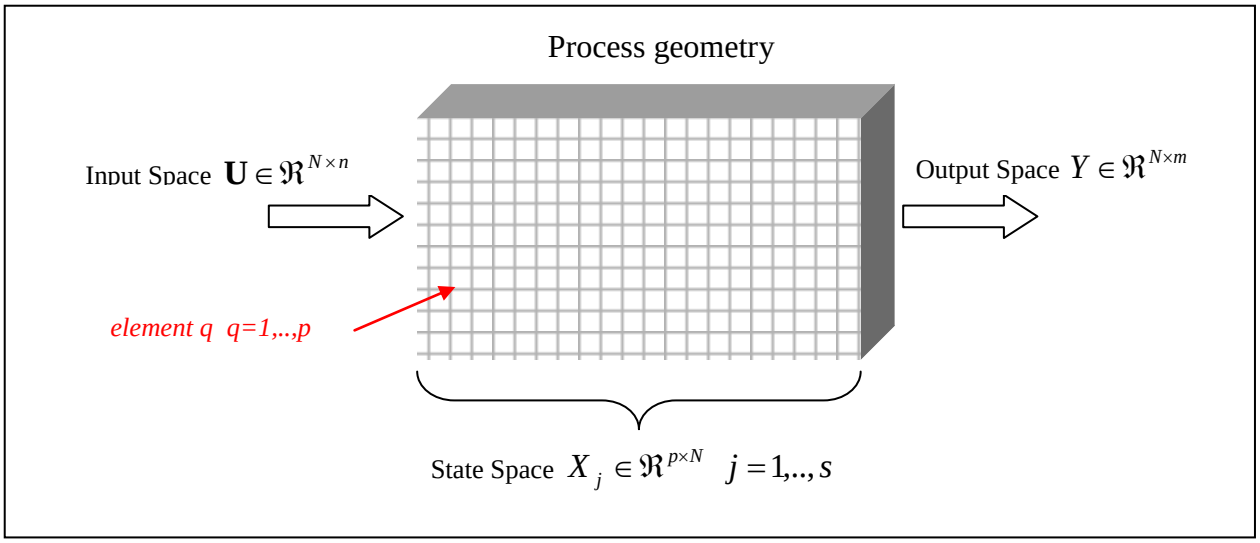


Figure 15 . Snapshot PCA variable spaces of distributed process

As it is shown in Figure 15, the process has *n* inputs and *m* outputs, while *s* state variables are monitored within the process in a total of *p* discretized elements. In order to develop the PCA decomposition model it is necessary to build mappings between the input and output space, as well as between the input and distributed state space. For this purpose initially a set of *m* simulations are performed in order to collect the necessary data based on a design of computer experiments. Subsequently, the input-output mapping may be performed based on any chosen surrogate-based modeling method (linear or non-linear) directly. Since the state variable space is multidimensional, PCA decomposition is performed to decompose the X matrices and identify the necessary PCs, loadings and scores. Finally a lower dimensional input to scores mapping is sufficient for the prediction of the distribution of the state variables inside the process at any new given input conditions. This data-based decomposition methodology is very convenient since it is independent of the process geometry, it is inexpensive and the associated computational cost does not increase with variable space dimensionality.

Discrete Element Methods (DEM) are becoming increasingly popular for modeling micro- and macroscopic properties of particulate/granular materials processes (Zhu, Zhou, Yang, & Yu, 2007). In order to understand the macroscopic properties and dynamics of processed granular materials, the key lies in understanding the particle-particle interactions as well as the particle-surrounding walls interactions at the microscopic level. This is achieved by DEM simulations which allow the tracking of individual particle motion and deformation due to elastic, plastic and frictional forces between particles and walls. Newton's laws of motion form the basic underlying system of equations and each particle has two types of motion: translational and rotational. During a simulation each particle may interact with other particles or surrounding walls and obstacles (i.e. blades) and exchange momentum and energy. Each particle movement is governed mainly by the forces and torques originating not only by the particles or walls with which there is immediate contact, but also by particles and walls that are in a neighboring region through non-contact effects such as Van der Waals and electrostatic forces. Due to the high complexity of this system, it is clear that an analytical solution approach is impossible. The numerical approach followed in DEM simulation software is to compute the distance of particles at every time step and identify all possible interactions and forces acting on each particle in order to calculate its movement and rotation. For this to accurately capture the dynamics of particulate systems, however, the time step must be very small, in the order of 10^{-6}sec. In addition, a DEM simulation must include a large number of particles to resemble reality. For all the aforementioned reasons, the computational cost of a single DEM run is very high, ranging from weeks to even months according to the computational capabilities and thus it becomes clear that

reduced order models can play a very beneficial role in the field of particulate process modeling.

The discrete nature of DEM simulations is the key aspect which differentiates them from continuum type simulation schemes, such as Computational Fluid Dynamic (CFD) or Finite Element Method (FEM) simulations, used to model fluid or gas processes. In addition, a discrete element approach is by nature more random since a contact detection engine step must be implemented prior to the Newtonian physics solution step which detects the number of nearby particles or walls for each discrete entity. Several attempts have been made to transition from this discrete nature to continuum through averaging techniques in the literature (A. B. Yu, Zhu, Zhou, & Yang, 2008; Zhu & Yu, 2004; Zhu, Yu, & Wu, 2006), in order to correlate macroscopic properties to DEM simulation results. Another interesting approach was developed in (Glosmann, 2010), where the authors develop a reduced hybrid-DEM model using Karhunen-Loeve transformation and clustering of particles for discrete systems driven solely by potential forces, disregarding contact forces. However, these techniques often require certain assumptions, i.e. the ignorance of the rotational motion of particles, which lead to inaccurate results and failure to capture the true dynamics of the system.

Due to the data-driven nature of the methodology introduced in this work, it becomes apparent that the quality of the data extracted from the DEM simulation is very critical to the model accuracy. This observation leads to the notion that the discretization of the process geometry for data extraction is a critical step towards ROM robustness. Due to the discrete nature of DEM simulation one has to face the problem of sufficient sampling size in order to decide upon the correct discretization density. The optimal

discretization scheme should be dense enough to capture the spatial distribution of the variables inside the process, but coarse enough to ensure that the average number of particles inside each element is large enough to provide reliable average information. Consequently, the right discretization must be chosen based on the actual total number of elements inside the process geometry, the relative size of the element compared to the size of the process and finally based on the total variance of the specific variable. For example, if the velocity profile is constant throughout the process, then most particles will move with this velocity, subsequently less samples would be sufficient to provide a reliable measurement. In order to handle all of the aforementioned challenges, the Discrete Element- Reduced-Order Modeling approach is proposed and described in the following section in order to maximize ROM accuracy using the minimum number of required simulations.

The application of a PCA based ROM follows the steps outlined in Figure 16. Initially a computer design of experiments must be identified in order to collect the necessary data for the ROM construction. This DoE can be a factorial design, fractional factorial, Latin Hypercube or a different design based on the number of inputs, number of desired levels of investigation and the nature of the process.

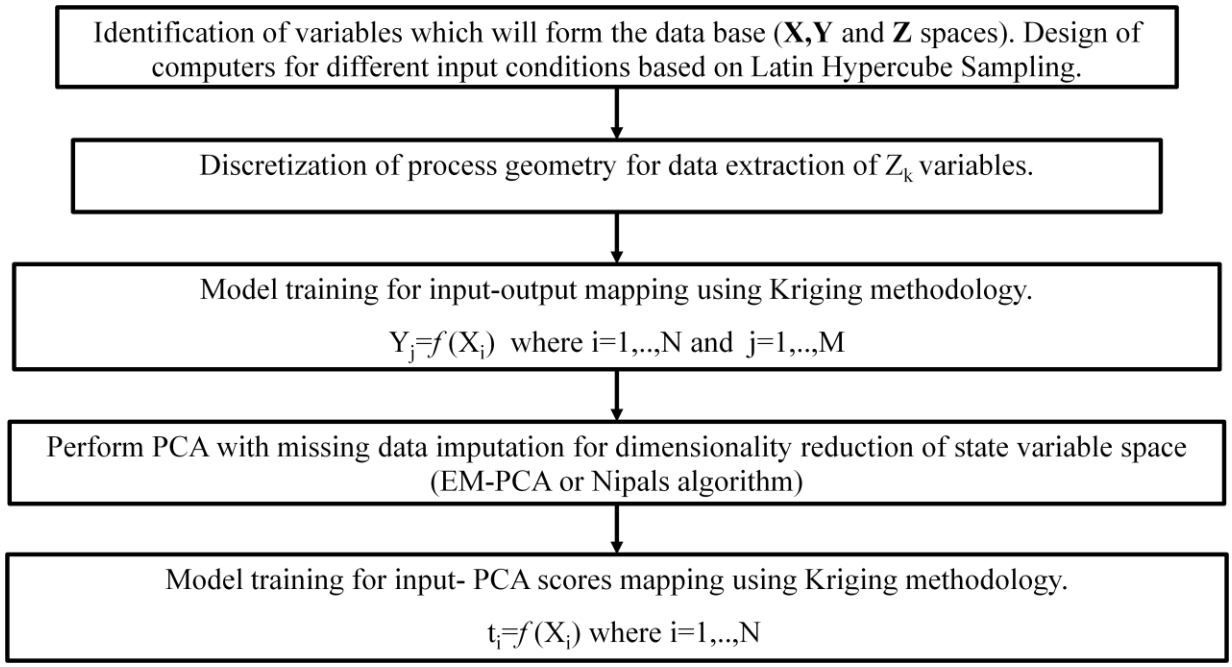


Figure 16. DE-ROM proposed steps

According to the desired use of the reduced order model, the three sets of variables must be specified: input space (inlet variables, design parameters, initial conditions), state space (distributed variables of interest within the process geometry) and output space (outlet variables). Subsequently, based on the chosen design of experiments, a number of simulations are performed in order to collect the output space values as well as snapshots of the state space values at different combinations of inputs. At this point, the data needs to be analyzed and a suitable discretization scheme must be chosen in order to ensure that a sufficient number of samples (particles) are located in the majority of the bins to be analyzed. However, it is realized that even if a satisfactory discretization is chosen, the presence of elements (or bins) which will contain very few number of particles is inevitable. Thus, prior to the implementation of the PCA decomposition, the extracted data is processed such that bins are categorized into three groups according to

the number of elements they contain. Group 1 is the subset of bins which contain a very small number of particles [*0, UB_{g1}*] for which all measurements are discarded and all values of the state variables are set equal to zero. A general rule for the calculation of the maximum number of particles for this category (UB_{g1}) is the one tenth of the average number of particles inside each bin which is equal to the total number of particles divided by the total number of bins ($Part_{tot}/p$). This group represents the subset of bins assumed to be empty of particles, which is a common observation in solids handling processes based on the fill-level of the process. The bins which contain a mid-range number of particles [$UB_{g1}+1$, UB_{g2}] belong in Group 2 in which it is assumed that the total number of solid particles is not sufficient for extracting average information and the value considered to be missing. The range of the number of particles which are classified in this category is decided based on the total average variance of the state variable, the variability of measurements (standard deviation) as well as the average number of particles in each element. Descriptive statistics of the data such as histograms are very helpful in identifying the optimal cut-off value (UB_{g2}), below which the sample does not follow a normal distribution and thus the size of the sample is insufficient. As a general guideline, in this work UB_{g2}~[20-30] particles, but this value may decrease or increase according to the variability in the extracted data. Subsequently, the value of the average state variable for this set of bins is imputed using the performance of the PCA decomposition with missing data. Finally, bins which contain a sufficient number of particles (>UB_{g2}) are classified in Group 3 and do not need any type of preprocessing. In the case where a fast first-principle relation is available for correlation of certain inputs to outputs, this is used; otherwise a direct input-output mapping is performed based on a

chosen surrogate method (i.e. Response Surface Method, Kriging, Neural Networks). Finally, PCA decomposition is performed to compress the high dimensional snapshot data and the identified scores are correlated to the inputs based, again, on a surrogate method.

Application to a continuous powder blender case study

The goal of this case study is to understand the significance of different operating and design variables on the continuous mixing performance using DEM simulations. Specifically, the purpose of the developed Reduced Order Model, may be direct use within a flowsheet simulation or incorporation to the already existing Population Balance Model in order to enhance its predictive range and develop a multiscale model which incorporates design aspects, is fast to simulate and predicts output bulk powder properties. Specifically, it is realized that the parameters of the PBM model, are the velocity fluxes within each discretization element of the model, which govern the movement of particles within the blender and affect mixing variability and output stream composition. These parameters, however, are extracted from a DEM simulation for a specific rotation rate and design and are used as constant in the current flowsheet simulation. An integration with a ROM model, which can predict the velocity flux distribution as a function of operating conditions and design parameters on the fly, may improve the performance of the PBM model. Thus the goal is to implement the described methodology, choose the set of inputs of interest, collect the samples from DEM, build the ROM and integrate it within the flowsheet simulation.

For this case study DEM data from a periodic section of a blender is used, based on previous work of our group (Y. Gao, M. G. Ierapetritou, & F. J. Muzzio, 2011b) which

has proven that the operation of a continuous mixer can be represented as a sequence of multiple batch-like periodic sections. The periodic section approach is used to minimize the computations of the DEM simulations, and this allows us to investigate the effect of more input variables.

The influence of six predictive variables (fill level, blade speed, width, and angle, shaft angle, and weir height) is characterized by simulating a set of periodic section-mixing samples based on a Latin hypercube sampling design (Table 3). In this case study, the total flowrate is not considered as a controlled variable since this would have an effect on the fill-level, making it impossible to distinguish the effects of each input. As opposed to modifying the total flowrate of particles, the fill-level is controlled based on the total number of simulated particles. Exploiting the advantages of DEM simulations, the effects of design aspects such as blade angle and width, weir design and shaft angle can be investigated. Specifically shaft angle has been investigated in a previous study experimentally (Portillo, Ierapetritou, & Muzzio, 2008) and was found to be critical for the mixing performance.

Table 3. Operation and geometry design variables- Input variable space

Design parameters	**Low bound**	**High bound**
Blade speed (RPM) (x_1)	40	250
Blade angle (deg) (x_2)	10	40
Blade width (mm) (x_3)	10	40
Weir height ratio *w/d* (-) (x_4)	0%	75%
Fill level (-) (x_5)	25%	75%
Shaft angle (deg) (x_6)	-30 (upward shaft)	30 (downward shaft)

.

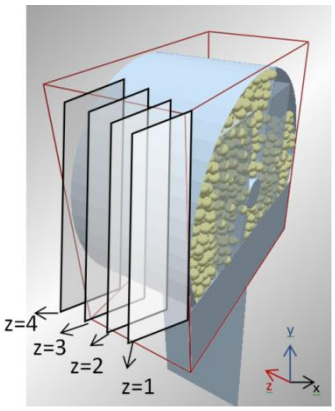


Figure 17. Periodic section geometry and planes in z dimension for which the state variables are plotted.

Up to now the input space (Table 3) and the state space (velocity fluxes in the x, y and z dimension) have been described, however, the output space has not been defined yet. Due to the use of the periodic section model, it is harder to select a set of outputs which will be correlated to the inputs which will be of any use to a flowsheet simulation. This is not a problem however, since the main purpose is to integrated it to the already existing PBM which is integrated within the flowsheet. The selected output is a set of indices developed by Gao et al. (Gao, et al., 2011b), which aim to connect the performance of a periodic section to a full blender: (1) the variance decay rate of the batch-like mixing in the cross-sectional directions (k_b), and (2) the mean particle velocity in the axial direction (v_x), which characterize mixing performance and are considered as outputs in this case study. The ratio of these two indices is called variance decay ratio based on which the exponential decay of the RSD along the axis of the mixer predicted. More details about this can be found in (Boukouvala, Muzzio, & Ierapetritou, 2012) however, the input-output mapping will not be discussed here further, since the main

attention is given to the improvement of the PBM model, which is only dependent on the input-state space mapping.

A total of 64 simulations are designed and performed based on a Latin Hypercube design, following the rule of approximately 10n samples for an n-dimensional space. The four additional samples are simulated in order to create a validation set of points of the produced ROM model. The periodic section is discretized into 4 x 8 x 8 number of bins in order to have a reasonable average number (~ 40) of particles inside each bin. Any element with particles fewer than 5 is categorized in Group 1 and its data is set to zero. Bins with a number of particles between 5 and 20 are considered as missing information (Group 2) and the rest of the data belong in Group 3 and is used without any modification. The monitored distributed parameter spaces are the velocity of the particles in the x,y,z dimension and for each of the state variables, a data matrix is formed ($X_i \in R^{256x64}$ where i=1,..,3), where each column represents one simulation and each row corresponds to one location (element) of the periodic section. Subsequently, PCA is performed in order to identify the optimal loadings and scores which represent the full data matrix in the reduced space. In table 4, the necessary number of principal components is given- which is decided upon based on the cross-validation error. Figure 18 represents the comparison of the predicted DE-ROM profiles with the actual DEM profiles for the sample with specifications:

$$\mathbf{x} = (x_1, x_2, x_3, x_4, x_5, x_6) = (177.16,\ 60.18,\ 10.47,\ 30.82,\ 39.45,\ -18.25)$$

which is left out from the data-base. The average error reported in Table 4, is the average leave-one-out cross-validation error out of the 64 different combinations of samples.

Table 4. Model prediction characteristics

Variable	Total # of PC's	Variance explained(%)	Variance explained in raw data (%)	Average error (out out 256 bins) (%)
U_x (m/s)	12	81	70	24
U_y (m/s)	7	89	86	13
U_z (m/s)	8	83	77	16

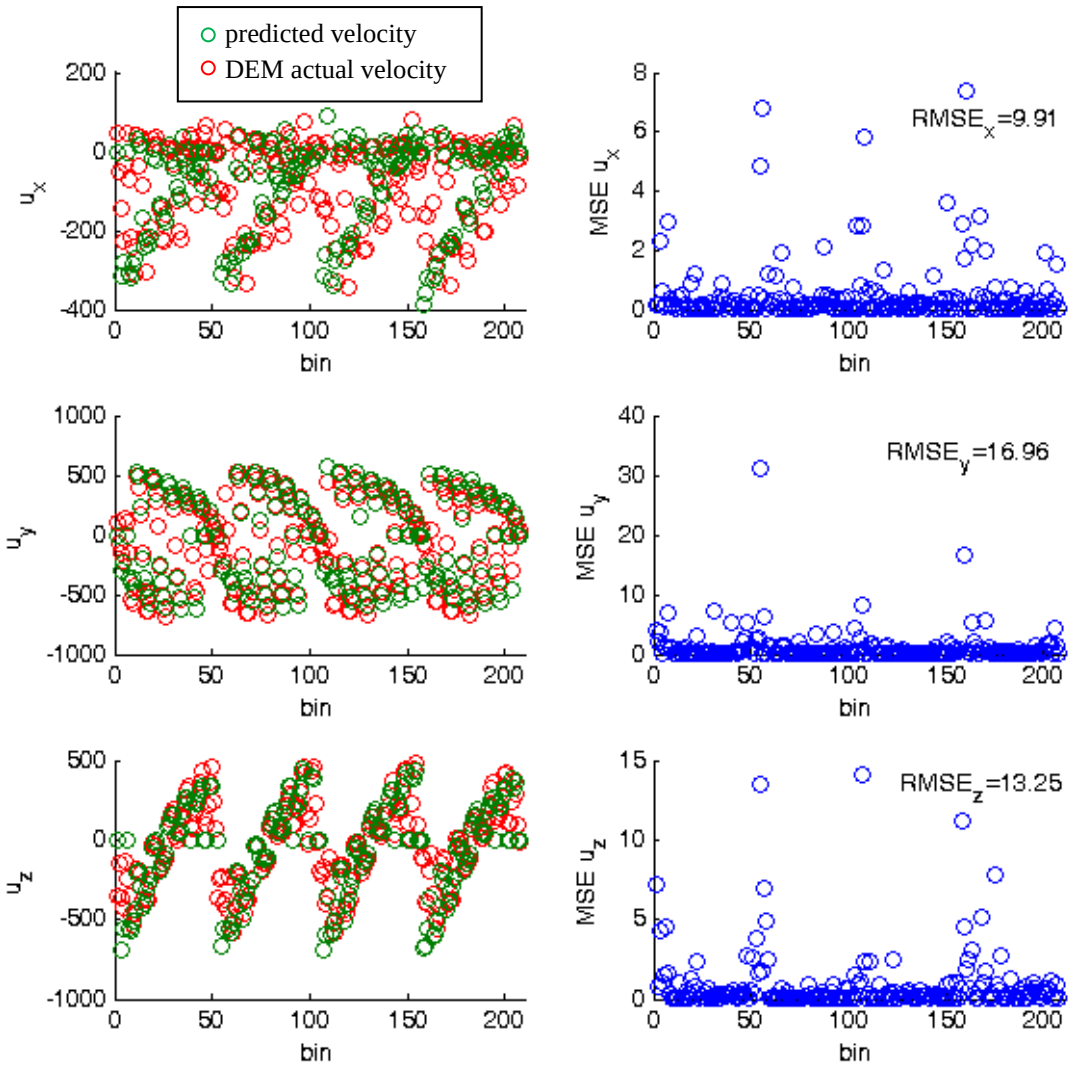


Figure 18. Comparison of predicted (green) to actual (red) velocity fluxes for the periodic section blender and their Root Mean Squared Error (blue).

The final step to the methodology is the integration of the model within the gPROMS flowsheet model. For this purpose, the PCA parameters of each of the three velocity fields need to be stored into a database in the flowsheet library, from which the necessary calculations of any new velocity flux distribution may be reconstructed. However, the kriging model which is built in order to correlate the set of inputs to the scores of each of the fields is implemented in MATLAB. Thus, the go:MATLAB

interface of gPROMS and MATLAB is necessary in order for the model, given any new design or set of operating conditions, to be able to call MATLAB, obtain the predicted scores for each of the velocity fluxes, and then using the PCA block within gPROMS reconstruct the velocity fields within the 256 bins and send them to the PBM model in order to use them as the new parameters of the model (Figure 19).

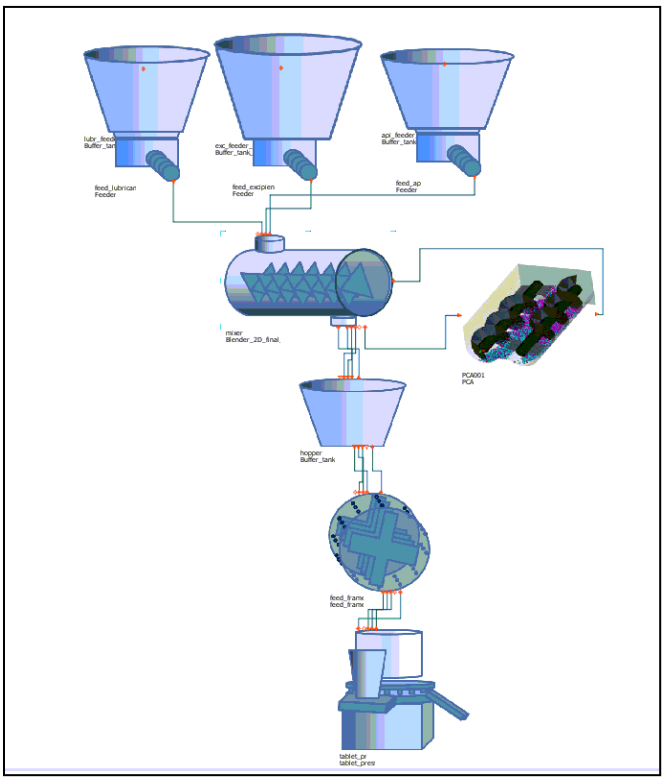


Figure 19. Integration of the DEM-ROM model within a gPROMS direct compaction process simulation

3.2.2. Material property integration

As a second example to the application of multivariate latent space models within a flowsheet simulation, the work which is described deals with the incorporation of a material property database into the flowsheet simulation. Surprisingly, this exercise is very similar to the work performed for the integration of reduced order DEM models.

This work has been performed in collaboration with the Powder Characterization project of the ERC-SOPS, where the aim is to identify the significant material property attributes which characterizes a set of powders. During the past years, the project team has worked on assessing already existing powder characterization techniques, but also developing new techniques which are now being used in industry. All of this knowledge, however, has not been put into any use within the current flowsheet simulation, due to the fact that the first-principles which correlate all of these properties to powder behavior and process operation are not known. In other words, there is no clear way of how to incorporate the effects of the variability of powder processes within operation. As it can be observed from chapter 2, solely particle size information and bulk density are used within the developed population balance models.

In addition, it is realized that powder blends are usually simulated within a given flowsheet, and mixture rules for powders are also not yet developed. However, a recent study which aims to develop a methodology for the acceleration of product development can be employed to solve both challenges. The proposed Weighted Scores PLS methodology (García-Muñoz & Polizzi, 2012; Polizzi & García-Muñoz, 2011) can solve the three following problems towards the final goal. Based on a set of data of material properties for a set of raw powder materials and a set of collected data for specific blends of these materials, the method can: (a) predict properties of untested blends of materials within the database, (b) predict properties of blends of materials which are not incorporated within the initial database, if they have similar properties to the already existing materials, (c) predict in-process properties of blends, as long as they have been measured for the designed blends as well.

Even though this project has yet to be completed, the main concepts will be described here. In addition, the main algorithmic components for the integration of the WSPLS method within the flowsheet simulation has been completed and tested on a database provided by Dr. Salvador Garcia-Munoz, thus they are ready to be used when the designed experiments for the database of materials and blends has been completed. The main steps of the WS-PLS method will be described next, along with results obtained so far for the specific case study.

Initially, a set of data must be collected for a number of m material properties and selected n raw materials (APIs, Excipients). This forms the X matrix of the problem. The method allows for a collection of different set of properties for APIs and Excipients, if this is necessary, however, in the current case a common set of properties for the characterization of all of the available materials is selected. The properties and materials are included in Table 5.

Table 5.Materials included in test database and material properties measured

Lactose Monohydrate	Conditioned Bulk Density
Compap L APAP	Compressibility at 0.5kPa
Vivapur MCC 102	Compressibility at 15kPa
Avicel 200	Pressure drop at 1kPa
Magnesiun Stearate	6kPa UYS
	6kPa MPS
	6kPa FFC
	6kPa Cohesion
	9kPa UYS
	9kPa MPS
	9kPa FFC
	9kPa Cohesion
	Flow Index
	d10
	d50
	d90
	Specific Energy
	Basic Flowability Energy

Next, PCA is applied on the obtained data, in order to identify the true dimensionality of the data set and obtain the scores, which are a reduced set of coordinates for the original data set, which are also orthogonal. The next step of the method requires the collection of the same set of properties for a selected number of designed blends for the raw materials. In order to minimize the required number of mixture experiments, but however, obtain valuable information about the blend properties of the materials used in this work, a subset of important materials may be chosen, which span the scores space well. This subset of materials, explains most of the variability in the data, thus would be optimal to use for blend experiments. This subset of materials can be chosen based on a mathematical mixed integer optimization problem, provided to us by Dr. Salvador Garcia-Munoz. Results for the PCA decomposition of the data set showed

that most of the variability can be explained by four PCs (Table 6), and based on this reduced model, the selected materials to perform mixture experiments on is shown in Table 7. The mixture DOE was designed based on a constrained D-optimal mixture design using the MODDE software. The constraints on the design were the following: $0 \le C_{MgSt} \le 0.01$ and $0 \le C_{APAP} \le 0.3$.

Table 6.Percentage of variance explained by 4 PC model

# Principal Component	% Explained Variance
1	53.59
2	18.33
3	9.63
4	5.45

Table 7. Total number of experiments performed for raw materials and blends

Experiment	**Lactose Monohydrate**	**Compap L APAP**	**Vivapur MCC 102**	**Avicel 200**	**Magnesiun Stearate**
1	1	0	0	0	0
2	0	1	0	0	0
3	0	0	1	0	0
4	0	0	0	1	0
5	0	0	0	0	1
6	**0**	**0.3**	**0.7**	**0**	**0**
7	**0**	**0**	**0.99**	**0**	**0.01**
8	**0**	**0**	**0**	**0.99**	**0.01**
9	**0**	**0.3**	**0**	**0.69**	**0.01**
10	**0.282**	**0.15**	**0.282**	**0.282**	**0.005**

These constraints were imposed simply because there is no possibility of using higher than this amounts of the specific material in any blend. The matrix R is the matrix

of formulations which has r rows (number of blend experiments) and n columns (number of materials). As soon as the same set of m properties present in the X matrix are collected for the blends (Y matrix), then a PLS model is constructed between the weighted scores (RT) matrix and Y. At this point, it should be mentioned that there is no restriction to only including the same set of properties of X in the Y data set. Once all of this information has been collected, the WS-PLS model is ready to be tested for the prediction of new blend properties of the materials present in the dataset. In the current setting, the properties in Y are the same as the properties in X, however, in future work a careful analysis of prospective downstream properties for which no available process model exists and which can be added as new columns to the Y dataset will be investigated.

The final step of this work, involves the integration of such a model into the flowsheet simulation. All of the above analysis (PCA, PLS) is performed in advance using available software such as ProMv (Prosensus) or phi (provided by Dr. Salvador Garcia Munoz), and the parameters of the PLS model are inputted as text documents within the gPROMS model library. This can be thought of as a physical property database, from which new blend properties are predicted. There are multiple possible uses of such a unit operation, such as the availability of the properties of the raw powders, which perhaps will affect the operation of the feeding units. However, after the blending stage, where the output stream is a blend with a predicted dynamic composition, the WS-PLS model can be used to provide any of the Y properties dynamically as a function of time and outlet concentration.

For the incorporation of the WS-PLS model within a flowsheet simulation, two additional input-output connection types are created, where the inlet stream reads mixture composition (R_i vector) and the outlet stream provides predicted Y properties (Y_i vector) (Figure 20). This unit operation may be connected to any other process stage to provide valuable information (since several of the Y properties may be inputs to specific unit operation models) . This implementation is tested on the materials database published in (Polizzi & García-Muñoz, 2011)where as it can be seen in Figure 21, the model can predict the varying property Y_i as a function of time at the outlet of the blender. The periodic fluctuations are due to upstream feeder refilling effects, which cause perturbations to the concentration of the outlet composition, which in turn affect the properties of the blend.

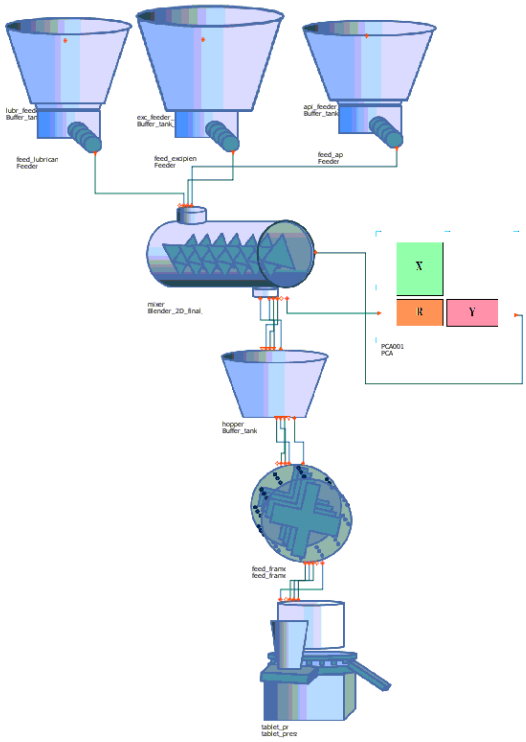


Figure 20. Integration of WS-PLS model within Direct Compaction flowsheet simulation

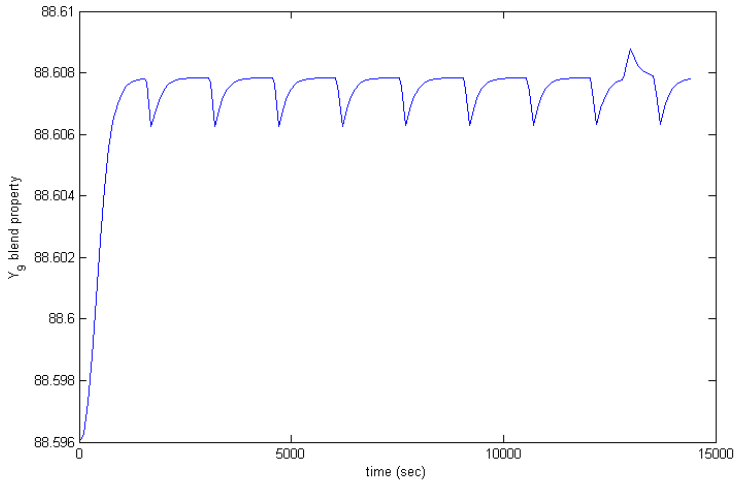


Figure 21. Dynamic prediction of property Y_9 at the outlet of the blender after a series of feeder refills

Chapter 4

4. Integrated process Design Space

4.1. Design Space of pharmaceutical processes

The concept of Design Space (DS) is a key aspect of the pharmaceutical Quality by Design (QbD) initiative and was formally introduced in the International Conference on Harmonization (ICH) Q8 guideline for pharmaceutical development ("Q8 Pharmaceutical Development," 2006). According to the formal definition it is “the multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality. Working within the design space is not considered as a change. Movement out of the design space is considered to be a change and would normally initiate a regulatory post approval change process”. The importance of the Design Space definition lies firstly in the “assurance of quality”, which is the key goal of QbD and secondly in the broadening from an acceptable operating set point to a collection of tolerable operating regions which make up the Design Space. Thus if a process DS is accurately identified, one can have the freedom to operate within this entire region without additional necessary regulatory approval and simultaneously, have higher confidence about the final product quality. For the same reasons, the DS is a critical component of the recent effort of introducing QbD in pharmaceutical development, and has attracted a lot of attention in literature.

In ICH Q8 two critical aspects of the DS are defined, namely the selection of variables and the final presentation of the DS. The first step to identifying a DS is the recognition of the critical variables that should be included. The presentation of a DS should provide the ranges of these critical variables that result to a product meeting a

desired quality, but should not be confused with a simple combination of proven acceptable ranges. As a result to this introductory report, two more elaborate guidance papers (Garcia, et al., 2008; Lepore & Spavins, 2008) were published from specialized teams which aim to provide a more detailed description of the Design Space. Specifically, the Design Space Task Team of the International Society for Pharmaceutical Engineering (ISPE) outlines the components and steps of a Design Space through the Product Quality Lifecycle Implementation initiative (PQLI).

In (Lepore & Spavins, 2008) it is emphasized that the sources for accurately defining the Design Space of a process may include the available literature, experience and knowledge of the process, first principles, experimental data, empirical models or- most often- some combination of all these methods. However, the choice of the tools used to characterize the Design Space depends on the availability of these resources. For example, when the process is well characterized and the underlying principles are known, the Design Space should be obtained by predictive first-principle models. The main advantage of the existence of a first- principle model is the ability to simulate the process under any given conditions for predicting the outputs of interest. Recently researchers aim to develop such models for pharmaceutical processes, and the advantages of computer-aided process design and simulation has become an essential tool for pharmaceutical process development and optimization of manufacturing (L. Yu, 2008).On the other hand, if a process is new and its scientific principles are not well understood, it can be treated as a black-box process. In the later case, the Design Space dynamically evolves, since as additional knowledge and information about a process is obtained or as new raw materials, evolving specifications, and new technology become

available, the design space of the process is better characterized. In addition, the DS should be insensitive to the scale of a process even if this requires the performance of additional scale-up experiments.

Even though recent discussion has focused on general guidelines (Lepore & Spavins, 2008) and the types of variables that a DS should include (L. Yu, 2008), relatively little has been put forth in terms of defining methods to construct a Design Space of a process. Clearly, there is no unique approach in defining a design space, but once the methodology and tools to determine the Design Space are chosen, the Design Space has to be presented efficiently, in a way that can be easily interpreted. Figure 22 is a very common representation of the Design Space, used to explain graphically its relationship to the Knowledge Space and the Normal Operating Ranges. The larger Knowledge Space contains all the information about all regions of the process that have been investigated. Subsequently, the difference between the Knowledge and Design Space can be defined as the region of the Knowledge space that generates unacceptable product. If the process can be controlled by a set of manipulated variables, the DS may be enlarged when compared to an open-loop DS, where there are no controls in place.

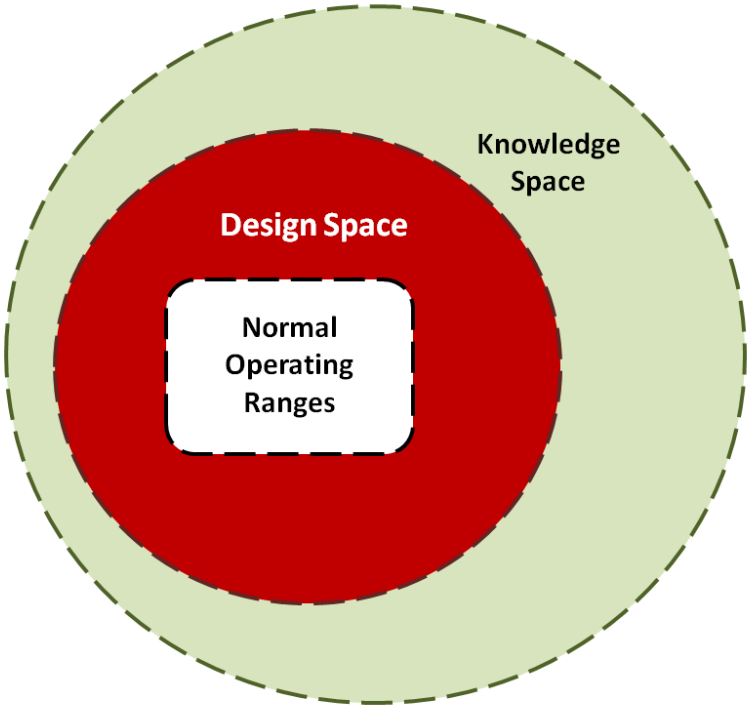


Figure 22. Link between Knowledge Space, Design Space and Normal Operating Ranges

The Design Space is interconnected with the two remaining PQLI key topics: criticality and control strategy (B. Davis, Lundsberg, & Cook, 2008; Garcia, et al., 2008; Nosal & Schultz, 2008). Criticality refers to the determination of which quality attributes and operating conditions of a process are necessary in order to predict a desired output so that they will only be included in the Design Space. A risk assessment procedure is defined that filters the variables through a series of questions in order to identify their effects on safety, quality and efficacy of a process output. As a result of this procedure, the variables are classified according to their level of criticality. The necessary set of controls for assurance of acceptable process performance and product quality is the final PQLI key topic, that is the required Control Strategy. Clearly, the three topics are highly connected since it is necessary to first identify the critical variables (Criticality) which

define a Design Space and then these critical variables should be manipulated based on an efficient Control Strategy in order for a process to remain within the Design Space at all times.

A framework for the development of the Design and Control Space was presented in (MacGregor & Bruwer, 2008) where the importance of simultaneously identifying the material property, process operating parameter and control space is discussed. The significance of raw material properties in the DS characterization is important since these are properties which cannot be easily manipulated or controlled- as opposed to operating conditions. In addition, the need to take account of control strategies in the DS can be illustrated by identifying the- possibly significant- changes in a Design Space if the applied control strategy is modified (García-Muñoz, Dolph, & Ward Ii; MacGregor & Bruwer, 2008). However, even by implementing a very efficient control strategy, specific raw materials cannot meet the acceptable final product specifications. Recently, the work of Peterson et al.(Peterson, 2008; Peterson & Lief, 2010) has focused on showing the limitations of a simplistic overlapping means approach for mapping a DS which fails to answer the question about how much assurance of quality such an approach can provide. Specifically a Bayesian approach for the identification of the ICH Q8 Design Space is adopted in order to calculate the probability of meeting specifications.

Several examples of specific experimental applications to define the Design Space of pharmaceutical processes can be found in the literature. For example in (Lipsanen, Antikainen, Räikkönen, Airaksinen, & Yliruusi, 2007), the operating window for the process of a fluidized bed granulation has been identified by assessing the impact of critical parameters, such as inlet air humidity, on fluidization behavior and granule size.

A similar procedure was followed in (Lebrun, et al., 2008) in order to define the confidence zone of chromatographic analytical methods. In (Boukouvala, Muzzio, & Ierapetritou) , three different data-based methodologies are used to generate the graphical representation of the region in which acceptable performance of unit black-box processes can be ensured. The efficiency of these approaches is compared through two pharmaceutical case studies: predicting the design space of a continuous powder mixer and a loss-in-weight feeder of a continuous tablet manufacturing process. Integration of feed- forward control in the design space of a high shear wet granulation process was studied by Garcia- Munoz et al. (García-Muñoz, Dolph, & Ward, 2010)where it is shown that a DS can be significantly enlarged if a correct control strategy is employed. This work signifies the advantage of a successful control sequence, which increases the flexibility of a process towards uncertainty and inevitable variations.

This work aims to propose concepts and algorithms which aim to map a feasible region of any process, given the input variables of interest and the output performance criteria that must be met in order to result to feasible operation. Specifically, a black-box Design Space mapping method is developed, which aims to locate boundaries of feasible operation of processes for which a first-principle model is either lacking, or for processes which rely on very computationally expensive models. The effects of uncertainty and control to a process DS are also discussed in this work.

4.2. PSE concepts useful for the concept of Design Space: feasibility, and flexibility

From a PSE perspective, the construction of the Design Space can be considered as the problem of determining the boundaries of a process where feasible, profitable and

acceptable performance is guaranteed. This problem, however, has been considered as a major concern in many process industries and a substantial amount of work has been performed in order to define concepts such as “operability”, “feasibility” and “flexibility” of processes that contain uncertain parameters. Uncertainty can occur for a variety of reasons, most commonly among them is the variability of certain process parameters during plant operation. In traditional chemical engineering industries, the design space may be defined by objectives related mostly to cost and product quality, however, the pharmaceutical design space should also be guided by safety and efficacy of the produced product. The purpose of this work, however, is to discuss the tools which will aid one to map the process design space, while having the flexibility of defining any desired objective.

Optimal process design under uncertainty was defined as a rigorous formulation in the 1980’s (Halemane & Grossmann, 1987), where the effects of parameters that contain considerable uncertainty on the optimality and feasibility of a chemical plant were studied. The objective of solving such problems was to ensure optimality and feasibility of operation for a given range of uncertain parameter values, by identifying a measure of the size of the feasible region of operation. According to the methodology introduced in this work, the problem was represented as a max-min-max formulation, where, for a given design and fixed values of the uncertain parameters, the feasible region was calculated. In (Swaney & Grossmann, 1985), the flexibility of such processes was defined and quantified by the flexibility index (FI), which represented the maximum allowed deviation of uncertain parameters from their nominal values, such that feasible operation could be guaranteed by changing the control variables. A series of papers

dealing with flexibility analysis and the formulation and optimization of processes under uncertainty were published in the following years for cases where the process model is known in closed form (Floudas & Gumus, 2001; Grossman & Floudas, 1987; Vishal & Marianthi, 2002, 2003) as well as for cases where the process is treated as a black- box (Banerjee & Ierapetritou, 2002, 2003; Banerjee, Pal, & Maiti; Boukouvala, et al.).

A general optimization problem has the following form:

$$\begin{aligned} &\min_{d,z,x} f(d,z,x,\theta) \\ &s.t. \\ &h(d,z,x,\theta) = 0 \\ &g(d,z,x,\theta) \le 0 \\ &d \in R^n, z \in R^q, x \in R^q, \theta \in \mathrm{T} \end{aligned} \tag{25}$$

where *d* corresponds to the design variables, *z* and *x* represent the control and state variables respectively, θ correspond to the uncertain parameters of the process, *h* are process equations describing the system, *g* correspond to inequality constraints which are either bounds on variables, design specifications or logical constraints, *f* is the objective function to be minimized, and $T = \{\theta | \ \theta^L \le \theta \le \theta^U\}$. Eliminating the equality constraints *h* by expressing all state variables in terms of *d*, *z* and θ, the objective function of problem 25 becomes:

$$\begin{aligned} &\min_{d,z} f(d,z,\theta) \\ &s.t. \\ &y(d,z,\theta) \le 0 \\ &d \in R^n, z \in R^q, \theta \in \mathrm{T} \end{aligned} \tag{26}$$

Solving problem 26 determines whether for a given design *d* and values of uncertain parameters θ the control variables z can be adjusted to satisfy all the necessary constraints and attain feasibility. This can be accomplished if for a given value of θ, all constraints

$y_j \leq 0$ are satisfied. By defining the feasibility function $\psi(\mathrm{d},\theta)=\min_{z}\max_{j\in J}\{y_j(d,z,\theta)\}$, where J is the set of inequality constraints, the controls are selected such that the maximum y_j is minimized. This optimization problem can be further transformed into the following form by introducing the scalar parameter u such that:

$$\begin{aligned}&\psi(\mathrm{d},\theta)=\min_{\mathrm{u,z}} u\\ &s.t. y_j(d,z,\theta)\leq u, \quad j\in J\end{aligned} \tag{27}$$

In order to determine whether feasible operation can be attained in the parameter uncertainty range T, it is clear that $\psi(\mathrm{d},\theta)\leq 0$ for all $\theta\in \mathrm{T}$. In its most compact form, the flexibility test problem can be represented as a *max-min-max* formulation, since it is sufficient to ensure whether the maximum value of the feasibility function is less or equal to zero in order to maintain feasible operation within the uncertain experimental range.

$$\chi(d)=\max_{\theta\in \mathrm{T}}\min_{z}\max_{j\in J} g_j(d,z,\theta) \tag{28}$$

where $\chi(d)$corresponds to the flexibility function of design d over the range T of uncertain parameters θ. The above formulations of the feasibility problem aim to identify the feasible space of a process over a range of uncertain parameters θ. In this work, it is assumed that the uncertain parameters θ are the critical input variables which are varied during a process operation, or may be varied to optimize a process, in order to investigate which are the feasible regions of operation over the entire investigated space. For this reason, whenever uncertain parameters are referred to in the current DS work, these are parameters which cause the predictions of a model to be stochastic, or to the uncertainty associated with any experimental measurement used to identify the DS.

Undoubtedly, effects of noise and control highly affect a Design Space. Noise may come from errors in the experimental measurements used to identify the DS, or model parameter uncertainties which will affect the model prediction. On the other hand, the effect of control in the DS is usually beneficial since it enlarges the range of feasible operation. This is due to the fact that controllers aim to intervene in order to keep the process within its design space. The effect of control variables is already within the formulation of Problems 26-28, thus it is straightforward to use the same ideas presented in this chapter to map the feasible region of an open-loop process and a closed-loop process. Methods for taking into account the effect of noise have also been developed in the PSE literature through the formulation of the Stochastic Feasibility/ Flexibility concepts, where the probability of operating within the DS is evaluated. These will play an important role especially in the uncertain world of pharmaceutical manufacturing, and this has been pointed out by recent publications. Specifically, the work of Peterson et al.(John Peterson, Ronald Snee, Paul McAllister, Timothi Schofield, & Carelia, 2009) points out the incessant increase in the role of statistics for the identification of design spaces. Through a Bayesian framework the quantification of the amount of reliability that a DS provides can be achieved through a clearly defined figure of merit (Equation 17) (Gregory W Stockdale & Cheng, 2009; John Peterson, Guillermo Miro- Quesada, & Castillo, 2009; Peterson, 2008; Peterson & Lief, 2010).

$$\Pr(\mathrm{Y} \in \mathrm{A} \,|\, \mathbf{x}, data) \geq R \qquad (29)$$

where for a given acceptance region A, the probability of a response Y to belong in A, given the input vector x and a set of data should be greater or equal to a predefined reliability level R. For this calculation a posterior predictive distribution for Y must be

assumed and all the uncertainty of the process model parameters should be taken into account. Several applications of Bayesian design spaces of different pharmaceutical and biopharmaceutical processes are presented in (Gregory W Stockdale & Cheng, 2009; Peterson & Lief, 2010).

The Bayesian framework for DS identification is similar with a concept that has been introduced in the process systems design community, namely Stochastic Flexibility (SF) (Pistikopoulos & Mazzuchi, 1990; Sahinidis, 2004; Straub & Grossmann, 1993). Stochastic flexibility is a probabilistic measure that was established in the 1990's to measure the system's ability to tolerate continuous uncertainty (Acevedo & Pistikopoulos, 1998; Bansal, Perkins, & Pistikopoulos, 1998; Samsatli, Papageorgiou, & Shah, 1998; Straub & Grossmann, 1993). In other words, it can be defined as the probability of a given process design to operate feasibly. In this problem, the uncertain parameters are assumed to follow a distribution function, thus the determination of the SF involves the calculation of a multiple integral of the joint probability of all the uncertain parameters over the parametric space.

During the past few decades, the issue of explicitly handling uncertainty in process design and optimization has attracted a lot of attention leading to a large number of methods that quantify feasibility. The methods developed for solving process design under uncertainty problems can be categorized into three groups: (a) deterministic, (b) stochastic and (c) black-box programming techniques (Banerjee & Ierapetritou, 2002; Sahinidis, 2004). In deterministic parametric programming approaches a complete profile of optimal solutions as a function of uncertain parameters is given, while stochastic approaches provide optimal solutions based on expected performance criteria for known

probability distributions of the uncertain parameters. All proposed methodologies however rely heavily on the nature and existence of the explicit form of the model equations. In addition, they are restricted by assumptions of convexity of the feasible regions of the process. In other words, years of research has focused on formulating rigorous optimization programming techniques to identify the feasible region around a nominal point, without turning to the solution of the calculation of feasibility over a dense set of points, or else complete enumeration. If a feasible region is compact and linear around the nominal point, these methods are very effective. However, when a region is non-linear, non-convex and even disjoint, these methods either underestimate or overestimate the feasible region. Previous work of our group introduced a methodology that treats the process as a black-box and the feasibility problem is solved through the use of input-output mappings (Banerjee & Ierapetritou, 2002, 2003; Banerjee, et al.). This work is the first attempt to touch upon the problem of design under uncertainty from a completely different perspective, where the effect of parameter uncertainty is predicted using only a set of input uncertainties irrespective of the nature or complexity of the underlying process model. High Dimensional Model Representation (HDMR) is used as a method of input-output mapping of black-box processes and this method is used to model the effect of parameter uncertainty on process design, and consequently identify the feasible operation of the process. In (Banerjee & Ierapetritou, 2003), the authors extend their work to handle more complex cases where the variability of the optimal solution of Mixed Integer Nonlinear Programming (MINLP) models with parameter uncertainty is captured through the use of HDMR.

Based on the promising results of the work of Banerjee et al. (Banerjee & Ierapetritou, 2002, 2005; Banerjee, Pal, & Maiti, 2010) and based on the advantages of Kriging as a modeling technique identified in the current work, a new approach for performing black-box feasibility has been proposed, which follows an adaptive sampling strategy and minimizes the necessary sampling (Boukouvala & Ierapetritou, 2012). The basis of the current method has the same goal as rigorous feasibility analysis problems, namely avoidance of complete enumeration, which is a brute force solution and would lead to high computational cost. However, it is no longer aimed to identify maximum deviations around a nominal point, however, it is attempted to use an approximation model to predict feasibility, which is guaranteed to be accurate, if the samples used to build it are located at critical points within the region. These will be points which provide information about the true feasibility boundaries.

The steps of the proposed methodology and results on a variety of different design spaces are described in the next sections of this chapter. This parallelization of Design Space- as it is defined in the QbD guidance- and feasibility, aims to point out once again that well-established process system engineering tools can be very useful in developing tools to facilitate QbD goals.

4.3. Black-box feasibility analysis

Even though black-box feasibility as a concept can be used for any problem, there are two cases, where black-box feasibility is required and greatly beneficial, (a) when the first-principle model is not available and one needs to rely on time consuming and expensive experimental data to characterize the feasible region, or (b) when the process model is so computationally expensive, so it is desired to minimize the number of

function calls to the process model, while obtaining the required information to identify the boundaries of the feasible region. In both cases, a surrogate or data-based response surface model is built to approximate the underlying process model and aid the sampling strategy. As pointed out in chapter 2, when using a surrogate response surface to approximate a real process, the accuracy of the predicted output is highly dependent on the quality and quantity of the sampling set. Even though quantitatively increasing the data set will guarantee a more accurate predicted surface, this will also increase the sampling cost which is undesirable. Hence, there is a need of developing strategies which use information obtained from the model itself to identify sampling locations that provide maximum information about the process in order to avoid any redundant sampling.

In literature, the problem of minimization of sampling cost and time for the optimization of expensive or black-box models has attracted a lot of attention. In (Donald R. Jones, et al., 1998) a figure of merit that directs the search for a global minimum is introduced, as the Expected Improvement (EI) function. Based on this criterion, the next sampling point that is chosen, is the one that has the highest Expected Improvement. To obtain the EI, the expected value of the following expression is calculated:

$$E[I(x)] = E[\max(f_{min} - Y), 0] \tag{30}$$

where f_{min} is the current optimal (minimum) value based on the model at the current iteration, and Y is the predicted value by the Kriging model. In simple terms, this equation identifies either a region that has the largest possible expected difference to the current optimum, or a region that has a high uncertainty of prediction. This adaptive sampling concept has been proven to be very successful in identifying promising locations for sampling, but also sampling globally within the entire investigated space. It

has been a breakthrough in the optimization literature for expensive models, and has been widely used in many applications. The EI maximization concept, makes use of the advantage of Kriging to provide an error mapping of the predicted response surface which is a function of the density of sampled points within the investigated space- described in Chapter 2- in order to calculate a statistical EI criterion in closed form and enable its global optimization. Throughout this chapter, it is advised for the reader to keep in mind that the prediction of Kriging is assumed to be a distribution (normal) with a mean Y and standard deviation s^2. Further details of the method can be found in the original paper by Jones et al. (Donald R. Jones, et al., 1998).

In feasibility analysis, however, it is important to identify boundaries of feasible operation within the entire range of input variables, and not a single minimum point. Employing the concept of feasibility, however, the problem DS is quantified to a collection of points which are either negative, positive or zero. Boundaries can be regarded as the points for which the value of the feasibility function is equal to zero, whereas all feasible regions are characterized by negative feasibility function values and non-feasible regions should be described by positive feasibility function. In a first attempt to develop an adaptive sampling methodology (Boukouvala & Ierapetritou, 2012) proposed identifying next promising samples by comparing all pairs of the existing samples and identifying directions between points which have a negative product, since this signifies the change in feasibility from negative to positive between these two points. The next sample would be located in the midpoint between the distance of the two points with a negative product. Further details of this method and results can be found in (Boukouvala & Ierapetritou, 2012). However, the disadvantage of this approach is that it

does not provide information regarding the actual spatial location of the change in sign which signifies the location of a boundary. This may lead to increased sampling requirements, which should be minimized further in order to avoid performing unnecessary expensive simulations or costly experiments. In addition, this approach does not take into account the uncertainty of prediction.

For this reason, and following the concept of formulating an expected improvement criterion, a **modified Expected Improvement** function is designed (Equation 31), which aims to balance the sampling between regions which have not been sampled enough (high uncertainty) and regions which the probability of the predicted response to be equal to zero is maximized.

$$E_{feas}[I(x)] = E_{feas}\{\max(0 - Y), 0\} = -Y\Phi\left(\frac{-\mathrm{Y}}{s}\right) + s\phi\left(\frac{-\mathrm{Y}}{s}\right) \tag{31}$$

where Y is the predicted kriging response and s is the kriging predicted error. Φ represents the distribution function and φ the standard normal density. The first term of Equation 31 represents the probability of Y to be less than zero, while the second term represents the probability of Y to be equal to zero. In other words, when $E_{feas}I$ is maximized due to a large value of the first term, this will have identified a point with predicted negative feasibility value (inside the feasible region), while if it is maximized due to a higher second term, this will mean that a point with a value close to zero (on a boundary) or a high s value has been located. By using this criterion to locate next sampling points, it is found that the search is directed mostly to the center of the feasible region. This is due to the fact that the fist term is very dominant, and the sampling is directed to points with the most negative predicted feasibility value. For this reason, it is advised that the first term is eliminated, in order to account for points with either high

uncertainty or a predicted value close to zero. Using this approach, the search is directed mostly on boundaries and once any new point is sampled, the Kriging response surface is updated. As a consequence, when the Kriging model is updated, the uncertainty (s) in the vicinity of the sampled point will decrease, thus the algorithm will avoid sampling in points which are close to each other, which is another advantage of the adaptive sampling strategy. The explicit form of the modified EI function with only the second term can be derived as:

$$E_{fea}[I(x)] = s \cdot \left(\frac{1}{\sqrt{2\pi}} \right) \exp\left(-0.5 \cdot \left(\frac{Y^2}{s} \right) \right) \tag{32}$$

This approach shares the convergence characteristics of the EI type methods, which state that at infinite number of samples, the method will have sampled the desired region densely (D. R. Jones & Law, 1993; Donald R. Jones, et al., 1998) However, it is not desired to oversample the space, since the initial goal is in fact to minimize sampling. Thus an efficient stopping criterion needs to be formulated. Three criteria are used to terminate the sampling: (a) a maximum number of function calls, (b) an average error between the predicted feasibility space of the previous iteration and the current, or (c) a low value of expected improvement. If either of these three criteria are met, the algorithm stops sampling further. After careful evaluation of the performance of the algorithm on a variety of complex feasible regions, it is observed that as the complexity of the feasible region increases, the value of the EI does not usually reach the tolerance value, however the feasible region has been identified with great accuracy. Thus the criterion of the average sum of squared differences between the predicted feasibility value of the previous iteration and the current over a grid of points within the investigated space, is a

very good way of indentifying whether a new sample will improve the prediction. In addition, it should be mentioned that the ability of kriging to provide error s and its incorporation into the sampling criterion is very important, because the algorithm chooses samples which are on the boundary and far away from each other at the same time, during the initial iterations and improves the predicted feasibility space drastically. Next, the steps of the black-box feasibility analysis are described and shown in Figure 23.

Initially, a small sampling set is chosen to produce the initial Kriging response surface based on a space filling experimental design. Specifically, the performance of two experimental designs can be used, the equally spaced grid and random Latin Hypercube sampling design. The experimental design should cover the entire range of uncertain parameters. The general rule of thumb of 10^k design points, where k equals to the number of uncertain parameters, is often found to be a good starting point. Next, the $E_{feas}I$ criterion is maximized in order to identify the next promising sample, which is collected and the kriging response surface is updated. Any global optimization algorithm can be used to optimize this function, since it is known in closed form and it is cheap to evaluate. In the current implementation, the TOMLAB LGO algorithm is employed, which is a MATLAB based multi-start global optimization algorithm. This procedure is iteratively performed until any of the aforementioned stopping criteria is met.

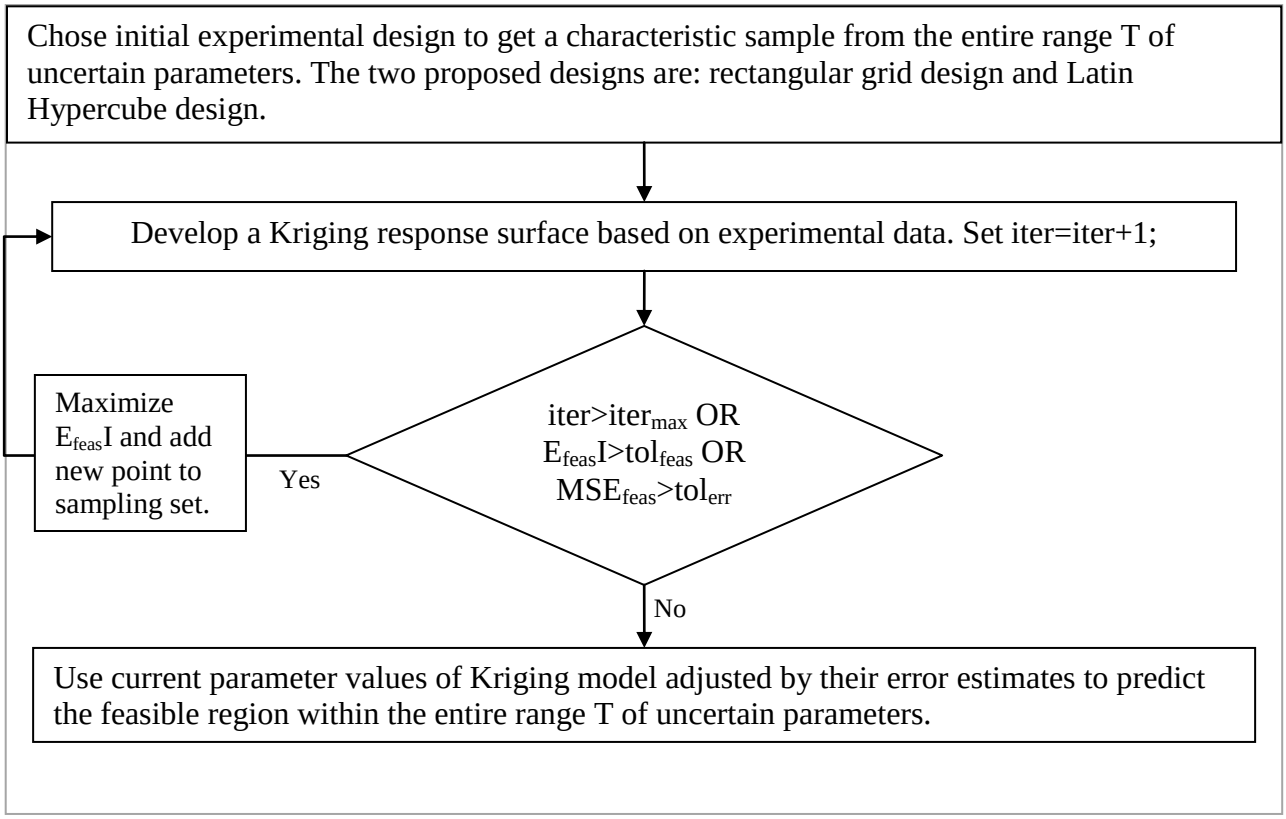


Figure 23. Adaptive Sampling Methodology for Black Box Feasibility Analysis

4.3.1. Results of performance of adaptive sampling strategy for black-box feasibility analysis

In order to test the performance of the proposed black-box feasibility approach, its ability and sampling requirements is tested on a variety of different known feasible regions, ranging from linear and convex regions to highly non-linear, non-convex and even disjoint regions.

Example 1: Non-linear convex feasible region

$$\begin{aligned} &\min u \\ &s.t. \\ &x_1^2 + x_2^2 - a^2 \le u \\ &x_1, x_2 \in [-1,1] \end{aligned} \tag{33}$$

This first example contains one non-linear constraint which forms a circular region within which the process is feasible. An initial Latin Hypercube Design of 21 samples is collected along with 7 additional sampling points through the adaptive sampling strategy. The real feasible region is shown in Figure 24a, along with the location of all the collected points. The initial Kriging response surface of the feasible region is formed and updated to become the more accurate one shown in Figure 24c. Finally, a grid of points is tested for feasibility based on the final kriging model and the predicted feasible points all lie within the actual feasible region shown in Figure 24d. As it is shown in Figure 24a, the red points which are collected through the maximization of the EI function all lie on the real boundary.

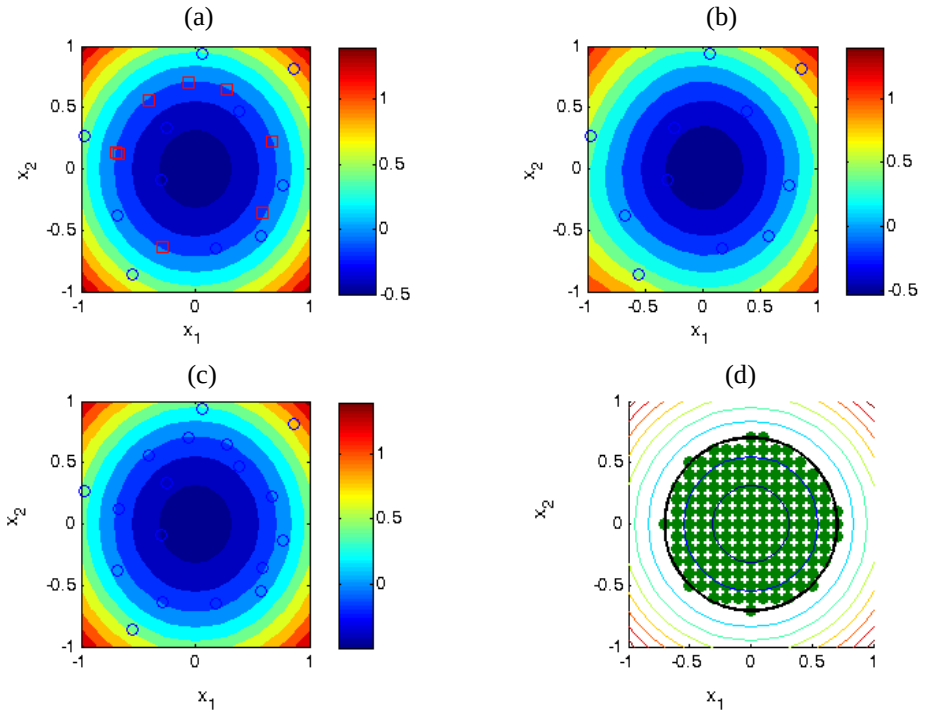


Figure 24. (a) Actual contours of feasible region, initial LHS points (blue) and points of adaptive black box feasibility sampling (red), (b) contours based on initial sampling set,

(c) contours based on final sampling set, (d) predicted feasible region points over a grid of test data for problem 33.

Example 2: Non-linear convex problem

The current example was first presented by Ierapetritou (M.G. Ierapetritou, 2001) and it involves two input parameters, and both linear and nonlinear convex constraints. The feasibility problem of this case study is defined by the set of inequalities in Problem 34. For this problem, 21 initial points were designed based on a LHS design, and 25 were additionally collected based on the adaptive sampling strategy. From the Figure 25 it can be noticed that the sampled points lie on the boundary of the feasible region. The final prediction of the feasibility mapping is not accurate for regions far away from boundaries (Figure c),however, this is not an undesirable result since the objective is to minimize sampling. From the comparison of figure 25b with Figure 25c, the improvement to the feasibility region after the additional samples are collected can be clearly observed. The final plot of Figure 25d shows the predicted feasible points (green points) over a fine grid of samples based on the final Kriging model.

$$
\begin{aligned}
&\min u \\
&s.t. \\
&x_1^2 - x_1 + x_2 - 40 \le u \\
&x_1^2 + x_1 - x_2 - 2 \le u \\
&-4x_1 + x_2 - 30 \le u \\
&x_1 \in [-5,5] \\
&x_2 \in [-50,50]
\end{aligned}
\qquad (34)
$$

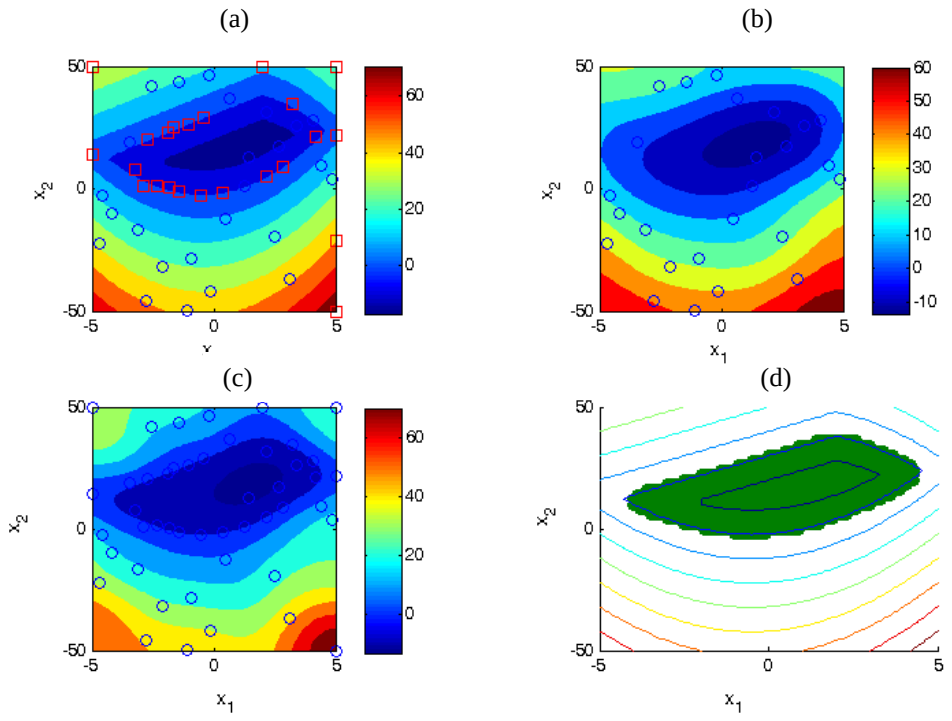


Figure 25 (a) Actual contours of feasible region, initial LHS points (blue) and points of adaptive black box feasibility sampling (red), (b) contours based on initial sampling set, (c) contours based on final sampling set, (d) predicted feasible region points over a grid of test data for problem 34.

Example 3: Non-linear non-convex feasible region.

$$\begin{aligned} &\min u \\ &s.t. \\ &x_1^2 + x_2^2 - a^2 \le u \\ &b^2 - x_1^2 - x_2^2 \le u \\ &x_1, x_2 \in [-1,1] \end{aligned} \tag{35}$$

This problem deals with a non-linear non-convex feasible region which is defined by two inequalities. From this example, it can be seen that the sampling required for mapping the feasible region increases with the complexity of the region. In this case a total of 21

initial samples, plus 91 points based on the adaptive sampling (Figure 26a). The improvement of the initial predicted feasible region (b) to the final prediction (c) is dramatic, since the collected samples all lie on the boundaries. Through the demonstration of this problem, one of the main advantages of the proposed approach is shown, namely the fact that as long as there are regions of high uncertainty and/or boundaries, the adaptive search will identify them.

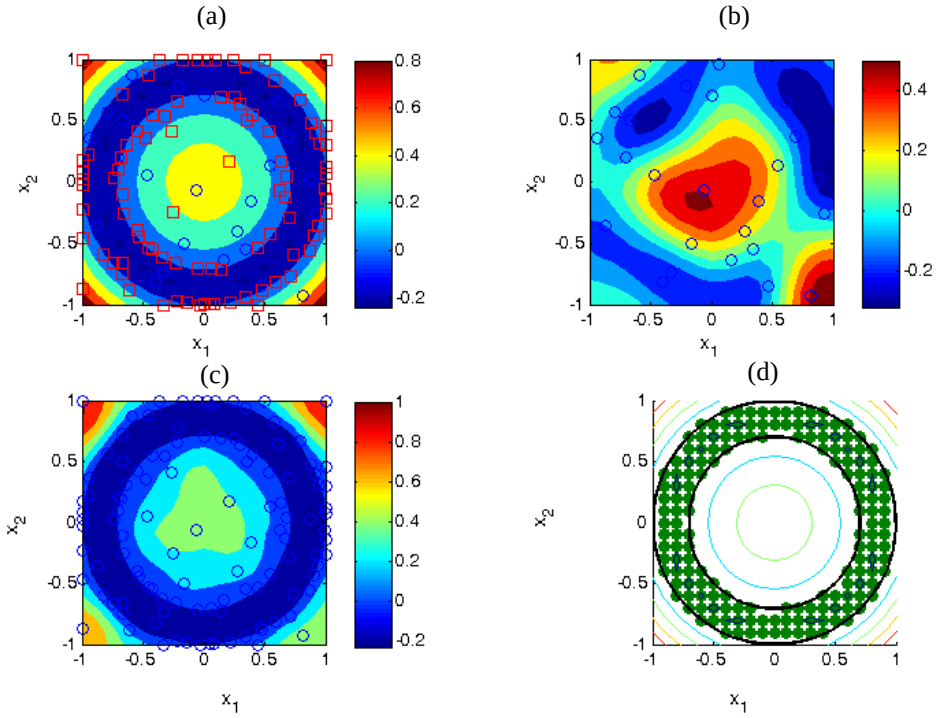


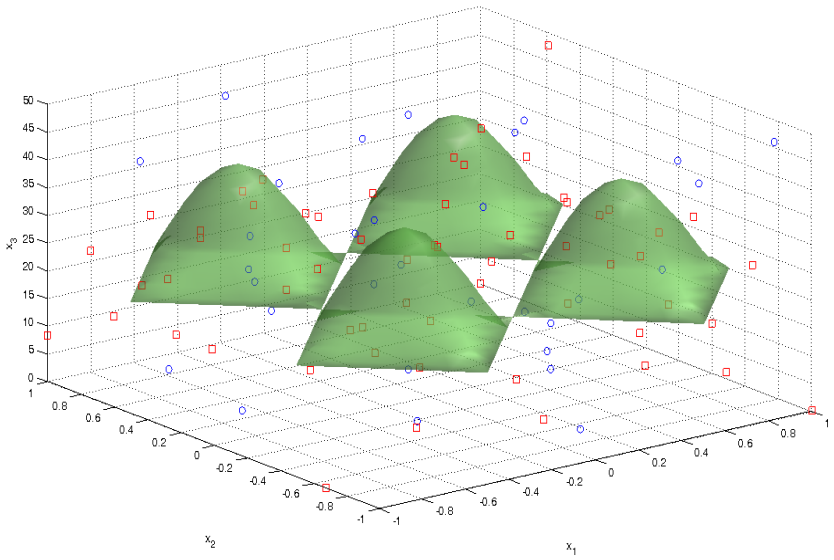
Figure 26. (a) Actual contours of feasible region, initial LHS points (blue) and points of adaptive black box feasibility sampling (red), (b) contours based on initial sampling set, (c) contours based on final sampling set, (d) predicted feasible region points over a grid of test data for problem 35.

Example 4: 3D non-linear, non-convex disjoint feasible region

The performance of the algorithm is tested on a problem of higher dimensionality, which is expected to be more demanding in terms of the sampling requirements. The formed feasible region is very complex, as it is non-linear, non-convex and is composed by four disjoint feasible sets (Problem 36).

$$
\begin{aligned}
&\min\ u \\
&s.t. \\
&x_3 - \left[20 + \left(x_1^2 - 10\cos(2\pi x_1)\right) + x_2^2 - 10\cos(2\pi x_2)\right] \le u \\
&20 - x_3 \le u \\
&x_1, x_2 \in [-1,1] \\
&x_3 \in [0,50]
\end{aligned}
\tag{36}
$$

A total number of 31 points is designed for the initial kriging fitting, while an additional 55 points are required (Figure 27) to reach to the final prediction as is shown in Figure 28



.

Figure 27. Actual surfaceof feasible region, initial LHS points (blue) and points of adaptive black box feasibility sampling (red) for problem 36

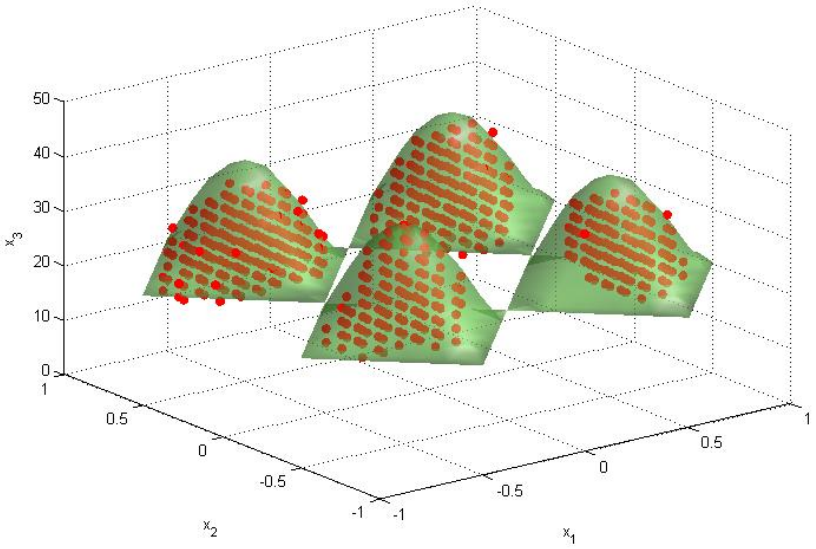


Figure 28. Predicted feasible region points over a grid of test data for problem 36.

The developed black-box feasibility approach can be further used in various other problems, other than feasibility mapping and DS identification. As it will be discussed in chapter 5, this method is the first step of a global optimization algorithm for solving problems which are constrained and rely on expensive simulations.

4.4. Feasibility analysis for mixed integer non-linear problems

So far, black-box feasibility concepts for continuous input variables have been described, however, when solving process design under uncertainty and feasibility problems two different types of decisions are often necessary: 1) selection of components such as processes, type of equipment, etc. which are represented by discrete variables, and 2) selection and determination of the operating conditions (Banerjee & Ierapetritou, 2003; Grossmann, 1990) which are usually defined by continuous variables. In these cases, one deals with a Mixed Integer Non-Linear Problem (MINLP) or a Mixed Integer

Linear Problem (MIP), depending of the form of the constraints and objective. In these problems, discrete/ binary variables are introduced for representing existence of units and composing different design configurations.

Modeling and optimization of pharmaceutical unit operations often includes modeling with discrete- sometimes even non-numerical input variables. These parameters often represent design variables such as the use, size or configuration of a specific part of a piece of equipment (e.g. screw size of a feeder, design of a nozzle). Up to now, however, data-driven modeling of pharmaceutical processes not often treats these types of variables as integer decision variables. Conversely, discrete or non-numerical variables are usually represented in a multivariate data set by coded values based on which the final response surface is fitted (Raymond H. Myers, 1990; Raymond H. Myers & Montgomery, 1995).

What is proposed in this work is that individual models are produced for alternative process designs, which are complemented by the assignment of a decision variable for each design. A statistical analysis of the design variables is always performed initially to identify the variables that are statistically significant. Once variables are found to be significant, using the proposed modeling approach enables the formulation of the mathematical optimization problem that can lead to the optimum design for different values of operating conditions. This is the basic advantage of the proposed method since this cannot be achieved if the design is modeled as an additional input variable. Consequently, the Design Space can be constructed separately and more accurately for each different design configuration while the optimum design is identified for each combination of operating conditions.

The steps to identify the Design Space using any type of model are outlined here in Figure 29. The basic modification to the analysis performed in the previous section is the fact that a surrogate response surface is built for each of the k available designs, and the black-box feasibility analysis is performed separately for each one. In order to identify the optimal DS over the range of the continuous input variables, an optimization problem is solved in order to identify which design is optimal for a given combination of input variables. In this optimization problem, the competing designs are compared and the one with the lowest feasibility function is chosen, by imposing an additional constraint to ensure that only one design can be chosen for each combination of operating parameters. A lower feasibility function value signifies lower constraint violation, which can be translated to operating further away from the DS boundaries, which is preferable. One additional reason for operating far away from a boundary is that- as will be discussed in the next section- DS mapping is always associated with certain amount of uncertainty coming from the noise in physical experimental measurements or uncertainty of model parameters.

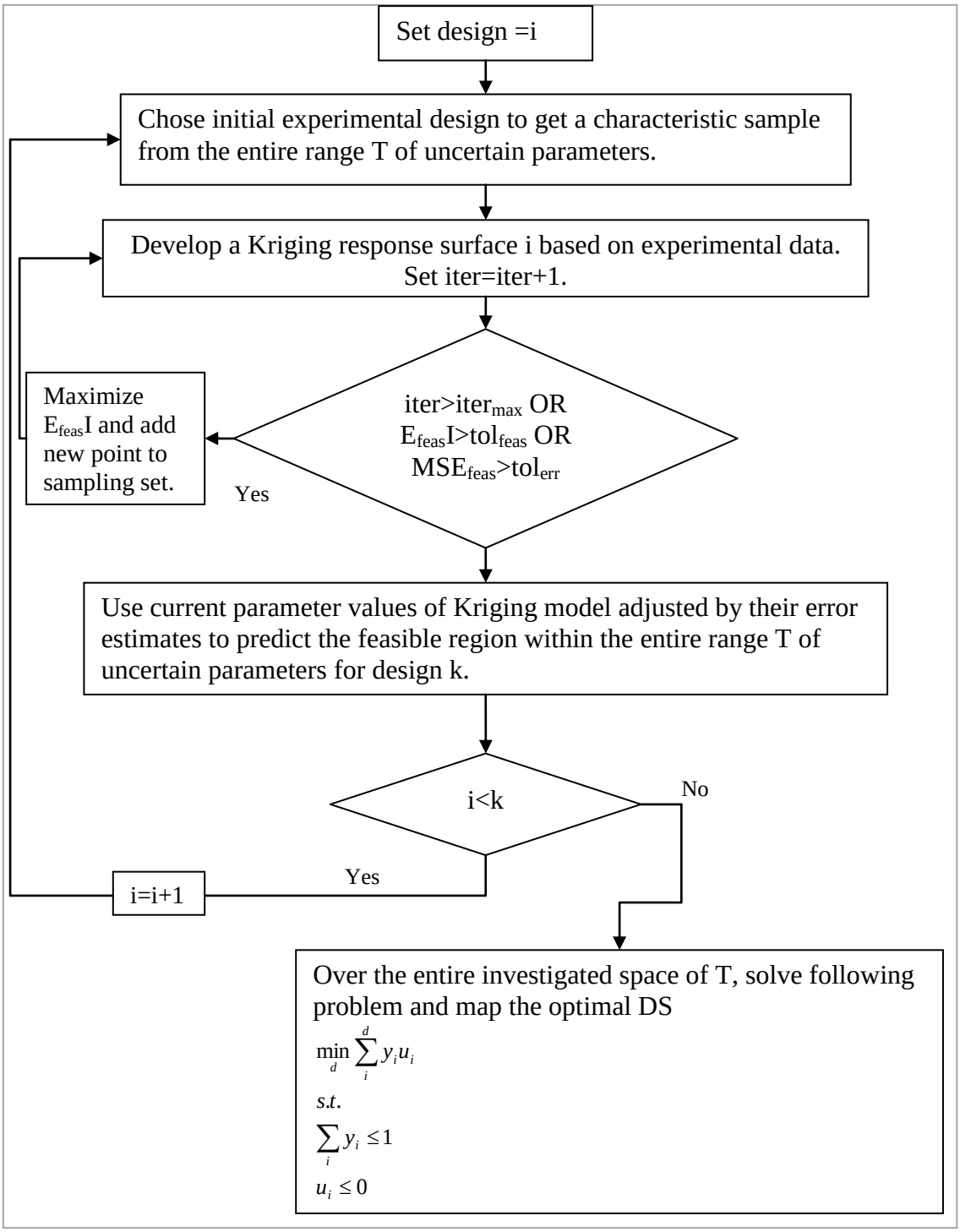


Figure 29. Steps towards identification of Design Space for processes with discrete type decisions.

4.4.1. Application of black-box feasibility concepts for pharmaceutical processes

Example 1: Continuous blender Design Space

The first application of the described concepts is tested on a process for which a first principle model is not available and experimental measurements are performed in order to map the DS. The objective of this case study is to identify regions within which the mixing performance of the blender is acceptable, which is quantified by the final relative standard deviation of the API concentration of the outlet stream of the mixer. The blender has two different design configurations, all-forward blades and alternate which can be used for mixing powder materials. The continuous conditions which are identified as significant for the mixing performance of the process (RSD) are the flowrate and mixing rotation speed. Through this case study, the aim is to show that under different operating conditions, a different design may be optimal.

The results obtained by the methodology (Figure 30) agree with the expected results based on the existing knowledge of mixing process. In general, rotation rate is an important factor in continuous mixing. Increasing the rotation rate leads to a higher degree of dispersion of powder in the mixer. However, higher impeller rotation rate also decreases the time available for mixing (lower residence time). Understanding these opposing effects is the key to achieving optimal mixing and can explain the fact that lowest RSD is achieved in mid-range rotation rates. In contrast, flow rate is found to be the least significant factor using ANOVA and this can explain the fact that for the investigated rotation rates, feasible operation can be achieved for all flow rates in the range of 30-45 kg/hr. However, this variable must be included in the design space since it is a significant factor for feeder performance. Specifically, increasing the flow rate can

decrease the output flow variability of the feeder. The fact that this variable does not affect the variability in concentration of this specific case study implies that the mixer can efficiently filter out the feeder variability, but this might not be the case for other feeding-mixing integrated systems. For almost the entire design space, better performance is achieved when using the 'Alternate Blades' configuration, while for high flow rates and low rotation rates, the 'All forward' design results to better mixing performance. More details about the specifics of the experimental data can be found in the original publication (Boukouvala, et al., 2010).

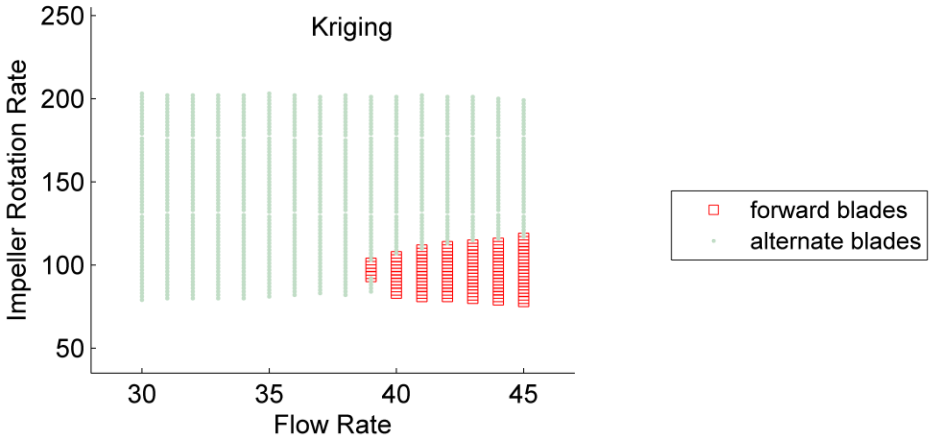


Figure30. Design space produced by Kriging approach showing the optimum design configurations within the investigated space defined by Impeller Rotation Rate (rpm) and Flow rate (kg/hr).

Example 2: Roller compaction Design Space

For the roller compaction case study, there are no discrete variables, however there are two simultaneous specifications which must be met in order to assess feasibility. The model which connects the input variables of interest to the two outputs which are considered in the DS is available for this case study, and it is based on Johanson's

theory(Dec, et al., 2003). The existence of a model for this case study, allows us to evaluate feasibility using the existing model not relying on experimental data, as well as simulate a stochastic case study in order to demonstrate the concepts of stochastic feasibility, and finally examine the effect of control of certain variables on the final DS. The three independent and significant input variables which are considered in the DS are the roll rotation speed (ω) and the compaction pressure (P_h) and inlet feed speed (u_{in}). Upper and lower bounds of the two output variables of interest- ribbon density and thickness are chosen, based on the desired quality specifications of the product (Equations 37), and if both are satisfied, then the process is operating within its DS.

$$\rho_{exit}^{lb} \le \rho_{exit} \le \rho_{exit}^{ub}$$
$$h_0^{lb} \le h_0 \le h_0^{ub} \qquad (37)$$

Following the black-box feasibility approach, Problem 38 can be defined through the introduction of the feasibility function u. If u is negative or zero, this implies that all constraints are satisfied and thus the process is within the Design Space. A positive value of u implies that even the minimum possible value of u, violates at least one constraint and thus the process is not in the DS.

$$\min_{\rho_{exit},h_0,u} u$$

s.t.

$$\rho_{in}\cos\theta_{in}(1+{h_0}/{R}-\cos\theta_{in})({u_{in}}/{\omega h_0})=\rho_{exit}$$

$$Ph=\frac{W}{A}\frac{\sigma_{exit}R}{1+\sin\delta}\int_0^{\alpha}\left[\frac{{h_0}/{R}}{(1+{h_0}/{R}-\cos\theta)\cos\theta}\right]^{K}\cos\theta d\theta$$

$$h_0-h_0^{ub}\le u$$

$$h_0^{lb}-h_0\le u$$

$$\rho_{exit}-\rho_{exit}^{ub}\le u$$

$$\rho_{exit}^{lb}-\rho_{exit}\le u \qquad (38)$$

The Design Space in Figure 31 is obtained by using black-box feasibility concepts for solving Problem 38 in the three dimensional space. The blue bounded region represents the set of conditions that lead to acceptable product in terms of its ribbon density, while the operating conditions within the green bounded region satisfy the thickness quality constraints, and their intersection defines the region where both specifications are met. The results suggest operating the process at low pressures, high inlet flowrates and low angular velocities, if the specific quality specifications must be met. This DS suggests that the process can produce ribbons of desired properties for a range of flow rates and angular velocities, however, as ω increases, u_{in} should also be increased.

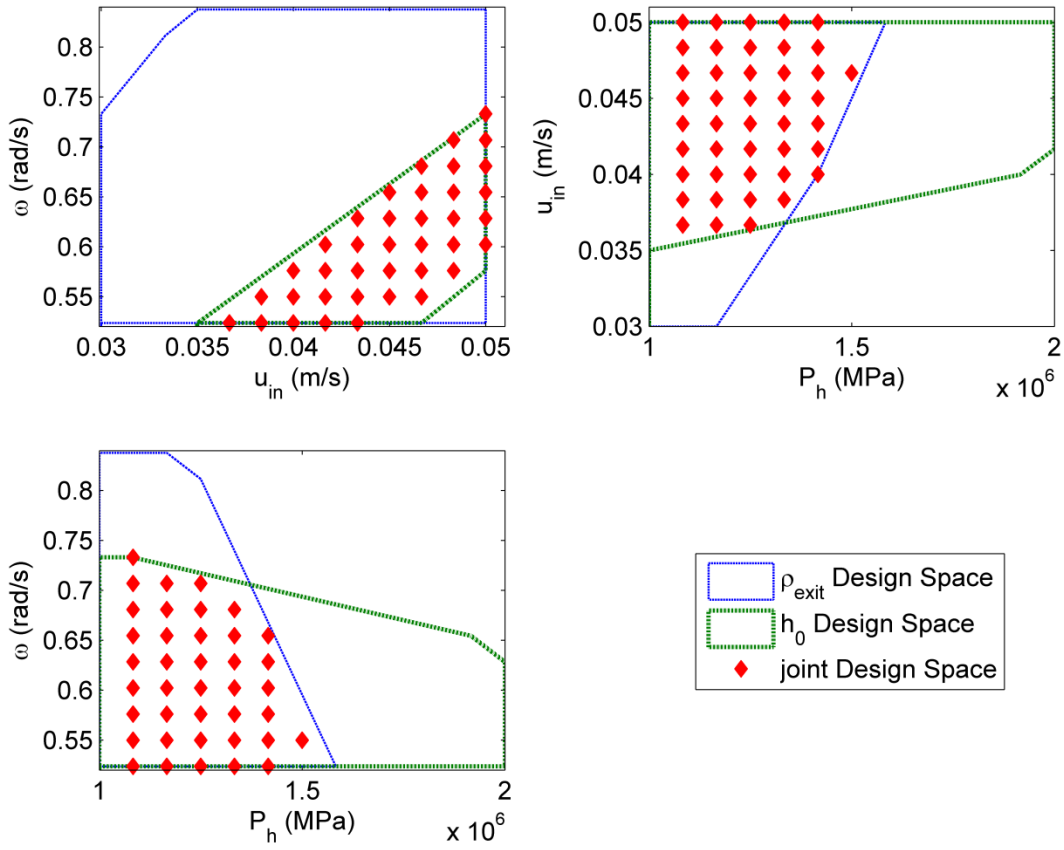
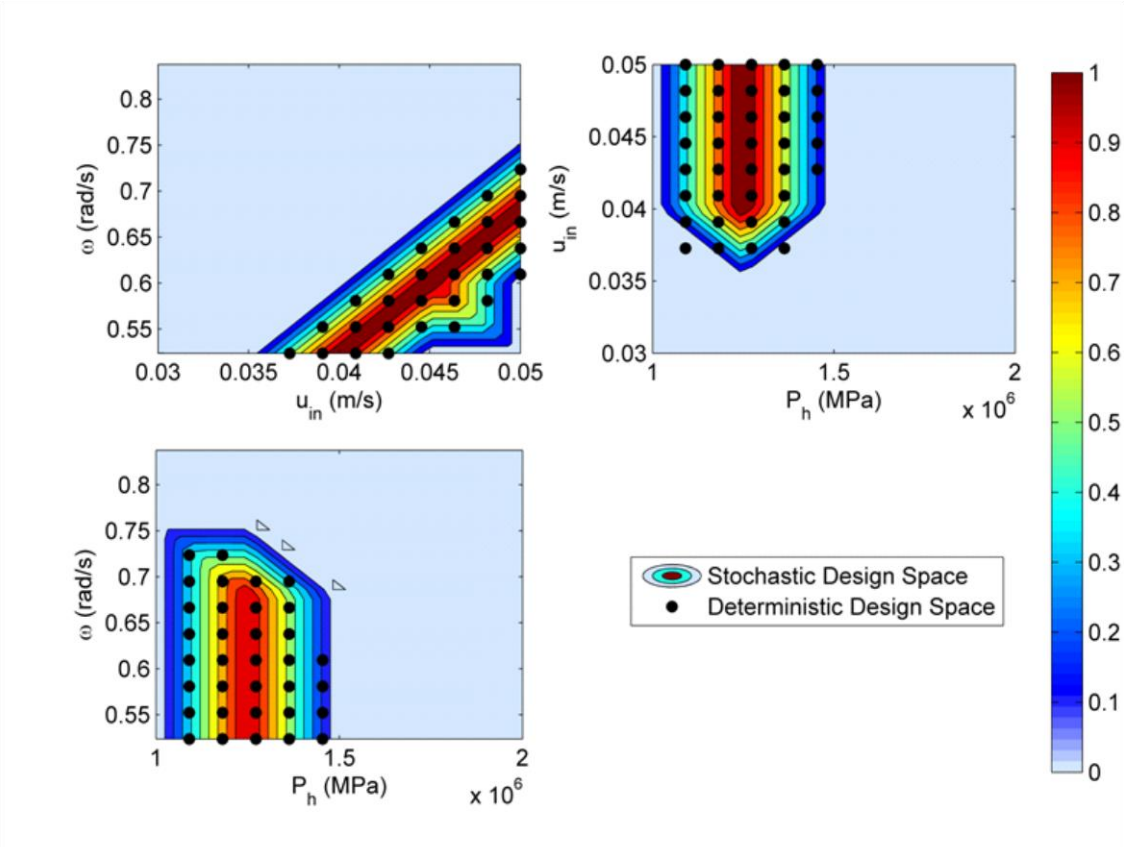


Figure 31. Design Space of Roller compactor

Following a stochastic feasibility approach- similar to the methodology proposed in (Peterson, 2008)- uncertainty of the inlet density is incorporated into the DS. In the roller compaction case study (Figure 31) MATLAB is used for simulating the process under different operating conditions in the presence of uncertainty in the material properties. Specifically, it is assumed that the inlet powder density fluctuates as powder enters in the process and follows a normal distribution of mean 0.3 g/cm^3 and standard deviation of 0.05. The ability to simulate the process by randomly drawing values from the above uncertain parameter from its assigned distributions at each simulation allows

the evaluation of the level of uncertainty that has propagated to both roll gap and ribbon density. Thus the process is simulated under steady state conditions and by studying the predicted density and ribbon throughput, their distribution and the associated parameters can be identified. Based on the obtained steady state values of ribbon density and throughput for a large amount of random simulations, their distribution type is identified as normal and thus the mean and standard deviation can be easily calculated. Based on this information, the probability of each output to lie within the predefined acceptable limits can be calculated using the cumulative distribution. Finally, the joint probability of both outputs belonging within their acceptable limits is the product of the two calculated probabilities. Consequently, for each point within the Knowledge space, the value of the joint probability can be the figure of merit for defining the reliability or the quantification of the assurance of quality of the Design Space based on Equation 29.

In Figure 32, the deterministic Design Space is compared to the stochastic DS , where the effects of the uncertainty of the inlet ribbon density are now taken into account.



It can be observed that the first deterministic approach results in an overestimated and optimistic DS which suggests one can produce ribbons with the desired characteristics operating at regions where the probability of meeting the specifications is only 20- 40%, due to the uncertainty of the inlet density.

Figure 32. Comparison of Deterministic and Stochastic DS with respect to the operating conditions

Finally the effect of an efficient control strategy is assessed for the same case study. Specifically, it is of interest to investigate whether the effects of the variability caused by the uncertainty of the raw material properties, can be reduced through control. For this analysis, an effective control strategy- which is identified as the manipulation of the angular velocity is evaluated. When solving the black-box feasibility problem, the same optimization problem shown in Problem 38 is solved, however, this time ω is also a decision variable to the problem, since it is the control variable and thus an additional degree of freedom is added to the problem. Following this approach, the optimal control variable ω over the investigated space is chosen such that feasibility is minimized. Comparing Figure 33 with Figure 32, it can be seen that the reliability of the process to meet quality specifications is higher in a larger fraction of the Knowledge Space, when control is employed. Thus, it becomes clear that a successful control strategy can alleviate the effects of possible uncertainty of the process input variables, allowing the process to have a larger Design Space.

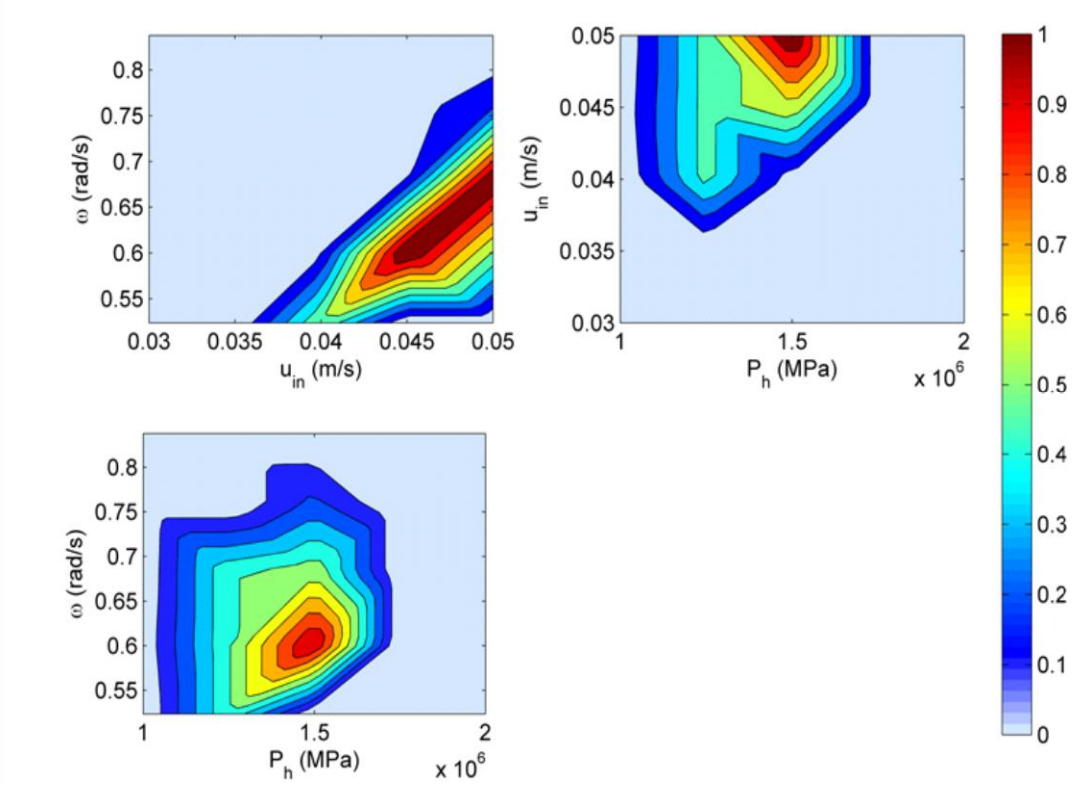


Figure 33. Stochastic Design Space when ω is considered a manipulative variable

4.5. Design Space mapping for integrated systems

The methods for identifying the Design Space described above will be very useful for problems of higher dimensionality and multiple discrete type decisions, coming from integrated systems. For example, identifying the Design Space of an integrated continuous direct compaction process, described in chapter 2, will include multiple input variables, multiple sources of uncertainty and perhaps multiple design aspects, which will lead to large MINLP feasibility problems. Moreover, as discussed in the final chapter of future work, once the work on flowsheet modeling of many production scenarios reaches a state where all of the scenarios can be integrated to form a flowsheet synthesis problem, an optimal integrated Design Space can be identified following the approach of black-box feasibility identifying which flowsheet design is more robust for producing a product of

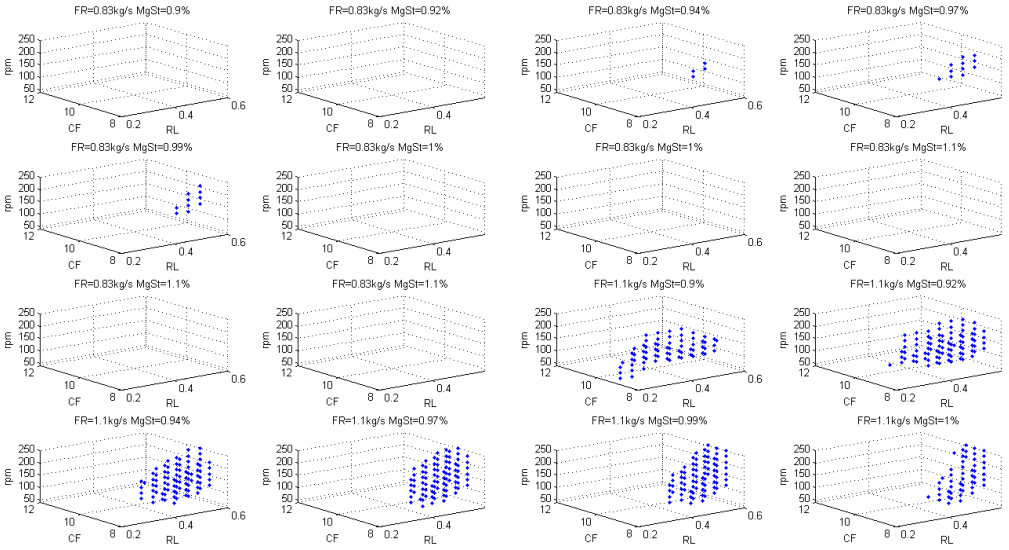
defined specifications for a given set of materials. However, unfortunately, the current form of the flowsheet work and the integration of models is not yet at this point, where flowsheet synthesis is possible.

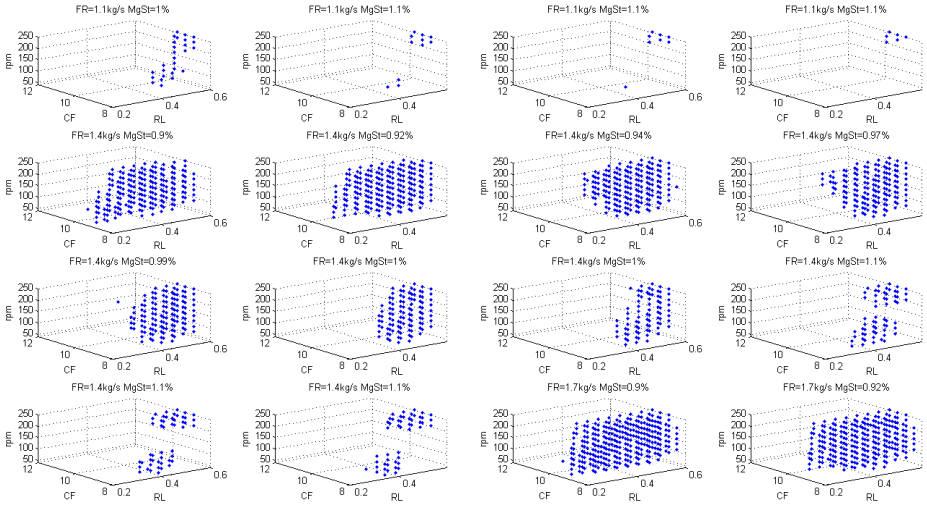
However, the concept of black box feasibility is tested on an integrated case study of the production of tablets using a direct compaction flowsheet, where three constraints form the Design space are related to the final tablet dissolution properties, API concentration and total production count. The identified critical inputs to these outputs based on the current flowsheet model are considered as the total throughput of operation (kg/h), the mixing rotation speed (rpm), the feeder refilling frequency (%), the MgSt concentration, and the compaction force (kPa) of the tablet press, which form the investigated space. In a continuous operation of an open-loop direct compaction line, the feeder hoppers would need to be refilled in order to supply the mixer with the desired components. Based on an extensive study of feeder operation and refilling (Engisch & Muzzio, 2010), it has observed that refilling causes a pulse of material to exit the feeders, leading to a perturbation in the concentration of the mixture which is fed to the mixer. If the total flowrate of operation is high and the residence time of the material in the mixer is not sufficient, these perturbations might not be completely filtered out by the mixer and affect the quality of the final tablets. In addition, the magnitude of the pulse of material is a function of the frequency of sampling, due to the fact that when the feeder is refilled more often, the hopper already contains some material and this causes the loss-in-weight operation to perform better. When refilling is less frequent, however, less total number of perturbations are caused, but if the feeder is refilled when the hopper is almost empty, the magnitude of the pulse of material exiting the feeder is very large. Moreover, the

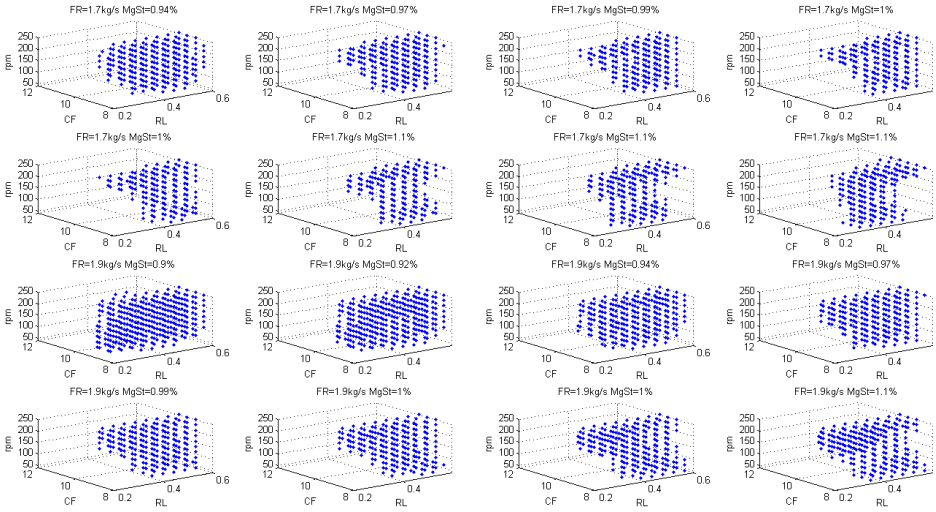
concentration of MgSt affects the RSD of the material in the blender, and this is captured by an empirical correlation based on experimental data. A constraint added for the satisfaction of demand, is enforced such that a minimum number of tablets is produced for a time period of four hours (Problem 39). If at any given time point, the properties of the produced tablets do not satisfy any of the constraints, they are subtracted from the number of produced tablets. In this 5dimensional DS, the application of the black-box feasibility concepts is quite useful since the simulation of the direct compaction line for a four hour production, during which feeders are refilled and perturbations are imposed, is an expensive simulation.

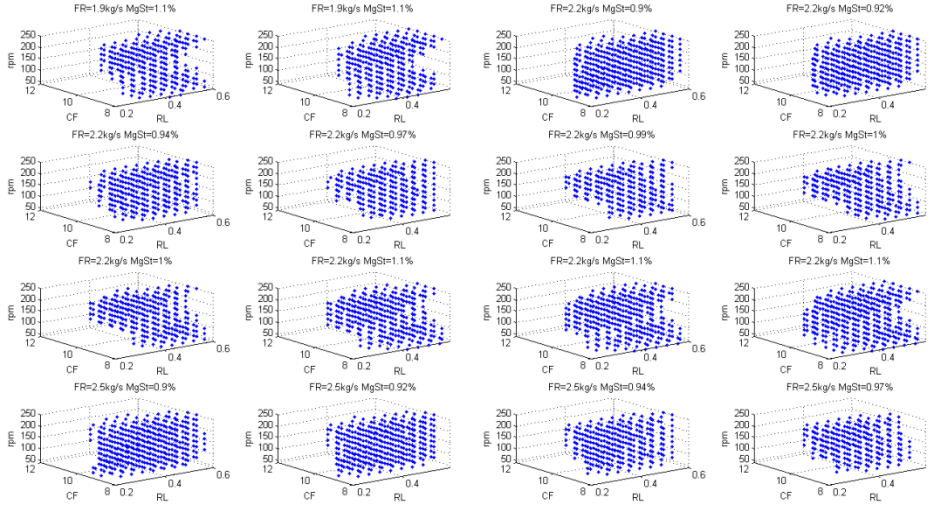
$$
\begin{aligned}
&N_{demand} - N_{tablets} \le u \\
&Dis - Dis^{up} \le u \\
&Dis^{lo} - Dis \le u \\
&C_{API} - C_{API}^{up} \le u \\
&C_{API}^{lo} - C_{API} \le u \\
&RSD - RSD^{up} \le u \\
&RSD^{lo} - RSD \le u \\
&d_{50,j} = normal(d_{50,j}, sd_j) \\
&\rho_j^{bulk} = normal(\rho_{mean,j}^{bulk}, sd_j^{\rho bulk}) \\
&j \in \{APAP, Avicel, MgSt\} \\
&x_i \in \left[x_i^{lo}, x_i^{up}\right] \text{ for } i = 1,..,5
\end{aligned}
\tag{39}
$$

The initial experimental design consists of 51 samples based on a Latin Hypercube Design. The adaptive sampling stage for mapping of feasibility requires an additional 100 samples (reaches the maximum number of iterations), for which the DS is shown in Figure 34.


FR=0.83kg/s MgSt=0.9%
FR=0.83kg/s MgSt=0.92%
FR=0.83kg/s MgSt=0.94%
FR=0.83kg/s MgSt=0.97%
FR=0.83kg/s MgSt=0.99%
FR=0.83kg/s MgSt=1%
FR=0.83kg/s MgSt=1%
FR=0.83kg/s MgSt=1.1%
FR=0.83kg/s MgSt=1.1%
FR=0.83kg/s MgSt=1.1%
FR=1.1kg/s MgSt=0.9%
FR=1.1kg/s MgSt=0.92%
FR=1.1kg/s MgSt=0.94%
FR=1.1kg/s MgSt=0.97%
FR=1.1kg/s MgSt=0.99%
FR=1.1kg/s MgSt=1%
rpm
CF
RL


FR=1.1kg/s MgSt=1%
FR=1.1kg/s MgSt=1.1%
FR=1.1kg/s MgSt=1.1%
FR=1.1kg/s MgSt=1.1%
FR=1.4kg/s MgSt=0.9%
FR=1.4kg/s MgSt=0.92%
FR=1.4kg/s MgSt=0.94%
FR=1.4kg/s MgSt=0.97%
FR=1.4kg/s MgSt=0.99%
FR=1.4kg/s MgSt=1%
FR=1.4kg/s MgSt=1%
FR=1.4kg/s MgSt=1.1%
FR=1.4kg/s MgSt=1.1%
FR=1.4kg/s MgSt=1.1%
FR=1.7kg/s MgSt=0.9%
FR=1.7kg/s MgSt=0.92%
rpm
CF
RL


FR=1.7kg/s MgSt=0.94%
FR=1.7kg/s MgSt=0.97%
FR=1.7kg/s MgSt=0.99%
FR=1.7kg/s MgSt=1%
FR=1.7kg/s MgSt=1%
FR=1.7kg/s MgSt=1.1%
FR=1.7kg/s MgSt=1.1%
FR=1.7kg/s MgSt=1.1%
FR=1.9kg/s MgSt=0.9%
FR=1.9kg/s MgSt=0.92%
FR=1.9kg/s MgSt=0.94%
FR=1.9kg/s MgSt=0.97%
FR=1.9kg/s MgSt=0.99%
FR=1.9kg/s MgSt=1%
FR=1.9kg/s MgSt=1%
FR=1.9kg/s MgSt=1.1%
rpm
CF
RL


FR=1.9kg/s MgSt=1.1%
FR=1.9kg/s MgSt=1.1%
FR=2.2kg/s MgSt=0.9%
FR=2.2kg/s MgSt=0.92%
FR=2.2kg/s MgSt=0.94%
FR=2.2kg/s MgSt=0.97%
FR=2.2kg/s MgSt=0.99%
FR=2.2kg/s MgSt=1%
FR=2.2kg/s MgSt=1%
FR=2.2kg/s MgSt=1.1%
FR=2.2kg/s MgSt=1.1%
FR=2.2kg/s MgSt=1.1%
FR=2.5kg/s MgSt=0.9%
FR=2.5kg/s MgSt=0.92%
FR=2.5kg/s MgSt=0.94%
FR=2.5kg/s MgSt=0.97%
rpm
CF
RL

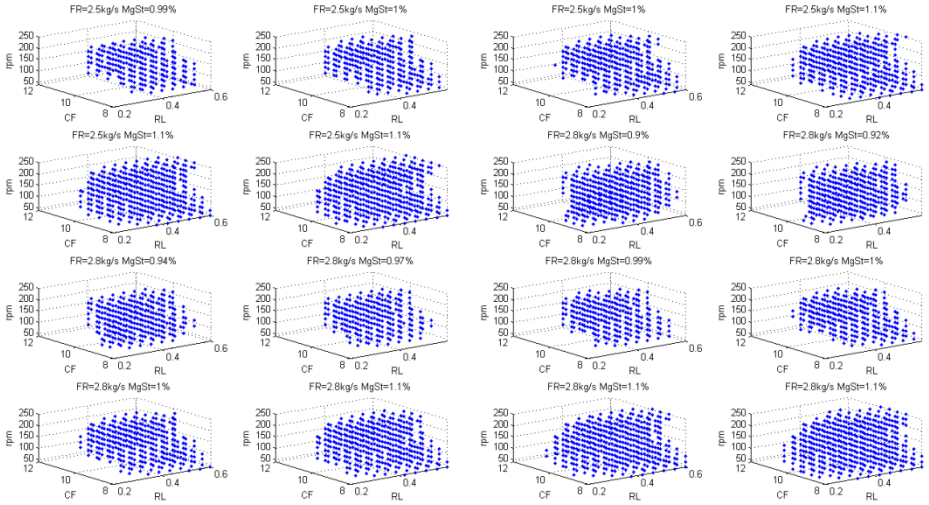


Figure 34. Integrated Design Space for Direct Compaction process.

Figure 34 can be used to point out one great challenge of the efficient representation of multidimensional design spaces, since as the number of critical attributes increases, the graphical representation is no longer suitable. Thus, the need for a metric which could potentially quantify the DS would be a more efficient and beneficial approach. However, the feasibility function proposed in this work, which is used to make this plot, is a very efficient quantification metric, since it is an aggregated value of constraint violation and it summarizes the DS by one single value (negative indicates within DS, while positive indicates outside DS).

Chapter 5

5. Integrated simulation-based optimization of expensive stochastic flowsheet models

5.1. Simulation-based optimization using surrogate models with effects of noise and black-box constraints

Simulation based optimization using surrogates is a research area that has attracted attention to a great extent recently in many industrial applications, where an expensive simulator of a product or process must be optimized (Booker, et al., 1999; Fu, 2002; Gray, Fowler, & Griffin, 2010; D. Huang, 2005; Donald R. Jones, 2001). Examples of expensive simulators used to approximate, design and optimize real systems come from many different engineering fields (T. Chen, Hadinoto, Yan, & Ma; Fowler, Jenkins, & LaLonde, 2010; Horowitz, Guimarães, Dantas, & Afonso, 2010; Husain & Kim, 2010; Kim, von Spakovsky, Wang, & Nelson; Wan, Pekny, & Reklaitis, 2005; Yuan, Wang, Yu, & Fang, 2008) and can range from complex CFD models to large-scale integrated flowsheet models. However, common straight-forward optimization techniques cannot be applied due to the lack of knowledge of analytic derivatives, possible discontinuities caused by if-then operations, high computational cost which prohibits the realization of multiple function evaluations, and finally due to the presence of various types of disturbances (Bertsimas, Nohadani, & Teo, 2010). Specifically, in these types of problems, possible perturbations are classified into two categories: implementation

errors, which refer to suboptimal realizations of the decision variables, and parameter uncertainty errors, which are caused by modeling inaccuracies, noise, deviations between computer simulation and actual model (Jakobsson, Patriksson, Rudholm, & Wojciechowski, 2010).

In order to overcome the problem of computational cost, surrogate based optimization techniques have been developed (Jakobsson, et al., 2010; Jack P. C. Kleijnen, 2009; Queipo, et al., 2005; Regis, 2011; Wan, et al., 2005), which also introduce an additional source of uncertainty, namely that caused by the differences between the computer model and the surrogate model (or metamodel). Surrogate-based optimization, on the other hand, has raised another significant research topic- design and analysis of computer experiments- which aims not only to efficiently plan computer experiments for global optimization of surrogate model approximations, but also to minimize sampling cost (Baldi Antognini & Zagoraiou, 2010; Crary, 2002; E. Davis & Ierapetritou, 2010; D. Huang, Allen, Notz, & Zeng, 2006; Jin, Chen, & Sudjianto, 2005; Donald R. Jones, et al., 1998; Papalambros, Goovaerts, & Sasena, 2002; Pedone, Romano, & Vicario, 2009; Pistone & Vicario, 2009; Sacks, Welch, Toby, & Wynn, 1989; Villemonteix, Vazquez, Sidorkiewicz, & Walter, 2009; Villemonteix, Vazquez, & Walter, 2009).

Surrogate simulation-based optimization belongs into a greater category of algorithms which are derivative-free methods, referring to methods which do not use derivative information to guide the search for an optimum. A comprehensive review of such algorithms is presented in (Rios & Sahinidis, 2009) where the history and performance of such approaches is described and compared. In addition, recently, the first text book for derivative-free optimization methods has been published and is

available online (Conn, et al., 2009). One of the main categorization of such algorithms is ones which search globally, and others which follow a local search approach. The global search methods use a set of samples which are spanned within the entire investigated region in order to fit a global response-surface model, which is then refined in regions of interest. This methodology is similar to the one followed by the proposed black-box feasibility analysis proposed in chapter 4. On the other hand, local derivative-free optimization methods focus within a small trust region of the space, which is updated based on a set of rules which aim to drive the size of the trust region to zero, when a local optimum has been identified.

In order to reach to the final proposed algorithm, both types are assessed and compared in terms of convergence and ability to locate a global optimum. Global search approaches have the advantage of not being dependent on the initial estimate, and have greater chance of locating the global optimum, however, this comes at high computational cost. On the other hand, convergence of global approaches states on the fact that at infinite number of samples, the method will have sampled the space densely, thus it will have definitely located the global optimum. Local trust region framework methods have been studied for optimality extensively. All of the theoretical proof of convergence to a stationary point for trust-region methods can be found in (Conn, et al., 2009). In summary, however, the three aspects which lead to assurance of optimality will be briefly described here.

When employing a surrogate-based trust region framework the three following aspects must be included into the developed algorithm:

(1) **good control of geometry of the sampling sets**: the collected samples within a trust region must have good spanning properties, or else, they must be located in such a way within the trust region, such that they are a representative sample of the space. A full factorial design, for example, has very good spanning properties.

(2) **mechanism to impose descent direction:** if the sample set used to build the surrogate model is based on a positive basis set, then it is guaranteed that a descent direction will be found (if in fact there is one). A set of vectors whose positive span is $\Re^n$, is called a positive spanning set, while a positive spanning set whose vectors are independent is a positive basis. Thus a positive basis is the minimum number of vectors whose convex cone is $\Re^n$, thus one can guarantee to find a descent direction if there is one.

(3) **mechanism to drive step size to zero**: It can be proven that as the step size of the trust region decreases when a descent direction is not found, then the algorithm will converge to a stationary point where the trust region size is very small. For this mechanism to work, it is required to ensure that the accuracy of the surrogate model is good within the trust region, or in other words, ensure that a descent direction is not found because it really does not exist and not because of the model error. Local trust region methods provide a higher guarantee of convergence to a true stationary point, however, they are highly dependent to the initial guess of the algorithm. It is often the case, that trust-region methods will find the closed local optima and get trapped there.

Clearly, many opportunities lie in the research for simulation-based optimization techniques in (a) designing an appropriate computer experiment, (b) identifying an efficient surrogate approximation method which requires the minimum number of

function evaluations and can also handle uncertainty caused by various sources of noise, and (c) distinguishing the suitable gradient-free optimization technique which can successfully identify the global optimum with a limited number of function evaluations in the presence of noise and possible constraints. Overall, such approaches are of interest because they could be greatly beneficial towards the optimization of solid-based flowsheet models which are computationally expensive and contain a large amount of noise due to the high uncertainty introduced by the powder raw materials.

In the surrogate-based optimization literature, the majority of the developed methods deal with unconstrained optimization problems (Bertsimas, et al., 2010), or with box-constrained regions defined by upper and lower bounds of the input variables (D. Huang, et al., 2006; Jakobsson, et al., 2010; Donald R. Jones, 2001; Donald R. Jones, et al., 1998). However, in real systems it is very common that the feasible region is defined by a set of linear, or even non- linear constraints which may be known in advance. This knowledge can be exploited once it is taken into account in later stages by limiting the sampling region for locating the global optimum. Several approaches for avoidance of sampling in infeasible regions have been proposed in the literature. In (D. Huang, 2005), the criterion used for identifying promising sampling locations is multiplied by an assigned probability of a design point to lie within the feasible space, while in (Papalambros, et al., 2002) the authors propose a penalty type approach where a large penalty constant value is added to the function used for identification of new infill samples in order to avoid sampling when the constraints are violated. However, in some cases the closed- form expression of the feasibility constraints may not be available in advance. In such cases, black-box feasibility techniques (Banerjee & Ierapetritou, 2002,

2005; Banerjee, et al., 2010) – described in detail in Chapter 4- are used to approximate the feasibility boundaries. Such techniques require the sampling of the experimental region for accurate approximation of the feasible space, and this information can be then incorporated in the design of computer experiments for optimization, to avoid unnecessary sampling.

Up to now, however, the effects of uncertainty have not been handled yet. Specifically, the interpolating nature of kriging is undesired in this work since it is not observed in the types of real processes which will be handled. In this work, the process is assumed to have a random error or noise which affect the kriging model (non interpolating- Chapter 2)

Proposed algorithm

The proposed approach aims to take advantage of the both global and local search aspects, by combining them within a framework which searches globally initially, and then focuses within small trust regions of interest. Within this framework, the concept of black-box feasibility can be used in the initial stages of global search to provide a good approximation of the feasible regions and direct sampling within only feasible regions. The steps of the algorithm are outlined in Figure 35.

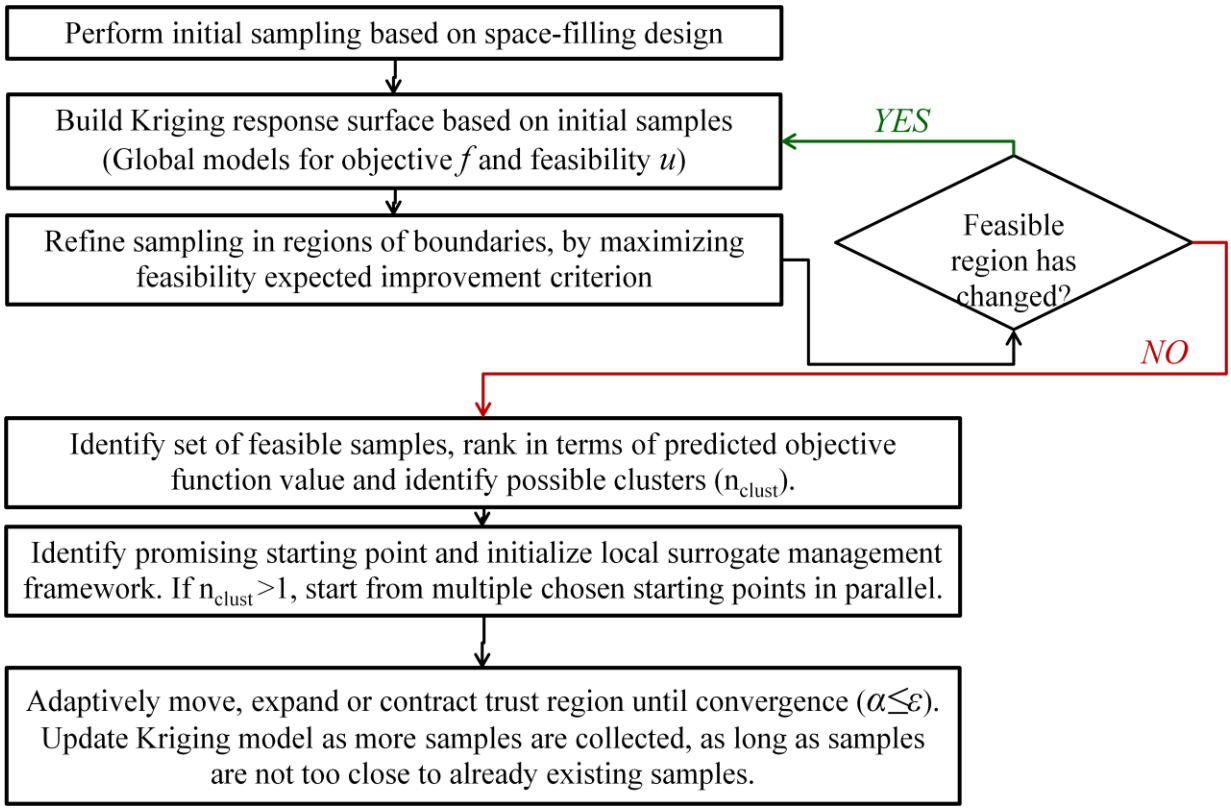


Figure 35. Steps of proposed optimization algorithm for simulation based optimization

Initially a global space filling design is collected, based on which black-box feasibility analysis is performed in order to identify the form of the feasible region. Two kriging models are developed, one for the objective function and one for the feasibility mapping. Once this initial stage has been completed, a large number of feasible pool of points is identified by the prediction of the feasibility function for grid of points within the investigated space. These points are ranked in terms of their objective function value and cluster analysis is performed to identify possible clusters of warm-start initial points. If more than one cluster is identified, then the cluster centers are chosen as initial points for the transition to the next stage of local sampling. This approach aims to minimize the decrease the sampling cost of random multistart methods, which choose random initial

values in order to increase the probability of locating a global optimum. It should be noted that a new kriging model is recalibrated every time a new unique point is added to the sample set.

The final local trust region framework is governed by the following set of rules, based on work described in (Conn, et al., 2009):

(1) An initial trust region size is chosen and the location of the possible next samples are calculated based on a positive basis (n+1 - 2n points). The current iteration kriging model is used to predict the value of these points, and starting from the minimum predicted value the true function is called. If at any point an improved and feasible objective function is obtained, then the search is designated as successful and no further search is performed around this iterate. The new improved sample now becomes the center point of the next trust region. If the predicted value for this point is accurate compared to the true result, then the poll size is increased. If not, then the trust region size remains as is.

(2)If the entire set of positive basis points around a center point are calculated and no improved solution is obtained, the current kriging model is used to optimize the function within the current trust region, subject to the constraint of negative feasibility. If such a point is found, then the search is considered successful and the new point is used as the next center point. The real function is called for this location and if the prediction to the actual point is found to be accurate, then the trust region size is increased. Otherwise it remains the same.

(3) If no improvement in the objective function is obtained after the optimization of the local kriging model, then the trust region size is decreased. The new optimum is now

chosen as the next central point. This procedure is continued as long as the size of the trust region is larger than a small tolerance value. Since all of the aspects of the original trust-region mechanisms are followed here, it is proven that the decrease in the size of the trust region is associated with the drive to the gradient to zero value.

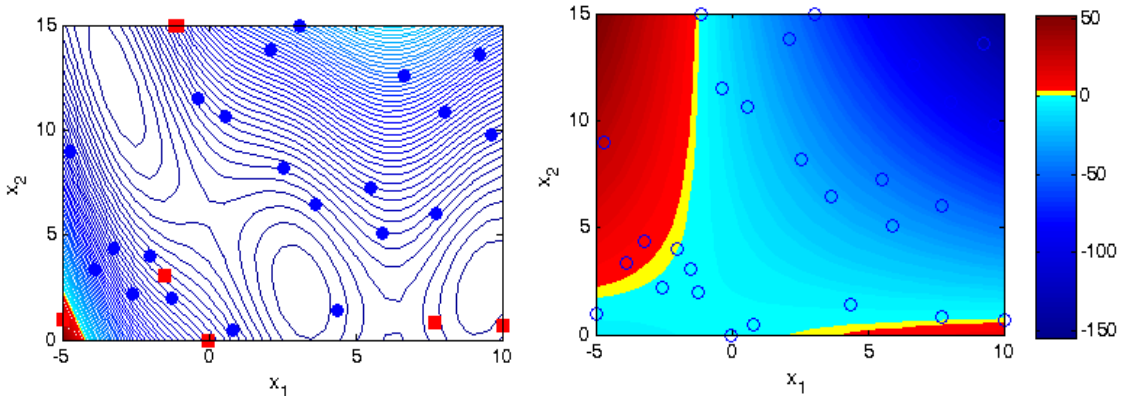
5.2. Testing of developed methodology on optimization benchmark examples

The developed approach is shown to perform quite well in identifying multiple promising locations, in the case of multiple optima. This is shown through Figure 36, where the performance of the method on a constrained and noisy Branin function (Problem 40) is shown.

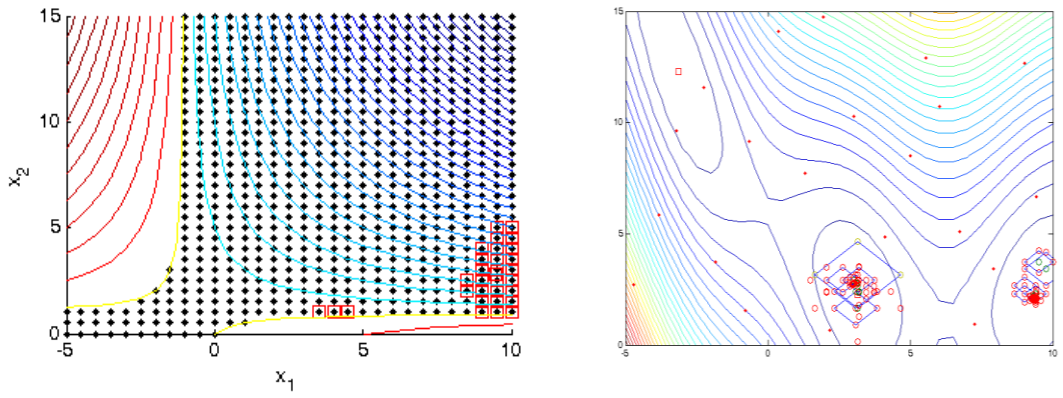
$$
\begin{aligned}
&\min_{x_1,x_2} \ \alpha(x_2 - bx_1^2 + cx_1 - d)^2 + h(1-e)\cos x_1 + h + \varepsilon \\
&s.t. \\
&x_1(1-x_2) - x_2 \le 0 \\
&-5 \le x_1 \le 10 \\
&0 \le x_2 \le 15 \\
&\varepsilon = uniform(-0.01, 0.01) \\
&a = 1, b = {5.1}/{4\pi^2}, c = {5}/{\pi}, d = 6, h = 10, e = {1}/{8\pi}
\end{aligned}
\tag{40}
$$

The algorithm requires 21 samples for the initial model approximation, an additional 4 for the feasibility mapping and two clusters of promising regions are located. The local search is then initiated from these two cluster centers to converge to two similar global optima. The algorithm is further tested on a set of benchmark examples which are modified to contain constraints and noise as well. The performance for locating the optimum (or optima) is compared to a commercially available multistart algorithm which can handle constraints, in terms of the required function evaluations. Table 8 includes the

comparison between the number of function calls required from the proposed approach and the multistart algorithm, to show that similar accuracy can be obtained with far less required calls to the expensive simulation. It should be noted that noise is a very significant factor to convergence. In order to verify this, the algorithm is tested without the effect of noise, and by using an interpolating kriging approach. The number of required function calls is minimized and the algorithm converges to an optimum 100% of the times it is called. However, as the amount of noise increases, the performance of the algorithm during the last steps of small trust regions is very much affected by the noise and fails to converge to the true optimum after a very large number of iterations. However, in the presence of noise and by using a non-interpolating kriging, the performance is more stable (Table 8). This is a very significant conclusion, that implies that performance can be improved by smoothing the regression function, however, in the presence of stochastic effects, only replication would give more accurate estimates of the true response. In the current cases, however, replication is not desired due to the fact that it would increase computational cost tremendously.



(a) (b)



(c) (d)

Figure 36. (a) Contours of Problem Branin, initial samples (blue) and feasibility samples (red), (b) feasibility u, (c) predicted feasible region and clusters of promising initial points, (d) Local trust region performance.

Table 8.Statistics of proposed methodology compared to Tomlab LGO

f	**Current method**	**Found all solutions?**	**Function calls LGO(multistart)-Tomlab**
Branin	197	27/30	4038
Camel	186	28/30	3823
Rosenbrock	206	23/30	3907
Hartman 3	191	28/30	4143
Hartman 6	584	24/30	10071

5.3. Optimization of direct compaction using developed methodology

The approach described in the previous section is implemented for the optimization of a 4 hour day operation of a direct compaction flowsheet model by satisfying all of the constraints of Problem 39. Specifically, the objective is to minimize the cost of operation which includes: operating cost (utilities cost, raw material cost) and cost of waste (cost of produced tablets which are off-spec and should be discarded). The optimization problem has the following form:

$$
\begin{aligned}
&\min_{x} \; c_{feed} + c_{waste} \\
&s.t. \\
&N_{tablets} \geq 1e6 \\
&Dis^{lo} \leq Dis \leq Dis^{up} \\
&C_{API}^{lo} \leq C_{API} \leq C_{API} \\
&RSD^{lo} \leq RSD \leq RSD^{up} \\
&d_{50,j} = normal(d_{50,j}, sd_j) \\
&\rho_j^{bulk} = normal(\rho_{mean,j}^{bulk}, sd_j^{\rho bulk}) \\
&j \in \{APAP, Avicel, MgSt\} \\
&x_i \in \left[x_i^{lo}, x_i^{up}\right] \text{ for } i = 1,..,5
\end{aligned}
\tag{41}
$$

where the decision variables considered with their upper and lower bounds are given in Table 9. The computer experiment is based on a LH design composed initially of 51 samples. Based on this set of points and based on the methodology described in Chapter 4, the feasible region is approximated by 60 additional samples, and the black-box constraints are approximated with a closed-form using a kriging model. Subsequently, the search for the next sampling location can be limited within the feasible space using

constrained optimization approaches, starting from three identified cluster centers. This entire procedure is implemented in MATLAB, which communicates with gPROMS to exchange information. The optimal values are included in Table 9, where it is observed that the minimum cost occurs at mid range throughputs and mid range refilling frequency, since this balances between the often need for refilling and the magnitude of perturbation. Specifically it is observed that at higher throughputs the discarded product is very high and this increases the objective function significantly and leads to the violation of the minimum demand constraint. In addition to this, the optimal mixing rotation speed is at mid range levels which agrees with experimental evidence. The tablet compaction pressure is directly and solely related to tablet hardness, so its optimal value is solely dependent by the desired hardness characteristics. The amount of MgSt, is a little bit higher than the nominal value of 1%, which is also expected since the RSD correlation to MgSt imposes the effect of a decrease in RSD with an increase in MgSt (based on experimental data). Thus, due to the small amount of MgSt, the cost is not significantly affected if more amount is included in the formulation, as long as the product specifications allow for this modification.

Table 9.Decision variables of optimization problem, bounds and optimal values

Variable	**Lower Bound**	**Optimal value**	**Upper bound**
Throughput (kg/h)	10	54	100
MgSt (%)	0.99	1.006	1.01
Refill ratio (%)	20	52	60
Compression Force (kPa)	8	8.9	12
Mixing Rate (rpm)	40	81	250

It is realized that this is a simple optimization problem, however, feeder refilling strategies are an important aspect of continuous operation of powder flowing processes. As the flowsheet models are improved, through the incorporation of the complete set of effects of interest, larger and even more interesting optimization problems will be formed for the design of the pilot plant. However, the surrogate-based approaches have not been tested by us for large dimensionality problems. It is predicted, that the initial global sampling overhead will be a limiting step, as the dimensionality of the problem becomes very high. For this reason it is suggested that, if the decoupling of expensive and non-expensive models is possible, then surrogate models are used only for the components which are expensive to evaluate (Caballero & Grossmann, 2008; Henao & Maravelias, 2011). This way, the dimensionality of the problems will not drastically increase, and a fraction of the values will come from the actual true model, which will naturally be more accurate.

Chapter 6

6. Conclusions and future perspectives

The initial aim of this work was the integration of various forms of knowledge in order to develop tools for the simulation and optimization of powder handling processes. However, the nature of the systems which have been studied has also given rise to a variety of different problems such as the mapping of the feasibility of complex black-box processes and the optimization of constrained expensive simulation models. Overall, this work aims to introduce certain PSE tools which can be beneficial towards process and product design to the pharmaceutical industry, but also develop more generally applied computational tools for the design and optimization of products with high uncertainty and quality constraints.

The major challenges when dealing with pharmaceutical solid processes can be summarized as the lack of knowledge of the properties that characterize the powder mixtures, critical operating conditions and how these affect the final product properties. Through the development of integrated flowsheet models, the aim is to capture the insofar studied and known correlations which govern the behavior of powders as they are processed as well as their dynamics. The developed flowsheet configurations describe the projected real life typical operation of a continuous plant for the production of pharmaceutical tablets. Multiple feeders continuously supply raw materials into a blender, from which the exiting blend can be processed through direct compaction, dry granulation or wet granulation. The required inputs to the simulation are the total throughput, the composition of each individual material, as well as the particle size

distribution and bulk density of each material. Through the simulation of different step changes and expected perturbations, it is observed that the unit operation models can capture certain desired effects, such as the effects in the concentration of the mixture, the bulk density and particle size. However, there are several significant effects which the developed flowsheet models do not capture efficiently, which should be incorporated in order to develop an even better tool for simulation, design and optimization. Several methods for allowing the incorporation of additional effects are proposed in chapter 3, through reduced order models; however, significant missing correlations are also described as future work.

The lack of first principle models and the use of extremely expensive simulations in powder processing, lead us to the development of reduced order data-based models. Through the integration of reduced-order models within the flowsheet simulations, it is aimed to enlarge the set of effects which are captured by the unit operation models, such as design and material property effects.

In addition, process design space is described in this work as the concept of process feasibility and novel methods to identify the boundaries feasible operation are developed. Specifically it is shown that the representation of all the process constraints through one aggregated feasibility test value, allows the quantification of a boundary and the search for its exact location, which is very critical towards the design of uncertain processes.

Finally, the need for optimization of the developed flowsheet models, which are computationally expensive and noisy, has directed us towards the use and development of a simulation based optimization approach, which employs surrogate models as

approximations of the actual process models. The developed methodology employs concepts of global search for both good objective function values and feasibility, and a local trust region framework as a final stage to converge to local optima. The developed methodology has been applied for the optimization of a direct compaction flowsheet in order to minimize cost of 4 hour continuous production of a 3 component formulation tablet. The goal of this optimization is the identification of the most efficient refilling strategy of the powder feeders, which is found to be at mid fill level of the hopper volume. In addition, the optimal operating compaction pressure of the tablet press, rotation speed of mixer, MgSt concentration and total throughput of the process are identified through this optimization procedure.

This work aims to point out the great opportunities which arise from merging all the insofar available knowledge, experience, experimental and modeling work available for the development of computer aided tools for the design and optimization of the production of pharmaceutical solid dosage forms. Several new directions to be pursued in the future have arised from this work.

Firstly, the flowsheet modeling library must be refined in order to include a complete set of unit operation models which have a common set of inputs and outputs and can capture the entire set of correlations of interest. The predicted outputs should be correlated with significant and measurable inputs, in order to allow for the system to be represented efficiently, designed, controlled and optimized. Specifically, the unit operation models which necessarily need further improvement are the hopper and the tablet press. The behavior of powders within a hopper is governed by complex phenomena which are dependent on the particle size distribution and flow characteristics

of a powder, as well as hopper geometry. In addition, all of the above characteristics lead to various phenomena in the mixer, such as segregation or further mixing of the material. All of this phenomena may be described well either by a DEM study, a Population Balance Model or a hybrid DEM/PBM model, such as the one described in chapter 3 for the mixer. This processing step should definitely be considered as the next step, since hoppers are necessary units which highly affect the behavior of processed powders. In addition, the tablet press is another major gap in the integrated flowsheets, since an experimental data set must be collected in order to estimate the parameters of the empirical correlation. In order to improve this model, a reduced-order model based on the work of (Gonzalez & Cuitiño, 2012) is proposed, since this is a very detailed and validated simulation of the performance of the tablet press, correlating input parameters which can be provided for any material to properties of the actual produced tablet. Overall, through this work, a general framework is developed for the extraction of DEM data and the methodology required to build a reduced order model from this data. This framework has been integrated in gPROMS as an additional unit operation and it can act as a sensor connected to the output of each process exit port of interest.

In terms of the future of materials property database, the future project team should focus on connecting in-house pure components to the formulation data, which is not completed so far. In addition, the variables which represent the performance of the powder mixture downstream within various processes, should be designed and added as new columns to the output data, such that they can be correlated to process behavior. Towards the decision for which variables should be included in this dataset, every process step must be studied at a time, identifying which variables are significant within

each process unit, and not yet predicted by the existing models. Finally the necessary DOE for the collection of such properties for the set of mixture blends must be decided upon. The existence of the continuous pilot plant within Rutgers University should really help in performing such integrated experiments, by collecting information from all units during one run.

Flowsheet synthesis, which is a steady-state optimization problem that aims to identify the optimal process configuration of an integrated system, is the final goal of this Specific Aim. Aspects which can be identified by the solution of a synthesis problem are optimal operating conditions, but also optimal design configurations, and need for recycles based on a given objective. The optimization problem formed enumerates possible feasible designs in order to identify the optimal sequence and this implies the formation of a large scale mixed-integer non-linear optimization problem. The envisioned synthesis framework for tablet manufacturing will be able to identify the most suitable configuration (direct compaction, dry or wet granulation) based on the raw material properties and final product objectives. A heuristic approach for making such decisions is outlined by Ng et al. (Ng & Fung, 2003), however the aim of this work will be to set the foundations for the development of a computational tool in order to make these decisions in a systematic way. The objective of the optimization synthesis problem may be the product variability as well as cost minimization for the production of a specific formulation. Following the optimization and feasibility algorithms developed in this work, an integrated optimal design space may also be mapped, showing the feasible operating or uncertain variable ranges in order to satisfy all of the desired constraints and objectives.

Bibliography

Acevedo, J., & Pistikopoulos, E. N. (1998). Stochastic optimization based algorithms for process synthesis under uncertainty. *Computers & Chemical Engineering, 22*(4-5), 647-671.

Akande, O. F., Rubinstein, M. H., & Ford, J. L. (1997). Examination of the compaction properties of a 1:1 acetaminophen:Microcrystalline cellulose mixture using precompression and main compression. *Journal of Pharmaceutical Sciences, 86*(8), 900-907.

am Ende, M. T., Moses, S. K., Carella, A. J., Gadkari, R. A., Graul, T. W., Otano, A. L., et al. (2007). Improving the Content Uniformity of a Low-Dose Tablet Formulation Through Roller Compaction Optimization. *Pharmaceutical Development and Technology, 12*(4), 391-404.

Anttonen, J. S. R., King, P. I., & Beran, P. S. (2003). POD-based reduced-order models with deforming grids. *Mathematical and Computer Modelling, 38*(1-2), 41-62.

Aquino, W. (2007). An object-oriented framework for reduced-order models using proper orthogonal decomposition (POD). [doi: 10.1016/j.cma.2007.05.009]. *Computer Methods in Applied Mechanics and Engineering, 196*(41-44), 4375-4390.

Auger, A., Hansen, N., Perez Zerpa, J., Ros, R., & Schoenauer, M. (2009). Experimental Comparisons of Derivative Free Optimization Algorithms. In J. Vahrenhold (Ed.), *Experimental Algorithms* (Vol. 5526, pp. 3-15): Springer Berlin / Heidelberg.

Baldea, M., & Daoutidis, P. (2006). Model reduction and control of reactor–heat exchanger networks. *Journal of Process Control, 16*(3), 265-274.

Baldi Antognini, A., & Zagoraiou, M. (2010). Exact optimal designs for computer experiments via Kriging metamodelling. [doi: DOI: 10.1016/j.jspi.2010.03.027]. *Journal of Statistical Planning and Inference, 140*(9), 2607-2617.

Banerjee, I., & Ierapetritou, M. G. (2002). Design Optimization under Parameter Uncertainty for General Black-Box Models. *Industrial & Engineering Chemistry Research, 41*(26), 6687-6697.

Banerjee, I., & Ierapetritou, M. G. (2003). Parametric process synthesis for general nonlinear models. *Computers & Chemical Engineering, 27*(10), 1499-1512.

Banerjee, I., & Ierapetritou, M. G. (2005). Feasibility Evaluation of Nonconvex Systems Using Shape Reconstruction Techniques. [doi: 10.1021/ie049294d]. *Industrial & Engineering Chemistry Research, 44*(10), 3638-3647.

Banerjee, I., Pal, S., & Maiti, S. Computationally efficient black-box modeling for feasibility analysis. *Computers & Chemical Engineering, In Press, Corrected Proof*.

Banerjee, I., Pal, S., & Maiti, S. (2010). Computationally efficient black-box modeling for feasibility analysis. [doi: DOI: 10.1016/j.compchemeng.2010.02.016]. *Computers & Chemical Engineering, 34*(9), 1515-1521.

Bansal, V., Perkins, J. D., & Pistikopoulos, E. N. (1998). Flexibility analysis and design of dynamic processes with stochastic parameters. *Computers & Chemical Engineering, 22*(Supplement 1), S817-S820.

Barron, M. G., Duvall, S. E., & Barron, K. J. (2004). Retrospective and current risks of mercury to panthers in the Florida Everglades. *Ecotoxicology, 13*(3), 223-229.

Berkooz, G., Holmes, P., & Lumley, J. L. (1993). The Proper Orthogonal Decomposition in the Analysis of Turbulent Flows. *Annual Review of Fluid Mechanics, 25*(1), 539-575.

Bertrand, F., Leclaire, L. A., & Levecque, G. (2005). DEM-based models for the mixing of granular materials. *Chemical Engineering Science, 60*(8-9), 2517-2531.

Bertsimas, D., Nohadani, O., & Teo, K. M. (2010). Robust Optimization for Unconstrained Simulation-Based Problems. *Oper. Res., 58*(1), 161-178.

Betz, G., Junker-Bürgin, P., & Leuenberger, H. (2003). Batch And Continuous Processing In The Production Of Pharmaceutical Granules#. *Pharmaceutical Development and Technology, 8*(3), 289-297.

Biegler, L. T., & Grossmann, I. E. (2004). Retrospective on optimization. [doi: 10.1016/j.compchemeng.2003.11.003]. *Computers & Chemical Engineering, 28*(8), 1169-1192.

Biegler, L. T., Grossmann, I. E., & Westerberg, A. W. (1997). *Systematic methods of chemical process design*. Upper Saddle River, N.J.: Prentice Hall PTR.

Biegler T L, Grossmann E I, & Westerberg W.A. (1997). *Systematic Methods of Chemica Process Design*. New Jersey: Prentice Hall.

Booker, A. J., Dennis, J. E., Frank, P. D., Serafini, D. B., Torczon, V., & Trosset, M. W. (1999). A rigorous framework for optimization of expensive functions by surrogates. *Structural and Multidisciplinary Optimization, 17*(1), 1-13.

Boukouvala, F., Dubey, A., Vanarase, A., Ramachandran, R., Muzzio, F. J., & Ierapetritou, M. (2011a). Computational Approaches for Studying the Granular Dynamics of Continuous Blending Processes, 2 – Population Balance and Data-Based Methods. *Macromolecular Materials and Engineering*, In Press.

Boukouvala, F., Dubey, A., Vanarase, A., Ramachandran, R., Muzzio, F. J., & Ierapetritou, M. (2011b). Computational Approaches for Studying the Granular Dynamics of Continuous Blending Processes, 2 – Population Balance and Data-Based Methods. *Macromolecular Materials and Engineering*, n/a-n/a.

Boukouvala, F., & Ierapetritou, M. G. (2012). Feasibility analysis of black-box processes using an adaptive sampling Kriging-based method. [doi: 10.1016/j.compchemeng.2011.06.005]. *Computers & Chemical Engineering, 36*(0), 358-368.

Boukouvala, F., Muzzio, F., & Ierapetritou, M. Design Space of Pharmaceutical Processes Using Data-Driven-Based Methods. *Journal of Pharmaceutical Innovation, 5*(3), 119-137.

Boukouvala, F., Muzzio, F., & Ierapetritou, M. (2010). Design Space of Pharmaceutical Processes Using Data-Driven-Based Methods. *Journal of Pharmaceutical Innovation, 5*(3), 119-137.

Boukouvala, F., Muzzio, F., & Ierapetritou, M. (2012). Reduced order DEM modeling. *Submitted to Chemical Engineering Science, under review*.

Boukouvala, F., Muzzio, F. J., & Ierapetritou, M. G. (2011). Dynamic Data-Driven Modeling of Pharmaceutical Processes. *Industrial & Engineering Chemistry Research, 50*(11), 6743-6754.

Briesen, H., & Marquardt, W. (2000). Adaptive model reduction and simulation of thermal cracking of multicomponent hydrocarbon mixtures. *Computers & Chemical Engineering, 24*(2-7), 1287-1292.

Caballero, J. A., & Grossmann, I. E. (2008). An algorithm for the use of surrogate models in modular flowsheet optimization. *AIChE Journal, 54*(10), 2633-2650.

Chen, L., Hontoir, Y., Huang, D., Zhang, J., & Morris, A. J. A. J. (2004). Combining first principles with black-box techniques for reaction systems. [doi: 10.1016/j.conengprac.2003.09.006]. *Control Engineering Practice, 12*(7), 819-826.

Chen, T., Hadinoto, K., Yan, W., & Ma, Y. Efficient meta-modelling of complex process simulations with time-space-dependent outputs. [doi: DOI: 10.1016/j.compchemeng.2010.05.013]. *Computers & Chemical Engineering, In Press, Corrected Proof*.

Christofides, P. D. (2001). *Nonlinear and robust control of PDE systems : methods and applications to transport-reaction processes*. Boston: Birkhäuser.

Cizmas, P. G., Palacios, A., O'Brien, T., & Syamlal, M. (2003). Proper-orthogonal decomposition of spatio-temporal patterns in fluidized beds. [doi: 10.1016/S0009-2509(03)00323-3]. *Chemical Engineering Science, 58*(19), 4417-4427.

Conn, A. R., Scheinberg, K., & Vicente, L. N. (2009). *Introduction to derivative-free optimization*. Philadelphia: Society for Industrial and Applied Mathematics/Mathematical Programming Society.

Contou-Carrere, M. N., & Daoutidis, P. (2008). Model reduction and control of multi-scale reaction–convection processes. *Chemical Engineering Science, 63*(15), 4012-4025.

Crary, S. B. (2002). Design of Computer Experiments for Metamodel Generation. *Analog Integrated Circuits and Signal Processing, 32*(1), 7-16.

Cressie, N. (1993). *Statistics for Spatial Data (Wiley Series in Probability and Statistics)*: Wiley-Interscience.

Cunningham, J. C., Winstead, D., & Zavaliangos, A. (2010). Understanding variation in roller compaction through finite element-based process modeling. [doi: DOI: 10.1016/j.compchemeng.2010.04.008]. *Computers & Chemical Engineering, 34*(7), 1058-1071.

Davis, B., Lundsberg, L., & Cook, G. (2008). PQLI Control Strategy Model and Concepts. *Journal of Pharmaceutical Innovation, 3*(2), 95-104.

Davis, E., & Ierapetritou, M. (2007). A kriging method for the solution of nonlinear programs with black-box functions. *AIChE Journal, 53*(8), 2001-2012.

Davis, E., & Ierapetritou, M. (2008). A Kriging-Based Approach to MINLP Containing Black-Box Models and Noise. [doi: 10.1021/ie800028a]. *Industrial & Engineering Chemistry Research, 47*(16), 6101-6125.

Davis, E., & Ierapetritou, M. (2009). A kriging based method for the solution of mixed-integer nonlinear programs containing black-box functions. *J. of Global Optimization, 43*(2-3), 191-205.

Davis, E., & Ierapetritou, M. (2010). A centroid-based sampling strategy for kriging global modeling and optimization. *AIChE Journal, 56*(1), 220-240.

Deane, A. E., Kevrekidis, I. G., Karniadakis, G. E., & Orszag, S. A. (1991). *Low?dimensional models for complex geometry flows: Application to grooved channels and circular cylinders* (Vol. 3): AIP.

Dec, R. T., Zavaliangos, A., & Cunningham, J. C. (2003). Comparison of various modeling methods for analysis of powder compaction in roller press. [doi: DOI: 10.1016/S0032-5910(02)00203-6]. *Powder Technology, 130*(1-3), 265-271.

Engisch, W., & Muzzio, F. J. (2010). *Hopper Refill of Loss-in-Weight Feeding Equipment*. Paper presented at the AIChE Annual Conference.

Faqih, A. N., Alexander, A. W., Muzzio, F. J., & Tomassone, M. S. (2007). A method for predicting hopper flow characteristics of pharmaceutical powders. [doi: DOI: 10.1016/j.ces.2006.06.027]. *Chemical Engineering Science, 62*(5), 1536-1542.

Floudas, C. A., & Gumus, Z. H. (2001). Global Optimization in Design under Uncertainty: Feasibility Test and Flexibility Index Problems. *Industrial & Engineering Chemistry Research, 40*(20), 4267-4282.

Forrester, A. I. J., Sobester, A., & Keane, A. J. (2008). *Engineering design via surrogate modeling: A practical guide*: Wiley.

Forrester, A. I. J., Sóbester, A., & Keane, A. J. (2008). Engineering Design via Surrogate Modelling - A Practical Guide: John Wiley & Sons.

Fowler, K., Jenkins, E., & LaLonde, S. (2010). Understanding the effects of polymer extrusion filter layering configurations using simulation-based optimization. *Optimization and Engineering, 11*(2), 339-354.

Frouzakis, C. E., Kevrekidis, Y. G., Lee, J., Boulouchos, K., & Alonso, A. A. (2000). Proper orthogonal decomposition of direct numerical simulation data: Data reduction and observer construction. [doi: 10.1016/S0082-0784(00)80197-6]. *Proceedings of the Combustion Institute, 28*(1), 75-81.

Fu, M. C. (2002). Feature Article: Optimization for simulation: Theory vs. Practice. *INFORMS JOURNAL ON COMPUTING, 14*(3), 192-215.

Galbally, D., Fidkowski, K., Willcox, K., & Ghattas, O. (2010). Non-linear model reduction for uncertainty quantification in large-scale inverse problems. *International Journal for Numerical Methods in Engineering, 81*(12), 1581-1608.

Gao, Y., Ierapetritou, M., & Muzzio, F. (2011a). Investigation on the effect of blade patterns on continuous solid mixing performance. *The Canadian Journal of Chemical Engineering, 89*(5), 969-984.

Gao, Y., Ierapetritou, M. G., & Muzzio, F. J. (2011b). Periodic section modeling of convective continuous powder mixing processes. *AIChE Journal, In Press*, DOI: 10.1002/aic.12348.

García-Muñoz, S., Dolph, S., & Ward, H. W. (2010). Handling uncertainty in the establishment of a design space for the manufacture of a pharmaceutical product. *Computers & Chemical Engineering, 34*(7), 1098-1107.

García-Muñoz, S., Dolph, S., & Ward Ii, H. W. Handling uncertainty in the establishment of a design space for the manufacture of a pharmaceutical product. *Computers & Chemical Engineering, 34*(7), 1098-1107.

García-Muñoz, S., & Polizzi, M. (2012). WSPLS — A new approach towards mixture modeling and accelerated product development. *Chemometrics and Intelligent Laboratory Systems*, 1-6.

Garcia, T., Cook, G., & Nosal, R. (2008). PQLI Key Topics - Criticality, Design Space, and Control Strategy. *Journal of Pharmaceutical Innovation, 3*(2), 60-68.

Gay, D. H., & Ray, W. H. (1995). Identification and Control of Distributed-Parameter Systems by Means of the Singular-Value Decomposition. *Chemical Engineering Science, 50*(10), 1519-1539.

Gernaey, K. V., & Gani, R. (2010). A model-based systems approach to pharmaceutical product-process design and analysis. [doi: DOI: 10.1016/j.ces.2010.05.003]. *Chemical Engineering Science, 65*(21), 5757-5769.

Glasser, B. J., Remy, B., Canty, T. M., & Khinast, J. G. (2010). Experiments and simulations of cohesionless particles with varying roughness in a bladed mixer. *Chemical Engineering Science, 65*(16), 4557-4571.

Glasser, B. J., Remy, B., & Khinast, J. G. (2009). Discrete Element Simulation of Free Flowing Grains in a Four-Bladed Mixer. *Aiche Journal, 55*(8), 2035-2048.

Glasser, B. J., Remy, B., & Khinast, J. G. (2010). The Effect of Mixer Properties and Fill Level on Granular Flow in a Bladed Mixer. *Aiche Journal, 56*(2), 336-353.

Glosmann, P. (2010). Reduction of discrete element models by Karhunen-LoSve transform: a hybrid model approach. *Computational Mechanics, 45*(4), 375-385.

Gluba, T., Obraniak, A., & Gawot-Mlynarczyk, E. (2004). The effect of granulation conditions on bulk density. *Physiochemical Problems of Mineral Processing*.

Gonnissen, Y., Gonçalves, S. I. V., De Geest, B. G., Remon, J. P., & Vervaet, C. (2008). Process design applied to optimise a directly compressible powder produced via a continuous manufacturing process. [doi: DOI: 10.1016/j.ejpb.2007.09.007]. *European Journal of Pharmaceutics and Biopharmaceutics, 68*(3), 760-770.

Gonzalez, M., & Cuitiño, A. M. (2012). A nonlocal contact formulation for confined granular systems. [doi: 10.1016/j.jmps.2011.10.004]. *Journal of the Mechanics and Physics of Solids, 60*(2), 333-350.

Gorsek, A., & Glavic, P. (1997). Design of Batch Versus Continuous Processes: Part I: Single-Purpose Equipment. *Chemical Engineering Research and Design, 75*(7), 709-717.

Gray, G. A., Fowler, K., & Griffin, J. D. (2010). Hybrid optimization schemes for simulation-based problems. [doi: DOI: 10.1016/j.procs.2010.04.150]. *Procedia Computer Science, 1*(1), 1349-1357.

Gregory W Stockdale, & Cheng, A. (2009). Finding Design Space and Reliable Operating Region Using a Multivariable Bayesian Approach with Experimental Design. *Quality Technology and Quantitative Management, 6*(4), 391-408

Grossman, I. E., & Floudas, C. A. (1987). Active Constraint Strategy for Flexibility Analysis in Chemical Processes. *Computers & Chemical Engineering, 11*(6), 675-693. Retrieved from http://cepac.cheme.cmu.edu/pasi2008/slides/ierapetritou/library/reading/GrossFlouFlexCACE.pdf

Grossmann, I. E. (1990). Mixed-integer nonlinear programming techniques for the synthesis of engineering systems. *Research in Engineering Design, 1*(3), 205-228.

Gruhn, G., Werther, J., & Schmidt, J. (2004). Flowsheeting of solids processes for energy saving and pollution reduction. *Journal of Cleaner Production, 12*(2), 147-151.

Halemane, K. P., & Grossmann, I. E. (1987). Optimal Process Design Under Uncertainty. from http://hdl.handle.net/1903/4569

Hamad, M. L., Bowman, K., Smith, N., Sheng, X., & Morris, K. R. (2010). Multi-scale pharmaceutical process understanding: From particle to powder to dosage form. [doi: DOI: 10.1016/j.ces.2010.01.037]. *Chemical Engineering Science, 65*(21), 5625-5638.

Hassanpour, A., Tan, H. S., Bayly, A., Gopalkrishnan, P., Ng, B., & Ghadiri, M. (2011). Analysis of particle motion in a paddle mixer using Discrete Element Method (DEM). *Powder Technology, 206*(1-2), 189-194.

Haware, R. V., Tho, I., & Bauer-Brandl, A. (2009). Application of multivariate methods to compression behavior evaluation of directly compressible materials. [doi: DOI: 10.1016/j.ejpb.2008.11.008]. *European Journal of Pharmaceutics and Biopharmaceutics, 72*(1), 148-155.

Hein, S., Picker-Freyer, K. M., & Langridge, J. (2008). Simulation of Roller Compaction with Subsequent Tableting and Characterization of Lactose and Microcrystalline Cellulose. *Pharmaceutical Development and Technology, 13*(6), 523-532.

Henao, C. A., & Maravelias, C. T. (2011). Surrogate-based superstructure optimization framework. *AIChE Journal, 57*(5), 1216-1232.

Horowitz, B., Guimarães, L. J. d. N., Dantas, V., & Afonso, S. M. B. (2010). A concurrent efficient global optimization algorithm applied to polymer injection strategies. [doi: DOI: 10.1016/j.petrol.2010.02.002]. *Journal of Petroleum Science and Engineering, 71*(3-4), 195-204.

Hsu, S.-H., Reklaitis, G., & Venkatasubramanian, V. (2010). Modeling and Control of Roller Compaction for Pharmaceutical Manufacturing. Part I: Process Dynamics and Control Framework. *Journal of Pharmaceutical Innovation, 5*(1), 14-23.

Huang, D. (2005). Experimental planning and sequential kriging optimization using variable fidelity data. from http://etd.ohiolink.edu/view.cgi?acc_num=osu1110297243

Huang, D., Allen, T., Notz, W., & Zeng, N. (2006). Global Optimization of Stochastic Black-Box Systems via Sequential Kriging Meta-Models. *Journal of Global Optimization, 34*(3), 441-466.

Huang, J., Kaul, G., Cai, C., Chatlapalli, R., Hernandez-Abad, P., Ghosh, K., et al. (2009). Quality by design case study: An integrated multivariate approach to drug product and process development. [doi: DOI: 10.1016/j.ijpharm.2009.07.031]. *International Journal of Pharmaceutics, 382*(1-2), 23-32.

Husain, A., & Kim, K.-Y. (2010). Enhanced multi-objective optimization of a microchannel heat sink through evolutionary algorithm coupled with multiple surrogate models. [doi: DOI: 10.1016/j.applthermaleng.2010.03.027]. *Applied Thermal Engineering, 30*(13), 1683-1691.

Ierapetritou, M., Gao, Y. G., Y. J., & Muzzio, F. (2011). Characterization of Feeder Effects on Continuous Solid Mixing Using Fourier Series Analysis. *Aiche Journal, 57*(5), 1144-1153.

Ierapetritou, M., Gao, Y. J., Vanarase, A., & Muzzio, F. (2011). Characterizing continuous powder mixing using residence time distribution. *Chemical Engineering Science, 66*(3), 417-425.

Ierapetritou, M. G. (2001). New approach for quantifying process feasibility: Convex and 1-D quasi-convex regions (Vol. 47, pp. 1407-1417).

Ierapetritou, M. G., Boukouvala, F., & Muzzio, F. J. (2010). Predictive Modeling of Pharmaceutical Processes with Missing and Noisy Data. *Aiche Journal, 56*(11), 2860-2872.

Ierapetritou, M. G., Portillo, P. M., & Muzzio, F. J. (2006). Characterizing powder mixing processes utilizing compartment models. *International Journal of Pharmaceutics, 320*(1-2), 14-22.

Ierapetritou, M. G., Portillo, P. M., & Muzzio, F. J. (2007). Hybrid DEM-compartment modeling approach for granular mixing. *Aiche Journal, 53*(1), 119-128.

Immanuel, C. D., & Doyle Iii, F. J. (2005). Solution technique for a multi-dimensional population balance model describing granulation processes. [doi: DOI: 10.1016/j.powtec.2005.04.013]. *Powder Technology, 156*(2-3), 213-225.

Jain, S. (1999). Mechanical properties of powders for compaction and tableting: an overview. [doi: DOI: 10.1016/S1461-5347(98)00111-4]. *Pharmaceutical Science & Technology Today, 2*(1), 20-31.

Jakobsson, S., Patriksson, M., Rudholm, J., & Wojciechowski, A. (2010). A method for simulation based optimization using radial basis functions. *Optimization and Engineering, 11*(4), 501-532.

Jin, R., Chen, W., & Sudjianto, A. (2005). An efficient algorithm for constructing optimal design of computer experiments. [doi: DOI: 10.1016/j.jspi.2004.02.014]. *Journal of Statistical Planning and Inference, 134*(1), 268-287.

John Peterson, Guillermo Miro- Quesada, & Castillo, E. d. (2009). A Bayesian Reliability Approach to Multiple Response Surface optimization with seemingly unrelated regression models. *Quality Technology and Quantitative Management, 6*(4), 353-369

John Peterson, Ronald Snee, Paul McAllister, Timothi Schofield, & Carelia, A. (2009). *Statistics in Pharmaceutical Development and Manufacturing*: GlaxoSmithKlein Pharmaceuticals.

Jones, D. R. (2001). A Taxonomy of Global Optimization Methods Based on Response Surfaces. *Journal of Global Optimization, 21*(4), 345-383.

Jones, D. R., & Law, C. (1993). Lipschitzian Optimization Without the Lipschitz Constant. *October, 79*(1).

Jones, D. R., Schonlau, M., & Welch, W. J. (1998). Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization, 13*(4), 455-492.

Kahrs, O., & Marquardt, W. (2007). The validity domain of hybrid models and its application in process optimization. [doi: 10.1016/j.cep.2007.02.031]. *Chemical Engineering and Processing: Process Intensification, 46*(11), 1054-1066.

Kim, K., von Spakovsky, M. R., Wang, M., & Nelson, D. J. A hybrid multi-level optimization approach for the dynamic synthesis/design and operation/control under uncertainty of a fuel cell system. [doi: DOI: 10.1016/j.energy.2010.08.024]. *Energy, In Press, Corrected Proof*.

Klatt, K.-U., & Marquardt, W. (2009). Perspectives for process systems engineering--Personal views from academia and industry. [doi: DOI: 10.1016/j.compchemeng.2008.09.002]. *Computers & Chemical Engineering, 33*(3), 536-550.

Kleijnen, J. P. C. (2009). Kriging metamodeling in simulation: A review. [doi: DOI: 10.1016/j.ejor.2007.10.013]. *European Journal of Operational Research, 192*(3), 707-716.

Kleijnen, J. P. C., & Beers, W. C. M. v. (2004). Application-Driven Sequential Designs for Simulation Experiments: Kriging Metamodelling. *The Journal of the Operational Research Society, 55*(8), 876-883.

Kushner, J. t., & Moore, F. (2010). Scale-up model describing the impact of lubrication on tablet tensile strength. *Int J Pharm, 399*(1-2), 19-30.

Lang, Y.-d., Malacina, A., Biegler, L. T., Munteanu, S., Madsen, J. I., & Zitney, S. E. (2009). Reduced Order Model Based on Principal Component Analysis for Process Simulation and Optimization†. [doi: 10.1021/ef800984v]. *Energy & Fuels, 23*(3), 1695-1706.

Lang, Y., Zitney, S. E., & Biegler, L. T. (2011). Optimization of IGCC processes with reduced order CFD models. [doi: DOI: 10.1016/j.compchemeng.2011.01.018]. *Computers & Chemical Engineering, In Press, Corrected Proof*.

Lebrun, P., Govaerts, B., Debrus, B., Ceccato, A., Caliaro, G., Hubert, P., et al. (2008). Development of a new predictive modelling technique to find with confidence equivalence zone and design space of chromatographic analytical methods. *Chemometrics and Intelligent Laboratory Systems, 91*(1), 4-16.

Lepore, J., & Spavins, J. (2008). PQLI Design Space. *Journal of Pharmaceutical Innovation, 3*(2), 79-87.

Leuenberger, H. (2001). New trends in the production of pharmaceutical granules: batch versus continuous processing. [doi: DOI: 10.1016/S0939-6411(01)00199-0]. *European Journal of Pharmaceutics and Biopharmaceutics, 52*(3), 289-296.

Liang, Y. C., Lee, H. P., Lim, S. P., Lin, W. Z., Lee, K. H., & Wu, C. G. (2002). PROPER ORTHOGONAL DECOMPOSITION AND ITS APPLICATIONS--PART I: THEORY. [doi: 10.1006/jsvi.2001.4041]. *Journal of Sound and Vibration, 252*(3), 527-544.

Lionberger, R., Lee, S., Lee, L., Raw, A., & Yu, L. (2008). Quality by Design: Concepts for ANDAs. *The AAPS Journal, 10*(2), 268-276.

Lipsanen, T., Antikainen, O., Räikkönen, H., Airaksinen, S., & Yliruusi, J. (2007). Novel description of a design space for fluidised bed granulation. *International Journal of Pharmaceutics, 345*(1-2), 101-107.

Lophaven, S. N., Nielsen, H. B., & Sondergaard, J. (2008). DACE: A Matlab Kriging Toolbox.

Lucia, D. J., Beran, P. S., & Silva, W. A. (2004). Reduced-order modeling: new approaches for computational physics. [doi: 10.1016/j.paerosci.2003.12.001]. *Progress in Aerospace Sciences, 40*(1-2), 51-117.

MacGregor, J., & Bruwer, M.-J. (2008). A Framework for the Development of Design and Control Spaces. *Journal of Pharmaceutical Innovation, 3*(1), 15-22.

McKenzie, P., Kiang, S., Tom, J., Rubin, A. E., & Futran, M. (2006). Can pharmaceutical process development become high tech? *AIChE Journal, 52*(12), 3990-3994.

Mendez, R., Muzzio, F., & Velazquez, C. (2010). Study of the effects of feed frames on powder blend properties during the filling of tablet press dies. [doi: 10.1016/j.powtec.2010.02.010]. *Powder Technology, 200*(3), 105-116.

Michaut, F., Busignies, V., Fouquereau, C., de Barochez, B. H., Leclerc, B., & Tchoreloff, P. (2010). Evaluation of a rotary tablet press simulator as a tool for the characterization of compaction properties of pharmaceutical products. *Journal of Pharmaceutical Sciences, 99*(6), 2874-2885.

Mizonov, V., Berthiaux, H., Marikh, K., Ponomarev, D., & Barantzeva, E. (2004). Modeling continuous powder mixing by means of the theory of Markov chains. *Particulate Science and Technology, 22*(4), 379-389.

Muzzio, F. J., Dubey, A., Sarkar, A., Ierapetritou, M., & Wassgren, C. R. (2011). Computational Approaches for Studying the Granular Dynamics of Continuous Blending Processes, 1-DEM Based Methods. *Macromolecular Materials and Engineering, 296*(3-4), 290-307.

Muzzio, F. J., Portillo, P. M., & Ierapetritou, M. G. (2009). Effects of rotation rate, mixing angle, and cohesion in two continuous powder mixers-A statistical approach. *Powder Technology, 194*(3), 217-227.

Muzzio, F. J., Sudah, O. S., Chester, A. W., Kowalski, J. A., & Beeckman, J. W. (2002). Quantitative characterization of mixing processes in rotary calciners. *Powder Technology, 126*(2), 166-173.

My-Ha, D., Lim, K. M., Khoo, B. C., & Willcox, K. (2007). Real-time optimization using proper orthogonal decomposition: Free surface shape prediction due to underwater bubble dynamics. [doi: 10.1016/j.compfluid.2006.01.016]. *Computers & Fluids, 36*(3), 499-512.

Myers, R. H. (1990). *Classical and Modern Regression with Applications* (Second ed.). Boston, MA: PWS-KENT.

Myers, R. H., & Montgomery, D. C. (1995). *Response Surface Methodology: Process and Product in Optimization Using Designed Experiments*: John Wiley \& Sons, Inc.

Nedderman, R. M., Tüzün, U., Savage, S. B., & Houlsby, G. T. (1982). The flow of granular materials--I: Discharge rates from hoppers. [doi: DOI: 10.1016/0009-2509(82)80029-8]. *Chemical Engineering Science, 37*(11), 1597-1609.

Ng, K. M. (2002). Design and development of solids processes - a process systems engineering perspective. *Powder Technology, 126*(3), 205-210.

Ng, K. M., & Fung, K. Y. (2003). Product-centered processing: Pharmaceutical tablets and capsules. *AIChE Journal, 49*(5), 1193-1215.

Nosal, R., & Schultz, T. (2008). PQLI Definition of Criticality. *Journal of Pharmaceutical Innovation, 3*(2), 69-78.

Oh, M., & Pantelides, C. C. (1996). A modelling and simulation language for combined lumped and distributed parameter systems. *Computers & Chemical Engineering, 20*(6-7), 611-633.

Oksanen, C. A., & García Muñoz, S. (2010). Process modeling and control in drug development and manufacturing. [doi: 10.1016/j.compchemeng.2010.04.013]. *Computers and Chemical Engineering, 34*(7), 1007-1008.

Papalambros, P., Goovaerts, P., & Sasena, M. J. (2002). Exploration of Metamodeling Sampling Criteria for Constrained Global Optimization. *Engineering Optimization*, 263-278.

Pedone, P., Romano, D., & Vicario, G. (2009). New Sampling Procedures in Coordinate Metrology Based on Kriging-Based Adaptive Designs. In P. Erto (Ed.), *Statistics for Innovation* (pp. 103-121): Springer Milan.

Peterson, J. (2008). A Bayesian approach to the ICH Q8 definition of Design Space. *Journal of Biopharmaceutical Statistics, 18*(5),

Peterson, J., & Lief, K. (2010). The ICH Q8 Definition of Design Space: A Comparison of the Overlapping means and Bayesian predictive approaches. *Statistics in Biopharmaceutical Research, 2,*

Pistikopoulos, E. N., & Mazzuchi, T. A. (1990). A novel flexibility analysis approach for processes with stochastic parameters. *Computers & Chemical Engineering, 14*(9), 991-1000.

Pistone, G., & Vicario, G. (2009). Design for Computer Experiments: Comparing and Generating Designs in Kriging Models. In P. Erto (Ed.), *Statistics for Innovation* (pp. 91-102): Springer Milan.

Plumb, K. (2005). Continuous Processing in the Pharmaceutical Industry: Changing the Mind Set. [doi: DOI: 10.1205/cherd.04359]. *Chemical Engineering Research and Design, 83*(6), 730-738.

Polizzi, M. a., & García-Muñoz, S. (2011). A framework for in-silico formulation design using multivariate latent variable regression methods. *International Journal of Pharmaceutics, 418*(2), 235-242.

Poon, J. M. H., Ramachandran, R., Sanders, C. F. W., Glaser, T., Immanuel, C. D., Doyle iii, F. J., et al. (2009). Experimental validation studies on a multi-dimensional and multi-scale population balance model of batch granulation. *Chemical Engineering Science, 64*(4), 775-786.

Portillo, P., Ierapetritou, M. G., & Muzzio, F. J. (2008). Characterization of continuous convective powder mixing processes. *Powder Technology, 182*(3), 368-378.

Portillo, P., Muzzio, F., & Ierapetritou, M. (2008). Using Compartment Modeling to Investigate Mixing Behavior of a Continuous Mixer. *Journal of Pharmaceutical Innovation, 3*(3), 161-174.

Psichogios, D. C., & Ungar, L. H. (1992). A hybrid neural network-first principles approach to process modeling. *AIChE Journal, 38*(10), 1499-1511.

Q8 Pharmaceutical Development. (2006). from http://www.fda.gov/downloads/RegulatoryInformation/Guidances/ucm128029.pdf

Queipo, N. V., Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan, R., & Kevin Tucker, P. (2005). Surrogate-based analysis and optimization. *Progress in Aerospace Sciences, 41*(1), 1-28.

Ramachandran, R., Ansari, M. A., Chaudhury, A., Kapadia, A., Prakash, A. V., & Stepanek, F. (2012a). A quantitative assessment of the influence of primary particle size polydispersity on granule inhomogeneity. [doi: 10.1016/j.ces.2011.11.045]. *Chemical Engineering Science*(In Press).

Ramachandran, R., Arjunan, J., Chaudhury, A., & Ierapetritou, M. G. Model-based control-loop performance assessment of a continuous direct compaction pharmaceutical process. *Submitted to Journal of Pharmaceutical Innovation*.

Ramachandran, R., & Barton, P. I. (2010). Effective parameter estimation within a multi-dimensional population balance model framework. *Chemical Engineering Science, 65*(16), 4884-4893.

Ramachandran, R., & Chaudhury, A. (2012b). Model-based design and control of a continuous drum granulation process. [doi: 10.1016/j.cherd.2011.10.022]. *Chemical Engineering Research and Design*(In Press).

Ramachandran, R., Immanuel, C. D., Stepanek, F., Litster, J. D., & Doyle III, F. J. (2009). A mechanistic model for breakage in population balances of granulation: Theoretical kernel development and experimental validation. *Chemical Engineering Research and Design, 87*(4), 598-614.

Regis, R. G. (2011). Stochastic radial basis function algorithms for large-scale optimization involving expensive black-box objective and constraint functions. [doi: DOI: 10.1016/j.cor.2010.09.013]. *Computers & Operations Research, 38*(5), 837-853.

Rios, L. M., & Sahinidis, N. V. (2009). Derivative-free optimization: A review of algorithms and comparison of software implementation. *AICHE Annual Conference, Nashville, TN*.

Romijn, R., Özkan, L., Weiland, S., Ludlage, J., & Marquardt, W. (2008). A grey-box modeling approach for the reduction of nonlinear systems. [doi: 10.1016/j.jprocont.2008.06.007]. *Journal of Process Control, 18*(9), 906-914.

Sacks, J., Welch, W. J., Toby, J. M., & Wynn, H. P. (1989). Design and Analysis of Computer Experiments. *Statistical Science, 4*(4), 409-423.

Sahinidis, N. V. (2004). Optimization under uncertainty: state-of-the-art and opportunities. *Computers & Chemical Engineering, 28*(6-7), 971-983.

Salman, A., Hounslow, M., & Jonathan, S. (2007). *Granulation*: Elsevier Science.

Samsatli, N. J., Papageorgiou, L. G., & Shah, N. (1998). Robustness metrics for dynamic optimization models under parameter uncertainty. *AIChE Journal, 44*(9), 1993-2006.

Sarkar, A., & Wassgren, C. R. (2009). Simulation of a continuous granular mixer: Effect of operating conditions on flow and mixing. *Chemical Engineering Science, 64*(11), 2672-2682.

Savage, S. B. (1965). The mass flow of granular materials derived from coupled velocity-stress fields. *British Journal of Applied Physics, 16*(12), 1885.

Savage, S. B., & Sayed, M. (1981). Gravity flow of coarse cohesionless granular materials in conical hoppers. *Zeitschrift für Angewandte Mathematik und Physik (ZAMP), 32*(2), 125-143.

Schaber, S. D., Gerogiorgis, D. I., Ramachandran, R., Evans, J. M. B., Barton, P. I., & Trout, B. L. (2011). Economic Analysis of Integrated Continuous and Batch Pharmaceutical Manufacturing: A Case Study. [doi: 10.1021/ie2006752]. *Industrial & Engineering Chemistry Research, 50*(17), 10083-10092.

Seitz, J. A., & Flessland, G. M. (1965). Evaluation of the physical properties of compressed tablets I. Tablet hardness and friability. *Journal of Pharmaceutical Sciences, 54*(9), 1353-1357.

Sinka, I. C., Cunningham, J. C., & Zavaliangos, A. (2003). The effect of wall friction in the compaction of pharmaceutical tablets with curved faces: a validation study of the Drucker-Prager Cap model. [doi: DOI: 10.1016/S0032-5910(03)00094-9]. *Powder Technology, 133*(1-3), 33-43.

Soh, J. L. P., Wang, F., Boersen, N., Pinal, R., Peck, G. E., Carvajal, M. T., et al. (2008). Utility of Multivariate Analysis in Modeling the Effects of Raw Material Properties and Operating Parameters on Granule and Ribbon Properties Prepared in Roller Compaction. *Drug Development and Industrial Pharmacy, 34*(10), 1022-1035.

Sotiropoulos, V., Contou-Carrere, M.-N., Daoutidis, P., & Kaznessis, Y. N. (2008). Model reduction of multiscale chemical langevin equations: a numerical case study. *IEEE/ACM transactions on computational biology and bioinformatics / IEEE, ACM, 6*(3), 470-482.

Stephanopoulos, G., & Reklaitis, G. V. (2011). Process systems engineering: From Solvay to modern bio- and nanotechnology. A history of development, successes and prospects for the future. *Chemical Engineering Science, 66*(19), 4272-4306.

Straub, D. A., & Grossmann, I. E. (1993). Design optimization of stochastic flexibility. *Computers & Chemical Engineering, 17*(4), 339-354.

Swaney, R. E., & Grossmann, I. E. (1985). An index for operational flexibility in chemical process design. Part I: Formulation and theory (Vol. 31, pp. 621-630).

Tabib, M. V., & Joshi, J. B. (2008). Analysis of dominant flow structures and their flow dynamics in chemical process equipment using snapshot proper orthogonal decomposition technique. [doi: 10.1016/j.ces.2008.04.046]. *Chemical Engineering Science, 63*(14), 3695-3715.

Turkoglu, M., Aydin, I., Murray, M., & Sakr, A. (1999). Modeling of a roller-compaction process using neural networks and genetic algorithms. [doi: DOI: 10.1016/S0939-6411(99)00054-5]. *European Journal of Pharmaceutics and Biopharmaceutics, 48*(3), 239-245.

Varshney, A., & Armaou, A. (2008). Reduced order modeling and dynamic optimization of multiscale PDE/kMC process systems. *Computers & Chemical Engineering, 32*(9), 2136-2143.

Varshney, A., Pitchaiah, S., & Armaou, A. (2009). Feedback control of dissipative PDE systems using adaptive model reduction. *AIChE Journal, 55*(4), 906-918.

Verkoeijen, D., Pouw, G. A., Meesters, G. M. H., & Scarlett, B. (2002). Population balances for particulate processes - a volume approach. *Chemical Engineering Science, 57*(12), 2287-2303.

Villemonteix, J., Vazquez, E., Sidorkiewicz, M., & Walter, E. (2009). Global optimization of expensive-to-evaluate functions: an empirical comparison of two sampling criteria. *Journal of Global Optimization, 43*(2), 373-389.

Villemonteix, J., Vazquez, E., & Walter, E. (2009). An informational approach to the global optimization of expensive-to-evaluate functions. *Journal of Global Optimization, 44*(4), 509-534.

Vishal, G., & Marianthi, G. I. (2002). Determination of operability limits using simplicial approximation (Vol. 48, pp. 2902-2909).

Vishal, G., & Marianthi, G. I. (2003). Framework for evaluating the feasibility/operability of nonconvex processes (Vol. 49, pp. 1233-1240).

Wan, X., Pekny, J. F., & Reklaitis, G. V. (2005). Simulation-based optimization with surrogate models--Application to supply chain management. *Computers & Chemical Engineering, 29*(6), 1317-1328.

Wang, J., Wen, H., & Desai, D. (2010). Lubrication in tablet formulations. [doi: DOI: 10.1016/j.ejpb.2010.01.007]. *European Journal of Pharmaceutics and Biopharmaceutics, 75*(1), 1-15.

Wassgren, C., Freireich, B., Li, J. F., & Litster, J. (2011). Incorporating particle flow information from discrete element simulations in population balance models of mixer-coaters. *Chemical Engineering Science, 66*(16), 3592-3604.

Wassgren, C., & Sarkar, A. (2010). Continuous blending of cohesive granular material. *Chemical Engineering Science, 65*(21), 5687-5698.

Weir, G. J. (2004). A mathematical model for dilating, non-cohesive granular flows in steep-walled hoppers. [doi: DOI: 10.1016/j.ces.2003.09.031]. *Chemical Engineering Science, 59*(1), 149-161.

Werther, J., Reimers, C., & Gruhn, G. (2008). Flowsheet simulation of solids processes - Data reconciliation and adjustment of model parameters. *Chemical Engineering and Processing, 47*(1), 138-158.

Werther, J., Reimers, C., & Gruhn, G. (2009). Design specifications in the flowsheet simulation of complex solids processes. *Powder Technology, 191*(3), 260-271.

Werther, J., Toebermann, J. C., Rosenkranz, J., & Gruhn, G. (2000). Block-oriented process simulation of solids processes. *Computers & Chemical Engineering, 23*(11-12), 1773-1782.

Winkel, M. L., Zullo, L. C., Verheijen, P. J. T., & Pantelides, C. C. (1995). Modeling and Simulation of the Operation of an Industrial Batch Plant Using Gproms. *Computers & Chemical Engineering, 19*, S571-S576.

Wold, S., Martens, H., & Wold, H. (1983). The multivariate calibration problem in chemistry solved by the PLS method

Matrix Pencils. In B. Kågström & A. Ruhe (Eds.), (Vol. 973, pp. 286-293): Springer Berlin / Heidelberg.

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. [doi: 10.1016/S0169-7439(01)00155-1]. *Chemometrics and Intelligent Laboratory Systems, 58*(2), 109-130.

Wold, S., Sjöström, M., & Eriksson, L. (2002). Partial Least Squares Projections to Latent Structures (PLS) in Chemistry *Encyclopedia of Computational Chemistry*: John Wiley & Sons, Ltd.

Wu, H., Heilweil, E. J., Hussain, A. S., & Khan, M. A. (2007). Process analytical technology (PAT): Effects of instrumental and compositional variables on terahertz spectral data quality to characterize pharmaceutical materials and tablets. *International Journal of Pharmaceutics, 343*(1-2), 148-158.

Yin, J., Ng, S. H., & Ng, K. M. (2011). Kriging metamodel with modified nugget-effect: The heteroscedastic variance case. [doi: 10.1016/j.cie.2011.05.008]. *Computers & Industrial Engineering, 61*(3), 760-777.

Yu, A. B., Standish, L., & Lu, L. (1995). Coal agglomeration and its effects on bulk density. *Powder Technology*.

Yu, A. B., Zhu, H. P., Zhou, Z. Y., & Yang, R. Y. (2008). Discrete particle simulation of particulate systems: A review of major applications and findings. *Chemical Engineering Science, 63*(23), 5728-5770.

Yu, L. (2008). Pharmaceutical Quality by Design: Product and Process Development, Understanding, and Control. *Pharmaceutical Research, 25*(4), 781-791.

Yuan, J., Wang, K., Yu, T., & Fang, M. (2008). Reliable multi-objective optimization of high-speed WEDM process based on Gaussian process regression. [doi: DOI: 10.1016/j.ijmachtools.2007.07.011]. *International Journal of Machine Tools and Manufacture, 48*(1), 47-60.

Zhu, H. P., & Yu, A. B. (2004). Steady-state granular flow in a three-dimensional cylindrical hopper with flat bottom: microscopic analysis. *Journal of Physics D-Applied Physics, 37*(10), 1497-1508.

Zhu, H. P., Yu, A. B., & Wu, Y. H. (2006). Numerical investigation of steady and unsteady state hopper flows. *Powder Technology, 170*(3), 125-134.

Zhu, H. P., Zhou, Z. Y., Yang, R. Y., & Yu, A. B. (2007). Discrete particle simulation of particulate systems: Theoretical developments. [doi: 10.1016/j.ces.2006.12.089]. *Chemical Engineering Science, 62*(13), 3378-3396.

Zinchuk, A. V., Mullarney, M. P., & Hancock, B. C. (2004). Simulation of roller compaction using a laboratory scale compaction simulator. [doi: DOI: 10.1016/j.ijpharm.2003.09.034]. *International Journal of Pharmaceutics, 269*(2), 403-415.