PROCESS SYSTEMS ENGINEERING METHODS FOR THE DEVELOPMENT OF CONTINUOUS PHARMACEUTICAL MANUFACTURING PROCESSES

by

AMANDA J. ROGERS

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

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May, 2015

ABSTRACT OF THE DISSERTATION

Process Systems Engineering Methods for the Development of Continuous Pharmaceutical Manufacturing Processes

By AMANDA J. ROGERS

Dissertation Director:

Marianthi G. Ierapetritou, PhD

The objective of this dissertation is to develop mathematical methods to assist in the development of continuous processes for the manufacture of pharmaceutical drug products. Specifically, modeling, analysis and optimization techniques will be developed in order to apply concepts from the field of process systems engineering to pharmaceutical development. These concepts, including sensitivity analysis, feasibility and flexibility analysis and optimal design under uncertainty, can contribute to the development of robust and well-understood manufacturing processes. This is consistent with the aim Quality-by-Design in the pharmaceutical process development. Throughout this dissertation, case studies related to continuous solids-based manufacturing will be used to demonstrate the role that process systems engineering can play in process development. The mathematical methods introduced in this dissertation will emphasize the use of reduced-order and surrogate models to address the unique challenges associated with modeling, analysis and optimization of solids-based processes.

Acknowledgement and Dedication

I have been extremely fortunate to have a wonderful group of mentors, colleagues, friends and family supporting me for the last three years. Without all of your guidance and support none of my work would have been possible.

I would first like to express my sincerest gratitude to my advisor, Dr. Marianthi Ierapetritou, whose guidance has helped me grow academically, professionally and personally. Thank you for always encouraging me to do my best, for knowing when I wasn't there yet, and for believing that I could get there even when I wasn't so sure. Working with you has been a privilege and a pleasure.

I would also like to thank the members of my committee, Dr. Fernando Muzzio, Dr. Rohit Ramachandran and Dr. Jose Tabora, for their constructive and thoughtful feedback. I would especially like to thank Dr. Fernando Muzzio, whose honest advice has greatly improved my ability to communicate my research effectively.

Had it not for my amazing mentors and colleagues at Bristol-Myers Squibb, I might never have pursued a PhD. I would especially like to thank Dr. Jean Tom for encouraging me to go back to graduate school and supporting me in choosing to do so. I would also like to thank Dr. Jose Tabora, for his technical and professional guidance both before and since I began my doctoral studies. I have really appreciated the encouragement I've received from many of my BMS colleagues over the last 3 years, in particular Dr. Brenda Remy and Sabrina Ivy.

I was fortunate be advised by Dr. Larry Biegler at Carnegie Mellon University for the first 8 months of my doctoral studies. I learned a great deal in that short time and am grateful to Dr. Biegler for his guidance and support. I would also like to acknowledge Dr. Mauricio Futran, who together with Jean and Marianthi helped me to transfer to Rutgers University.

My friends and colleagues in the Laboratory for Optimization and Systems Analysis have made it a pleasure coming to work every day. Thank you Fani, Sebastian, Parham, Zhaojia, Chaitali,

Sohyun, Charles, Ravendra, Nihar, Jinjun, Nikisha, Jun and Xu. I have also had a great time working and traveling with my colleagues in the ERC-SOPS, including Dana, Sara, Sarang, Pallavi, Krizia, Abhishek, Bill, Juan Yifan, Anwesha and Maitraye.

I also need to thank my friends and colleagues at NJ-STEP and the prison teaching initiative, especially Dr. Jill Knapp, and all of our students at EJSP, whose hard work and dedication is truly inspiring.

Last, but most certainly not least, I need to thank my wonderful family. Without their love and support I would never have made it out of high school, let alone through graduate school. To my parents Michael and Margie Rogers, thank you for inspiring me by your example, for always believing that I could do anything, for always reminding me that it would take hard work, and for encouraging me in all of my endeavors. To my siblings Emily and Harry, you may be my younger siblings but you've always given me so much to look up to. Thanks for always listening when I needed to talk, talking when you knew what I needed to hear and having my back no matter what. To my husband Matt, words are completely inadequate to express how grateful I am to you. Thank you for being my best friend and confidante, my partner in all things, for making my goals your goals, for encouraging me to take on tough challenges and never doubting that I am up to the task, for helping me keep sight of what really matters and for making me smile every day no matter what happens. I am so lucky to have you and Queenie in my life.

Table of Contents

List of Tables

List of Illustrations

Chapter 1

1. Introduction

1.1. The current state of pharmaceutical manufacturing

In recent years, the pharmaceutical industry has experienced significant changes in the prevailing economic and regulatory environments. Increased global competition, particularly from manufacturers of generic products, has decreased the competition-free lifespan of products and reduced profit margins as drugs come off-patent. (Shah, 2004; Suresh & Basu, 2008) Regulatory agencies worldwide, including the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA), have adopted the Quality-by-Design (QbD) paradigm. QbD, introduced by the ICH Q8 guidance on pharmaceutical development (ICH, 2009), advocates science-based process development and emphasizes assurance of quality through the implementation of robust manufacturing processes. (L. X. Yu, 2008) Despite economic and regulatory pressure to enhance robustness and efficiency, innovation in pharmaceutical manufacturing has been limited relative to that in other chemical process industries. (P. McKenzie et al., 2006) Sequential scale-up of batch processes remains the norm in pharmaceutical development. (Reinhardt, 2001) This approach does not yield cost effective manufacturing processes, as evidenced by the fact that manufacturing costs consume as much as 27% of revenue for pharmaceutical companies by some estimates. (Basu et al., 2008) In addition, product quality issues can occur due to incomplete process understanding or lack of robustness for batch processes. (Buchholz, 2010; Plumb, 2005)

Several of the aforementioned issues can be addressed in part through the implementation of continuous manufacturing processes. Continuous processing has the potential to reduce capital cost, as well as utility requirements, through the use of smaller equipment. (Seifert et al., 2012) Continuous processes are also amenable to scale-up via increased throughput or parallelization. (Plumb, 2005) As a result fewer scale-up studies are required, thereby expediting process

development and decreasing time to market. (Schaber et al., 2011; Shah, 2004) Continuous processing also lends itself to on-line process control, thereby enhancing process robustness. (Buchholz, 2010; Plumb, 2005)

1.1.1. Continuous tablet manufacturing processes

The case studies presented in this work will focus on continuous manufacturing processes for the production of pharmaceutical tablets, which are among the most prevalent dosage forms in the US market. (Muzzio et al., 2002) The exact sequence of unit operations may change depending on the compound being formulated (P. McKenzie et al., 2006), but manufacturing processes for continuous tablet production can be classified into three broad categories: direct compaction, processes involving wet granulation and processes involving dry granulation. (M. A. Järvinen et al., 2012) A general schematic summarizing each of these manufacturing routes is depicted in Figure 1. Tableting processes begin with the feeding of raw materials to the process. These generally include an active pharmaceutical ingredient (API), excipients and lubricants. In many cases the concentration of excipient in the formulation is greater than that of the API or lubricant. The material may then be passed through a comill for delumping prior to being blended in a continuous mixer. Optionally the blend may be granulated via wet or dry granulation (roller compaction). Wet granulation necessitates a granule drying step prior to further processing. For processes involving granulation a milling step may be needed to reduce the granule size before tableting. (Vervaet & Remon, 2005) In the absence of a granulation step material may proceed directly to tableting after blending.

The case studies in this work will focus primarily on continuous direct compaction, as this is the most straightforward route for tablet production.(M. A. Järvinen et al., 2012) However unit operations related to dry granulation, such as roller compaction, are also evaluated. The modeling and implementation of these unit operations is discussed in greater detail in Chapter 2.

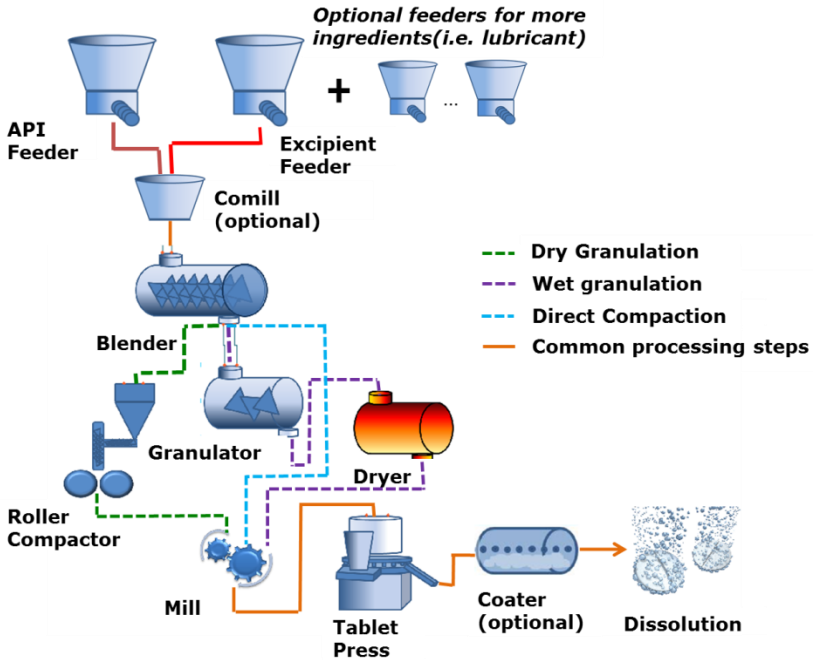

Figure 1: Schematic for continuous tablet production via direct compaction, wet granulation and dry granulation are summarized

1.2. Regulatory framework for pharmaceutical process development: Quality by Design

Regulatory guidance for the development and validation of pharmaceutical processes within the Quality by Design (QbD) framework has been provided by International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals in Human Use (ICH). (ICH, 2009) QbD advocates a science-based approach to process development, wherein the quality of a pharmaceutical product is ensured through fundamental understanding of the relationship between process design and operating conditions and product quality attributes. (ICH, 2005a; P. McKenzie et al., 2006; L. X. Yu, 2008) ICH has also developed guidance describing best practices for quality risk assessment (Q9) (ICH, 2005a) and the implementation of Pharmaceutical Quality Systems (Q10) (ICH, 2005b). These guidelines include a discussion of process modeling and other mathematical tools that can be used to develop and understand pharmaceutical processes with the goal of achieving quality by design.

ICH guidelines emphasize the importance of determining process design space, defined as the "multidimensional parametric space within which acceptable quality product is obtained". (Chatterjee, 2008; ICH, 2009) In order to implement a QbD approach it is necessary to have quantitative indicators of product quality. These are referred to as critical quality attributes (CQAs), which are defined as physical or chemical characteristics that are controlled directly or indirectly to ensure product quality. (L. X. Yu, 2008) CQAs should be correlated in some way with acceptable clinical performance of a drug. In the context of tablet manufacturing, potential critical quality attributes include tablet hardness, tablet API composition, tablet weight and tablet dissolution performance. The design space can be described with respect to specific process parameters that affect these critical quality attributes, referred to as critical process parameters (CPPs). CPPs are formally defined as process inputs which must be controlled within a specific range in order to ensure product quality. (ICH, 2009) Generally these include raw material properties, process design parameters, and operating conditions that significantly affect product quality when varied within the ranges anticipated during normal process operation. (L. X. Yu, 2008) Identifying critical process parameters is an important aspect of quality risk assessment. (ICH, 2005a)

The quality by design approach to pharmaceutical process development has important regulatory and financial implications. Under the previously implemented quality by testing approach, the effect of processing conditions on product quality was not always well understood. As a result, conservative quality specifications could be set, potentially causing clinically acceptable materials to be discarded. (L. X. Yu, 2008) It is also difficult to implement strategies for robust process design and control in a quality by testing environment. (Degerman et al., 2009) The understanding of design space developed through a QbD approach can create opportunities for more robust process implementation. Conditions can be adjusted within the design space to compensate for disturbances and to achieve optimal process performance from both a quality and cost perspective. (Leopore & Spavins, 2008) It is beneficial to develop processes that have a

relatively large design space, as this facilitates a greater deal of operational flexibility. It can also avoid the need to seek further regulatory approval related to changes in a filed manufacturing process. (FDA, 1995, 1997a, 1997b; Peterson, 2008) A comprehensive understanding of critical quality attributes, critical process parameters and design space is necessary to achieve quality by design and can save both time and money in the development of pharmaceutical products. (L. X. Yu, 2008)

1.3. Process systems engineering tools for pharmaceutical process development

For both economic and regulatory reasons it is necessary for pharmaceutical manufacturing processes to be robust, efficient and sufficiently understood to ensure product quality. Process systems engineering (PSE) tools can play an important role in developing such processes. (K. V. Gernaey et al., 2012) In the context of this work, process systems engineering will refer to mathematical techniques for process modeling, analysis and optimization. PSE concepts such as equation-oriented process simulation (Tolsma et al., 2002), flexibility analysis (Morari, 1983; Swaney & Grossmann, 1985a), and constrained global optimization (Floudas et al., 2001) have long been used in other industries to design resilient, cost effective manufacturing processes. (Bahri et al., 1997; I. E. Grossmann, 1996; Saboo & Morari, 1984) The pharmaceutical industry has lagged behind other industries in the widespread implementation of these tools. (P. McKenzie et al., 2006; Troup & Georgakis, 2013) This can be attributed in part to the difficulties associated with modeling pharmaceutical processes, particularly the solids-based processes associated with the manufacture of drug product. (Muir Wood, 2008; Muzzio et al., 2002) Fortunately recent advances in solids-based process modeling, which are discussed in Chapter 2, have made it possible to apply concepts from process systems engineering to pharmaceutical manufacturing processes.

Process systems engineering tools can be used to enhance and expedite process development in ways that are consistent with the quality by design philosophy. For instance, process modeling

can supplement experimental studies during process design and scale-up. Integrated flowsheet models can be used to evaluate design alternatives and identify robust control architectures. (Schaber et al., 2011; Singh et al., 2012) These models can also be used for sensitivity analysis. (F. Boukouvala et al., 2012; Saltelli, Tarantola, et al., 2000) The results of sensitivity analysis can inform quality risk assessment, as sensitivity metrics can identify potential critical process parameters. Sensitivity analysis can also be used to develop control strategies for product quality attributes by indicating the process operating conditions that most significantly influence quality. (F. Boukouvala et al., 2010a; Rogers et al., 2014) The concept of feasibility analysis can be applied to pharmaceutical process models to aid in the determination of design space. (F. Boukouvala & M. G. Ierapetritou, 2012; F. Boukouvala et al., 2010a) Feasibility and flexibility analysis can also be incorporated to algorithms for process design under uncertainty. These algorithms can be used to suggest process design alternatives that are robust and cost effective. (Pistikopoulos & Ierapetritou, 1995) Table 1 summarizes concepts from process systems engineering and the potential role they can play in pharmaceutical process development.

Table 1: Summary of concepts from process systems engineering and their potential applications in the context of pharmaceutical process development.

PSE concept	Process development applications
Integrated process modeling (flowsheet modeling)	Expedite scale-up Compare design alternatives Develop and evaluate control strategies
Reduced-order modeling	Efficient process simulation Improved accuracy in flowsheet models Model-predictive control
Sensitivity analysis	Identify CPPs Develop and evaluate control strategies
Feasibility analysis	Evaluate design space
Flexibility analysis/ optimal design under uncertainty	Quantify process robustness Compare design alternatives Select robust and cost effective process design

1.4. Motivation: development and implementation of mathematical tools for pharmaceutical process development

Despite the many potential applications for process systems engineering concepts in the development of pharmaceutical manufacturing processes, modeling and optimization tools are underutilized in the industry. (K. V. Gernaey et al., 2012; P. McKenzie et al., 2006) This is due in part to the difficulties associated with modeling pharmaceutical processes (Muzzio et al., 2002) and to the unique challenges presented by optimizing complex and dynamic process models. (Biegler, 2007; Rios & Sahinidis, 2012) The objective of this dissertation is to develop and demonstrate mathematical tools that will allow process systems engineering concepts to be applied to pharmaceutical manufacturing processes. An emphasis will be placed on methods that can be used to model, evaluate and optimize drug product manufacturing processes. These include, but are not limited to, reduced-order modeling, sensitivity analysis, and process feasibility and flexibility analysis techniques. Each of these methods will be discussed in the context quality by design and demonstrated using case studies related to continuous solids-based manufacturing processes.

1.5. Outline of dissertation

Each of the five main chapters in this dissertation will emphasize a specific concept or tool from process systems engineering and its role in the development of robust pharmaceutical manufacturing processes. Chapter 2 will introduce some of the basic modeling techniques and mathematical models used in this work. These include unit operation models for solids-based processes. The concept of integrated process modeling, or flowsheet modeling, will also be introduced in this chapter. Chapter 3 will describe methods for reduced-order modeling, which have applications in process simulation and optimization. The use of reduced-order models to bridge the gap between high-fidelity process simulations and more computationally efficient flowsheet models will be demonstrated in this chapter. Chapter 4 will focus on the use of

sensitivity analysis, which has potential applications in the identification of critical process parameters and development of control strategies for pharmaceutical processes. Several sensitivity analysis techniques will be introduced and these will be applied to a continuous tablet manufacturing process. In Chapter 5, the concept of feasibility analysis will be discussed in the context of design space development. A novel method for surrogate-based feasibility analysis of both steady state and dynamic processes will be introduced. This method will be illustrated with a number of case studies, including a roller compaction process. Chapter 6 will focus on flexibility analysis and process design under uncertainty, which can be used to select process design alternatives that are both robust and cost effective. A novel method for surrogate-based flexibility analysis will be introduced. This method will be applied to a number of case studies, including a continuous feeding and blending process. Finally Chapter 7 will summarize major conclusions and also discuss directions for future work in the area of process systems engineering for pharmaceutical applications.

Chapter 2

2. Integrated modeling for solids-based processes

The development of process models is critical to the use of process systems engineering tools such as sensitivity analysis and optimal design under uncertainty. (F. Boukouvala & Ierapetritou, 2013) Integrated process models, which describe the sequence of unit operations in a manufacturing process, have many potential applications in the context of pharmaceutical process development. Simulation results can supplement experimental studies and reduce the experimental burden associated with process development and scale-up. (K. V. Gernaey, Cervera-Padrell, C., Woodley, J.M., 2012) Models can be used to explore process design space (F. Boukouvala & M. G. Ierapetritou, 2012), conduct sensitivity and uncertainty analysis (F. Boukouvala et al., 2012; Rogers et al., 2014), and evaluate control strategies for continuous processes (Ramachandran et al., 2011; Singh et al., 2013). Several of these applications will be discussed in greater detail in Chapters 4 through 6. In this chapter, several methods used to model solids-based processes in the pharmaceutical industry are introduced and mathematical models for pharmaceutical unit operations are described. In addition, the use of integrated process modeling, or flowsheet modeling, for pharmaceutical applications is discussed

2.1. Modeling solids-based processes

Solids-based processes can be challenging to model because the bulk behavior of particulate flows is influenced by particle level phenomena. This is known as the continuum duality of particulate materials, and it makes it difficult to develop constitutive models describing the response of granular flows to applied stresses and strains (Muir Wood, 2008). A variety of modeling techniques have been developed that address the influence of particle-level phenomena on bulk behavior in varying degrees of detail. These are summarized in Figure 2, in which methods are arranged according to the level of detail with which they represent the particulate nature of granular flows. These methods are briefly introduced in this section, with the exception

of reduced-order models which are described in greater detail in Chapter 3. Specific applications of these methods for modeling pharmaceutical unit operations are described in section 2.2.

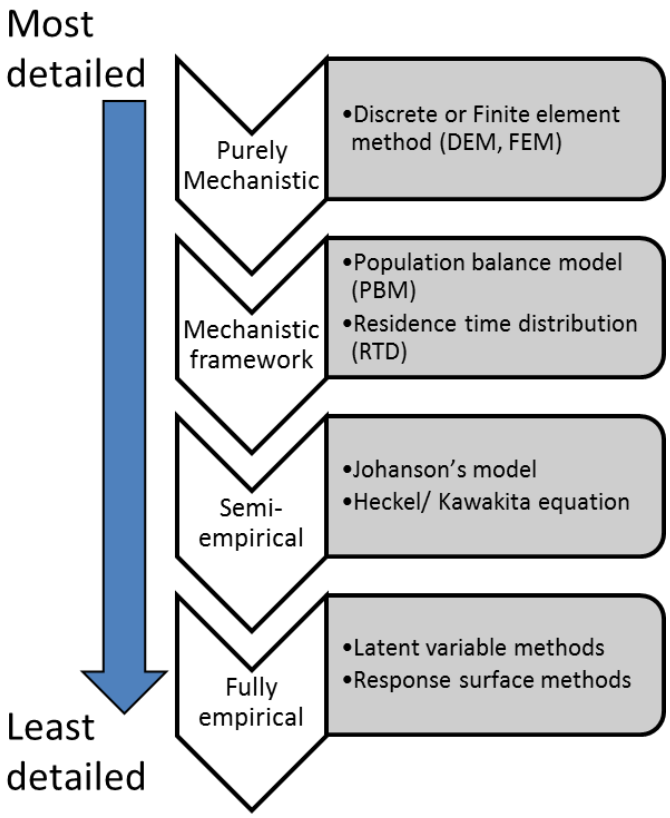

Figure 2: Modeling approaches for solids-based processes

Discrete Element Method (DEM)

One way of addressing the issue of continuum duality is to explicitly consider particle-level phenomena, as in discrete element method (DEM) models. In DEM simulations, Newton's equations of motion are evaluated for all particles in a system, considering all particle-particle and particle-geometry interactions. (Zhu et al., 2008) Particle level phenomena such as dispersion, segregation, particle packing and bed densification can be accurately represented using DEM. (Zhu et al., 2008) DEM has therefore been used to model a variety of pharmaceutically relevant operations including powder conveying, granular mixing, powder flow into dies and compaction of powders into tablets. (Ketterhagen et al., 2009) However, DEM simulations have a number of drawbacks, the most significant being their computational cost.

Particle-level calculations tend to be computationally expensive, as the number of equations to be solved at each time step scales with the number of particles in the simulation. For this reason it may not be practical to implement DEM simulations in applications requiring multiple model evaluations or rapid simulation. However, information from DEM models can be incorporated into flowsheet simulations through the use of certain types of reduced-order models, which will be discussed in Chapter 3. DEM simulations also require information about particle-level properties including friction coefficients and coefficients of restitution for interparticle and particle-geometry interactions. These cannot be directly measured for many pharmaceutically relevant solids and therefore need to be estimated using indirect methods. (Ketterhagen et al., 2009)

Population Balance Models

Population balance models (PBM) consider the evolution of a population of particles over time with respect to internal and external coordinates. Internal coordinates include particle-level properties like size (length, volume or mass) and composition (liquid, solid and vapor content). External coordinates describe location with respect to the process geometry, like axial or radial position within a continuous blender. While not as detailed as DEM, PBM models can represent distributed properties such as particle/granule size, moisture content or active ingredient composition. PBM can represent size-dependent phenomena, and are often used to model processes involving size change like granulation and milling. PBM may also be used to model distributed composition processes like continuous mixing and drying. (K. V. Gernaey et al., 2012) While not as computationally expensive as DEM, solving PBM can be numerically intensive for higher-order models like those representing wet granulation processes. Hierarchal solution strategies and order-reduction techniques can be used to expedite the solution of PBM in these cases. (Barasso & Ramachandran, 2012)

Residence Time Distribution

The concept of residence time distribution (RTD), introduced in the 1950s (Danckwerts, 1953), describes the notion that in continuous flow systems it is not entirely accurate to assume plug flow behavior. In fact some portions of the flow (particles or fluid elements) will remain in the system longer than others due to differences in their velocities. For granular flow processes, this implies that a collection of particles entering a system at the same time will exit the system at a variety of different times, distributed around some mean residence time. This collection of different residence times can be thought of as the residence time distribution. The residence time distribution in a continuous flow system can be determined experimentally via tracer pulse experiments (Danckwerts, 1953; P. M. Portillo et al., 2008; Vanarase & Muzzio, 2011).

RTD approaches can be applied to model continuous powder blending processes.(Gao, Vanarase, et al., 2011; Marikh et al., 2006; Pernenkil & Cooney, 2006; P. M. Portillo et al., 2008) Many studies to date have focused on the use of experimentally obtained residence time distributions to characterize mixing processes.(P. M. Portillo et al., 2008; Vanarase & Muzzio, 2011) However RTD models can also be used to predictively model powder blending processes. For instance, Marikh et al. (Marikh et al., 2006) have reported the use of a Markov chain model to simulate the residence time distribution in a continuous powder mixer. Process parameters like powder flow rate and blender rotation rate are related to the parameters of the RTD model via a correlation with the steady state powder holdup (Marikh et al., 2006). Gao et al. (Gao, Vanarase, et al., 2011) have implemented an RTD model based on experimentally determined parameters including the Peclet number and the mean residence time for the blender. The relative standard deviation (RSD) is then determined based on the fluctuations in the RTD via an empirically estimated dimensionless coefficient (Gao, Vanarase, et al., 2011). An RTD model developed using velocity information obtained from a DEM simulation is implemented in the flowsheet model described in Chapter 3.

Semi-empirical Models

Although constitutive models for particulate processes are not widely available, semi-empirical models have been developed for a number of pharmaceutical processes. These models combine equations developed based on fundamental process understanding with empirical parameters that can calibrate the model for a specific combination of material and equipment. These parameters are estimated using experimental data collected from the process of interest. (Kremer & Hancock, 2006) Examples of such models include Johanson's theory of rolling granular solids (Johanson, 1965), which has been used to describe roller compaction processes. In addition, tableting processes can be described by the Heckel (Heckel, 1961a, 1961b) or Kawakita (Kawakita et al., 1977; Kawakita & Lüdde, 1971) compression equations. These are semi-empirical relations describing the pressure-density relationship in powder compaction processes. Semi-empirical models for a number of unit operations are described in detail in section 2.2.

2.2. Unit operation models

The case studies presented in this work will focus primarily on direct compaction processes for continuous tablet manufacture, as direct compaction is the most straightforward route for the production of tablets. (M. A. Järvinen et al., 2012) Case studies based on roller compaction, which is used in dry granulation processes, are also presented. In this section the various unit operations involved in direct compaction and roller compaction processes are briefly described and the mathematical models used to represent them are introduced.

Loss-in-weight Feeder

Accurate powder feeding is critical for pharmaceutical processes, particularly for active pharmaceutical ingredients (API), in order to ensure content uniformity. Loss-in-weight (LIW) feeders can provide the required accuracy for these processes, even when low flow rates are required as is sometimes the case for APIs and lubricants. (W. E. Engisch & Muzzio, 2012; Yang & Evans, 2007) LIW feeders typically consist of a feed hopper into which powders are charged and a conveying mechanism such as a screw. Some feeders also contain an agitator at the base of

the feed hopper in order to convey powder into the screw region. LIW feeders contain a load cell that is integrated with a gravimetric controller which can be operated in closed-loop control to provide a user-specified mass flow rate of powder. Optionally, a discharge screen can be included at the exit of the feeder to break up any static lumps. Tooling for a feeder (screws, discharge screens) should be selected based on the powder properties of the material to be fed and the desired flow rate. Engisch and Muzzio (W. E. Engisch & Muzzio, 2012) have described a method that can be used to characterize LIW feeders and help determine appropriate tooling for a particular application.

The key operating parameter of interest for loss in weight feeders is the screw speed, which dictates the powder flow rate. Loss-in-weight feeders are typically operated in closed loop control, where a controller manipulates the screw speed to achieve the desired flow rate set point. This means that the screw speed may vary depending on the hopper fill level, typically increasing as the hopper empties in order to achieve a constant mass flow rate. (Berthiaux et al., 2008) During hopper refilling, feeders must switch to volumetric operation. As a result disturbances in feed rate are often observed when LIW feeders are refilled.(W.E. Engisch & Muzzio, 2010)

Loss-in-weight feeders can be model using a transfer function relating powder feed rate to the feeder screw speed (ω^f), as shown in equation (1). The dynamic response of the feeder to a change in set point can be accounted for through the time constant τ^f. The delay differential equation (2) is used to represent the mean residence time in the feeder, which is indicated by the time delay factor θ_d^f. The model parameters including the process gain k_g^f, time constant τ^f and a time delay factor θ_d^f can be estimated based on data collected from step change experiments, wherein the feed rate set point is changed and the corresponding change in the total flow rate is monitored over time. (F. Boukouvala et al., 2012)

In practice LIW feeders exhibit periodic high-frequency noise due to the nature of the screw conveying mechanism. A noise term v is included in the model to account for this effect, as shown in equation (3). This term is sampled from a uniform distribution once per second. The

bounds on this distribution may be set to as much as five percent of the target mass flow rate, which is consistent with the noise levels observed in commercial LIW feeding applications. (W. E. Engisch & Muzzio, 2012) Both closed and open-loop implementations of the loss in weight feeder are considered in the current work. In prior simulation studies (F. Boukouvala, Chaudhury, et al., 2013; F. Boukouvala et al., 2012) open-loop implementations have been used, in which the user manually sets the screw speed (ω^f) to achieve the desired flow rate. Closed-loop implementations in which the screw speed is set by a ratio controller have also been demonstrated. (Singh et al., 2012, 2013) In the current work a closed-loop implementation is used. The user specifies a desired flow rate set point (y_d) and the corresponding screw speed is determined by the model based on the feeder gain.

$$\tau^f \frac{dm_{out}(t)}{dt} + m_{out}(t) = y_d$$
$$y_d = k_g^f \omega^f \quad (1)$$

$$\theta_d^f \frac{\partial m_{out}(t,z)}{\partial t} = -\frac{\partial m_{out}^{delayed}(t,z)}{\partial t}$$
$$I.C. m_{out}^{delayed}(t, z=0) = m_{out}(t) \quad (2)$$

$$m_{out}^{noisy}(t) = m_{out}(t)\left(1-v\right) \quad (3)$$

For the purposes of this model, material bulk density and particle size are assumed to be constant throughout the feeder. This assumption has been experimentally verified for common pharmaceutical powders like microcrystalline cellulose, magnesium stearate and acetaminophen. (F. Boukouvala et al., 2012) However during hopper refill it is possible for changes in bulk density to occur for the powder in the feed hopper, contributing to feed rate disturbances. The effect of powder densification during refill on a continuous feeding and blending process is considered in Chapter 6.

Continuous Blender

The goal of a blending process is to reduce spatial and, for continuous systems, temporal variations in mixture composition. A well-designed continuous mixer should produce uniform blends with active ingredient compositions consistent with the desired drug loading for the process. Continuous convective mixers, which have been studied extensively for pharmaceutical mixing processes, are considered in this work. (P.M. Portillo et al., 2009; Sarkar & Wassgren, 2009; Vanarase & Muzzio, 2011) In these mixers, the primary mixing mechanism is convective mixing induced by rotating blades. Therefore design parameters related to the agitator size, configuration and geometry can significantly influence mixing performance, as can the vessel geometry (length and diameter). Operating parameters including the rotation rate of the agitator and the powder flow rate, also play an important role in continuous blending operations. (Vanarase & Muzzio, 2011)

One common metric used to describe homogeneity of a blend is the relative standard deviation (RSD), which describes the variability in the mixer outlet composition over space and/or time. (Pernenkil & Cooney, 2006) RSD is calculated as shown in equation (4), where *N* is the number of samples taken to measure the RSD, *Ci* is the API composition of an individual sample and $\bar{C}$ is the average API composition over all *N* samples. For a continuous blending process RSD can be determined experimentally from concentration data at the outlet of the mixer. Mixture composition can be measured using process analytical technology (PAT) tools like near infrared (NIR) spectroscopy. (Vanarase et al., 2010) Throughout this dissertation, the RSD and the mean concentration of API at the outlet of the mixer are used to characterize blending performance in process simulations.

$$RSD = \frac{\sigma}{\bar{C}}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}\left(C_i - \bar{C}\right)^2}{N-1}} \qquad (4)$$

Continuous blenders can be modeled using a variety of approaches including DEM simulations (Dubey et al., 2011), population balance models (F. Boukouvala, Dubey, A., Vanarase, A., Ramachandran, R., Muzzio, F.J., Ierapetritou, M., 2012), hybrid models combining DEM and PBM (M. Sen, Dubey, et al., 2013) and residence time distribution (RTD) models(Gao, Vanarase, et al., 2011). In this work both PBM and RTD models are used to represent continuous blending processes.

A continuous blending process can be modeled using a multidimensional population balance model like the one shown in equation (5). In this model *n* is an index that indicates an individual component in the mixture, z_1 and z_2 are external (axial and transverse) spatial coordinates that indicate position in the blender and *r* is an internal coordinate that describes the particle size. In the case of a mixing process, the formation and depletion terms on the right hand side of the population balance refer to material entering or leaving different axial and radial segments of the blender due to powder movement. This PBM model can be informed by powder velocity information obtained directly from a DEM simulation (M. Sen, Dubey, et al., 2013) or from a reduced-order model (F. Boukouvala, Gao, et al., 2013; Rogers & Ierapetritou, 2014). Reduced-order models based on DEM data will be described in greater detail in Chapter 3.

$$\frac{\partial}{\partial t}F\left(n,z_1,z_2,r,t\right)+\frac{\partial}{\partial z_1}\left[F\left(n,z_1,z_2,r,t\right)\frac{dz_1}{dt}\right]+\frac{\partial}{\partial z_2}\left[F\left(n,z_1,z_2,r,t\right)\frac{dz_2}{dt}\right]+\frac{\partial}{\partial r}\left[F\left(n,z_1,z_2,r,t\right)\frac{dr}{dt}\right]=R_{formation}\left(n,z_1,z_2,r,t\right)-R_{depletion}\left(n,z_1,z_2,r,t\right) \quad (5)$$

$$x_{out}(t)=\sum_{n\in N}y_{out}^{n}(t)x^{n} \quad (6)$$

The properties of materials exiting the mixer are estimated based on a linear mixing rule relating individual component concentrations and raw material properties to bulk stream properties, as shown in equation (6). In this equation *x* indicates a material property such as mean particle size or bulk density and x^n denotes the pure component property for component *n.* All streams

following the mixer are treated as single entities with bulk properties which propagate to subsequent unit operations.

The RTD model for continuous blenders used in this dissertation is based on a reduced-order method and is therefore described in Chapter 3.

Hopper

Hoppers are ubiquitous in solids processing as a means of primary or intermediate storage and of conveying particulate materials gravimetrically. In tablet manufacturing process hoppers are often associated with other equipment like feeders or a tablet press. It is desirable for hoppers to operate in the mass-flow regime, providing steady and consistent flow. Hopper geometry, including its shape (wedge or cone), outlet diameter and wall angle can be selected to ensure operation in this regime. Inserts may also be added to the flow path in a hopper in order to address problems like arching and bridging. (Gremaud et al., 2006; Weir, 2004) In continuous processes it is important to monitor the holdup in hoppers to avoid overfilling or, in the case of feed hoppers, to avoid running empty during continuous production.

Flow and holdup in a hopper can be modeled using a straightforward mass balance, as shown in equations (7)-(8). It is assumed that the hopper is operating in the mass-flow regime and that stream properties, including API concentration, particle size distribution and material bulk density, are constant throughout the hopper. Thus the streams exiting the hopper have the same properties as those entering it, albeit with a time delay indicated by the mean residence time in the hopper θ_{rt}. This is shown in equation (9), where *x* can be any stream property.

$$\frac{dm}{dt} = m_{in} - m_{out} \tag{7}$$

$$\dot{m}(t) = H\left(t\right) A \rho_{bulk} \tag{8}$$

$$\theta_{rt} \frac{\partial x_{out}(t,z)}{\partial t} = -\frac{\partial x_{out}(t,z)}{\partial z} \qquad I.C.\ x_{out}(t, z=0) = x_{out}(t) \tag{9}$$

For the relatively small hoppers typically associated with tablet presses in continuous pharmaceutical processes this assumption is reasonable. However for highly compressible materials or for large feed hoppers this assumption may need to be reevaluated. In Chapter 6 a case study is described in which densification of powder in feed hoppers and its effect on subsequent processing is considered.

Tablet Press

A tablet press contains several components integrated as a single processing unit. A hopper collects material to be conveyed into the tablet press. This material is fed into a series of dies using a feed frame. The shape of the die and depth of the die cavity determine the shape and size of the resulting tablet. Material within the dies is compressed by a punch, to form a hard compact or tablet. The dies move continuously, guided by cam tracks, so that they can be filled, compacted, and discharged in sequence. (M. A. Järvinen et al., 2012)

Tablet presses may differ in terms of their feed frames, number of compression stages, and selected tooling. Often the tooling, which includes the die and punch size and geometry, can be changed according to the desired tablet weight and shape. The die can be filled via force feeding or suction filling. (Jackson et al., 2007) The quality attributes of interest for a tablet include active ingredient content, content uniformity, weight variability, and physical properties such as friability, hardness and dissolution performance. For some products drug content can be measured nondestructively using near infrared (NIR) spectroscopy. (Corredor et al., 2011) Hardness, friability and dissolution must be measured offline using destructive techniques. However models may be used to predict hardness and dissolution performance based on operating conditions or spectroscopic measurements. (Gentis & Betz, 2012; M. Kuentz, Lunenberger, H., 2000) The quality attributes of tablets are influenced by a number of operating parameters, including the powder feed rate and the compression force applied to the tablets as well as the rate of tablet production. Tableting speed and powder feed rate can affect tablet weight and weight variability.

Compression force influences tablet hardness and porosity, which can in turn affect dissolution. (Wu et al., 2005)

The tablet press feed frame can be modeled as a time delay, similar to the hopper. The mean residence time in the feed frame is calculated using a response surface model of the form shown in equation (10), where the model parameters are estimated using data reported in the literature (Mendez et al., 2010). The feed frame residence time influences the propagation of stream properties in the same way that the mean residence time in the hopper does. Thus the feed frame can be modeled as shown in equation (9), but in this case the residence time is indicated by RT^{ff} rather than θ_{rt}.

$$RT^{ff} = b_1 + b_2 u^{disk} + b_3 u^{ff} + b_4 u^{disc} u^{ff} \tag{10}$$

The full tablet press model implemented in the current work consists of a system of 23 equations and has been previously described in the literature (Singh et al., 2010). Therefore not all of the model equations will be presented here, only a subset related to tablet quality attributes. The compression behavior of the powder in the die is described by the Kawakita compression equation, shown in equation (11). (Kawakita et al., 1977; Kawakita & Lüdde, 1971) In this equation *C_P* and *C_F* represent the compression pressure and the compression force respectively, while V_0 and V_f represent the powder volumes in the die before and after compression. ε_0 is the porosity of the material prior to compression and *b* is the Kawakita parameter, which can be determined empirically for a specific powder blend.

$$C\_P = \frac{b\left(V_0 - V_f\right)}{b\left(V_0\left(\varepsilon_0 - 1\right) + V\right)} \tag{11}$$
$$C\_F = 10^6 C\_P$$

Tablet hardness is modeled using the approach described in Kuentz and Leuenberger (M. L. Kuentz, H., 2000). The tablet hardness is a function of the relative density (ρ_r) of the final compact as well as the empirically determined critical relative density (ρ_{rc}) for the given powder

blend. The model also includes an empirically determined maximum hardness term (H_{max}), corresponding to the tablet hardness at the critical relative density. The empirical model for tablet hardness is shown in equation (12).

$$H = H_{\max}\left(1 - e^{\rho_r - \rho_{rc} + \lambda_H}\right)$$
$$\lambda_H = \ln\left(\frac{1-\rho_r}{1-\rho_{rc}}\right) \quad (12)$$

Roller Compactor

Roller compactors are used in dry granulation processes for continuous tablet manufacture. In dry granulation, granules are formed through compression rather than through the addition of a liquid binder as in wet granulation. Powders with good flowability and compressibility tend to be better candidates for roller compaction. (Lecompte et al., 2005; S. Yu et al., 2012) Dry granulation can be less forgiving than wet granulation, so it may be more challenging to develop a dry granulation process. However, dry granulation does not require a granule drying step and therefore has shorter processing time and lower capital investment and utility costs than wet granulation. (Kleinebudde, 2004) Roller compaction can also be used for moisture-sensitive APIs, which are not amenable to wet granulation. (Faldu et al., 2012; Vervaet & Remon, 2005)

Roller compaction creates ribbons of compacted powder which can subsequently be milled to provide granules. (Faldu et al., 2012) These ribbons are created when powder is compressed between a set of counter rotating rolls. The ribbon is then conveyed forward and released from the rolls. Powder fed to the rolls is initially in the *slip* region, wherein particles slip at the surface of the rolls and relatively little pressure is exerted on the powder. As the powder approaches the point where the rolls are closest together, the wall velocity of the powder matches that of the rolls and the pressure exerted on the powder increases substantially. The powder is compacted into a ribbon in at this point, in what is known as the *nip* region. The transition between the slip and nip (no-slip) regions is characterized by the nip angle, α. (Reynolds et al., 2010) The ribbon enters the release region once the roll gap begins to increase again. (Bindhumadhavan et al., 2005)

The behavior of powders in a roller compaction process can be described by the rolling theory of granular solids, proposed by Johanson (Johanson, 1965). This model makes several simplifying assumptions, including isotropic, cohesive material, no-slip between the powder and the roll surface in the nip region (the material is frictional) and that the ribbon thickness is exactly equal to the roll gap. (Hsu et al., 2010b) These assumptions are reasonable for gravity fed roller compacters with smooth rollers of relatively large diameter (>500 mm) and for powders with a high enough friction coefficient that the no-slip condition holds. In these cases this model agrees well with experimental data (Dec et al., 2003) and has been shown to accurately predict pressure profiles in the nip region of a roller compacter (Bindhumadhavan et al., 2005). Required parameters for Johanson's model include effective angle of internal friction and angle of friction, which can be found experimentally. In addition the relationship between pressure and density for the powder of interest must be determined empirically. This can be accomplished using a punch-die system similar to that found in a tablet press.

In this dissertation the dynamic Johanson's model introduced by Hsu et al.(Hsu et al., 2010b) is used to model roller compaction processes. This model consists of a system of differential and algebraic equations shown in equations (13) -(15). Equation (13) describes the change in the roll gap, and thereby the ribbon thickness, due to changes in the powder feed rate. Equation (14) can be used to calculate the pressure required to densify the powder between the rolls in the nip region. The density of the ribbon can be determined using the empirical correlation in equation (15). This relates the stress profile in the nip region to the ribbon density via two empirical parameters C_1 and K, which are functions of the material properties. The controller dynamics are described by equation (16), which can be used to determine the response of the controllers to a change in set point over time based on the time constants τ. (Hsu et al., 2010b) The dynamic roller compaction model is discussed in greater detail in Chapter 5, in which it is used to evaluate the feasible operating region for a roller compaction process.

$$\frac{d}{dt}\left(\frac{h_0}{R}\right)=\frac{\omega\left[\rho_{in}\cos\theta_{in}(1+{}^{h_0}\!/_{R}-\cos\theta_{in})({}^{u_{in}}\!/_{\omega R})-\rho_{exit}({}^{h_0}\!/_{R})\right]}{\int_0^{\theta_{in}}\rho(\theta)\cos(\theta)d\theta} \tag{13}$$

$$P_h=\frac{W}{A}\frac{\sigma_{exit}R}{1+\sin\delta}\int_0^{\alpha}\left[\frac{{}^{h_0}\!/_{R}}{(1+{}^{h_0}\!/_{R}-\cos\theta)\cos\theta}\right]^{K}\cos\theta d\theta \tag{14}$$

$$\sigma_{exit}=C_1\rho_{exit}{}^{K} \tag{15}$$

$$\begin{aligned}\tau_p\frac{dP_h}{dt}+P_h&=P_d\\ \tau_\omega\frac{d\omega}{dt}+\omega&=\omega_d\\ \tau_u\frac{du_{in}}{dt}+u_{in}&=u_d\end{aligned} \tag{16}$$

2.3. Integrated flowsheet modeling

Individual unit operation models like those described in the previous section can be connected sequentially in process simulation environments like gPROMS™ ModelBuilder or Aspen Plus®. Process simulation software facilitates dynamic simulation of blocks of equations, in which each block corresponds to an individual unit operation. Blocks may be connected in such a way that variables are passed between them, mimicking the transfer of material from one unit to another in a manufacturing process. Modern flowsheeting environments are equipped with robust dynamic solvers that can readily resolve complex process models containing tens of thousands of differential, algebraic and integral equations. (Tolsma et al., 2002) However it is still important to ensure that these problems are well posed. In the case of continuous process models this means that, among other things, valid initial conditions must be provided for all differential equations and that the information transferred from one block to another must contain valid operating conditions for both systems of equations.

While flowsheet modeling has long been established for petrochemical and other fluid-based processes, its use for solids-based applications is a relatively recent development. Several

publications have described the use of flowsheet modeling tools to represent continuous tablet manufacturing processes (F. Boukouvala, Chaudhury, et al., 2013; F. Boukouvala et al., 2012; M. Sen, Chaudhury, et al., 2013) as well as integrated upstream and downstream pharmaceutical processes, which incorporate both wet chemistry and solids-handling applications (Benyahia et al., 2012; Lakerveld et al., 2013; Schaber et al., 2011; M. Sen, Rogers, et al., 2013). Commercially available flowsheet modeling environments now include model libraries specifically focused on solids-based processes. For instance, gPROMS™ by Process Systems Enterprise Ltd. has introduced gSOLIDS™, which contains unit operations models for milling, granulation, powder blending and tablet compaction (Gavi & Reynolds, 2014). Aspen Plus® by Aspen Technology Inc. also contains solids-handling unit operation models such as granulation and fluidized bed drying (Dyment et al., 2013). Both gPROMS™ and Aspen Plus® also offer custom modeling environments wherein users can specify their own process models using differential, algebraic and integral equations (PSE, 2012; Tremblay & Peers, 2014).

In this work the gPROMS™ ModelBuilder's custom modeling environment is used to implement flowsheet models for continuous blending and tablet manufacturing processes. Within this environment it is possible to combine the models described in section 2.2 with reduced-order models, like those which will be described in Chapter 3. Specific flowsheet simulations will be presented within the chapters where they are implemented. In Chapter 3 a model for a continuous feeding and blending process is discussed. This model incorporates unit operation models described in section 2.2 as well as reduced-order models described in Chapter 3. In Chapter 4 a continuous direct compaction flowsheet model is presented and used to conduct a process sensitivity analysis.

Chapter 3

3. Reduced-order modeling for solids-based processes

The term reduced-order model (ROM) can refer to any model that is lower in dimensionality or cost than the original process it is intended to approximate. (Shvartsman et al., 2000) There are a number of reasons to employ ROM in pharmaceutical modeling applications. To begin with, reduced-order models may be the only models available in some instances. The level of fundamental process understanding for many solids-based processes is limited, particularly for continuous pharmaceutical manufacturing applications as this technology is still being developed. (Plumb, 2005) Therefore it may necessary to use data-based or reduced-order models to describe processes for which first principles understanding is incomplete. Reduced-order models can also be used to bridge the gap between high fidelity simulations (e.g. DEM) and flowsheet models, which generally evaluate much faster than DEM simulations.(F. Boukouvala, Gao, et al., 2013; Rogers & Ierapetritou, 2014) Because reduced-order models tend to be computationally inexpensive, they can be used for applications requiring rapid or repeated model evaluation, like model predictive control(Ramachandran et al., 2011) or optimization (F. Boukouvala & Ierapetritou, 2013).

In this dissertation, reduced-order models will be used to approximate computationally expensive models like DEM simulations or flowsheet simulations. Reduced-order models based on data from DEM simulations will be used to incorporate distributed parameter information, like particle velocity trajectories, into a flowsheet simulation without incurring the computational cost of DEM. Surrogate representations of flowsheet models will be used for feasibility analysis and optimization applications. In this chapter the reduced-order modeling techniques used throughout this dissertation are introduced. These include latent variable methods, response surface techniques and hybrid approaches that combine the dimensionality reduction of latent variable methods with the black-box modeling capabilities of response surface methods.

3.1.1. Latent variable models

Multivariate regression methods, including those based on latent variables, have been used extensively in pharmaceutical applications to identify and model relationships between material properties and process performance. (Gabrielsson et al., 2012; Garcia-Munoz, 2014; García-Muñoz & Polizzi, 2012; Haware et al., 2009) Latent variable methods like principal component analysis (PCA) and partial least squares regression (PLS) are well-suited to these types of problems, as they can identify underlying correlation structure in multivariate datasets. (Timm, 2002) Distributed parameter information obtained from discrete element (DEM) simulations tends to contain correlated data as well. The correlated nature of this data can be exploited to express high dimensional datasets from DEM in lower dimensional space without significant loss of information. (Liberge & Hamdouni, 2010; Lieu et al., 2006)

Principal component analysis (PCA) involves the projection of a dataset (X) into a reduced space on a set of orthogonal axes. The objective of this transformation is to identify an orthogonal basis that optimally represents the variance in the original process data using a limited number of variables. (Burnham et al., 1999; López-Negrete de la Fuente et al., 2010) The results of PCA can be expressed in terms of the transformed variables, or scores (T), and a corresponding matrix of loadings (P) as shown in equation (17). (Wold et al., 2001) The principal component scores represent the original variables projected into latent space. These transformed variables are obtained by multiplying the original variables with the principal component loadings. The dimensionality of the scores and loadings matrices depends on the number of principal components in the model (α), which must be less than or equal to the number of factors in the original dataset. The original data matrix (X) can be approximated based on the scores and loadings as shown in equation (17), where ε is a matrix of residuals. (Timm, 2002; Westerhuis et al., 1998)

$$X = TP' + \varepsilon \tag{17}$$

A number of algorithms are available to determine the principal component scores and loadings. The selection of an appropriate algorithm depends in part on the nature of the dataset *X*. Singular value decomposition is the most straightforward approach and is used in the current work.(Lang et al., 2009) However alternative algorithms like expectation-maximization PCA (EM-PCA)(Roweis, 1998) and a nonlinear programming approach (NLP-PCA)(López-Negrete de la Fuente et al., 2010) may be preferred in applications where *X* has missing elements.

3.1.2. Response surface models

Latent variable methods often rely on linear transformations and therefore may not be ideal for modeling nonlinear processes. Response surface techniques can be used to model processes with varying degrees of nonlinearity by expressing input-response relationships in the form of a multivariate response surface. The ability of response surface techniques to accommodate nonlinear and dynamic process data has proven useful in modeling a variety of pharmaceutical processes including loss-in-weight feeding, continuous mixing and roller compaction. (F. Boukouvala et al., 2010b, 2011) The form of a response surface can be simple or complex depending on the technique used to create the mapping. Response surface techniques include response surface methodology(Box & Wilson, 1951), high dimensional model representation (HDMR) (G. Li et al., 2001), artificial neural networks (ANN) (Basheer & Hajmeer, 2000) and kriging (Krige, 1951).

The selection of a reduced-order modeling method should be made based on the type and quality of data available and the intended purpose of the model. Response surface methodology (RSM) and high-dimensional model representation (HDMR) have the advantage of providing some insight into the relative importance of factors in a process model. For instance, model coefficients from HDMR may be used to carry out variance-based sensitivity analysis on a process. (Ziehn & Tomlin, 2008, 2009) Model parameters in artificial neural networks and kriging are selected to optimally fit the training data for the model and may not be interpretable in terms of the effect of

individual factors on process performance. However these methods can effectively model highly nonlinear and dynamic processes. (Basheer & Hajmeer, 2000; Kleijnen, 2009) Recent works comparing reduced-order modeling techniques for pharmaceutical manufacturing applications have identified kriging as a method that performs well relative to other reduced-order modeling approaches. (F. Boukouvala et al., 2010a, 2010b, 2011) Kriging-based reduced order models have also been shown to perform well in surrogate-based feasibility analysis (F. Boukouvala & M. G. Ierapetritou, 2012) and optimization applications (F. Boukouvala & Ierapetritou, 2013; Davis & Ierapetritou, 2007). For these reasons, kriging is the primary response surface technique used in the current work and further discussion will focus on this method. HDMR is also used for applications related to sensitivity analysis and will be discussed in greater detail in Chapter 4.

Kriging Methodology

Originally developed to predict mineral distributions in mining applications (Krige, 1951), kriging has recently become popular for a variety of applications due to its ability to model nonlinear and dynamic processes. (Kleijnen, 2009) Kriging is a black-box interpolating technique in which it is assumed that the process response at a previously untested point can be estimated as a weighted sum of function values at nearby points. (Calder & Cressie, 2009; Metheron, 1963) This is shown in equation (18), where x_k denotes a previously untested point, x_i represents a point in the set of N previously sampled points, also known as design sites, and w_i is the weight corresponding to x_i. The term ε_k in equation (18) represents the variance of the kriging predictor at x_k, which can be estimated based on the process variance at neighboring design sites. Estimation of the variance is possible due to the nature of the kriging predictor $\hat{f}$, which is assumed to be the realization of a Gaussian process F whose current distribution depends on previous function evaluations f. The kriging predictor can thus be viewed as the conditional distribution of F at a previously untested point. The mean and variance of this distribution can be calculated based on

prior knowledge, in this case data collected at previously sampled points *f*. (Calder & Cressie, 2009; Kleijnen, 2009)

$$\hat{f}(x_k) = \sum_{i=1}^{N} w_i f(x_i) + \varepsilon_k \quad (18)$$

Equation (18) can be expressed in terms of a regression model and a correlation model, which are used to calculate the mean and the variance of the kriging predictor respectively. This is shown in equation (19), where $f(x_i)^T$ is a vector of known process responses at the design sites and $r(x_i)^T$ is a vector of correlations between the response variances at the design sites.

$$\hat{f}(x_k) = f(x_i)^T \beta + r(x_i)^T \gamma \quad (19)$$

The regression model generally takes the form of a polynomial. In the simplest case a zero-order regression model is used and the first term in equation (18) is simply the mean process response. (Kleijnen, 2009) In the current work regression models that are either zero, first or second order polynomials are considered. The prediction variance ε_k calculated using a correlation model with arguments d_j and θ_j. d_j indicates the distance between the test point x_k and a design site x_i. The parameter θ_j influences the relationship between distance and correlation. The greater the value of θ_j, the faster the correlation between points x_k and x_i will decrease as a function of increasing distance between them. The correlation model can take a variety of forms. (Lophaven et al., 2002) The types of correlation models considered in this work are summarized in Table 2.

Table 2: Correlation models used for kriging model development

Model	Form of $r(\theta_j, d_j)$
Linear	$\max(0, 1-\theta_j\lvert d_j\rvert)$
Exponential	$\exp(-\theta_j\lvert d_j\rvert)$
Gaussian	$\exp(-\theta_j d_j^2)$

The coefficient vectors β and γ as well as the correlation model parameter θ for a kriging model can be fitted to a given set of design data using a maximum likelihood estimation (MLE) approach. (Davis & Ierapetritou, 2007; Lophaven et al., 2002; J. D. Martin, 2009) In this work the MATLAB toolbox DACE (Lophaven et al., 2002), which implements this procedure, is used to fit kriging model parameters.

3.1.3. PCA-based reduced –order modeling

PCA-based reduced-order modeling is a hybrid approach that combines the dimensionality reduction capabilities of PCA with the nonlinear modeling abilities of response surface techniques. The concept of PCA-based reduced-order modeling was initially introduced for applications involving computational fluid dynamics (CFD) simulations.(Lang et al., 2009) The goal of this work was to facilitate co-simulation of CFD models with flowsheet models for a gasification-based power plant. (Lang et al., 2011) Recently this approach has been extended to information obtained from DEM simulations (F. Boukouvala, Gao, et al., 2013; Rogers & Ierapetritou, 2014) with the goal of incorporating distributed parameter information into flowsheet simulations for pharmaceutical manufacturing processes.

PCA-based reduced-order modeling is specifically intended to model high-dimensional state data that is obtained from finite element simulations. Figure 3 depicts the type of datasets for which this method can be applied. This includes three relevant information matrices, the input space $\left(X \in \Re^{Nxn}\right)$, the state space $\left(Z \in \Re^{Nxk}\right)$ and the output space $\left(Y \in \Re^{Nxm}\right)$. The input space consists of design and operating parameters that can be controlled by the user or defined by the developer of the DEM model. The state space consists of distributed parameters (e.g. particle velocities), and its dimensionality is defined by the discretization of the process geometry. Finally, the output space consists of process responses at the outlet of the unit (e.g. blend RSD). The goal of this modeling methodology is to develop predictive models for both the state variables and the process outputs as a function of the variables in the input space. In this case

those variables include time because the dynamic response of the system to changes in operating conditions are of interest.

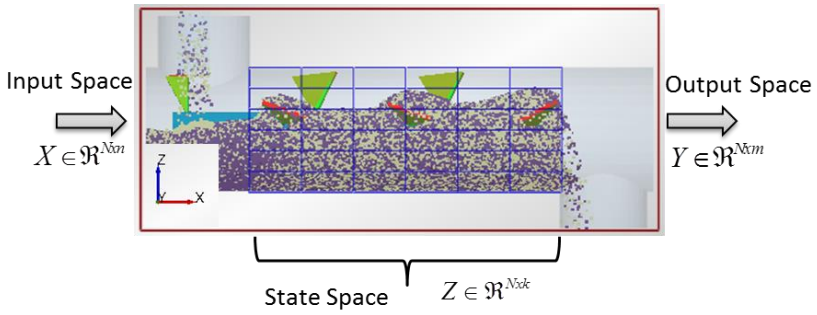

Figure 3: Space definitions for PCA-based reduced-order modeling method

The use of PCA is motivated by the high dimensionality of the state space, which depends on the discretization of the process geometry. In general the rank of the state space is significantly greater than that of the input space. For previously reported pharmaceutical case studies the state space has contained 200-260 elements while the input space contains fewer than 10. (F. Boukouvala, Gao, et al., 2013; Rogers & Ierapetritou, 2014; M. Sen, Dubey, et al., 2013) PCA can explain the high dimensional state data using a limited number of orthonormal basis vectors, or principal components, by taking advantage of the correlated nature of the state data.(Lang et al., 2009) The state data can then be expressed using principal component scores, which substantially reduces the dimensionality of the dataset without significant loss of information. (F. Boukouvala, Gao, et al., 2013; Lang et al., 2009; Rogers & Ierapetritou, 2014) A response surface modeling technique, such as kriging (F. Boukouvala, Gao, et al., 2013; Rogers & Ierapetritou, 2014) or ANN (Lang et al., 2009), can then be used to map the principal component scores to the process design and operating variables in the input space *X*. For a set of previously untested conditions, the state data can be estimated by first predicting the principal component scores from the response surface model and then multiplying the predicted scores by the corresponding principal component loadings, as shown in equation (20).

$$X_{new} \xrightarrow{Kriging} T_{pred}$$
$$Z_{new} = T_{pred} P' \quad (20)$$

It should be noted that the predicted state variables may not be valid if the predicted scores do not fall within the score space of the original model, as in this case the loadings from the original model are not valid. (Lang et al., 2009) Therefore when developing a PCA based reduced-order model it is important to collect samples that span the anticipated input space.

3.2. Dynamic discrete element reduced-order modeling (dynamic DE-ROM)

As discussed in Chapter 2, discrete element method (DEM) simulations can provide detailed information about particle-level phenomena. This type of information has the potential to improve the predictive ability of unit operation models for particulate processes. (M. Sen, Dubey, et al., 2013; Wassgren et al., 2011) However the computational cost associated with DEM simulations can make it impractical to run these concurrently with a flowsheet model. In this section a PCA-based reduced-order approach is introduced that can be used to incorporate information from DEM into flowsheet models for efficient process simulation.

In this section a PCA-based modeling method for dynamic particulate processes is presented and applied to a continuous blending process. The algorithm for this dynamic discrete element reduced-order modeling (dynamic DE-ROM) technique is presented in Figure 4. Design of experiments can be used to sample from the set of design and operating conditions for the process to generate the input space. (Kroonenberg, 2008; Timm, 2002) For dynamic DE-ROM the design parameters are held constant and only the operating parameters are varied over time. The high-fidelity (DEM) model is simulated for each condition specified in the input space. The process geometry is discretized prior to the extraction of the state data. The selection of the discretization is a nontrivial decision. Too fine a mesh can result in noisy data due to a small number of particles in each discrete element. Conversely, the shape of the distributed parameter profiles may be distorted by an overly broad discretization. It is also important to consider the sampling

frequency for the state space. Sampling too frequently can result in noisy data, similar to using too fine of a mesh for the discretization of the process geometry. However infrequent sampling could fail to adequately capture the process dynamics. The discretization and sampling frequency should ultimately be selected based on the intended application for the model. For instance, if the velocities obtained from the reduced-order model are to inform a unit operation model in a flowsheet simulation then the discretization of the process geometry should be consistent with the number of elements considered in the corresponding unit operation model.

In some cases, pre-processing may be required to smooth the state and response data after it is extracted from the simulation. State data is also mean-centered and scaled to unit variance prior to principal component analysis. (Burnham et al., 1999; López-Negrete de la Fuente et al., 2010) PCA is then used to reduce the dimensionality of the state space. For the case studies in this work the number of principal components in the reduced state space is selected such that at least 95 percent of the variance in the original state data is explained. A percent variance explained criterion is used to select the number of components because the purpose of employing PCA is to reduce the dimensionality of the state data without significant loss of information. It should be noted that in general the number of components in a PCA model can be selected through cross validation, such that the desired prediction accuracy is achieved by the PCA model. (Arlot & Celisse, 2010; Timm, 2002) In dynamic DE-ROM, cross validation is carried out on the PCA-based reduced-order model. This incorporates the principal component decomposition and subsequent response surface modeling. Once the PCA model has been developed, the reduced state space can be obtained. The reduced state space is given by the principal component scores (T). These scores are mapped to the process inputs using a kriging response surface to provide the PCA-based reduced order model.

A separate reduced-order model can also be developed to predict the process outputs (Y). In this case a kriging model is used to relate the relative standard deviation (RSD) for the blend to operating parameters like the agitator rotation rate.

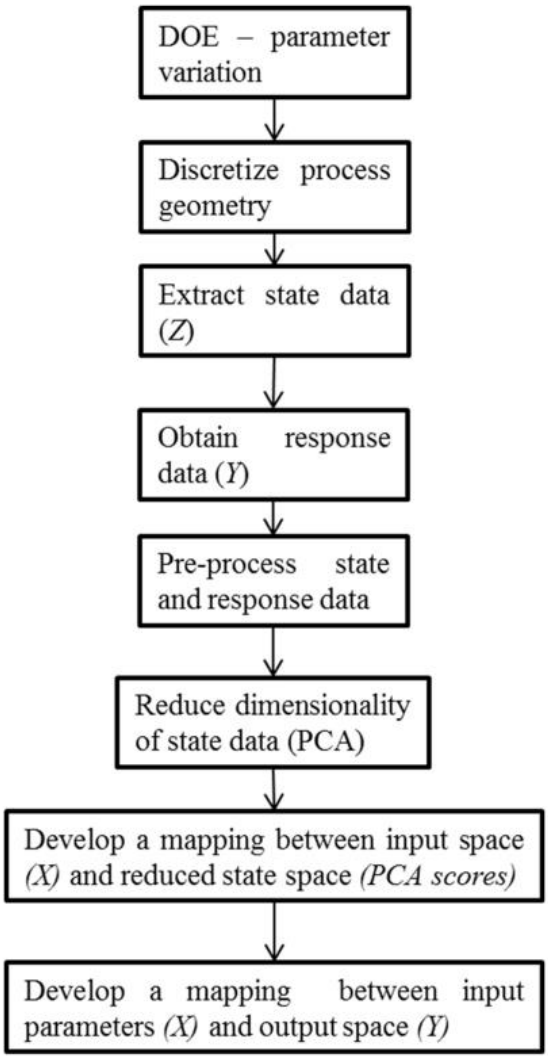

Figure 4: Algorithm for dynamic discrete element reduced-order modeling

3.2.1. Continuous blender case study

The dynamic discrete-element reduced-order modeling approach is demonstrated on a series of case studies for a continuous convective blender. These include two periodic section case studies as well as a full blender case study. A detailed discussion of the DEM models used to simulate the process will not be included here, as these models have been previously described in the literature. Interested readers are referred to Gao et al. (Gao, Ierapetritou, et al., 2012; Gao, Muzzio, et al., 2012) and Boukouvala et al. (F. Boukouvala, Gao, et al., 2013) for a discussion of the periodic section DEM simulations and to Dubey et al. (Dubey et al., 2012) for more

information on the full blender DEM simulations. Details pertaining to all DEM simulations are also provided in Rogers and Ierapetritou (Rogers & Ierapetritou, 2014).

Periodic Section case studies

Periodic section models of a continuous blender were used to test the dynamic DE-ROM methodology. These DEM simulations evaluate quickly relative to models of the entire blender as they contain fewer particles. They are thus preferable for preliminary evaluation of the proposed modeling approach. Periodic section modeling is well established in the literature as a means of characterizing and optimizing continuous blending processes. (Gao, Ierapetritou, et al., 2012; Gao, Muzzio, et al., 2012; P. M. Portillo, Muzzio, F.J., Ierapetritou, M.G., 2008; Sarkar & Wassgren, 2009) The underlying principle of periodic section modeling is that the continuous mixing process consists of two phenomena; powder flow, which is characterized by the residence time distribution (RTD) and powder mixing, which can be viewed as a batch mixing process occurring in a single periodic section of the mixer. (Gao, Ierapetritou, et al., 2012) In this case study, mixing performance was characterized using the relative standard deviation (RSD), which can be calculated as described in equation (4).

Two different periodic section case studies were considered in this work, each with a different shaft, blade and weir configuration. The specific design and operating conditions for each periodic section are shown in Table 2. The two periodic sections are hereafter referred to as Case 1 and Case 2, respectively.

Table 3: Design parameters for periodic section case studies

	Fill Level (%)	**Blade Angle (deg)**	**Weir Height Ratio (w/d, %)**	**Blade Width (mm)**	**Shaft Angle (deg)**
Case 1	40.39	13.29	67.41	29.04	-7.38
Case 2	27.26	36.31	61.71	25.7	-19.34

The periodic sections were simulated at 19 different agitator rotation rates, randomly sampled from a uniform distribution with a lower bound of 40 rpm and upper bound of 250 rpm. These

correspond to the operating limits for the Gericke™ GCM 250, a commercially available continuous blender which has been previously studied using DEM simulation. (Dubey et al., 2012) The simulation was carried out for 4 seconds at each blade speed, as it is was found that the performance metric (RSD) reached steady state within 4 seconds over the range of design and operating conditions considered. The periodic section was divided into 4 axial slices, which were further discretized into an 8 by 8 grid in the radial direction, resulting in a total of 256 elements. This discretization is depicted in Figure 5 and was selected such that each element contained, on average, more than 10 particles. A minimum of 10 particles was selected based on previous work (F. Boukouvala, Gao, et al., 2013) in which it was found that state data from elements containing fewer than 10 particles tends to be noisy. State data corresponding to the average particle velocities in the x, y and z directions (u_x, u_y and u_z respectively) was extracted from the periodic section simulation every 0.1 seconds. This sampling frequency was observed to capture the rapid change in particle velocities and RSD occurring immediately after a change in shaft rotation rate. The dynamic velocity profiles were smoothed using a Savitsky-Golay (Savitzky & Golay, 1964) filter prior to state space reduction to ensure that the reduced-order model fits the velocity trends rather than noise in the process responses.

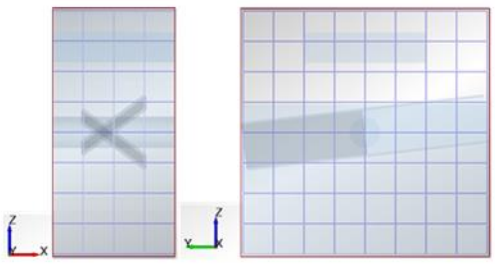

Figure 5: Discretization of periodic section for extraction of state data

The output space consisted of a single performance metric, the relative standard deviation (RSD). RSD was calculated as in equation (4), using only elements in the periodic section which contained more than 30 particles. This corresponded to 50 elements in each case study. Principal

component analysis of the state space data was completed in MATLAB using singular value decomposition. 18 principal components were required to capture 95 percent of the variance in u_x, while only 3 components were needed for u_y and u_z. This implies that the observed variability in axial velocity is significantly greater than that observed in the radial velocity for this process. The input space in this work consists of the agitator rotation rate (rpm) and time, in seconds, since a change in rotation rate was initiated. In this context, *t=0* corresponds to the moment when the agitator speed set point is changed. The principal component scores were mapped to the input space using kriging, with a unique kriging response surface developed for each score vector. A kriging response surface was also fitted to predict RSD as a function of shaft rotation rate and time for each periodic section.

The accuracy of the velocity predictions was examined using k-fold cross validation with *k=10*. (Arlot & Celisse, 2010) The percent mean square error of prediction, hereafter abbreviated as PMSE, for the state data in Case 1 and Case 2 is summarized in Table 4. The PMSE is less than 1.3 percent for the velocity predictions in both case studies. This compares favorably with prior work in which steady-state reduced-order models were developed to predict particle velocities in a periodic section of a continuous mixer as a function of design and operating parameters. These steady-state models for distributed parameters reported mean square prediction errors from 7 to 24 percent, as determined by leave-one-out cross validation. (F. Boukouvala, Gao, et al., 2013) The lower PMSE achieved in this work is likely due to the larger datasets considered in developing the dynamic reduced-order models.

Table 4: Percent mean square error of prediction for velocities in continuous blender case studies

Percent mean square error of prediction	**U_x**	**U_y**	**U_z**
Case 1	0.13%	1.21%	0.48%
Case 2	0.22%	1.12%	0.65%

A visual representation comparing the velocities obtained from the reduced-order model with those extracted from DEM simulations is shown in Figure 6 for Case 1 and in Figure 7 Case 2.

Each figure corresponds to a snapshot in time, and shows velocity profiles for each of the 4 axial slices in the discretization shown in Figure 5. Figure 6 shows velocity profiles for the z component (u_z) for Case 1 while Figure 7 depicts velocity profiles for the x component (u_x) in Case 2. It can be seen from Figure 6 and Figure 7 that the contours of the velocity profiles predicted using the reduced-order model are comparable to those obtained from the DEM simulation.

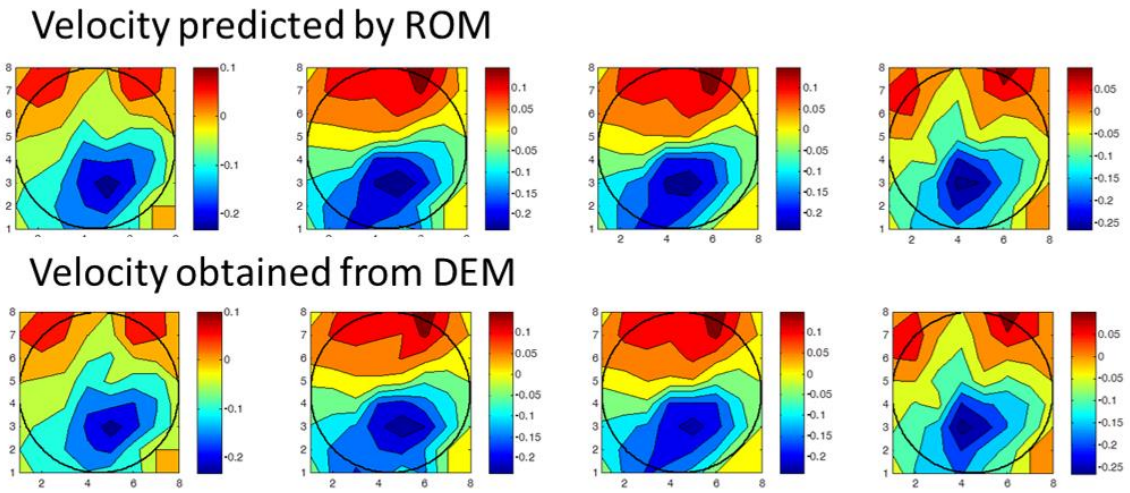

Figure 6: Predicted u_z from reduced-order model vs. u_z from DEM simulation for Case Study 1. Snapshot at t=3 seconds after a change in agitator rotation rate from 144 to 44 rpm

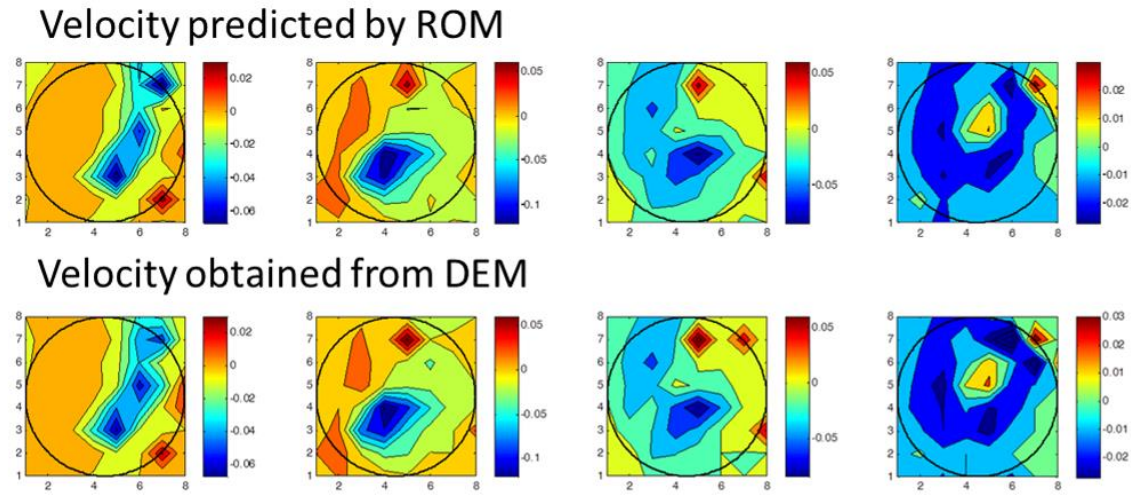

Figure 7: Predicted u_x from reduced-order model vs. u_x from DEM simulation for Case Study 2. Snapshot at t=0.7 seconds after a change in agitator rotation from 170 to 115 rpm

The performance of the mixer, as indicated by the RSD, was modeled dynamically as a function of the blade rotation rate. Kriging was used to develop a response surface relating RSD to blade rotation rate over time. These response surface models were also validated using k-fold cross validation with *k=10*, and the percent mean square error of prediction for the RSD was found to be less than 1.6 percent for both cases studies. This compares favorably to the accuracy previously reported for PCA based reduced-order models in DEM and CFD applications. In Lang et. Al (Lang et al., 2009), predicted process performance was generally within 1 percent of the value obtained from the detailed process model, as determined via leave-one-out cross validation. However for some process responses prediction errors as high as 5 percent were reported. In prior work on DEM-based reduced-order modeling for continuous blenders, a prediction error for blend RSD of 1 percent is reported at steady state. (F. Boukouvala, Gao, et al., 2013) However the prediction error is reported only for a single set of design and operating conditions, so a mean square error of prediction is not available for direct comparison. Prediction errors for the kriging response surface models are summarized in Table 5.

Table 5: Percent mean square error of prediction for RSD in a periodic section of a continuous convective mixer

	Percent mean square error of prediction
Case 1	0.51%
Case 2	0.55%

Figure 8 shows the predicted RSD values for Case 1 as a function of time on the same axes as the RSD values obtained from DEM. These are generally shown to be in good agreement, though the model tends to under predict RSD outliers. These outliers occur immediately after a change in agitator rotation rate, when the RSD is greatest. Kriging is an interpolating method, and as such prediction variance tends to be greatest when the gradient of the response surface is steepest. The gradient of the RSD with respect to time is steepest immediately after a change in operating

conditions, and as a result the prediction error is greatest in this region. Specifically the model under predicts in this case because a majority of data fall below these RSD outliers.

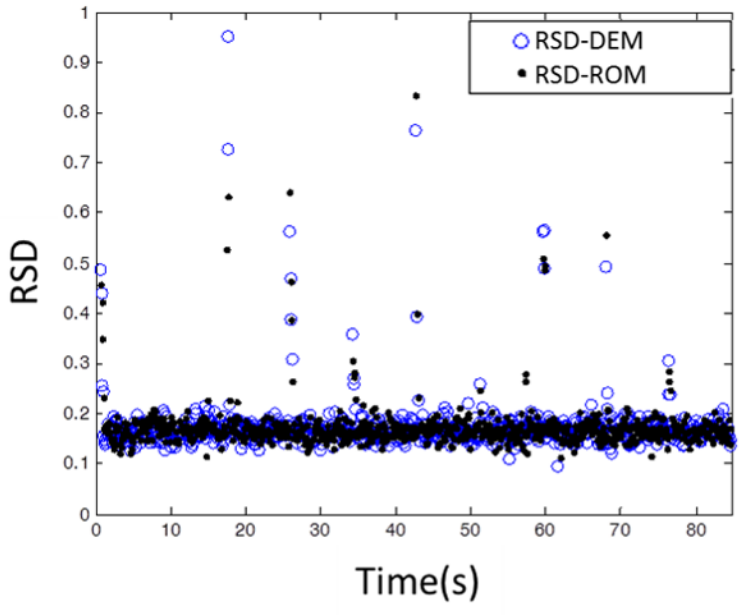

Figure 8: Predicted RSD from reduced-order model (RSD-ROM) and RSD from DEM (RSD-ROM) as a function of time for Case Study 1

Full-scale Blender case study

Periodic section modeling is useful for characterizing mixing performance with respect to design and operating parameters. However in process simulation applications velocity profiles within the entire blender are of interest. Therefore the proposed methodology is also applied to a full-scale blender case study. The DEM simulation used for the blender is a model of a Gericke™ GCM 250 continuous mixer with an alternating blade configuration. The shaft rotation rates are sampled from a uniform distribution between 40 and 250 rpm, as these are the operating limits for this particular equipment. The process geometry is divided into 6 axial sections and each of these is further discretized into a 6 by 6 grid in the y-z plane. The discretization of the process geometry is selected to be consistent with existing unit operation models for the continuous blender, so that the predicted velocities can be used directly in a flowsheet simulation if desired. (F. Boukouvala et al., 2012; M. Sen, Dubey, et al., 2013) The discretized geometry is depicted in

Figure 9. The blender DEM simulation was run for 90 seconds at each agitator rotation rate, as this was greater than the residence time for the blender even at the lowest rotation rate. Relative standard deviation was calculated at the outlet of the blender, the section corresponding to *x=6* in Figure 9, over a total of 36 elements. In this case 23 principal components were required to capture 95 percent of the variance in u_x, while 8 components were needed for u_y and u_z. This indicates greater variability in the axial velocity than in the radial velocity components.

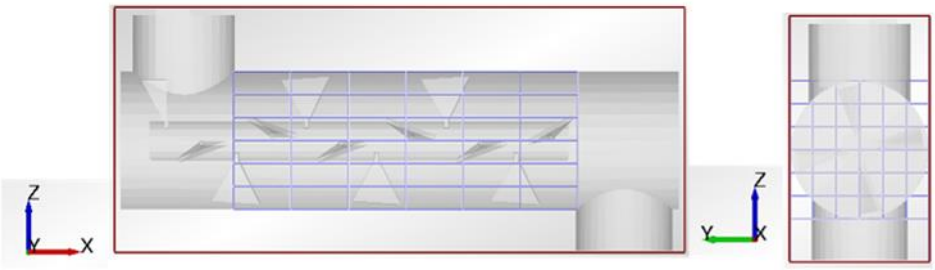

Figure 9: Discretization of process geometry for complete blender simulation

Model development and cross validation were carried out as in the periodic section case studies, using k-fold cross validation with *k=10*. Percent mean square error of prediction for the particle velocities and RSD are summarized in Table 6**.** The PMSE is less than 1.2 percent for prediction of all velocity components. This compares favorably to previously reported prediction accuracies for reduced-order modeling of state data using a PCA-based approach, which report PMSE ranging from 7 to 24 percent.(F. Boukouvala, Gao, et al., 2013) The MSE is 1.1 percent for the RSD predictions, which compares well to the prediction error of 1 percent reported for the RSD in Boukouvala et al. (F. Boukouvala, Gao, et al., 2013)

Table 6: Percent mean square error of prediction for velocity components and relative standard deviation for the complete blender simulation

	U_x	**U_y**	**U_z**	**RSD**
Percent mean square error of prediction	0.55%	0.96%	1.13%	1.07%

Good agreement between predicted RSD values and those obtained from DEM can also be seen in Figure 10, which shows the RSD predicted from the reduced-order model as a function of time on the same axes as the RSD values obtained from the DEM simulation. Figure 11 compares the velocity profiles for u_x predicted by the reduced-order model with those obtained from the DEM simulation. Each segment in Figure 11 corresponds to an axial slice of the mixer, as described in the discretization shown by Figure 9. As for the periodic section case studies, the velocity contours shown in Figure 11 correspond to a snapshot in time. Good agreement is observed between the velocity contours predicted by the reduced-order model and those obtained from the DEM simulation.

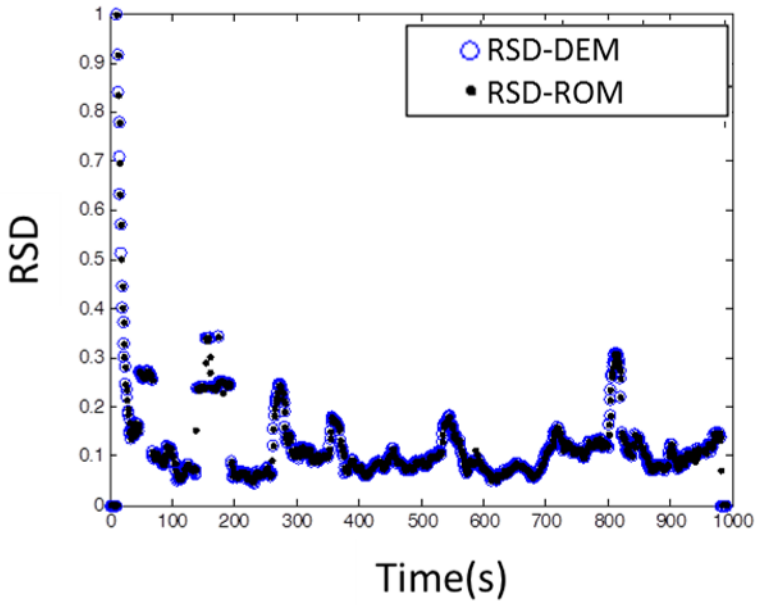

Figure 10: Predicted RSD from reduced-order model (RSD-ROM) and RSD from DEM (RSD-ROM) as a function of time for full blender case study

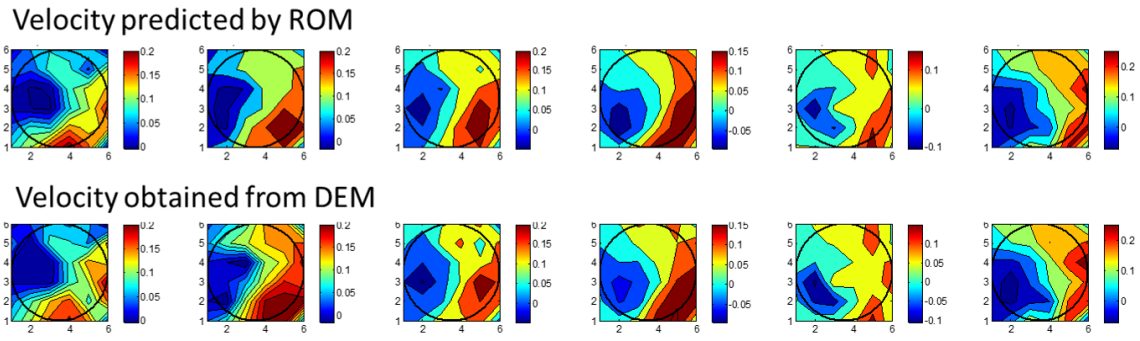

Figure 11: Predicted u_x from reduced-order model vs. u_x from DEM simulation for full-scale blender case study. Snapshot at t=23 seconds after a change in agitator rotation rate from 160 to 250 rpm

The prediction accuracy for the models developed in this case study are comparable to those reported in previous implementations of reduced-order models developed based on CFD or DEM simulations. (F. Boukouvala, Gao, et al., 2013; Lang et al., 2009) The developed ROM can thus be used in dynamic simulation applications to determine particle velocity profiles and relative standard deviation for a continuous blender. The reduced-order model can provide significant computational savings relative to the DEM simulation. The ROM evaluates in a matter of seconds while the DEM model for the Gericke^{TM} GCM250 takes 12 or more hours to run until steady state on the same computer (Dell OptiPlex 980 with Intel Corei7Vpro^{TM} and 16GB ram).

Despite the accuracy of the developed reduced-order models, it is important to note that the proposed approach does have several limitations. It is not recommended extrapolate using models developed with this method because kriging is an interpolating technique and as such is limited in its ability to extrapolate. In addition, the model for the state data relies on a principal component decomposition, which is only valid if the scores corresponding to a new data point fall within the score space of the dataset used to develop the model. In implementing this approach it is thus recommended to sample within the full range of operating conditions for the unit operation of interest so that the model is used for interpolation rather than extrapolation. In addition, the developed reduced-order models are fitted using information obtained by sampling a more detailed process simulation based on DEM. As such, the developed reduced-order models will only be as accurate as these detailed process simulations.

3.3. Integrating reduced-order models into flowsheet simulations

The discrete-element reduced-order modeling approach described in the previous section has been demonstrated to accurately represent state data, such as particle velocity information, from DEM. It is therefore possible to use DE-ROM models to incorporate detailed particle velocity information into a flowsheet simulation for a pharmaceutical manufacturing process. In this

section, a case study is presented in which discrete-element reduced-order models are incorporated into a flowsheet model of a continuous feeding and blending process. In addition, simulation results from the flowsheet model are compared with experimental data obtained for a comparable system. This is an important development, as verification of flowsheet simulation results against experimentally obtained data has not been reported extensively in the literature. Prior work in flowsheet modeling for continuous pharmaceutical processes has often focused on validation of models at the individual unit operation level. (F. Boukouvala et al., 2012; M. Sen et al., 2012) For flowsheet modeling applications, it is important to consider experimental validation for the integrated system as well. (K. V. Gernaey et al., 2012) The level of plant-model agreement desired for a particular flowsheet model will depend on its intended use. (Balci, 2010) While full quantitative agreement may not be required for process development applications, qualitative agreement is certainly useful. (Chatterjee, 2008; Kremer & Hancock, 2006) It is also important to compare simulated results with experimental findings in order to understand shortcomings in the existing model and identify areas where model improvement is required.

3.3.1. Case study: Continuous feeding and blending process

Continuous feeding and blending systems are common to numerous tablet manufacturing routes including direct compression, wet granulation and dry granulation. As such it is of particular interest to model integrated feeding and blending systems accurately. (Gao, Muzzio, et al., 2011; Rogers et al., 2013; Sarkar & Wassgren, 2009)

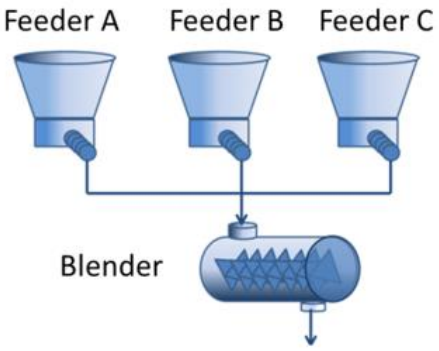

Figure 12: Flowsheet model of continuous feeding and blending system

The process modeled considered in this case study is shown in Figure 12. It consists of three loss-in-weight feeders and a single continuous convective mixer. In this case the component of interest, with respect to which relative standard deviation (RSD) will be calculated, is component A. The loss-in-weight feeders are modeled after Schenck AccuRate LIW feeders and the continuous blender is modeled based on a Gericke^{TM} Model GCM 250 with an alternating blade configuration. These units were selected so that the results from this study could be compared to prior experimental work conducted on a comparable system (Gao, Vanarase, et al., 2011; Vanarase & Muzzio, 2011).

Unit Operation Models

The loss-in-weight feeders are modeled as described in Chapter 2, with v set to 1% of the target mass flow rate. The continuous blending operation is modeled using a discretized residence time distribution approach, where residence time distributions are calculated from particle axial velocity information. The particle velocity information is obtained from the PCA-based reduced-order modeling approach described in Section 3.2. (Rogers & Ierapetritou, 2014) The discretized residence time distribution approach is based on the idea that a continuous residence time distribution can be expressed as a histogram, as shown in Figure 13**.** Each of the bins in the histogram corresponds to a radial segment in the discretized blender model, and the mean residence time for each bin is calculated based on the mean axial velocity in that segment.

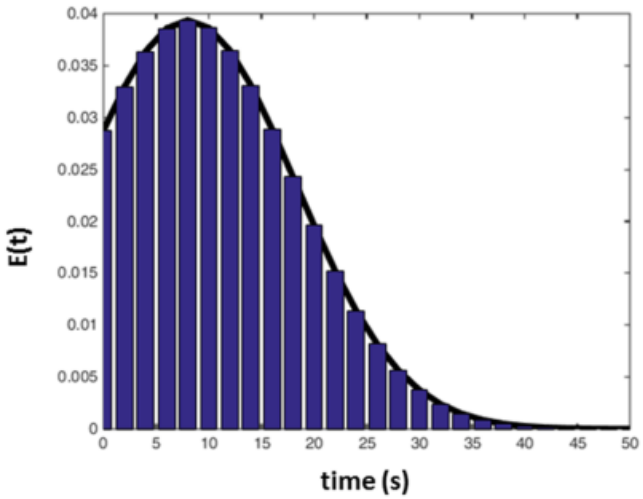

Figure 13: Discretized residence time distribution approach. The black line represents the continuous RTD with mean residence time τ and standard deviation σ. This can be approximated as a histogram, represented by the blur bars.

This approach shares some features with the Markov chain method described by Marikh et al. (Marikh et al., 2006) in that the flow rate entering the system is set to be the incoming feed rate from the upstream feeders and the powder flow is assumed to be perfectly mixed crosswise. This facilitates the use of a one dimensional model in which the residence time is governed by the axial motion of particles. It should be noted that the assumption of perfect crosswise mixing in this case means only that powder entering the system from each of the three feeders has the same probability of entering a particular radial segment of the discretized blender geometry.

The radial discretization of the blender used in this model is shown in Figure 9. The blender is discretized radially into 36 elements, or bins. For each bin a mean axial velocity ($\bar{u}_x^n$) is determined using a discrete element reduced-order model, as shown in equation (21). In this equation n is an index indicating the bin number and ω_b is the agitator rotation rate for the blender. This, in conjunction with the blender length (L), is used to calculate the residence time in segment n (τ_n) as shown in equation (22).

$$\bar{u}_x^n(t) = f^{DE-ROM}(\omega_b(t)) \tag{21}$$

$$\tau_n(t) = \frac{L}{\bar{u}_x^n(t)} \tag{22}$$

It is not realistic to assume that the mass of powder entering the blender distributes evenly among all 36 bins, as material may settle towards the bottom of the system due to gravity. This effect is particularly pronounced at lower rotation rates. For this reason, material entering the blender is distributed among the bins using the empirical model shown in equation (23). This kriging model relates the incoming flow rate to the mass of material in each radial segment at the entrance of the blender. The kriging model parameters are fitted using information from the DEM simulation. In this equation, $X_{in}^{\ n}$ represents the weight fraction of the incoming material that is assigned to bin *n*, which is determined as a function of the agitator rotation rate. $\dot{F}_n^{i,in}$ represents the mass flow rate of component *i* entering bin *n*.

$$\begin{aligned} X_{in}^{\ n}(t) &= f^{kriging}(\omega_b(t)) \\ \dot{F}_n^{i,in}(t) &= \dot{F}_{total}^{i,in}(t) X_{in}^{\ n}(t) \end{aligned} \tag{23}$$

A component mass balance is combined with a delay differential equation to determine the mass flow rate of material exiting the blender and the holdup of material in the blender. The delay differential equation is evaluated on a per component and per bin basis, as shown in equation (24).

$$\begin{aligned} &\tau_n \frac{\partial \dot{F}_n^{i,delayed}(t,z)}{\partial t} = -\frac{\partial \dot{F}_n^{i,delayed}(t,z)}{\partial t} \\ &\dot{F}_n^{i,delayed}(t, z=0) = \dot{F}_n^{i,in}(t) \\ &\dot{F}_n^{i,delayed}(t=0, z>0) = 0 \end{aligned} \tag{24}$$

τ_n is the mean residence time in bin *n*
the superscript *i* denotes a component in the mixture
z is the time delay domain

The mass holdup is calculated based on the relative flow rates of material into and out of the blender using equation (25). Finally the relative standard deviation with respect to component A at the exit of the blender is calculated as described in Chapter 2 using equation (4).

$$\frac{dm}{dt} = \sum_{i \in components} \dot{F}_n^{i,in}(t) - \dot{F}_n^{i,out}(t)$$
$$\text{where } \dot{F}_n^{i,out}(t) = \dot{F}_n^{i,delayed}(t, z=1) \quad (25)$$
$$m(t=0) = 0$$

It should be noted that the proposed discretized RTD approach only considers composition variability due to axial mixing in the blender. Micro mixing phenomena related to the radial dispersion is not explicitly accounted for in the current work. Methods for relating RTD fluctuations to blend uniformity are presented in the literature (Gao, Vanarase, et al., 2011), and can be added to this model in future work to enhance its predictive ability.

Comparison of Simulated and Experimental Results

The integrated feeder-blender system was simulated at experimental conditions intended to match those studied in prior work (Gao, Vanarase, et al., 2011; Vanarase & Muzzio, 2011) so that the simulation results could be compared with those from the experimental system. The conditions simulated in this study are shown in Table 7. The feed rate was kept approximately constant at 30 kg/hr, though the instantaneous feed rate varied by +/- 1% of the set point due to the random noise parameter *v*.

Table 7: Simulated conditions for continuous feeding and blending flowsheet

Feed Rate (kg/hr)	Rotation rate (rpm)	Blade configuration	Blender size (m)	Target Concentration (%w/w)
30	100, 160, 250	Alternating	Length = 0.3 Diameter = 0.1	3

A representative trajectory for the process simulation is depicted in Figure 14. This shows the dynamic response of the continuous blender to a feeder refill for component A occurring at *t=100* seconds. It can be seen that the blender dampens the disturbance in concentration of component A due to the residence time distribution in this unit operation. The response of the blend uniformity (RSD) to the disturbance is also shown in Figure 14b).

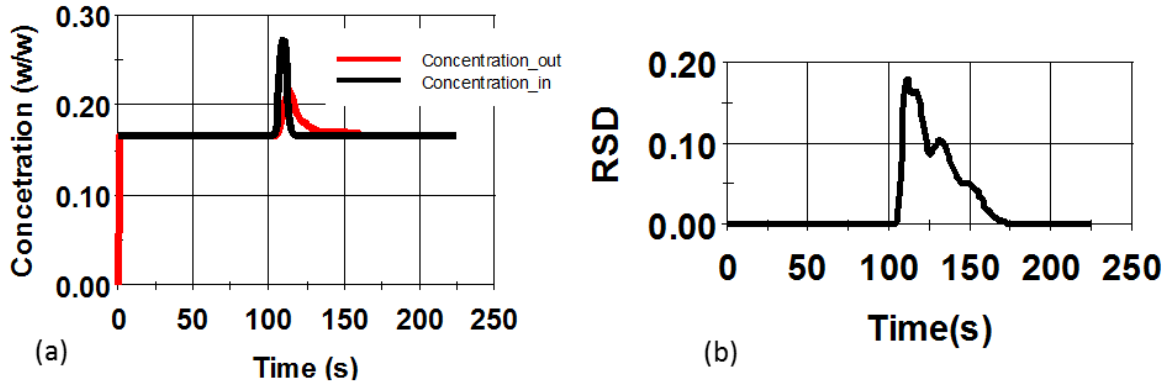

Figure 14: Simulation results for the integrated feeding and blending model shown in Figure 12. A disturbance introduced by refilling Feeder A at t=100s. a) composition of component A at the blender inlet and blender outlet respectively. b) dynamic response of the blend uniformity at the mixer exit

Figure 15 compares the mean residence time, steady state holdup and relative standard deviation as a function of the agitator rotation rate for the simulated system with previously reported experimental results. In Figure 15, Experiment 1 (Exp1) refers to the results of Vanarase and Muzzio (Vanarase & Muzzio, 2011) and Experiment 2 (Exp2) refers to the results reported in Gao et al. (Gao, Vanarase, et al., 2011). The residence time and holdup trends observed in the simulation agree well with those observed experimentally. The predicted holdup and mean residence time are somewhat lower than those observed experimentally at lower rotation rates. However the degree of variability between measurements for mean residence time reported in the two experimental datasets is on the same order of magnitude of the discrepancy between the simulated and experimentally observed values. It can be seen that at higher rotation rates the simulated holdup and mean residence times agree better with those observed experimentally. This may be due to the observed tendency for these responses to be less sensitive to the mass flow rate at higher rotation rates (Vanarase & Muzzio, 2011). The DEM simulation that was used to obtain velocity profiles for the flowsheet model was operated using a continuous particle flow rate that did not take into account the type of feed rate variability that is observed when using a loss-in-weight feeder. Therefore it is possible that the simulation does not agree as well with

experimental findings at lower rotation rates due to the influence of feed rate variability on blending performance which is not captured as effectively by the simulation.

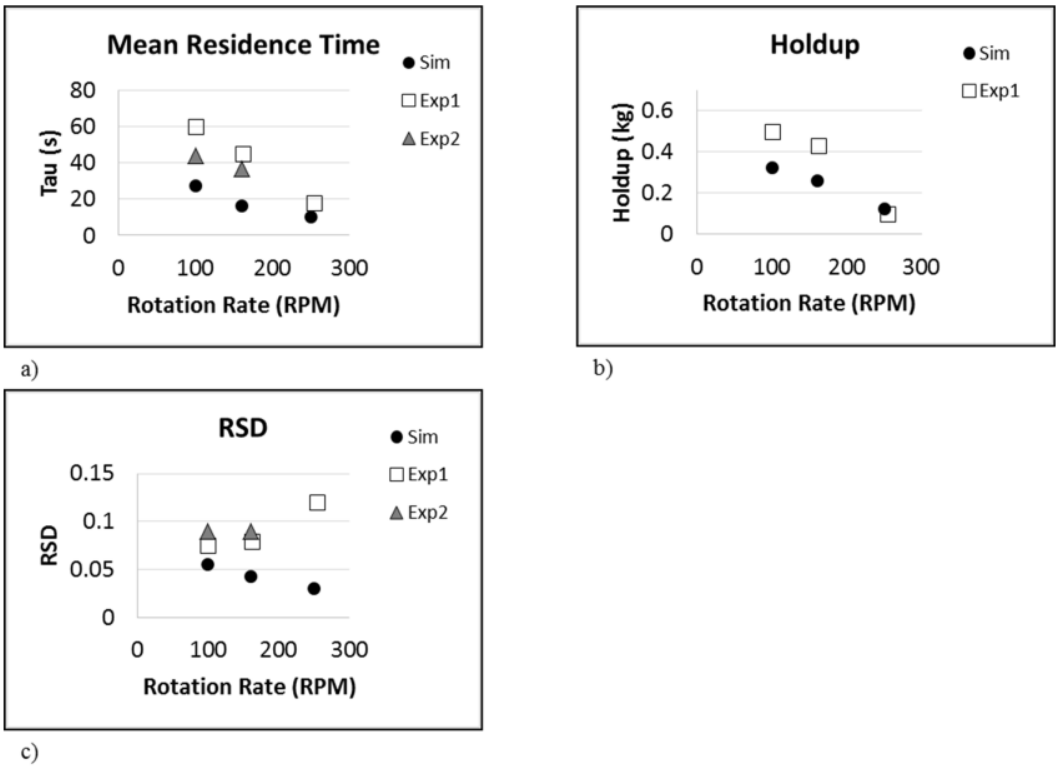

Figure 15: (a) Mean residence time (b) steady state holdup and c) relative standard deviation as a function of agitator rotation rate for continuous feeding and blending flowsheet simulation. Simulated results are compared with experimental findings indicated as Exp1 (Vanarase & Muzzio, 2011) and Exp2 (Gao, Vanarase, et al., 2011).

The simulated RSD tends to be lower than that observed experimentally, particularly at higher rotation rates. This could be due in part to the fact that an RTD model is not ideally suited to represent micro scale mixing behavior. (Gao, Vanarase, et al., 2011) In particular, at high rotation rates the ratio of axial to radial particle velocities may so high that the material experiences fewer blade passes in the blender. (Vanarase & Muzzio, 2011) Because the discretized RTD model does not explicitly incorporate axial particle fluxes, it may not fully describe this phenomenon. In the future, adding a radial dispersion model like that described in Gao et al (Gao, Vanarase, et al.,

2011) to the RTD model could enhance the ability of the simulation to represent the effect of rotation rate on blend homogeneity.

Under prediction of RSD relative to the experimentally determined values may also be due in part to the effect of measurement error. In the experimental studies, mixture concentration was evaluated using at-line samples which were subsequently analyzed by near infrared (NIR) spectroscopy. This process introduces some degree of sampling error and measurement error. The relative standard deviation of the NIR measurement technique was reported as 1% by Gao et al (Gao, Vanarase, et al., 2011) while Vanarase and Muzzio (Vanarase & Muzzio, 2011) report a root mean square error of prediction of 0.239 for the method. This measurement error is not accounted for in the simulated system. In future work, the effects of sampling or measurement error could be explicitly considered in the simulation to more accurately reflect experimental observations.

3.4. Summary of reduced-order modeling applications

This chapter has provided an overview of reduced-order modeling techniques that will be applied throughout this thesis. These include principal component analysis, kriging and PCA-based reduced-order modeling. These methods have been demonstrated in a case study for a continuous blending process. A PCA-based modeling approach known as dynamic discrete element reduced-order modeling has been introduced and it has been demonstrated that this method can accurately represent high dimensional state data from a DEM simulation. It has also been shown that kriging can successfully model highly nonlinear and dynamic data in the form of blend RSD information for a continuous blending process.

A flowsheet simulation for a continuous feeding and blending process that incorporates particle velocity information from a discrete element reduced-order model has been developed. This model demonstrates the integration of reduced-order models into flowsheet simulations for a continuous pharmaceutical process. Results from this flowsheet simulation have been compared

with experimental data from a comparable system previously reported in the literature (Gao, Vanarase, et al., 2011; Vanarase & Muzzio, 2011). This is a significant development, as validation of integrated process models against experimental data has been somewhat lacking in the area of continuous pharmaceutical manufacturing. The simulation results show order of magnitude agreement with experimental results for mean residence time and holdup in the system. At lower blending speeds reasonable agreement with experimentally observed relative standard deviation is also obtained. At higher blending speeds the simulation tends to under predict blend RSD. Additional model terms to better represent radial dispersion phenomena are needed to achieve better agreement with experimentally observed blend uniformity.

Chapter 4

4. Sensitivity analysis of integrated pharmaceutical processes

4.1. Role of sensitivity analysis in pharmaceutical process development

Sensitivity analysis is an important aspect of both process and model development. It can be used to quantitatively evaluate the significance of operating conditions and model parameters with respect to process variability. This information can direct experimental effort towards improving the estimation of important model parameters or identifying conditions that will enhance understanding of process design space. (Saltelli, Tarantola, et al., 2000; Sin et al., 2009) Sensitivity analysis can also be used to identify aspects of a process which are not sufficiently well understood and areas in which the corresponding model needs to be improved.

Lack of sensitivity to certain factors can also provide important information about a process. This can be used to justify reducing the uncertain input space (e.g. by holding these parameters constant in subsequent studies) or to suggest additional parameter variations to adequately capture the influence of these factors. (Ramachandran et al., 2009; Saltelli, Chan, et al., 2000; Saltelli, Tarantola, et al., 2000) Sensitivity analysis can also be used to identify critical process design and operating parameters. (Singh et al., 2010) This is particularly important in drug product processes, where input material properties can vary significantly on a product by product basis. (P. K. McKenzie, S. Tom, J. Rubin, E., Futran, M., 2006) Understanding the influence of active ingredient (API) physical properties on product quality can suggest API quality specifications and aid in the development of robust processes. (Zhang et al., 2011)

The use of sensitivity analysis for process development for pharmaceutical applications has been increasingly studied in recent years within the context of QbD. In solids-based pharmaceutical modeling, sensitivity analysis is often discussed in conjunction with parameter estimation for mechanistic and multi-scale models, particularly those incorporating population balance equations. (Cryer & Scherer, 2003; Gantt & Gatzke, 2006; Mortier et al., 2012; Ramachandran et

al., 2009; Vikhansky & Kraft, 2006) Sensitivity analysis has also been used to identify critical design and operating parameters for specific unit operations such as granulation, mixing and tablet coating. (Getaz et al., 2013; Sin et al., 2009) When applied to an integrated flowsheet model consisting of a series of sequential unit operations, sensitivity analysis can identify key relationships between raw material properties, design parameters, operating conditions and final product quality. The results of such an analysis can thus be used to guide equipment selection, contribute to quality risk assessment (e.g. by identifying critical process parameters) and aid in the development of appropriate control strategies to ensure product quality. (F. Boukouvala et al., 2010a; Chen et al., 2010; Ramachandran et al., 2011; Singh et al., 2012; Westerberg et al., 2013)

This chapter is concerned with the application of global sensitivity analysis techniques to an integrated model for a continuous direct compaction process. Several different methods will be applied and the resulting sensitivity metrics evaluated and compared. These metrics will then be used to identify potential critical process parameters and manipulated variables for control strategies. Sensitivity analysis will also be used to suggest model improvements in areas where the existing model is not sufficiently detailed to capture the effect of certain parameters.

4.2. Global sensitivity analysis methods

Sensitivity analysis (SA) describes the process of attributing variance in process responses to sources of uncertainty. (Saltelli, Chan, et al., 2000) Throughout this chapter the terms *input* or *uncertain parameter* will be used to describe these sources of uncertainty, which include raw material attributes, process design parameters and operating conditions. These inputs vary within some distribution $P(x)$, with lower and upper bounds based on anticipated levels of variability during normal operation. *Responses* or *outputs* will be used to describe model outputs which may be affected by variability in the inputs. Sensitivity analysis can be either local or global in nature. Local sensitivity analysis considers individual input-response relationships and often involves one at a time (OAT) variation of inputs. In order to evaluate the effect of multiple inputs acting in

concert on a particular response it is necessary to conduct a global sensitivity analysis, in which all inputs are varied simultaneously. Thus the effect of individual inputs as well as the cooperative effect of multiple inputs on a particular response can be evaluated. (Saltelli, Chan, et al., 2000; Saltelli, Tarantola, et al., 2000)

In this work, sensitivity analysis is applied to a continuous direct compaction process model. The goal of this analysis is to identify potential critical process parameters and control strategies with respect to tablet quality attributes including API concentration, concentration variability (RSD), weight and hardness. Interactions between process parameters, equipment configuration and material properties are known to influence these process responses, so a global sensitivity approach is selected. (F. Boukouvala et al., 2012; Gao, Muzzio, et al., 2011; M. A. Järvinen, Paaso, J., Paavola, M., Leivisk, K., Juuti, M., Muzzio, F., Järvinen, K., 2012; Martínez et al., 2013) The global sensitivity analysis methods considered include the partial rank correlation coefficient (PRCC), random sampling high dimensional model representation (RS-HDMR), the extended Fourier Amplitude Sensitivity Test (eFAST) and Sobol's method. A brief overview of each technique is given in the remainder of this section.

Partial Rank Correlation Coefficient

The partial correlation coefficient (PCC) is a regression-based metric that indicates the strength of the linear relationship between a specific input factor x_i and a response y, having corrected for the effect of all other inputs $x_{\sim i}$. By rank transforming data it is possible to apply this technique to nonlinear input - response data, thereby calculating the partial rank correlation coefficient (PRCC). The magnitude of the PRCC indicates the strength of the monotonic relationship between an input x_i and a response y. (Saltelli, Chan, et al., 2000) It is calculated by first expressing x_i and y in terms of all other inputs $x_{\sim i}$ using linear regression, as shown in equation (26).

$$\hat{y} = b_0 + \sum_{h \neq i} b_h x_h$$
$$\hat{x}_i = c_0 + \sum_{h \neq i} c_h x_h \qquad (26)$$

The partial correlation between x_i and y is then given by the correlation coefficient between $x_i - \hat{x}_i$ and $y - \hat{y}$.

The PRCC is an appealing metric for screening because it can be estimated using fewer samples than variance-based metrics. However it also has several limitations. For instance, it is most effective for processes where the inputs are not strongly correlated. In addition, the accuracy of the estimated sensitivities depends on the accuracy of the linear expressions fitted in equation (26). (Saltelli, Chan, et al., 2000) In this research partial rank correlation coefficients are calculated using SimLab (version 2.2), developed at the European Commission's Joint Research Centre.

4.2.1. Variance based sensitivity analysis

Unlike regression-based approaches (e.g. PRCC), variance based sensitivity analysis can be applied to processes with correlated inputs and can even identify the extent to which input interactions influence process responses. As their name implies, variance based metrics reflect the contribution of individual inputs or groups of inputs to variance in process responses. This contribution is calculated based on the conditional variance in a process response due to a specific input or group of inputs. An individual sensitivity metric (S_i) can be used to evaluate the contribution of a single input x_i to variability in a particular response y. An interaction index (S_{ij}) can be used to assess the cooperative effect of variables x_i and x_j on a process response. Finally, a total sensitivity index (S_{Ti}) can be used to evaluate the effect x_i and all of its higher order interactions with other inputs $x_{\sim i}$ on a response y. These metrics are determined as the ratio of partial variances (D_i, D_{ij} and $D_{\sim i}$) to total variance in the process response (D) as shown in equations (27) through (29). Variance based sensitivity analysis methods differ in how they estimate the conditional variances D_i, D_{ij} and $D_{\sim i}$. Estimating these quantities accurately can

require a large number of model evaluations, particularly if there are many uncertain inputs to consider. Therefore screening methods are recommended to reduce the number of variables considered prior to implementing variance based sensitivity analysis.

$$S_i = \frac{D_i}{D} = \frac{\mathrm{var}_{x_i}[\mathrm{E}(\mathrm{y} \mid \mathrm{x}_i)]}{\mathrm{var}(\mathrm{Y})} \tag{27}$$

$$S_{ij} = \frac{D_{ij}}{D} = \frac{\mathrm{var}_{x_i,x_j}[E(y \mid x_i, x_{,})] - \mathrm{var}_{x_i}[\mathrm{E}(\mathrm{y} \mid \mathrm{x}_i)] - \mathrm{var}_{x_j}[\mathrm{E}(\mathrm{y} \mid \mathrm{x}_j)]}{\mathrm{var}(Y)} \text{ for } i \neq j \tag{28}$$

$$S_{Ti} = 1 - \frac{D_{\sim i}}{D} = 1 - \frac{\mathrm{var}_{x_{\sim i}}[\mathrm{E}(\mathrm{y} \mid \mathrm{x}_{\sim \mathrm{i}})]}{\mathrm{var}(\mathrm{Y})} \tag{29}$$

Random Sampling High Dimensional Model Representation (RS-HDMR)

Sensitivity analysis using random sampling high dimensional model representation (RS-HDMR) is a model-based approach that can be used to estimate the various sensitivity metrics shown in equations (27)-(29). The number of samples required for the conditional variance estimates depends in part on the number of samples required to generate an accurate model by high dimensional model representation (HDMR). However in some cases this method can require significantly fewer samples than other variance based approaches like Sobol's method. (Sobol, 1993; Ziehn & Tomlin, 2009) High dimensional model representation (HDMR) is a reduced-order modeling approach that involves the use of a finite hierarchal correlated function expansion based on process inputs to express a system output. The general mathematical formulation for HDMR is shown in equation (30). This representation accounts for the contribution of individual inputs as well as higher order input interactions to a particular process response *f(x)*. (G. Li et al., 2001; G. Y. Li et al., 2008)

$$\begin{aligned} f(x) = f_0 + \sum_{i=1}^{n} f_i(\mathrm{x}_i) + \sum_{1 \le i < j \le n}^{n} f_{ij}(x_i, x_j) + \sum_{1 \le i < j < k \le n}^{n} f_{ijk}(x_i, x_j, \mathrm{x}_k) + \ldots \\ + \sum_{1 \le i_1 < \ldots < i_2 \le n}^{n} f_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l}) + \ldots + f_{12\ldots n}(x_1, x_2, \ldots, x_n) \end{aligned} \tag{30}$$

The term f_0 in equation (30) is the average process response over the input domain. The first-order terms $f_i(x_i)$ indicate individual contributions of variables x_i and the second order terms $f_{ij}(x_i,x_j)$ indicate the contribution of interactions between variables x_i and x_j to response variability. For most applications, equation (30) is limited to the second order expansion shown in equation (31). This limits the computational time for calculating the component functions and is justified by the fact that ternary and higher order interaction terms tend to contribute significantly less to overall response variability than first and second order terms.(G. Li et al., 2001; G. Y. Li et al., 2002)

$$f(x) \approx f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{1 \le i < j \le n}^{n} f_{ij}(x_i, x_j) \tag{31}$$

The form of the component functions in equation (31) can be selected to optimally represent the available process data. (G. Li et al., 2001; G. Li et al.; G. Y. Li et al., 2002) In order to expedite the fitting of these functions, analytical basis functions such as orthonormal polynomials or splines are often used. (G. Y. Li et al., 2002) In terms of these basis functions, equation (31) can be expressed as:

$$f(x) = f_0 + \sum_{i=1}^{n}\sum_{r=1}^{k} \alpha_r^i \varphi_r(\mathrm{x}_i) + \sum_{1 \le i < j \le n} \sum_{p=1}^{l}\sum_{q=1}^{l'} \beta_{pq}^{ij} \varphi_{pq}(\mathrm{x}_i, \mathrm{x}_j) \tag{32}$$

In equation (32) α_r^i and β_{pq}^{ij} are coefficients for the basis functions $\varphi_r(\mathrm{x}_i)$ and $\varphi_{pq}(\mathrm{x}_i, \mathrm{x}_j)$ respectively. (G. Y. Li et al., 2002)

HDMR component functions can be used to estimate the conditional variances required for the calculation of the sensitivity metrics, as shown in equations (33)-(35).

$$D = \int f^2(x)dx - f_0 \tag{33}$$

$$D_i = \int_0^1 f_i^2(x_i)dx_i \approx \sum_{r=1}^{k_i} (\alpha_r^i)^2 \tag{34}$$

$$D_{ij} = \int_0^1 \int_0^1 f_{ij}^2(\mathrm{x}_i, \mathrm{x}_j)\,\mathrm{dx}_i\,dx_j \approx \sum_{p=1}^{l_i} \sum_{q=1}^{l_j'} (\beta_{pq}^{ij})^2 \quad (35)$$

A MATLAB based platform for HDMR model development and sensitivity analysis called GUI-HDMR has been developed by Ziehn and Tomlin (Ziehn & Tomlin, 2009). This tool is used in the current work to fit HDMR models and estimate sensitivity indices.

Extended Fourier Amplitude Sensitivity Test

The extended Fourier Amplitude Sensitivity Test (FAST) is an alternative method for evaluating individual (S_i) and total (S_{Ti}) sensitivity indices. It can provide accurate estimates of these quantities using fewer samples than Sobol's method and for problems with a limited number of inputs may also be more efficient than HDMR. (Saltelli, 2002) It is based on the FAST method proposed in the 1970s for analyzing sensitivities in complex systems of chemical reactions. (R. I. Cukier et al., 1973; R.I. Cukier et al., 1978; R. I. Cukier et al., 1975; Schaibly & Shuler, 1973) The extended Fourier amplitude sensitivity test (eFAST) describes the extension of the FAST method to calculate total sensitivity indices. eFAST employs a frequency based method to sample the input space efficiently. The observed variance in the outputs can then be represented as a multiple Fourier series. (Saltelli et al., 1999) The spectrum of the i^{th} term of the Fourier series expansion indicates the contribution of the i^{th} input factor to the total variance in y. (Saltelli, Chan, et al., 2000) Taking advantage of the mutually orthogonal nature of terms in the multiple Fourier series, the partial variances attributable to each input can thus be obtained independently using a single set of samples. (Saltelli & Bolado, 1998; Saltelli, Chan, et al., 2000)

The eFAST sample strategy is based on search curves that systematically explore the input space. Each input is sampled along a curve which depends on some frequency ω_i as defined by equation (36). (Saltelli & Bolado, 1998; Saltelli et al., 1999)

$$x_i(s) = G_i(\sin(\omega_i s)) \quad (36)$$

In equation (36), s is a scalar that allows for systematic exploration of the input space and the form of the function G_i depends on the distribution from which the factor x_i is to be sampled. In the current work, the transformation proposed by Saltelli et al. (Saltelli et al., 1999) for uniform distributions has been used. This is shown in equation (37), where ϕ_i is a random phase shift on the interval $[0,2\pi)$. The phase shift facilitates resampling, a process by which several unique search curves can be generated for each input. For a process with multiple uncertain parameters, a set of search curves is generated for each parameter in which that input is varied at a frequency ω_i. This frequency is higher than the frequency $\omega_{\sim i}$ at which all other variables are sampled. The algorithm used to select the frequencies ω_i and $\omega_{\sim i}$ in this work can be found in Appendix A of Saltelli et al. (Saltelli, Chan, et al., 2000)

$$x_i = \frac{1}{2} + \frac{1}{\pi}\arcsin(\sin(\omega_i s) + \varphi_i) \tag{37}$$

The conditional variance attributed to the factor x_i can then be estimated using the first M harmonics of the spectrum of the Fourier series at ω_i. In this study M is set to 4 as recommended in Saltelli et al. (Saltelli et al., 1999) The conditional variance due to the complementary set of inputs $x_{\sim i}$ can be estimated similarly based on $\omega_{\sim i}$. These calculations are shown in equations (38) and (39) respectively.

$$\hat{D}_i = 2\sum_{p=1}^{M} \Lambda_{p\omega_i} \tag{38}$$

$$\hat{D}_{\sim \mathrm{i}} = 2\sum_{p=1}^{M} \Lambda_{p\omega_{\sim i}} \tag{39}$$

The spectrum Λ is calculated from the Fourier Coefficients as in equation (40). The coefficients themselves can be obtained by numerically evaluating the integrals in equations (41) and (42).

$$\Lambda_j = 2\sum_{j=1}^{\infty} A_j^2 + B_j^2 \tag{40}$$

$$A_j = \int_{-\pi}^{\pi} f(\mathrm{s}) \cos(\mathrm{js})\, \mathrm{ds} \tag{41}$$

$$B_j = \int_{-\pi}^{\pi} f(\mathrm{s}) \sin(\mathrm{js})\, \mathrm{ds} \tag{42}$$

The variances calculated in equations (38) and (39) can be used to determine sensitivity and total sensitivity indices in equations (27) and (29). (Saltelli, Chan, et al., 2000; Saltelli et al., 1999)

Sobol's Method

Unlike the extended FAST method, Sobol's method can be used to calculate all three of the variance based sensitivity metrics, S_i, S_{ij} and S_{Ti}. (Sobol, 1993) Sobol's method relies on a functional decomposition scheme similar to that shown for HDMR in equations (31) and (32). However Sobol's method does not rely on the use of component functions. Instead, the terms in the functional decomposition are calculated through numerical evaluation of multi-dimensional integrals as shown in equations (43) and (44). In (43) and (44), $dx_{\sim i}$ and $dx_{\sim ij}$ refer to integration over all variables but *i* and over all variables but *i* and *j* respectively.

$$f_i(x_i) = -f_0 + \int_0^1 \cdots \int_0^1 f(x)\, dx_{\sim i} \tag{43}$$

$$f_{ij}(\mathrm{x}_i, \mathrm{x}_j) = -f_0 - f_j(x_j) + \int_0^1 \cdots \int_0^1 f(x)\, dx_{\sim \mathrm{ij}} \tag{44}$$

The partial variances required for variance based sensitivity analysis can be calculated via Monte-Carlo integration of (43) and (44) as shown in equations (45)-(47).

$$\hat{f}_0 = \frac{1}{n} \sum_{m=1}^{n} f(x_m) \tag{45}$$

$$\hat{D} = \frac{1}{n} \sum_{m=1}^{n} f^2(x_m) - \hat{f}_0^2 \tag{46}$$

$$\hat{D}_i = \frac{1}{n}\sum_{m=1}^{n} f(x^{(1)}_{(\sim i)m}, x^{(1)}_{i\ m})\, f(x^{(2)}_{(\sim i)m}, x^{(1)}_{i\ m}) - \hat{f}_0^{\,2} \tag{47}$$

The most significant challenge associated with the calculation of sensitivity indices via Sobol's method is the potentially large number of samples (indicated by *n* in equations (45)-(47)) required to evaluate these multi-dimensional integrals accurately using Monte-Carlo integration. A variety of modifications to Sobol's method have been described to increase the accuracy of the estimated indices. (Homma & Saltelli, 1996; Saltelli, Chan, et al., 2000) In addition, sampling strategies have been suggested to reduce the number of samples needed to accurately estimate the Monte-Carlo integrals. (Dimov & Georgieva, 2010; Glen & Isaacs, 2012; Saltelli, 2002) In this work, the input space is sampled using quasi-random Sobol sequences. (Bratley & Fox, 1988; Joe & Kuo, 2003) The modified formula proposed by Saltelli et al. (Saltelli et al., 2010), shown in equation (48), is used to calculate the individual variance contribution D_i.

$$D_i = \frac{1}{N}\sum_{j=1}^{N} f(\mathrm{A})_j\,(f(B^{(\mathrm{i})}_A)_j - f(A)_j) \tag{48}$$

In equation (48), *A* and *B* are matrices of quasi-random numbers used to sample the input space. The matrix $B_A^{(i)}$ is one in which column *i* from matrix *A* replaces the same column of matrix *B*. The variance due to parameter interactions can be computed as shown in equation (49), which was originally proposed by Jansen (Jansen, 1999).

$$D_{ij} = \frac{1}{2N}\sum_{w=1}^{N} (f(A^{(i)}_B)_w - f(\mathrm{A}^{(\mathrm{j})}_B)_w)^2 \tag{49}$$

The total number of samples required depends in part on the number of inputs to be considered (*k*). In this case we follow the proposed sampling strategy of Saltelli et al. (Saltelli, 2002) and use *n(k+2)* model evaluations to estimate the partial variances, where *n* is a large number (200 in this work).

4.3. Sensitivity analysis of a continuous direct compaction process

Continuous Direct Compaction Flowsheet Simulation

Using the unit operation models described in Chapter 2, an integrated process model for continuous tablet production via direct compaction can be assembled. (F. Boukouvala et al., 2012) The continuous direct compaction flowsheet model used for this sensitivity study is depicted in Figure 16. The process consists of three loss-in-weight feeders, a continuous convective mixer, and a tablet press. In this case the continuous blender is modeled using a two dimensional population balance, as described by equation (5)-(6). Also shown in Figure 16 are the inputs to the sensitivity analysis and the process responses for which sensitivity indices are calculated. The uncertain parameters selected for the sensitivity analysis represent only a fraction of the approximately 120 user-specified inputs to the direct compaction simulation. These inputs include design parameters, operating conditions, material properties, and model parameters for the unit operation equations. The current analysis focuses on a subset of these parameters with an emphasis on design and operating conditions as well as material properties. Sources of model uncertainty are not considered in this analysis because the objective is to identify potential critical process parameters and control strategies. A summary of the varied parameters and responses considered is included in Figure 16, with the varied parameters and the distributions from which they are sampled on the left hand side and the process responses on the right hand side. A summary of the uncertain parameters, their bounds and the distributions from which they are sampled is also provided in Table 9 at the end of this chapter.

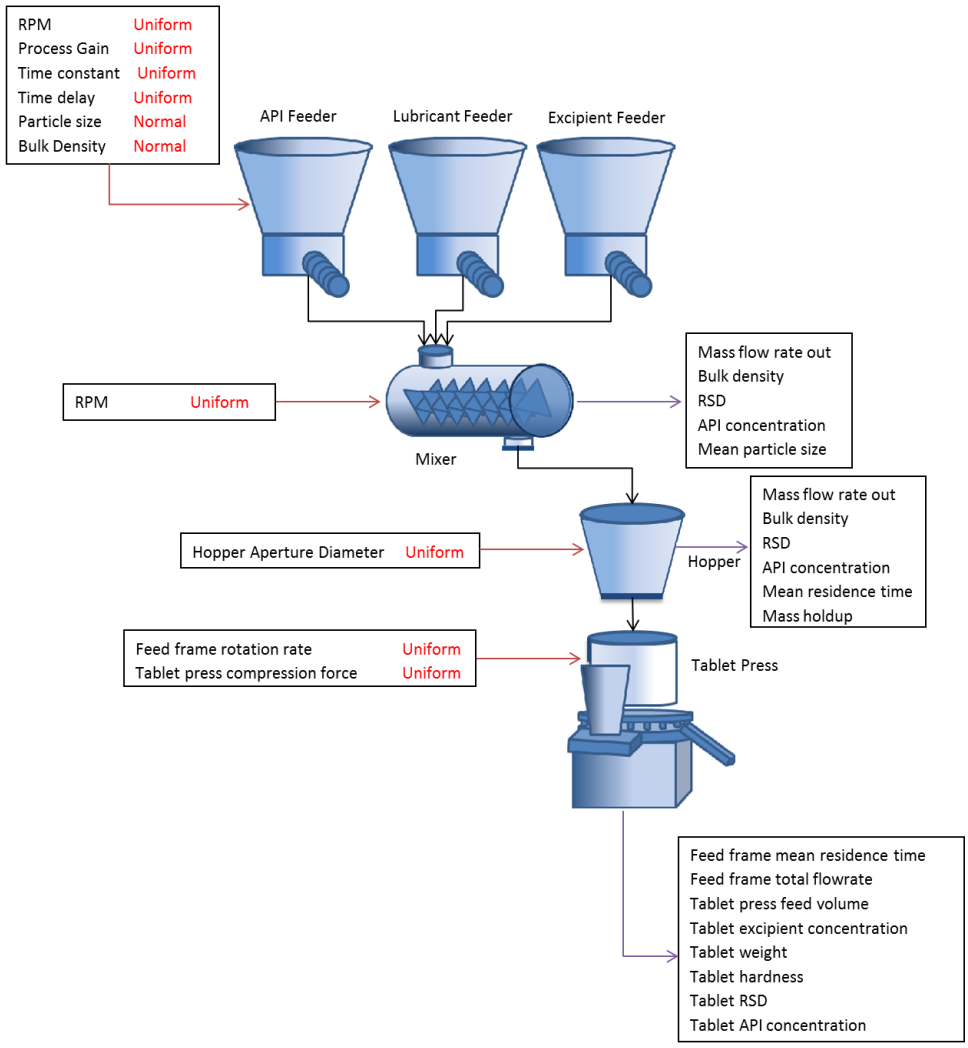

Figure 16: Flowsheet simulation for a continuous direct compaction process indicating inputs to the sensitivity analysis and their distributions as well as responses for sensitivity analysis

The direct compaction process model can be used to evaluate the effect of changes in the set of uncertain inputs on process outcomes. Variability in the tablet quality attributes is of particular interest, as inputs significantly affecting quality are potential critical process parameters (CPP) or manipulated variables for a control strategy. Figure 17 shows the trajectories of four quality attributes, tablet weight, hardness, active ingredient (API) concentration and relative standard deviation (RSD). Three different scenarios are shown: the case where all inputs are set to their mean (center point) values as well as the cases where all uncertain parameters are fixed at their

minimum (low) and maximum (high) values. The trajectories in Figure 17 indicate that tablet quality attributes vary as a function of at least some of the uncertain parameters. Identifying which inputs contribute most significantly to this variability is the goal of the sensitivity analysis.

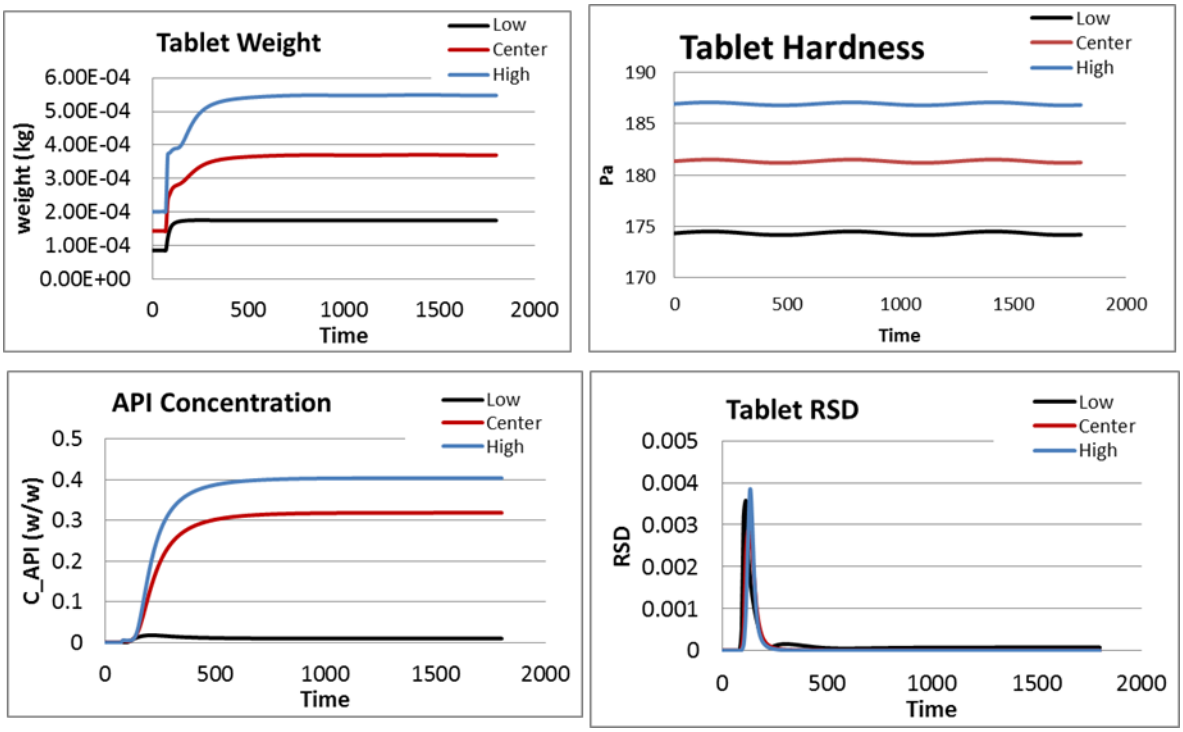

Figure 17: Tablet Properties over time for center point conditions as well as for extreme low and high values of all varied parameters

The direct compaction flowsheet is simulated for 1800 seconds, until the process variables of interest have reached steady state. Sensitivity analysis is then performed on the steady state responses.

Sensitivity Analysis of the Continuous Direct Compaction Process

An initial sensitivity analysis considering all 22 of the uncertain parameters shown in Figure 16 is conducted using the partial rank correlation coefficient and RS-HDMR methods. For these methods, the number of samples required to estimate the sensitivity indices is not a large multiple of the number of uncertain parameters, as is the case for eFAST and Sobol's method. These latter methods, while more sampling intensive, can provide more accurate estimates of sensitivity,

especially for processes with highly nonlinear input-response relationships.(Saltelli, 2002) The *PRCC* and RS-HDMR sensitivity metrics (S_{ij}) can be used to identify variables that contribute most significantly to response variance and to screen out those that do not. This information can be used to reduce the number of uncertain variables prior to conducting extended FAST and Sobol's sensitivity analysis.

The sampling methods for the eFAST approach and Sobol's method are specified by the methods themselves, as described in section 4.2. For the partial rank correlation coefficient and RS-HDMR methods random sampling can be used. In this work Latin Hypercube sampling (LHS) is used to sample the uncertain space for these methods. LHS was selected for its efficiency and demonstrated efficacy for sensitivity analysis in systems with a relatively large number of uncertain inputs. (Blower & Dowlatabadi, 1994; Helton & Davis, 2003; Helton et al., 2005; Olsson & Sandberg, 2002) A summary of the sensitivity analysis techniques used in this chapter and the corresponding sampling strategies and number of samples required is provided in Table 8.

Table 8: Summary of sensitivity analysis techniques applied to direct compaction process

Method	Sensitivity metrics	Sampling method	Number of variables considered	Number of samples
Partial correlation coefficient	*PRCC*	Latin hypercube	22	2048
RS-HDMR	S_i, S_{ij}	Latin hypercube	22	3048
Sobol	S_i, S_{ij}, S_{Ti}	Sobol quasi-random	9	2200
eFAST	S_i, S_{Ti}	Uniform eFast w/ resampling	8	1560

Partial Rank Correlation Coefficient

Partial rank correlation coefficients were determined for each pair of input-response relationships shown in Figure 16, resulting in a high dimensional (*22x19*) dataset. The magnitude of the *PRCC* are displayed as an intensity plot in Figure 18. The results indicate that several of the inputs are particularly significant contributors to response variability, while others have little to no effect on

product quality. The API and excipient material properties, as well as the feeder design and operating parameters for these two components are the factors found to be most influential according to the partial rank correlation coefficient. This is consistent with the fact that API and excipient are the predominant components most pharmaceutical blends. The relative standard deviation throughout the process is most sensitive to API feeder screw speed, which highlights this as a potential manipulated variable for the control of tablet RSD. The feeder gain parameter, which is comparable to the feed factor for loss-in-weight feeders, also significantly affects tablet quality attributes. Commercially available loss-in-weight feeders often determine the feed factor based on weight vs. time data collected from the gravimetric controller. When the feeder operates in volumetric mode (e.g. during feeder refill), this parameter is not updated and can either be set manually or maintained at the most recent value calculated prior to switching to volumetric mode. (W. E. Engisch & Muzzio, 2012) Given the importance of this parameter, it may be desirable to have the ability to set the feed factor (feeder gain) at a supervisory control level. This would be particularly useful during feeder refill. (W.E. Engisch & Muzzio, 2010)

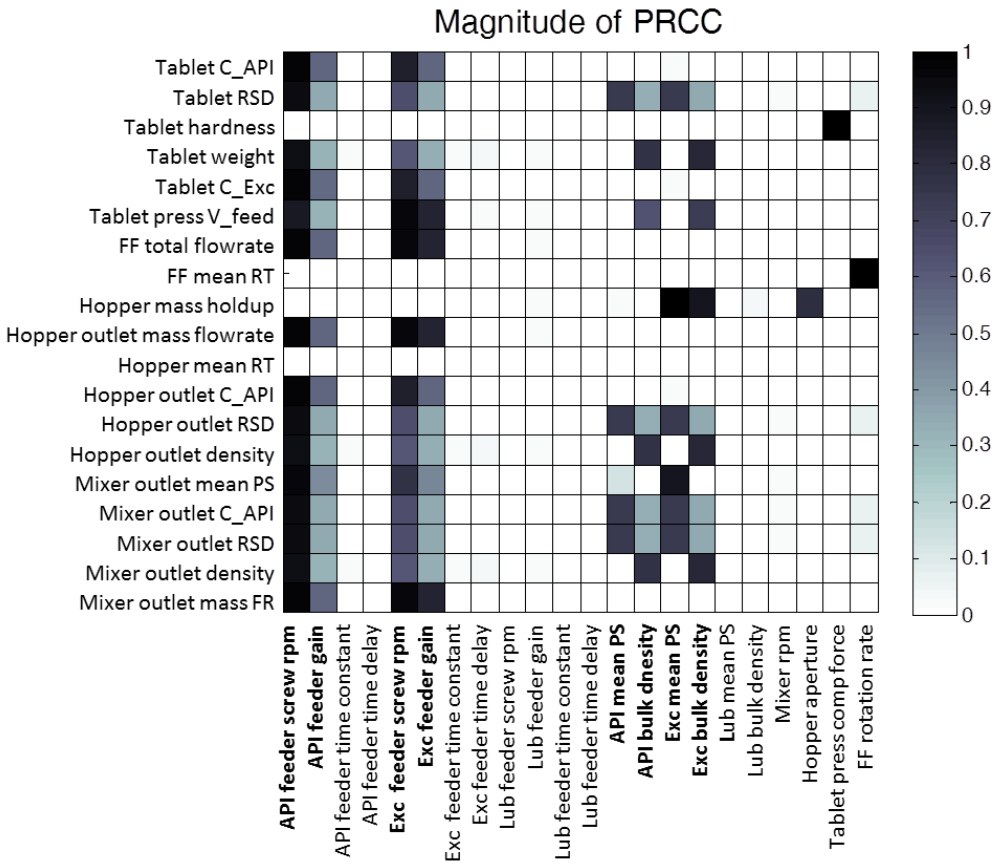

Figure 18: Intensity plot for magnitude of partial rank correlation coefficients

The *PRCC* is also useful in indicating which variables the process is relatively insensitive to. These include the lubricant properties and feeder parameters for the lubricant feeder, as well as the design parameters for the tablet press hopper. In addition, dynamic parameters for control of the feeder such as the time constant and time delay are not particularly influential in this steady state analysis. Based on the *PRCC*, the process also appears relatively insensitive to the agitator rotation rate for the blender. Before eliminating any of these parameters from subsequent consideration, it is important to determine if they interact with any other inputs to significantly influence process performance. This can be accomplished using the interaction indices (S_{ij}) obtained from RS-HDMR.

Random Sampling High Dimensional Model Representation

RS-HDMR has the advantage of being able to provide specific information about the cooperative effect of inputs on a particular response. HDMR models were constructed using 3048 samples generated using LHS (the 2048 sampled for the *PRCC* plus an additional 1000 for model cross validation). HDMR was carried out using the GUI-HDMR software tool developed by Ziehn and Tomlin(Ziehn & Tomlin, 2009). Prior to developing HDMR models the input data was mean-centered and scaled to unit variance. In addition, a variance threshold of 0.1% was applied to expedite HDMR component function fitting. Coefficients of determination for HDMR models are provided in the Table 10 at the end of the chapter.

The individual sensitivity metrics obtained from *RS-HDMR* largely confirm the findings from the *PRCC* and are therefore not presented here. The interaction indices (S_{ij}) are used to identify important variable interactions prior to removing any factors from the analysis. Intensity plots showing input interactions to which certain responses are highly sensitive are shown in Figure 19. For responses that are not significantly influenced by parameter interactions, the corresponding intensity plots are not shown.

The most significant parameter interaction giving rise to variability in process throughput, as indicated by of the interaction indices for the mixer and hopper outlet mass flow rates, is that between the excipient feeder gain and screw speed. The cooperative effect of the excipient feeder screw speed with API feeder parameters also contributes to variability production rate. Tablet weight variability throughout the process is attributed in part to the interaction between the API and excipient material properties as well as the cooperative effect of mixer agitation rate with excipient properties. This highlights the importance of considering univariate and interaction indices at the screening stage, as based on the *PRCC* alone the mixer parameters might have been removed from the set of inputs prior to subsequent analysis.

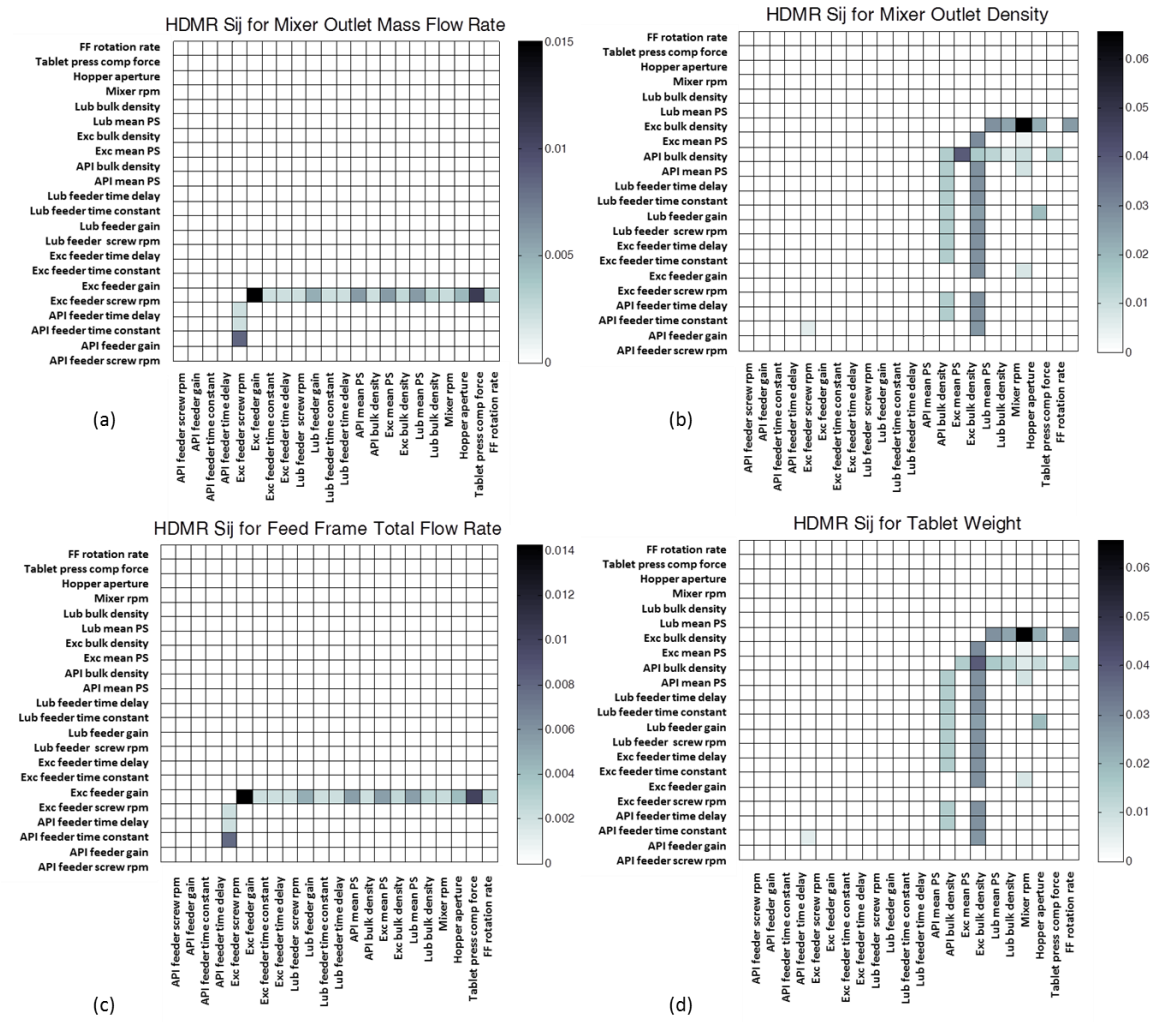

Figure 19: Binary Interaction Sensitivity Indices (S_{ij}) obtained using RS-HDMR for (a) Mixer outlet mass flow rate (b) mixer outlet density (c) Feed Frame Total Flow rate and (d) Tablet Weight

Based on the initial sensitivity results, parameters related to the lubricant properties and lubricant feeder are not considered in the remaining sensitivity analysis methods. This does not imply that these parameters are not significant for tableting processes in general, just that over the parameter ranges considered and for the process model used in the current study the predicted process responses are not sensitive to these parameters.

Sobol's Method

Having reduced the number of uncertain parameters under consideration, variance based methods that require a greater number of samples per variable can be implemented to further rank the most

significant parameters with respect to their impact on process outcomes. Sobol's sensitivity analysis was applied to the nine parameters indicated to be most significant based on the preliminary screening. These include API and excipient feeder parameters, API and excipient material properties, and agitator rotation rate in the blender (mixer rpm). One advantage of Sobol's sensitivity analysis is that it can provide individual, total and interaction sensitivity indices using the same set of samples. However the number of samples required scales by a factor of *n* (in this case 200) with the number of variables considered. In this case Sobol's method requires 2200 samples obtained via the Sobol quasi-random sequence. (Bratley & Fox, 1988; Joe & Kuo, 2003)

The first order and total sensitivities estimated using Sobol's method are shown in Figure 20. The Sobol metrics provide a more nuanced view of process sensitivities than the *PRCC* or RS-HDMR metrics. The API feeder screw speed clearly remains one of the most significant parameters. However it is also evident that excipient material properties are more significant than API properties in their influence on tablet weight. In addition it can be seen that API material properties affect tablet concentration and content uniformity, but not to the same extent that feeder parameters do. Finally it is clear that agitator rotation rate in the blender can affect process throughput and holdup in various unit operations, but that it does not significantly influence tablet quality attributes.

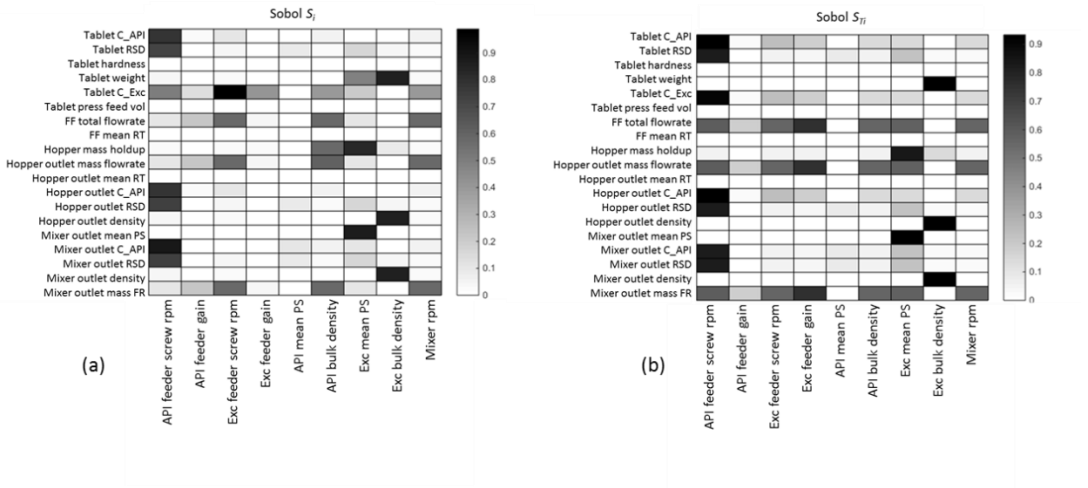

Figure 20: Sobol sensitivity indices for the continuous direct compaction process including a) individual sensitivity metrics S_i and b) total sensitivity metrics S_{Ti}

The interaction indices (S_{ij}) largely corroborate the findings from the individual and total sensitivity metrics. Intensity plots of interaction indices for process throughput (as indicated by flow rate exiting the blender) and the tablet quality attributes are shown in Figure 21. Tablet hardness is not included as this is affected almost exclusively by compression force so there are no significant parameter interactions to report among the subset of uncertain parameters considered. Based on the interaction indices the most significant source of variability affecting tablet weight is the excipient bulk density, which affects this response almost exclusively. Tablet API concentration and RSD are effected most significantly by the API feeder parameters, though interactions with the excipient feeder parameters also affect API concentration and interactions with excipient material properties influence content uniformity. The process throughput, as indicated by mass flow rate exiting the mixer, is affected by a variety of interactions, including those of the API and excipient feeder parameters with the mixer agitator rotation rate.

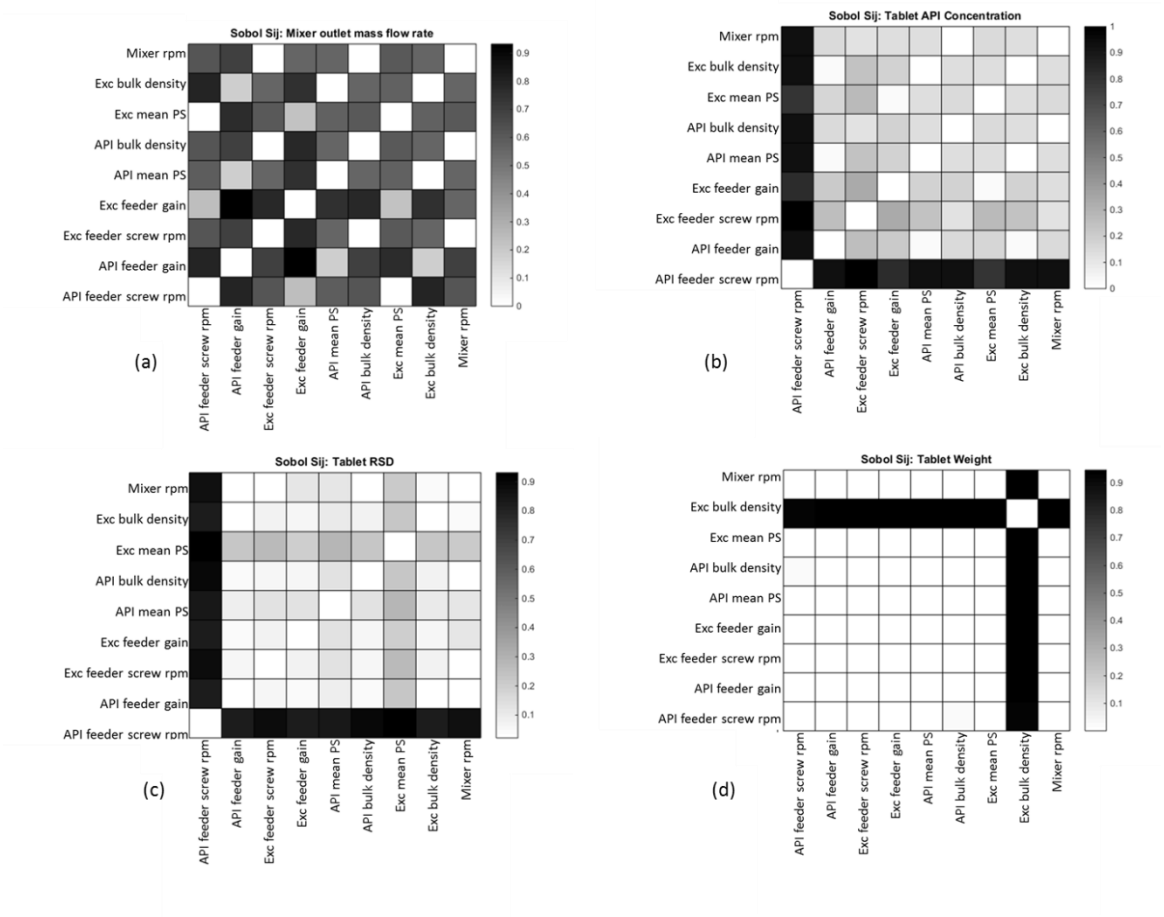

Figure 21: Sobol interaction sensitivity indices S_{ij} for a) mixer outlet mass flow rate b) tablet API concentration c) tablet relative standard deviation (RSD) and d) Tablet weight

Extended Fourier Amplitude Sensitivity Test

Based on the Sobol sensitivity analysis results it appears that the rotation rate of the agitator in the blender affects process throughput, but does not significantly influence tablet quality attributes. The extended FAST method is therefore used to study the eight inputs which were shown to contribute most to variance in tablet quality attributes. These include the API and excipient material properties as well as the feeder parameters (gain and screw speed) for these materials. Sampling was conducted as described in section 4.2, with three search curves per variable

consisting of 65 samples each as per recommendations by Saltelli et al. (Saltelli et al., 1999) This results in a total of 1560 samples.

Intensity plots for the sensitivity (S_i) and total sensitivity (S_{Ti}) indices obtained using extended FAST are given in Figure 22. The total sensitivity indices are generally greater than the individual sensitivity indices and the sum of individual sensitivity indices is less than one. This indicates the presence of higher order interactions between input parameters that contribute to output variability, which is consistent with the findings from Sobol's method. The two metrics S_i and S_{Ti} rank the relative importance of input factors similarly although their magnitudes differ.

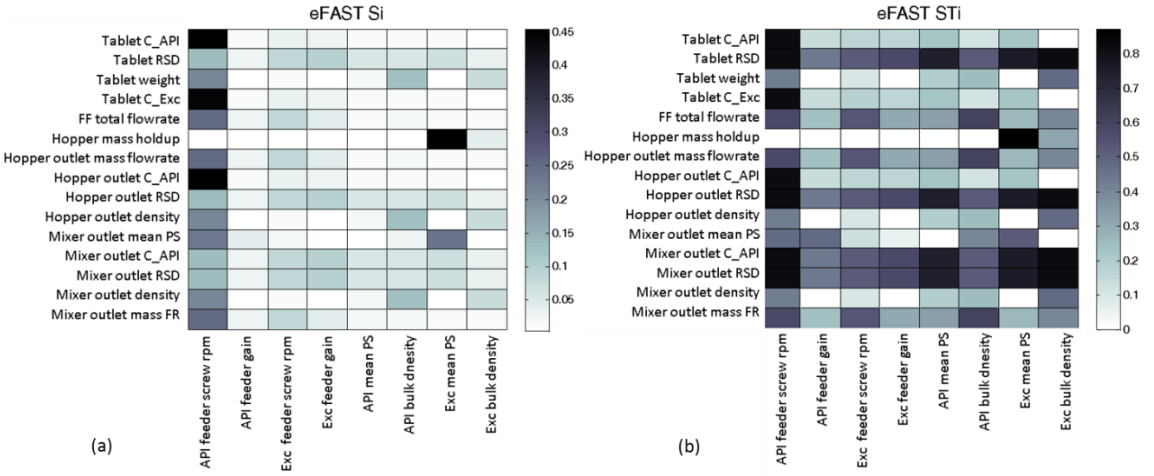

Figure 22: Sensitivity indices obtained from the extended FAST method including a) individual sensitivity metrics S_i and b) total sensitivity metrics S_{Ti}

The extended FAST results can be used to further distinguish between the set of variables that affect product quality. The total sensitivity indices obtained from eFAST highlight the API feeder screw speed as the most significant factor affecting process performance. However the effect of excipient properties on content uniformity, as indicated by RSD is also apparent based on the eFAST S_{Ti}. The total sensitivity indices obtained from eFAST also highlight sensitivities that were not immediately apparent based on the other metrics. For instance, based on the *PRCC* the hopper API concentration, mean residence time and total mass flow rate appeared independent of material properties. However based on the eFAST total sensitivity indices it is clear that material

properties do contribute to variability in mass flow rate exiting the mixer as well as hopper mean residence time and mass hold up. These results are more consistent with expectations, as the properties of the API and excipient, particularly bulk density, should affect total throughput and mass holdup. This highlights again the importance considering variance-based sensitivity metrics like those obtained from eFAST and Sobol's method.

4.4. Identification of potential critical process parameters and control strategies

The sensitivity metrics obtained in section 4.3 can be used to identify potential critical process parameters, as well as potential manipulated variables for control strategies. Based on the individual sensitivity metrics (e.g. *PRCC,* variance based S_i) it is clear that the feeder screw speed and feeder gain significantly contribute to variability in tablet API concentration. Therefore it is important to include these parameters when developing control strategies for tablet content uniformity. In the case of the feeder gain, this may require that the supervisory control system have access to set the feed factor when the feeder is not operating in gravimetric mode, for instance during hopper refill.(W.E. Engisch & Muzzio, 2010) This also highlights the importance of selecting an appropriate time interval for updating the feed factor when the feeder is operating in gravimetric mode.(W. E. Engisch & Muzzio, 2012) If it is updated too frequently, the feed factor calculations may be affected by high frequency noise in the feed rate. However if it is not updated frequently enough the gravimetric controller may not be able to accurately determine the appropriate setting for the screw speed to ensure the desired mass flow rate. In addition, it may be necessary to smooth the weight versus time data obtained from the load cell prior to calculating the feed factor.(F. Boukouvala et al., 2012)

Based on the total sensitivity indices, it is clear that tablet quality attributes including weight variability and content uniformity (RSD) are significantly influenced not only by the API material properties but also by the material properties of the excipients. In fact, the material properties of these two components interact to give rise to variability in product quality. This indicates that the

material properties of both the API and the excipient have the potential to be critical process parameters. It also suggests that the specifications for the API should be set with an awareness of the anticipated range of excipient properties. In fact it may be beneficial to consider a joint specification that accounts for the interaction between API and excipient properties to expand the design space for the process. (Duchesne & Macgregor, 2004; García-Muñoz, 2009) Based on the sensitivity results it can also be concluded that certain parameters are related predominantly to process throughput rather than product quality. For instance, the tablet press hopper geometry, feed frame rotation rate and blender agitator rotation rate contribute to variability in total process throughput and holdup in the various unit operations more so than to tablet quality. This does not imply that these parameters do not affect tablet quality at all, just that within the anticipated ranges for this process their effect on process throughput is more pronounced. As long as these variables can be maintained within the proposed ranges, it should not be necessary to consider them critical process parameters. It may still be beneficial to incorporate operating parameters like the blender rotation rate into a control strategy for tablet weight, as sensitivity indices for this parameter indicate some contribution to weight variability.

The sensitivity results also highlight potential areas for further model development. For instance, the sensitivity metrics indicate that lubricant feeder parameters and properties do not contribute to variability in tablet uniformity or hardness. In practice it is known that lubricant content can influence blending performance, tablet weight variability and tablet hardness.(Morin & Briens, 2013; Wang et al., 2010) The current unit operation models may need to be enhanced to incorporate more detailed information regarding the influence of lubricant content. This could be accomplished through data-based models like those described in Chapter 3. For instance, the maximum hardness parameter in the tablet hardness equation (12) could be modeled as an explicit function of lubricant content using a latent variable or response surface technique.

This study also highlights the importance of considering multiple sensitivity metrics when evaluating a large number of potential process parameters. Regression based metrics like the

PRCC are useful for screening, but can fail to adequately account for parameters that influence process performance through interaction with other variables. For example in this case the *PRCC* indicates almost no sensitivity to agitator rotation rate in the blender. However it is clear from interaction metrics and total sensitivity indices obtained from RS-HDMR and Sobol's method that this parameter can in fact influence process throughput. In addition, eFAST total sensitivity metrics indicate that material properties affect process throughput and mass holdup in various unit operations in a way that is not immediately apparent from screening methods like *PRCC*. By considering multiple sensitivity metrics in this work it is possible to have a more comprehensive understanding of input parameters and parameter interactions that significantly affect process outcomes.

Table 9: Uncertain inputs and their parameter distributions for global sensitivity analysis of the continuous direct compaction process

	Uncertain Input	Distribution	Mean Value	Bounds /SD	Units
1	API feeder screw rpm	Uniform		[0.01,1.5]	rpm
2	API feeder gain	Uniform		[5,7]	-
3	API feeder time constant	Uniform		[19,21]	-
4	API feeder time delay factor	Uniform		[19,21]	rpm
5	Excipient feeder screw rpm	Uniform		[1,2.2]	rpm
6	Excipient feeder gain	Uniform		[5,7]	-
7	Excipient feeder time constant	Uniform		[19,21]	-
8	Excipient feeder time delay factor	Uniform		[19,21]	-
9	Lubricant feeder screw rpm	Uniform		[0.0162,0.0172]	rpm
10	Lubricant feeder gain	Uniform		[5,7]	-
11	Lubricant feeder time constant	Uniform		[19,21]	-
12	Lubricant feeder time delay factor	Uniform		[19,21]	-
13	API mean particle size	Normal	3.00E-05	3.00E-06	m
14	API bulk density	Normal	600	60	g/cm3
15	Excipient mean particle size	Normal	2.00E-04	2.00E-05	m
16	Excipient bulk density	Normal	325	32.5	g/cm3
17	Lubricant mean particle size	Normal	2.00E-05	2.00E-06	m
18	Lubricant bulk density	Normal	160	16	g/cm3
19	Mixer rpm	Uniform		[5,15]	rpm
20	Hopper aperture diameter	Uniform		[0.065, 0.14]	m
21	Tablet press compression force	Uniform		[1415, 1930]	KPa
22	Feed frame rotation rate	Uniform		[0.66,0.67]	-

Table 10: Coefficients of determination for first and second order HDMR models

Process response	1st order R^2	2nd order R^2
Mixer outlet mass flow rate	86.91	83.99
Mixer outlet density	91.85	68.12
Mixer outlet RSD	75.37	75.37
Mixer outlet API concentration	75.36	75.36
Mixer outlet mean particle size	64.84	64.84
Hopper outlet density	91.85	68.12
Hopper outlet RSD	75.37	75.37
Hopper outlet API concentration	95.58	84.43
Hopper outlet mass flow rate	86.69	84.44
Hopper mass holdup	82.9	68.94
Feed frame mean residence time	0	0
feed frame total flow rate	86.89	84.44
Tablet press feed volume	0	0
Tablet excipient concentration	94.93	85.14
Tablet weight	91.85	68.12
Tablet hardness	99.66	55.23
Tablet RSD	75.37	75.37
Tablet API concentration	95.58	84.43

Chapter 5

5. Feasibility analysis of pharmaceutical processes

5.1. Feasibility analysis and the concept of design space

The concept of process feasibility can be useful in determining the design space for a pharmaceutical process. Feasibility describes the ability of a process to satisfy all relevant operating and production constraints in the presence of uncertainty. Feasibility analysis is concerned with identifying the range of conditions within which it a process can operate while satisfying these constraints. In other words, the goal of feasibility analysis is to identify the feasible operating region for a process. This feasible region is expressed in terms of the range of values for the uncertain parameters that the process can tolerate while maintaining feasible operation. (I. E. Grossmann et al., 2014) It should be noted that since feasibility analysis is conducted using a process model, uncertainty can also arise from plant-model mismatch. (Lima et al., 2010)

For pharmaceutical manufacturing processes, it is imperative that product critical quality attributes are within their specified limits in order to ensure regulatory compliance and ultimately patient safety. (L. X. Yu, 2008) Therefore the feasible operating region for a pharmaceutical process may be considered the region within which all constraints on product quality are satisfied. This region is defined with respect to sources of uncertainty in the process, such as variation in input material quality or disturbances in operating conditions. (Lima et al., 2010) The feasible operating region for a pharmaceutical manufacturing process is thus comparable to the design space, which is by definition the region within which material of acceptable quality can be produced. Therefore feasibility analysis can be applied to a pharmaceutical process to identify a region within which the process design space can be established.(F. Boukouvala et al., 2010a; Leopore & Spavins, 2008) It is important to realize that the boundaries of the feasible region correspond to the edges of failure for a process. This information can be used to guide the

selection of design space to ensure that it lies well within the feasible operating region. It is not recommended that the boundaries of the feasible operating region be established as the design space for a process.

5.2. Mathematical formulation of the feasibility test problem

Rigorous mathematical formulations for process feasibility analysis were introduced in the 1980s in the process systems engineering literature. (I. E. Grossmann, Floudas, C.A., 1987; Morari, 1983; Swaney & Grossmann, 1985a, 1985b) Mathematically, feasibility is defined according to the feasibility function $\Psi(d,\theta)$ shown in equation (50). In equation (50), d represents a given process design and z indicates a vector of control variables. θ represents a vector of uncertain parameters that vary on the interval $T = \left\{\theta \mid \theta^L \le \theta \le \theta^U\right\}$ where θ^L and θ^U are lower and upper bounds respectively. These upper and lower bounds must be selected based on prior knowledge about the anticipated level of variability in the process. A set of state variables, x, describing intermediate process variables such as material flow rates, stream compositions and mass holdup in equipment are implicit equation (50). These variables are explicit functions of the design, control and uncertain variables and as such do not directly influence the feasibility of the process. The state variables are therefore expressed in terms of the relevant design, control and operating variables as $x=x(d,z,\theta)$. With this substitution the feasibility problem can be expressed in terms of design, control and uncertain variables alone, as in equation (50). The functions g_j are constraints on the process of the form $g_j(d,z,\theta)\le 0$. These must be satisfied in order for the process to operate successfully. (Biegler et al., 1997; Swaney & Grossmann, 1985a, 1985b) A set of equality constraints $h_i(d,z,\theta)=0$ that constitute a model describing the process physics must also be satisfied.

$$\begin{aligned} &\Psi(d,\theta) = \min_{z} \max_{j\in J} \left\{ g_j(d,z,\theta) \right\} \\ &s.t.\ h_i\left(d,z,\theta\right) = 0 \ \forall i \in I \\ &\theta \in T = \left\{\theta \mid \theta^L \le \theta \le \theta^U\right\} \end{aligned} \tag{50}$$

The feasibility function $\Psi(d,\theta)$ will be less than or equal to 0 for a given design d if, for that design, it is possible to adjust the control variables z in such a way that the constraints g_j are satisfied for all $\theta \in T$. If $\Psi(d,\theta) \leq 0$ then the process is feasible for the given design over the range of uncertain parameters considered. If not, then the process is infeasible and cannot be made feasible through adjustment of the control variables. Figure 23 depicts the concept of feasibility for a process that has two uncertain variables (θ_1 and θ_2) and four constraints. The boundaries of the feasible region are shown as the region where the feasibility function $\Psi(d,\theta)$ is equal to 0. Figure 23a shows a process design that is feasible, because the feasibility function is less than or equal to 0 over the entire uncertain space T. Figure 23b depicts an infeasible process design. The range of the uncertain variables T exceeds the boundaries of the feasible region in this case.

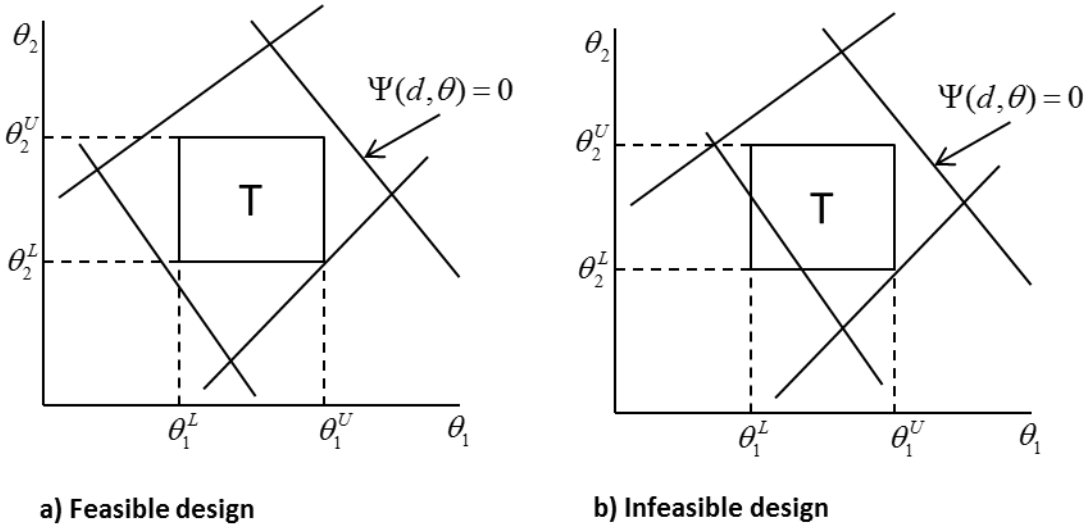

Figure 23: Geometrical representation of the concept of feasibility. a) depicts a process design that is feasible. b) depicts a design that is infeasible.

Problem (50) can be reformulated through the introduction of a scalar variable u, as shown in equation (51). This reformulation allows the problem to be solved using standard linear (LP) or nonlinear (NLP) programming solvers, depending on the nature of the constraints. (Biegler et al.,

1997; I. E. Grossmann et al., 2014) In this formulation a process is feasible for the given design *d* if *u* is less than or equal to 0.

$$\begin{aligned} &\Psi(d,\theta) = \min_{z,u} u \\ &s.t.\ \ g_j(d,z,\theta) \le u \quad \forall j \in J \\ &\ \ h_i(d,z,\theta) = 0 \quad \forall i \in I \\ &\theta \in T = \left\{ \theta \mid \theta^L \le \theta \le \theta^U \right\} \end{aligned} \tag{51}$$

5.2.1. Dynamic feasibility analysis

The concept of feasibility analysis can be extended to dynamic processes, wherein the process model incorporates differential equations with respect to time. In such processes it is possible for the state variables, the uncertain parameters and the control variables to change over time, thereby changing the feasible region. Initially formulations for dynamic feasibility analysis were very limited in terms of the type of dynamic behavior that could be evaluated (Holt & Morari, 1985a, 1985b; Skogestad & Morari, 1987). Then in the 1990s Dimitriadis and Pistikopoulos (Dimitriadis & Pistikopoulos, 1995) introduced formulations for the dynamic feasibility problem based on the steady state formulation of Swaney and Grossmann (Swaney & Grossmann, 1985a, 1985b). This approach can be used to determine process feasibility given time-varying uncertainty for processes described by differential and algebraic equations.

The concept of dynamic feasibility is illustrated in Figure 24, which shows the changing shape and size of the feasible region for a dynamic process over time. Figure 24a depicts the feasible region at some initial time $t=t_1$. Figure 24b illustrates the feasible region at some later point in time $t=t_2$. The value of one of the process constraints shifts during the time elapsed between t_1 and t_2 and as a result, the feasible operating region at $t=t_2$ differs from that at $t=t_1$.

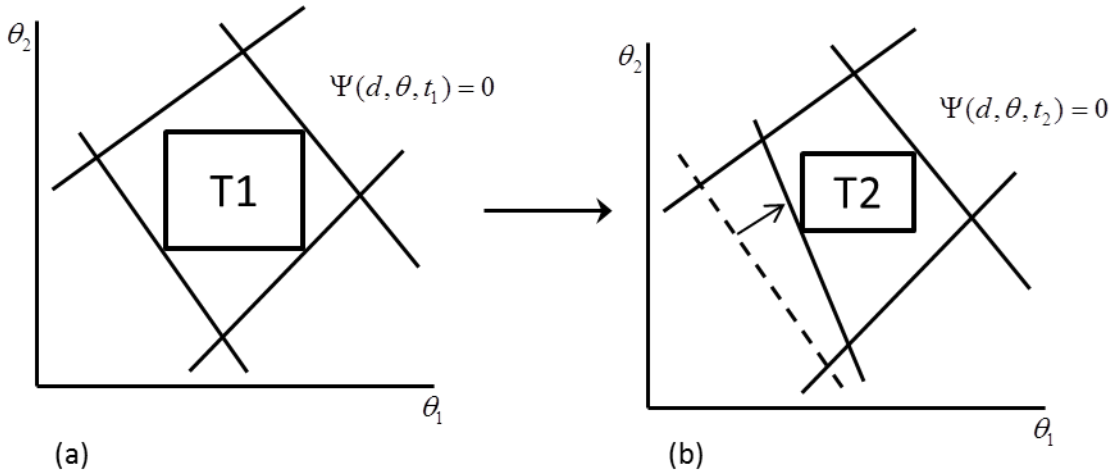

Figure 24: Geometrical representation of dynamic feasibility. a) shows the feasible region at some initial time $t=t_1$. b) shows the feasible region for the same process at some later point in time $t=t_2$. Between $t=t_1$ and $t=t_2$, one of the constraints shifts from its initial position along the dashed line to its final position, shown as a solid line in b.

In order to describe the mathematical formulation for the dynamic feasibility problem it is first important to describe two different types of constraints that may be imposed on a dynamic process. The first are path constraints, shown in equation (52), which must be satisfied at all times during process operation. The second are point constraints, shown in equation (53), which only need to be satisfied at specific time points during the operation. Common examples of point constraints are those imposed at the beginning or end of a batch process, like conversion at the end of a batch reaction. (Dimitriadis & Pistikopoulos, 1995) For the continuous processes discussed in this work we will be concerned primarily with path constraints.

$$g^{path}(d, x(t), z(t), \theta(t), t) \leq 0 \tag{52}$$

$$g_k^{point}(d, x(t^k), z(t^k), \theta(t^k), t^k) \leq 0 \quad k = 1, 2, ..., NP \tag{53}$$

In order to solve the dynamic feasibility problem it is also necessary to establish a time horizon (H) over which the process must remain feasible. Then the dynamic feasibility analysis problem can be formulated as shown in equation (54). This is comparable to the steady state formulation in (51), except that in this case the process model and the corresponding constraints are

differential functions of time. Note also that in this case the state variables are not expressed in terms of the design, control and uncertain parameters as it may not be straightforward to solve for them explicitly.

$$
\begin{aligned}
&\Psi(\mathrm{d},\theta(\mathrm{t})) = \min_{u,z(t)\in Z(t),t\in[0,\mathrm{H}]} u \\
&s.t.\quad h_i(d,\dot{x}(t),x(t),z(t),\theta(t),t)=0;\quad x(0)=x^0\ \ \forall i\in I \\
&g_j(d,x(t),z(t),\theta(t),t)\le u\ \ \forall j\in J \\
&T(t)=\{\theta(\mathrm{t})\,|\,\theta^L(\mathrm{t})\le\theta(\mathrm{t})\le\theta^U(\mathrm{t})\} \\
&Z(t)=\{z(\mathrm{t})\,|\,z^L(\mathrm{t})\le z(\mathrm{t})\le z^U(\mathrm{t})\}
\end{aligned}
\qquad (54)
$$

If $u\le 0$ in equation (54) then there exists at least one dynamic control strategy characterized by $z(t)$ that can satisfy the path and point constraints $g_j(d,x(t),z(t),\theta(t),t)$ for all possible profiles of the uncertain variables $\theta(t)$ within the time horizon $t\in[0,H]$.

The solution of problem (54) is much less straightforward than that of problem (51) due to the presence of differential equations. These can be addressed in a number of ways. One option is to embed a differential equation solver into the optimization algorithm. The differential equations can then be solved for each control profile and/or set of decision variables selected by the optimizer. This approach can result in long solution times because it requires the solution of all differential equations at each iteration. An alternative approach is to convert some or all of the differential equations into algebraic equations. This can be accomplished via control parameterization or full discretization. In control parameterization the vector of controls is partitioned along the time horizon (H) into a subset of control intervals that may vary in length using a set of finite parameters. (Vassiliadis et al., 1994a, 1994b) The resulting problem still contains both differential and algebraic equations since only the control variables are discretized. Alternatively, all of the equations in the dynamic feasibility problem can discretized with respect to time. This approach is known as direct transcription and results in a system of algebraic equations only. Both algorithms generate nonlinear programming (NLP) problems that can then

be solved using commercially available algorithms which may or may not be coupled with differential equation solvers. (Biegler, 2007)

5.3. Surrogate-based feasibility analysis

The feasibility and dynamic feasibility problems (51) and (54) can be formulated in such a way that they are able to be solved using commercially available deterministic optimization solvers. However these solvers have certain limitations which make them less than ideal for solving the types of problems frequently encountered in pharmaceutical process development. For instance, they require explicit, differentiable functions for all off the process constraints ($g_j(d,z,\theta)$). In pharmaceutical applications it is often necessary to use reduced-order or black-box models like those discussed in Chapter 3, which may not satisfy these requirements. In addition, deterministic solvers sometimes experience difficulty when dealing with large or complex optimization problems. (Biegler et al., 1997) Flowsheet models of pharmaceutical processes may become quite large, containing thousands of differential, algebraic and integral expressions.(F. Boukouvala, Chaudhury, et al., 2013; F. Boukouvala et al., 2012) Surrogate-based optimization methods can be employed to address some of these challenges. In the context of feasibility analysis, surrogate-based methods have the added benefit that they do not require inferences regarding the convexity of the feasible region. (F. Boukouvala & M. Ierapetritou, 2012; F. Boukouvala & M. G. Ierapetritou, 2012)

As their name implies, surrogate-based optimization methods involve building surrogate representations for the function to be optimized. These representations tend to be lower in computational cost than the complex process models they approximate. The optimization problem can then be solved using the surrogate model rather than the original complex process model. The surrogate-based feasibility problem can formulated as shown in equation (55), where g_k is a set of black-box constraints whose values are obtained from a surrogate model. In this

work, all constraints for a process will be represented using a single surrogate function, following the method of Boukouvala and Ierapetritou (F. Boukouvala & Ierapetritou, 2014).

$$\begin{aligned} &\Psi(d,\theta) = \min_{z,u} u \\ &s.t.\ \ g_k(d,z,\theta) \le u \\ &\theta \in T = \left\{ \theta \mid \theta^L \le \theta \le \theta^U \right\} \end{aligned} \tag{55}$$

Surrogate-based feasibility analysis involves sampling the uncertain space, evaluating the process model to determine the feasibility function value for each realization of the uncertain parameters, building a response surface for the feasibility function, and ultimately using this to identify the boundaries of the feasible region. Algorithms for surrogate-based feasibility analysis tend to be iterative in nature, with the surrogate representation being updated until it is sufficiently accurate to be used for feasibility analysis. A general overview of the iterative procedure for surrogate-based feasibility analysis is provided in Figure 25.

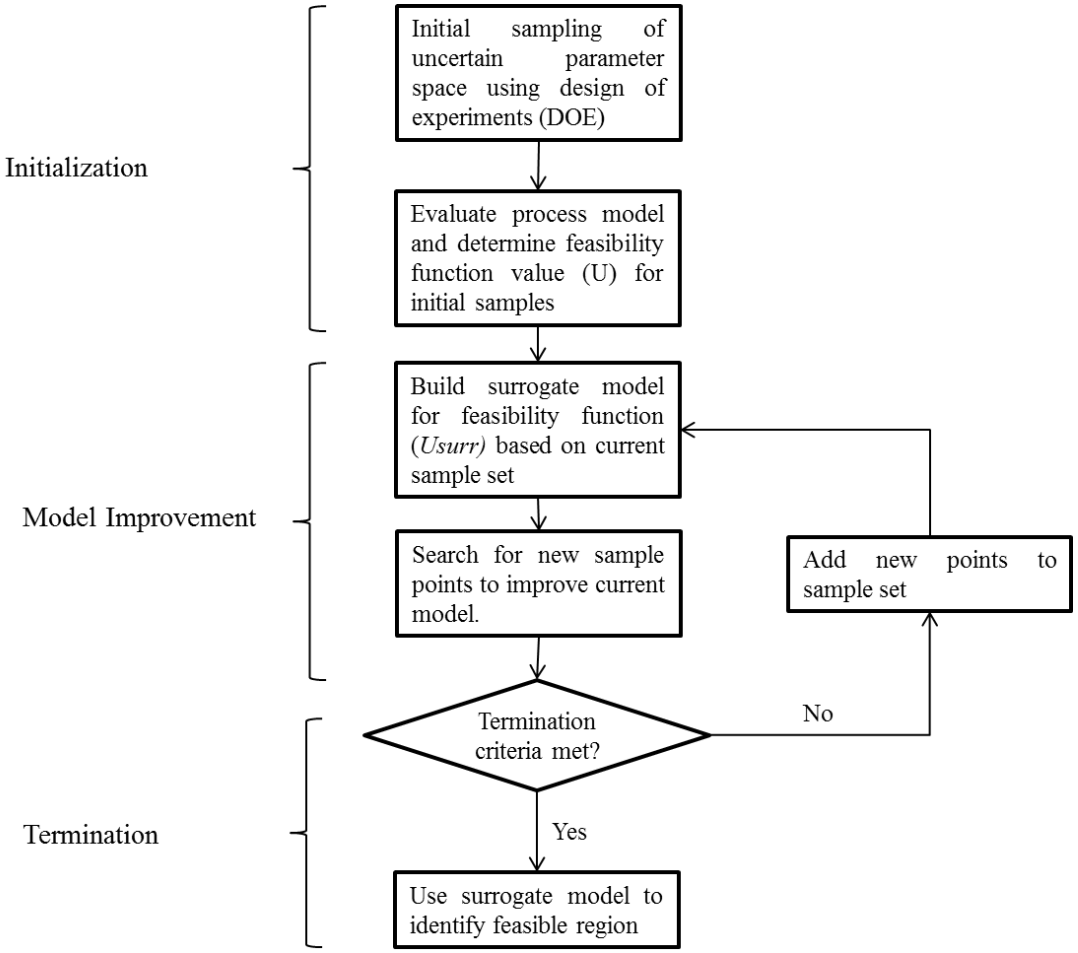

Figure 25: General overview of an iterative algorithm used for surrogate-based feasibility analysis

5.3.1. A novel method for surrogate-based feasibility analysis

Many of the challenges associated with implementing surrogate-based optimization methods arise in the development of the surrogate model. It is important to select a surrogate representation that can accurately represent the features of the objective function (e.g. smoothness, nonlinearity), or in the case of feasibility analysis the feasible region, which may not be known *ab initio*. It is also necessary to balance model accuracy requirements with the number of samples used to develop the surrogate model.(Jones et al., 1998) When evaluating process feasibility based on a surrogate model one must also acknowledge that the surrogate model is not a perfect representation, and it may therefore be desirable to explicitly account for prediction uncertainty when evaluating the feasible region. In this section a novel algorithm for surrogate-based feasibility that addresses

each of these challenges is presented. This method is then demonstrated using a series of case studies to demonstrate its ability to deal with nonconvex, nonlinear and disjoint problems. Finally it is applied to a pharmaceutical roller compaction process.

Selecting a surrogate model

Surrogate functions can be developed using any of the response surface techniques described in Chapter 3. Both High Dimensional Model Representation(I. Banerjee, Ierapetritou, M.G., 2002; I. Banerjee, Pal, S., Maiti, S., 2010) and kriging (F. Boukouvala & M. G. Ierapetritou, 2012) have previously been demonstrated in the literature as surrogate methods for black-box feasibility applications. For this work kriging is selected due to its demonstrated ability to model pharmaceutical processes effectively (F. Boukouvala et al., 2010b; Jia et al., 2009) and because the estimate of prediction variance provided by kriging is a useful feature for adaptive sampling strategies in surrogate-based optimization.(F. Boukouvala & M. G. Ierapetritou, 2012; Jones, 2001) When developing a kriging model, one can select from a variety of regression and correlation models, including but not limited to those enumerated in Table 2. In the presented algorithm, an initial model selection phase is implemented to find a combination of correlation and regression models that minimize prediction variance for the surrogate feasibility function. The model selection process is conducted using the samples collected in the initial design of experiments (DOE) step of the surrogate-based feasibility algorithm. These samples are used to build a model using each combination of regression and correlation models. The model combination providing the best prediction accuracy is selected for the surrogate representation. Because the initial number of samples is limited, model building is quick. The model selection phase was completed in less than 1 cpu-second for the case studies presented in this work.

Balancing model accuracy and sampling requirements

Adaptive sampling strategies can be employed to develop accurate surrogate representations without exhaustively sampling the original process model. (F. Boukouvala & M. G. Ierapetritou, 2012) After building an initial response surface based on a limited number of function

evaluations, additional samples can be added iteratively. The goal of an adaptive sampling strategy is to select these samples in such a way that they improve the current surrogate representation to the greatest extent possible. A variety of search methods to direct additional sampling have been reviewed by Jones (Jones, 2001). In this work, the modified expected improvement function introduced by Boukouvala and Ierapetritou (F. Boukouvala & M. G. Ierapetritou, 2012) is used. This criterion is specifically intended to identify samples that will improve the current representation of the feasible region boundary. The modified expected improvement function is shown in equation (56), where U represents the value of the feasibility function, s_u is the variance of the kriging predictor and φ is the standard normal distribution function. Thus equation (56) can be used to determine the probability that the value of the feasibility function evaluated at a particular point in the uncertain space is equal to 0 given the uncertainty associated with the feasibility prediction. By maximizing equation (56) over the uncertain variables θ it is possible to locate additional samples along the boundary of the feasible region.

$$E_{feas}[I(\theta)] = E_{feas}\{\max(0-U),0\} = -s_u\varphi\left(\frac{-U}{s_u}\right) = \frac{1}{\sqrt{2\pi}}\exp\left(-0.5\frac{U^2}{s_u}\right) \tag{56}$$

Maximizing the expected improvement (EI) function is a nontrivial problem, as this function is asymptotic and highly nonlinear. To illustrate these features, Figure 26 shows the feasible region and expected improvement values for the Branin function, a test problem for global optimization which has 3 disjoint feasible regions. (Jones, 2001; Sasena et al., 2002) The boundaries of these regions are indicated by the white contours in Figure 26. The expected improvement function is asymptotically 0 at a sufficiently large distance from the boundary of the feasible region. In the area surrounding the feasible region the expected improvement function is highly nonlinear. This nonlinearity arises from the fact that the prediction variance depends on the proximity of previously collected samples. Maximizing the expected improvement is also challenging because the quantities U and s_u are obtained from a black-box model. As such, analytical derivatives for

these functions are not available, making it difficult to use direct-search methods for optimization. In prior work, a global multi-start black-box optimization solver (TOMLAB/ LGO) has been used to evaluate the expected improvement function. This global search can slow down the model improvement phase of the surrogate-based feasibility algorithm. In addition, it requires a global black-box solver to be embedded within the black-box feasibility algorithm.

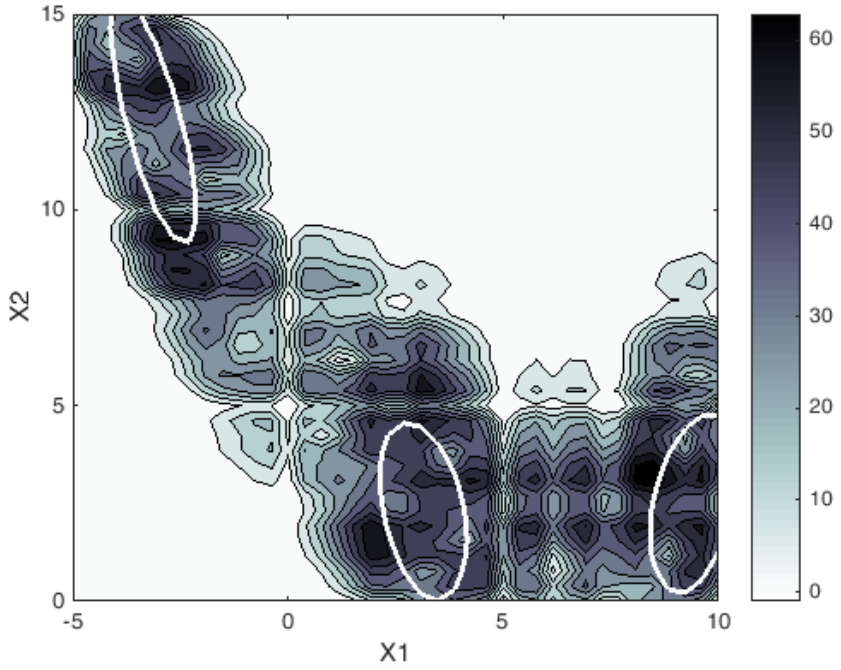

Figure 26: Contours of the expected improvement for the Branin function(Jones, 2001), which has 3 disjoint feasible regions. The boundaries of the feasible region are shown in white. The magnitude of the expected improvement function is indicated by the contour colors.

In this work, an alternative method for evaluating the expected improvement function is proposed that can reduce computational time and does not require a global black-box solver. This method involves first identifying a set of candidate points that have promising expected improvement values. Candidate points are evaluated by sampling the uncertain space using a uniform grid, and identifying points with an expected improvement value greater than a user specified tolerance. These candidate points are then clustered using k-means clustering. The cluster centers are used as starting points for a local optimization algorithm, in this case an interior-point method, which

identifies locally optimal values of the expected improvement function. If two local optima are less than some distance δ from one another or from an existing design site, then only the one of the two points is retained in the sample set. This is done to limit the number of samples added at each iteration of the algorithm. It is particularly important to limit sampling when the original process model is very computationally expensive to evaluate. (Müller et al., 2013) The proposed approach to evaluating the expected improvement function is illustrated visually in Figure 27, which depicts the process of identifying new sample points for the Branin function. The set of candidate points, shown in black, have expected improvement values greater than a tolerance of $1x10^{-5}$. The cluster centers after k-means clustering are shown in blue, and the locally optimal points to which they converge are shown in red. These points tend to lie on the boundaries of the feasible operating region, which have high expected improvement values as shown in Figure 26.

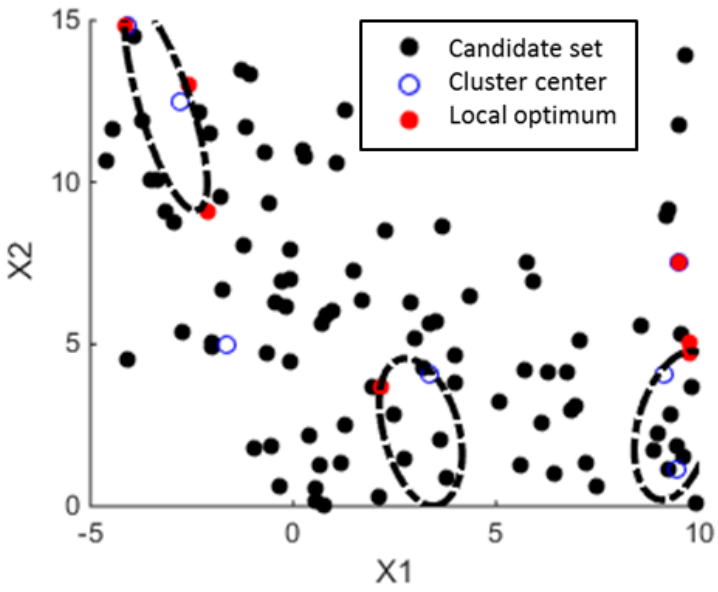

Figure 27: Visual representation of the local solution method for the expected improvement function applied to the Branin function. Boundaries of the feasible region are depicted by dashed black lines. The candidate points are shown in black. These are clustered using k-means clustering to obtain the points shown in blue. Cluster centers are used as initial conditions for a local optimization algorithm to obtain the local optima in red.

Remarks on model accuracy

It is important to recall that when implementing a surrogate-based approach for feasibility analysis, the surrogate model will have some degree of prediction error relative to the original process model. In the context of using feasibility analysis for design space evaluation, it is preferable for the surrogate model to be conservative relative to the original process model. In other words, it is better for the surrogate model to predict that the feasible region is slightly smaller than it actually is than to predict that it is larger than it actually is. The latter, not conservative, situation could result in setting specifications on the uncertain parameters that may result in unacceptable product quality. One way to ensure that the predicted feasible region is conservative, is to explicitly consider prediction uncertainty for the surrogate feasibility model. It is straightforward to account for prediction error when using a kriging response surface, as an estimate of the variance for the kriging predictor is provided every time the model is evaluated.

In this work two ways of accounting for model prediction error are considered. The first method, hereafter referred to as the probabilistic approach, evaluates the cumulative probability distribution function for each predicted value of the feasibility, as shown in equation (57). *P(U)* indicates the probability that the predicted feasibility function value *U* lies between negative infinity and 0 given the prediction variance s_u. It can be seen from equation (57) that *P(U)* depends exponentially on the negative inverse of the prediction variance. Thus as the prediction variance increases the probability decreases rapidly. Recall that the prediction variance for a kriging predictor depends on the number of nearby design sites. In the case of surrogate-based feasibility analysis, many samples are collected near the boundary of the feasible region, where $U=0$. However considerably fewer samples are collected well within the feasible region, where $U<<0$. Therefore it is possible for *P(U)* to evaluate to *0* at locations well within the feasible region due to less dense sampling in this area. For this reason, the approach in equation (57) is recommended only when seeking to identify the boundary of the feasible region. In this context, a

criterion can be specified for *P(U)* and the contours of the cumulative distribution function equal to this criterion can be selected as the boundary of the feasible region.

$$P(\mathrm{U}) = \Phi(0, \mathrm{U}, \sqrt{s_u}) = \frac{1}{\sqrt{s_u 2\pi}} \int_{-\infty}^{0} \exp\left(\frac{-(t-U)^2}{2 s_u} \right) dt \tag{57}$$

An alternative approach to considering model prediction error is to simply adjust the feasibility prediction by the variance, as shown in equation (58). This is hereafter referred to as the adjusted feasibility approach. The adjusted feasibility $\hat{U}$ is more conservative than the predicted feasibility *U* by an amount equal to the square root of the prediction variance. This is similar to the approach of dealing with "possible infeasibility" introduced by Caballero and Grossmann (Caballero & Grossmann, 2008). However in their work the kriging prediction error was used to make the predicted feasible region less conservative by permitting any constraint violation with a magnitude lower than the error in the kriging predictor.

$$\hat{U} = U + \sqrt{s_u} \tag{58}$$

Unlike the probabilistic feasibility metric, the adjusted feasibility does not vary exponentially as the prediction variance. For this reason it tends to work well even for points well within the boundaries of the feasible region. The adjusted feasibility metric can therefore be used to identify points as feasible or infeasible throughout the uncertain space $T = \left\{ \theta \mid \theta^L \le \theta \le \theta^U \right\}$.

The approaches described in equations (57) and (58) will both be demonstrated in section 0, where it is shown that both methods provide a conservative estimate of the feasible region relative to the original process model. Depending on the criterion selected for *P(U)*, the latter approach may be either more conservative or less conservative than the former approach when identifying the boundaries of the feasible region. For the purpose of classifying samples randomly drawn from the entire uncertain space *T* as feasible or infeasible, the adjusted flexibility metric is recommended, as the probabilistic metric is most accurate in densely sampled regions.

Surrogate-based Feasibility Analysis Algorithm

A novel algorithm for surrogate-based feasibility analysis is now presented. This algorithm takes advantage of the features described above to generate an accurate surrogate model in a computationally efficient fashion. This method explicitly accounts for prediction uncertainty in evaluating the feasible region.

The algorithm is summarized in Figure 28. Initially, the uncertain space is sampled based on a design of experiments (DOE). For the two dimensional case studies presented in this work, samples are collected from a uniform grid across the uncertain space with 7 samples in each dimension. This results in 49 initial samples. During the model selection step, the combination of regression and correlation model that provides the lowest prediction variance is selected. The regression and correlation models considered are summarized in Table 2. During the model improvement phase, a local optimization method is used to search for new sample points. The expected improvement function is evaluated for a uniform grid across the uncertain space consisting of 100 samples (10 in each dimension). Points with expected improvement values of greater than $1x10^{-5}$ are considered part of the candidate set. Starting points for local optimization of the expected improvement function are identified via k-means clustering. The number of cluster centers is specified to be less than one-tenth the size of the candidate set. The maximum number of clusters is set to 10, so that the number of clusters (N) can be determined as $N=min(floor(Nc/10),10)$, where Nc is the number of points in the candidate set. An upper bound on the number of cluster centers is needed to limit the number of samples added during each iteration of the algorithm. The N cluster centers are then used as initial points to find a set of local maxima for the expected improvement function using an interior-point algorithm. This provides a set of potential new design sites x_i, where $i=1,2,...,N$. In order to avoid adding samples that are very close to existing design sites, and in case multiple cluster centers converge to the same local optimum, any new points that are within a certain Euclidean distance (d_{ij}) of an existing design site are removed from the set. In this case the minimum Euclidean distance between points (δ) is selected to be $1x10^{-4}$. The algorithm model improvement phase will terminate if the maximum

expected improvement value is less than a certain tolerance (ε_1=1), or if the change in expected improvement less than a tolerance of ε_2 =0.1. If all of the new samples identified during the model improvement phase are within distance δ of existing design sites, then the set of new samples is empty and the model improvement phase will terminate. If none of these criteria are met, the new samples identified based on the expected improvement function are added to the existing set of design sites and the algorithm goes through another model-building iteration. Once the termination criteria are met, the final model can be used to evaluate the feasibility function. At this point the probabilistic feasibility metric and the adjusted feasibility metric can be employed to ensure that the predicted feasible region is conservative relative to the true feasible region for the process.

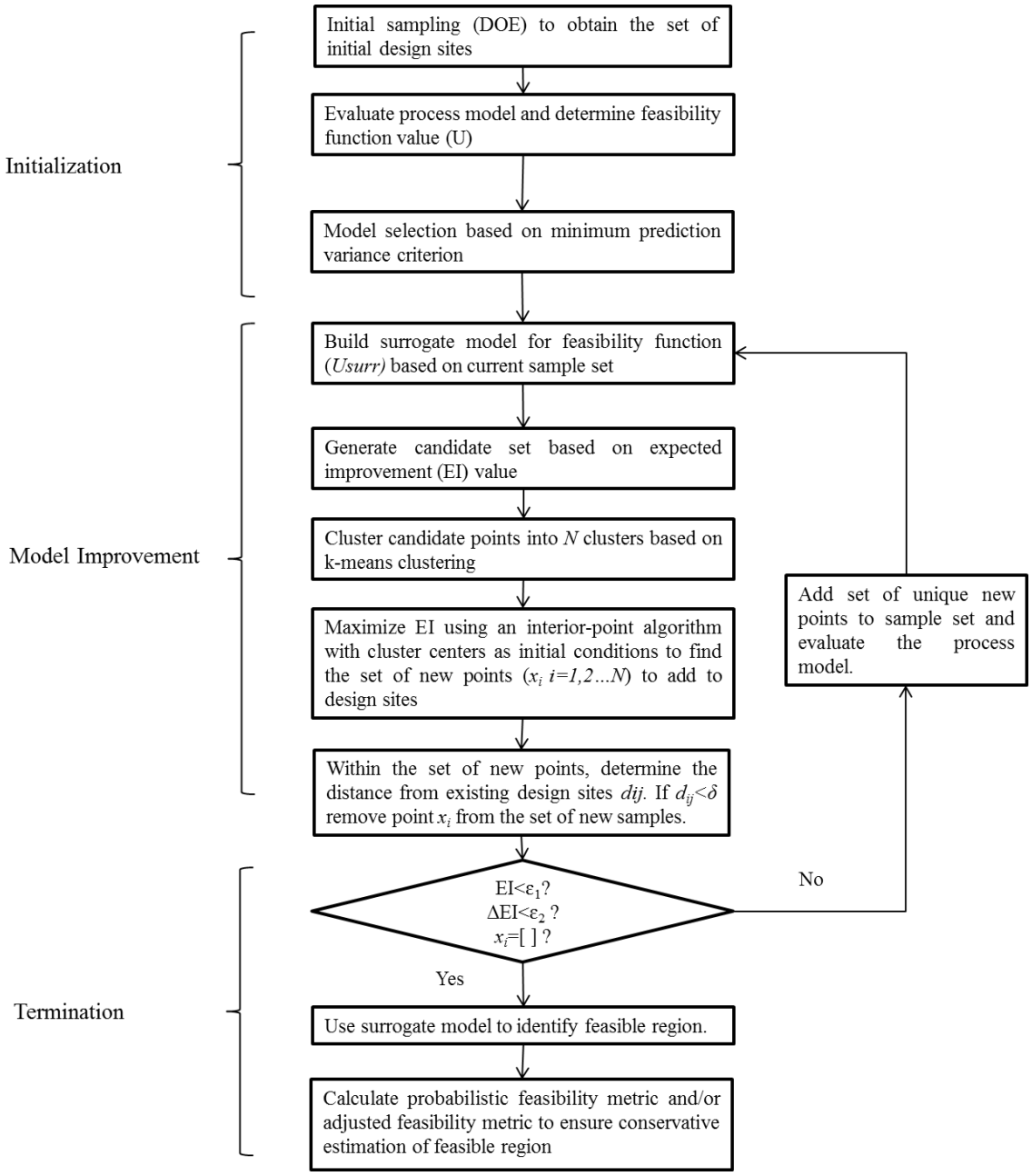

Figure 28: Algorithm for surrogate-based feasibility analysis, using a local search approach to select new sample points in model improvement phase and accounting for model prediction uncertainty when evaluating the feasibility function

Illustrative Case Studies

The surrogate-based feasibility method described in section 5.3.1 is first demonstrated using several benchmark problems from the literature. These include a nonlinear, nonconvex test

problem first introduced by Banarjee and Ierapetritou (I. Banerjee, Ierapetritou, M.G., 2002) and the Branin function, a well-known benchmark problem in optimization. (Jones, 2001; Sasena et al., 2002) In this case the modified Branin function introduced in Sasena et al. (Sasena et al., 2002) is used as a constraint to demonstrate the proposed method on a problem with disjoint feasible regions. Both the local approach described in section 5.3.1 and the global the approach previously described in the literature (F. Boukouvala & Ierapetritou, 2013; F. Boukouvala & Ierapetritou, 2014) are used to evaluate the expected improvement function for these test problems. This is done in order to compare the two methods and show that the local approach can provide comparably accurate surrogate models at reduced computational expense.

Nonlinear Nonconvex Test Problem

The nonlinear, nonconvex test problem includes two uncertain variables (x_1,x_2) and is defined by three constraints, as shown in equation (59). The contours of the feasibility function and the boundaries of the corresponding feasible region are depicted in Figure 29.

$$\begin{aligned} &-2x_1 + x_2 - 15 \le u \\ &\frac{x_1^2}{2} + 4x_1 - x_2 - 5 \le u \\ &\frac{-(x_1 - 4)^2}{5} - \frac{x_2^2}{0.5} + 10 \le u \\ &-10 \le x_1 \le 15 \\ &-15 \le x_2 \le 15 \end{aligned} \tag{59}$$

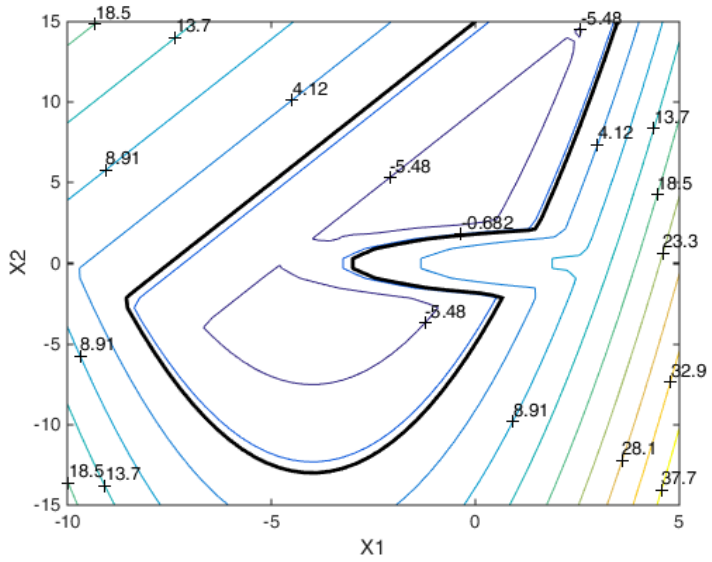

Figure 29: Contours of the feasibility function for the nonlinear nonconvex test problem shown in equation (59). The boundary of the feasible region is indicated by the bold, black contour.

Branin function

The Branin function also includes two uncertain variables (x_1,x_2) and is defined by a single constraint as shown in equation (60). The contours of the feasibility function and the boundaries of the corresponding feasible region are depicted in Figure 30.

$$
\begin{aligned}
&a\left(x_2 - bx_1^2 - cx_1 - d\right)^2 + h(1-ff)\cos x_1 - 5 + h \le u \\
&a = 1; b = \frac{5.1}{4\pi^2}; c = \frac{5}{\pi}; d = 6; h = 10; ff = \frac{1}{8\pi} \\
&-5 \le x_1 \le 10 \\
&0 \le x_2 \le 15
\end{aligned}
\tag{60}
$$

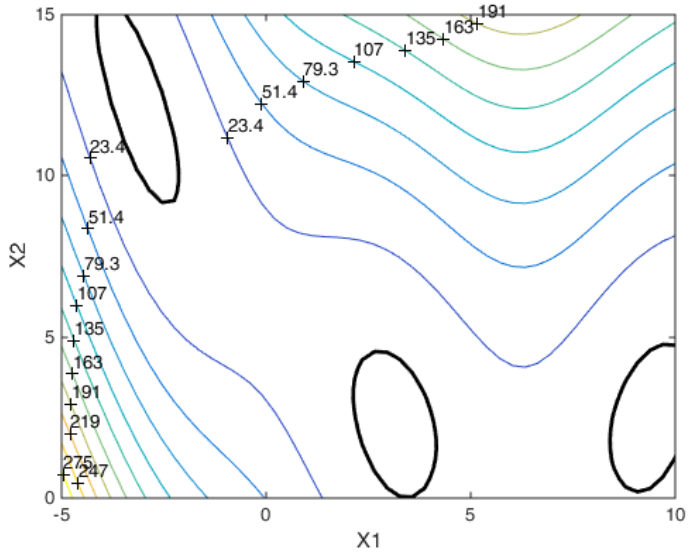

Figure 30: Contours of the feasibility function for the Branin function shown in equation (60). The boundary of the feasible region is indicated by the bold, black contour.

Comparison of local and global solution approaches for the expected improvement function

Table 11 compares the performance of the local and global approaches to evaluating the expected improvement function for the two test problems. In this study, the number of function calls is kept constant for both methods so that any differences in the cpu-time can be attributed to differences in the expected improvement phase alone. In addition to the cpu-time, the expected improvement (EI) value upon algorithm termination and the maximum prediction variance associated with the resulting model are reported. The ability of the model to correctly identify feasible or infeasible regions of the uncertain space is also evaluated over a set of 10,000 validation points. Based on the results reported in

Table 11 it is clear that the local expected improvement method can significantly reduce computational time for the surrogate-based feasibility algorithm. In addition, the resulting models are comparably accurate, as indicated by the prediction variance. The models produced using both the local and global EI methods can correctly characterize validation points as feasible or infeasible and tend to be conservative relative to the true feasible region.

The local and global approaches to evaluating the expected improvement for the Branin function are compared in Figure 31. The true contours of the feasible region are depicted along with the contours obtained using the surrogate model. The samples collected during the expected improvement phase are superimposed on the plot to show their location relative to the boundaries of the feasible region. Although the local approach tends to result in slightly denser clusters of data points along the boundary of the feasible region, the ability of the resulting model to accurately classify validation points as feasible or infeasible is comparable as shown in Table 11.

.

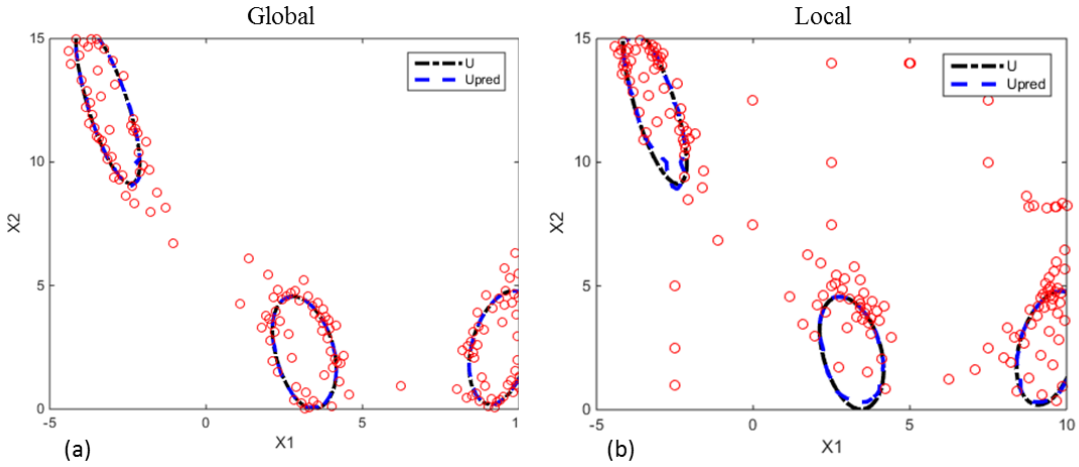

Figure 31: Comparison of a)Global and b)Local approaches to evaluating the expected improvement function for the Branin function (equation (60)). The true contours of the feasible region are shown in black. The contours of the predicted feasible region are shown in blue. The points added during the expected improvement phase are scattered in red.

Surrogate-based estimation of the feasible region

The feasible region identified by the surrogate-based approach agrees well with the analytically obtained feasible region for the two test problems considered. This is demonstrated by Figure 32, which shows the analytically obtained feasible region *(U)* as well as the surrogate-based feasible region *(Upred)* for both test problems. These are nearly identical in both cases. As Table 11 indicates, the surrogate model incorrectly classifies infeasible points as feasible less than one

percent of the time. However, in the case of design space evaluation it is important to ensure that the surrogate-based approach is conservative relative to the original model. In order to do this, the prediction variance is explicitly considered as described in section 5.3.1. By incorporating the prediction variance, the estimated feasible region is conservative relative to true feasible region for both test problems one hundred percent of the time. This is demonstrated by the green and red contours in Figure 32, which demonstrate the probabilistic approach (equation (57)) and the adjusted feasibility approach (equation (58)) of accounting for prediction uncertainty. Depending on the selected criterion for *P(U),* the probabilistic approach may be either more or less conservative than the adjusted feasibility approach. In this work criteria ranging from *P(U)=0.9* to *0.5* were evaluated and it was found that *P(U)=0.75* provided a conservative approximation of the feasible region without cutting off a significant portion of the feasible operating space. These criteria are depicted in Figure 33 for the nonlinear nonconvex test problem.

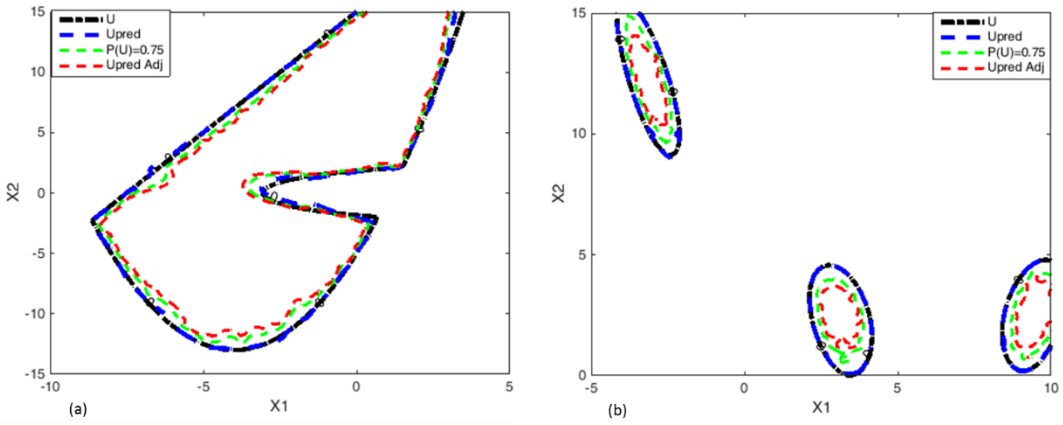

Figure 32: Feasible region for a) nonlinear nonconvex test problem and b) Branin function. Probabilistic metric (*P(U)*) and adjusted feasibility (*Upred Adj*) are more conservative relative to true feasible region (*U*) than surrogate model (*Upred*)

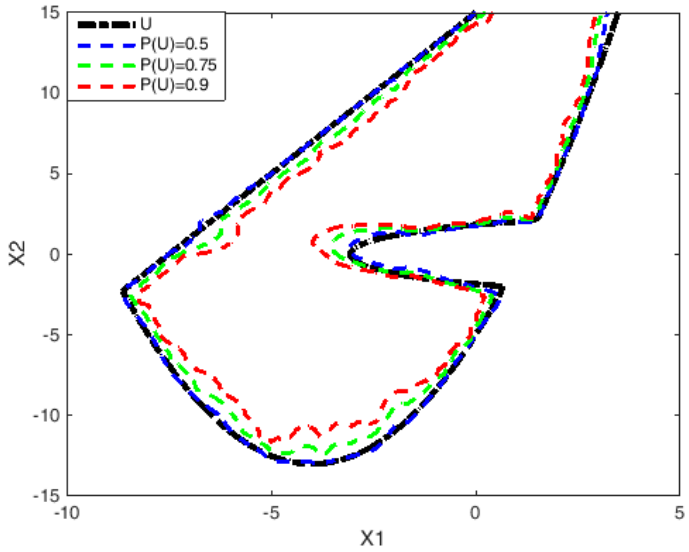

Figure 33: Comparison of different criteria for the probabilistic approach to considering prediction uncertainty. Different criteria for $P(\Psi(d,\theta))\leq 0$ are considered. $P(U)=0.75$ is conservative without excluding large portions of the feasible region.

5.3.2. Extension to dynamic surrogate-based feasibility analysis

The surrogate-based feasibility algorithm presented in Figure 28 can be readily extended to problems of dynamic feasibility analysis with a few modifications. First, it is necessary to generate a surrogate model that is an explicit function of time. Fortunately kriging can be readily applied to model dynamic systems, so the model selection approach can remain the same as for the steady state case. (F. Boukouvala et al., 2011) However there are several important sampling issues to consider in the dynamic case. The most significant of these is the frequency with which data will be sampled from the dynamic process model. High frequency sampling can result in a large number of design sites, especially if the time horizon (H) over which the dynamic problem will be solved is long. While fitting algorithms for kriging are generally computationally efficient, it is necessary to calculate the Euclidean distance between pairs of design sites in order to fit the correlation model. (Kleijnen, 2009) For software packages like DACE, which are implemented in MATLAB, fitting a model can become prohibitive when the number of design

sites exceeds 10,000. It is also important to consider the effect of dynamics during the model improvement stage. The samples that provide the best expected improvement may not be the same at all points along the time horizon. However it is not possible to selectively sample certain realizations of the uncertain parameters at a particular point in time t_k without first evaluating the process model from time 0 all the way to time t_k.

In order to address the first of these issues, an upper limit is placed on the number of samples that can be collected across the entire time horizon. These samples are uniformly spaced across the time horizon so that information is extracted from the simulation at t_k unique points in time, where $k=1,2,\ldots,N$. Thus for each realization of the uncertain parameters, the process model is evaluated along the time interval $t \in [0, H]$. The value of the feasibility function is stored at time points corresponding to $t=t_k$. The second issue is addressed by evaluating the expected improvement function at specific points in time to identify new sample points with respect to the uncertain parameters. The evaluation of the expected improvement function is also completed in such a way that new samples are obtained for points uniformly spaced along the time horizon. The process model is then evaluated for the full time horizon for each realization of the uncertain parameters identified during the model improvement phase.

The algorithm for dynamic surrogate-based feasibility analysis is presented in Figure 34. Only the steps which differ from the steady-state algorithm will be discussed in detail. During the initial sampling stage, the design of experiments is conducted with respect to the uncertain parameters. The sampling strategy is the same as that used for the steady state case, with 49 samples collected for the two parameter case studies used in this work. The process model is evaluated for each set of uncertain parameters in the initial DOE and data is collected at discrete time points, denoted t_k, uniformly sampled along the time horizon. For the case studies used in this work, which are restricted to problems with two uncertain variables, the value of N is selected such that data is collected every 2 seconds. During the model improvement phase, the previously presented local

optimization strategy is used to maximize the expected improvement at each discrete time point t_k. For each t_k, the local solution with the best expected improvement value is selected as the next model evaluation point. Thus a set of new design sites x_k where $k=1,2,\ldots,N$ is obtained. Each design site corresponds to the best expected improvement for a particular time point t_k. As in the steady state algorithm, new points that are too close to existing design sites are eliminated from the sample set prior to evaluating the process model. The same termination criteria are applied as in the steady state case as well. Once a final surrogate model is obtained, it can be used to evaluate the feasible region at different points in time. It can also be used to examine the trajectory of the feasibility function over time and determine how long it will take for a process to resume feasible operation following a disturbance.

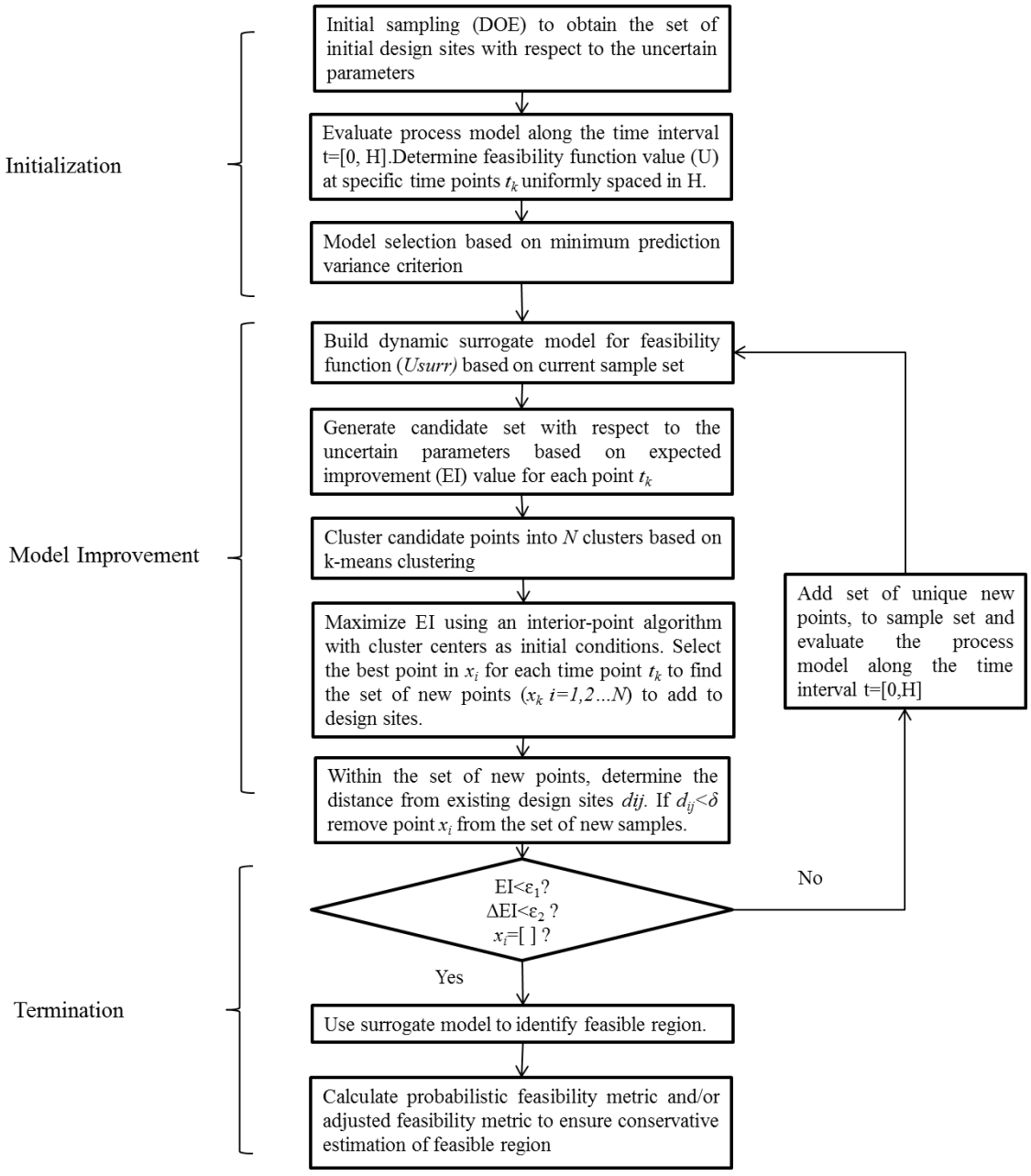

Figure 34: Algorithm for dynamic surrogate-based feasibility analysis

Case study: Dynamic Feasibility Test Problem

The dynamic surrogate-based feasibility algorithm is demonstrated using a test problem for which the shape and size of the feasible region change over time. This problem is specified by equation (61), and the contours of the feasible region are depicted in Figure 35. The feasible region is

initially non convex in nature, but over time becomes convex. The size of the feasible region also changes such that the feasible region at *t=15* is significantly smaller than that at *t=0*.

$$\begin{aligned}
&\frac{a}{5}x_1 + bx_2 + a \le 0 \\
&\frac{a}{10}x_2 - x_1 \le 0 \\
&\frac{da}{dt} = \exp\left(-0.1t\right) \\
&a(0) = -10 \\
&-5 \le x_1 \le 10 \\
&0 \le x_2 \le 15 \\
&t \in [0,15] \\
&b = -0.1
\end{aligned} \tag{61}$$

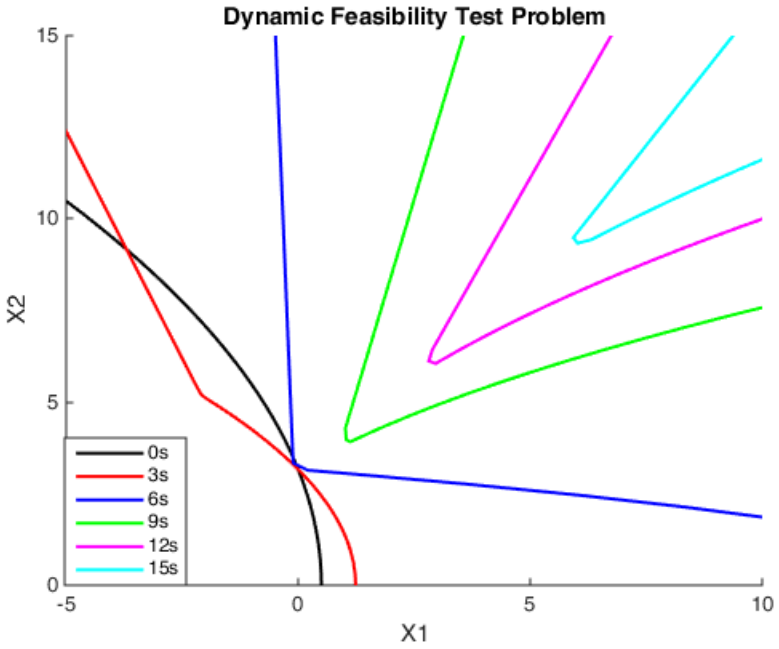

Figure 35: Contours of the feasibility function for the dynamic feasibility test problem (equation (61)). The shape and size of the feasible varies with time.

The local optimization approach to evaluating the expected improvement function provides significant computational savings for the dynamic feasibility test problem, as indicated by the significantly shorter cpu-times for the local method shown in Table 11.

A comparison of the performance for the local and global approaches to evaluating the expected improvement function for the dynamic test problem is shown in Figure 36. The true contours of the feasible region are depicted along with the contours obtained using the surrogate model. The samples collected during the expected improvement phase are superimposed on the plot to show their location relative to the boundaries of the feasible region. The local and global approaches both provide accurate surrogate models, but the local approach does so at approximately one third the computational cost. The additional samples collected by the local and global methods are also comparable, as indicated by Figure 36.

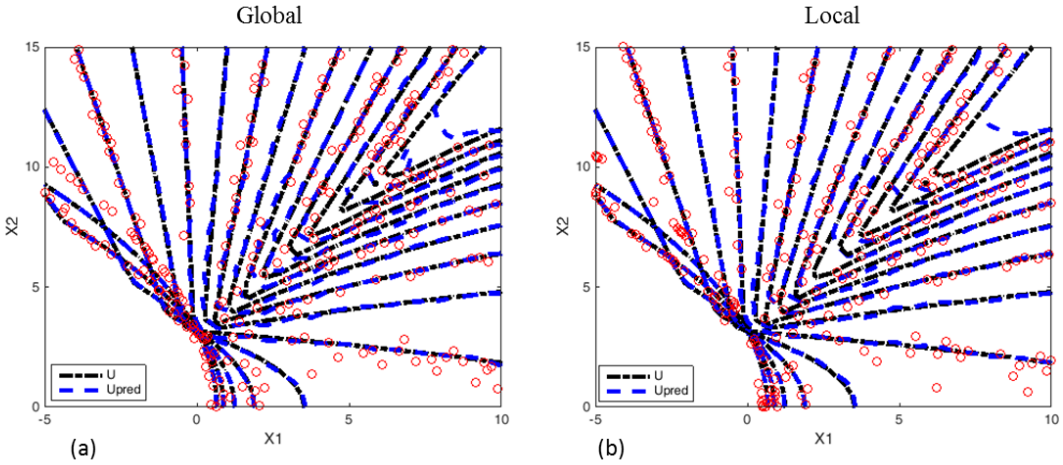

Figure 36: Comparison of a)Global and b)Local approaches to evaluating the expected improvement function for the dynamic feasibility test problem (equation (61)). The true contours of the feasible region are shown in black. The contours of the predicted feasible region are shown in blue. The points added during the expected improvement phase are scattered in red.

The surrogate model developed using the proposed algorithm correctly classifies feasible or infeasible points in the uncertain space over 99 percent of the time for a validation set of 10,000 data points. In addition, the surrogate model can accurately reflect the changing shape of the feasible region over the time horizon as indicated by Figure 36. It is important to note that even for the times between the discrete times points t_k the surrogate model can accurately represent the

feasibility contours. This is consistent with the interpolating nature of kriging metamodels.(Calder & Cressie, 2009)

The surrogate-based method for dynamic feasibility analysis offers significant advantages over the methods discussed section 5.2.1. These methods require explicit differentiation of the control trajectories and may also require full differentiation with respect to the uncertain parameters. Deterministic solvers also require explicit, differentiable functions for all constraints. Surrogate-based methods do not require any differentiation, and do not mandate closed-form functions for the constraints. Thus these methods can be applied to a much broader set of problems that might not be readily solved using traditional approaches to evaluating the dynamic feasibility problem. However, surrogate-based methods also suffer from several disadvantages. They are only valid over the dataset for which samples have been collected and should not be used to extrapolate beyond the tested uncertain space. In addition, it is difficult to prove global convergence for surrogate-based optimization problems, so there will always be some degree of uncertainty associated with the predicted feasible region.(Agarwal & Biegler, 2013)

Table 11: Comparison of local and global methods of evaluating the expected improvement function for two test problems. *Over a validation set of 10,000 samples, the percentage of samples correctly identified as feasible or infeasible by the surrogate model is reported. **In addition, the percentage of infeasible samples incorrectly classified as feasible is reported as the percent of samples for which the surrogate model is not conservative relative to the original function

Problem	Method	Iter	Func calls	Time (cpu-s)	EI	Maximum prediction variance	% correct*	% not conservative**
Nonlinear nonconvex example	Global	153	152	270	0.2	12	99.0	0.61
	Local	42	152	59	0.4	6.6	98.5	0.93
Branin function	Global	153	152	266	3.6	659	99.9	0.04
	Local	40	154	59	4.7	704	99.7	0.09
Dynamic feasibility test problem	Global	34	265	10183	0.01	1.1	99.5	0.15
	Local	34	265	3394	0.01	1.0	99.5	0.17

5.4. Surrogate-based feasibility analysis of a roller compaction process

Having demonstrated that the proposed steady state and dynamic feasibility algorithms can accurately generate surrogate models for the feasible region using various test problems, the proposed approach can be applied to a pharmaceutical case study. The methods described in section 5.3 are used to evaluate the feasible region for a pharmaceutical roller compaction process modeled by the dynamic Johanson's model (Hsu et al., 2010b) described in Chapter 2. The model equations used to simulate the roller compaction process are shown in problem (62). The specific parameter values used in this case study are provided in Table 12 at the end of this chapter. These values are based on the steady state conditions reported in the literature. (Hsu et al., 2010b)

$$\frac{d}{dt}\left(\frac{h_0}{R}\right) = \frac{\omega\left[\rho_{in}\cos\theta_{in}(1+{h_0}/{R}-\cos\theta_{in})({u_{in}}/{\omega R})-\rho_{exit}({h_0}/{R})\right]}{\int_0^{\theta_{in}}\rho(\theta)\cos(\theta)d\theta}$$

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

$$\sigma_{exit} = C_1\rho_{exit}{}^{\mathrm{K}} \qquad (62)$$

$$P_h \in \left[P_h^{\min}, P_h^{\max}\right]$$

$$u_{in} \in \left[u_{in}^{\min}, u_{in}^{\max}\right]$$

$$\omega \in \left[\omega^{\min}, \omega^{\max}\right]$$

$$\theta_{in} \in \left[\theta_{in}^{\min}, \theta_{in}^{\max}\right]$$

$$\rho_{in} \in \left[\rho_{in}^{\min}, \rho_{in}^{\max}\right]$$

The goal of this analysis is to identify the feasible operating region for the roller compaction process in the presence of plant-model mismatch and uncertainty in input material quality. Process feasibility is evaluated with respect to constraints on the quality attributes of the ribbon, namely the ribbon thickness (h_0) and the ribbon density (ρ_{exit}). These constraints are shown in

equation (63). Two sources of uncertainty are considered; variability in the powder density entering the system *(ρ_{in})* and uncertainty in the estimated inlet angle (θ_{in}), a parameter in Johanson's model. Three control variables, the hydraulic pressure applied by the rolls *(P_h)*, the roll speed *(ω)* and the powder feed speed *(u_{in})* can be manipulated in order to maintain feasible operation in the presence of disturbances. In order to account for the effect of control variables when simulating the process, a control calculation is embedded into the simulation of the roller compaction process. The controller set points (P_d, ω_d, u_d) are obtained by minimizing the violation of the constraints in equation (63) over the three control variables. The controller dynamics are given by equation (64). Additional information regarding model-based control schemes for this process are discussed at length in the literature (Hsu et al., 2010a, 2010b) and are therefore not presented here.

$$\begin{aligned} &\rho_{exit}^{L} \le \rho_{exit} \le \rho_{exit}^{U} \\ &h_0^{L} \le h_0 \le h_0^{U} \end{aligned} \tag{63}$$

$$\begin{aligned} &\tau_p \frac{dP_h}{dt} + P_h = P_d \\ &\tau_\omega \frac{d\omega}{dt} + \omega = \omega_d \\ &\tau_u \frac{du_{in}}{dt} + u_{in} = u_d \end{aligned} \tag{64}$$

The response of the roller compaction process to changes in the inlet density (ρ_{in}) and the inlet angle (θ_{in}) can be simulated using equations (62) and (64). In this case the process is simulated for 28 seconds, by which time the quality attributes have reached their new steady state values. A representative process trajectory for the dynamic simulation is shown in Figure 37. A disturbance in the uncertain variables is introduced at *t=0* seconds. Control action is taken to bring the process back to feasible operation such that all quality constraints are satisfied. The quality attributes have returned to feasible values by the end of the time horizon, *H=28* seconds.

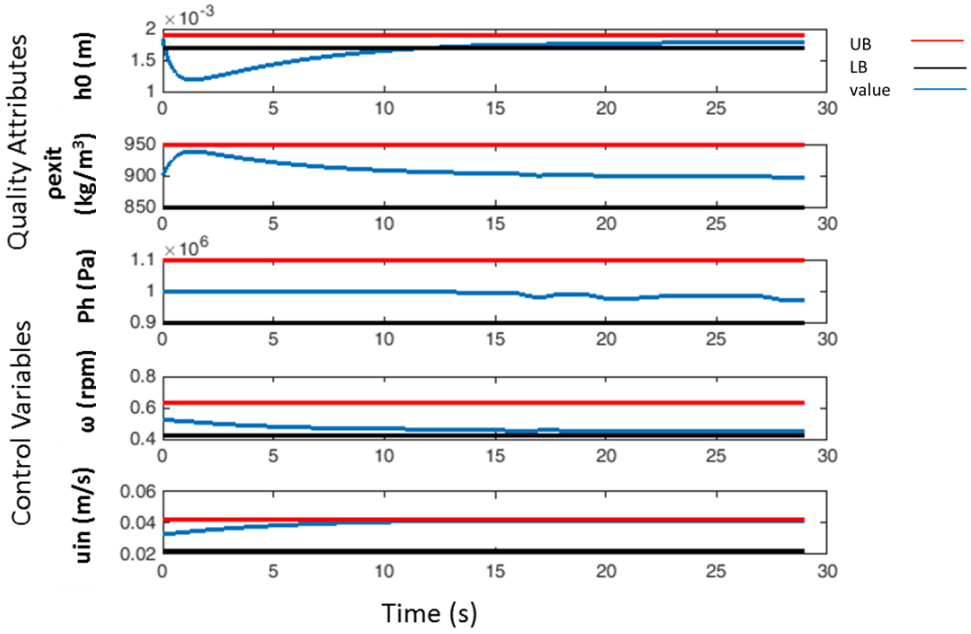

Figure 37: Representative process trajectory for a roller compaction process. The process is initialized at steady state conditions. At *t=0* a disturbance in the uncertain parameters (ρ_{in} and θ_{in}) is introduced. The controller acts to return the process to feasible operation.

In the absence of a surrogate-based approach, the steady state feasibility analysis problem for the roller compaction process would be formulated as problem (65). This has the potential to be computationally expensive to evaluate, and would require reformulation to express the integral terms as algebraic expressions.

$$\min_{P_h,u_{in},\rho_{exit},h_0,u} u$$

$$s.t.\ \rho_{in}\cos\theta_{in}\left(1+\frac{h_0}{R}-\cos\theta_{in}\right)\left(\frac{u_{in}}{\omega h_0}\right)=\rho_{exit}$$

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

$$\sigma_{exit}=C_1{\rho_{exit}}^K$$

$$h_0-h_0^U\le u$$

$$h_0^L-h_0\le u$$

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

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

$$P_h\in\left[P_h^{\min},P_h^{\max}\right]$$

$$\omega\in\left[\omega^{\min},\omega^{\max}\right]$$

$$u_{in}\in\left[u_{in}^{\min},u_{in}^{\max}\right]$$

$$\theta_{in}\in\left[\theta_{in}^{\min},\theta_{in}^{\max}\right]$$

$$\rho_{in}\in\left[\rho_{in}^{\min},\rho_{in}^{\max}\right] \qquad (65)$$

Reformulating problem (65) using surrogate-based feasibility analysis provides the far simpler optimization problem (66), where U_{pred} is the surrogate feasibility function. Note that the control variables P_h, ω and u_{in} do not appear explicitly in problem (66), as they are manipulated by the controller in the dynamic simulation. Therefore the control trajectories are an explicit function of the uncertain parameters.

$$\min_{\rho_{in},\theta_{in}} U_{pred}\left(\rho_{in},\theta_{in}\right)$$

$$s.t.\ \theta_{in}\in\left[\theta_{in}^{\min},\theta_{in}^{\max}\right] \qquad (66)$$

$$\rho_{in}\in\left[\rho_{in}^{\min},\rho_{in}^{\max}\right]$$

Problem (66) can be solved using the surrogate-based feasibility algorithm presented in Figure 28. This can be used to identify the contours of the feasible region with respect to the two uncertain parameters, the inlet density ρ_{in} and the inlet angle θ_{in}. The steady state feasible region

for the roller compaction process is shown in Figure 38. The feasible region obtained using the surrogate-based approach (problem (66)) is compared with that obtained by solving the original feasibility problem formulation (problem (65)). The surrogate-based approach provides a very accurate representation of the steady state feasible region. If desired, the probabilistic feasibility metric or the adjusted feasibility metric can also be calculated. For the steady state roller compaction case study, these metrics provide a conservative estimate of the feasible region relative to the original process model without significantly reducing the size of the feasible operating space.

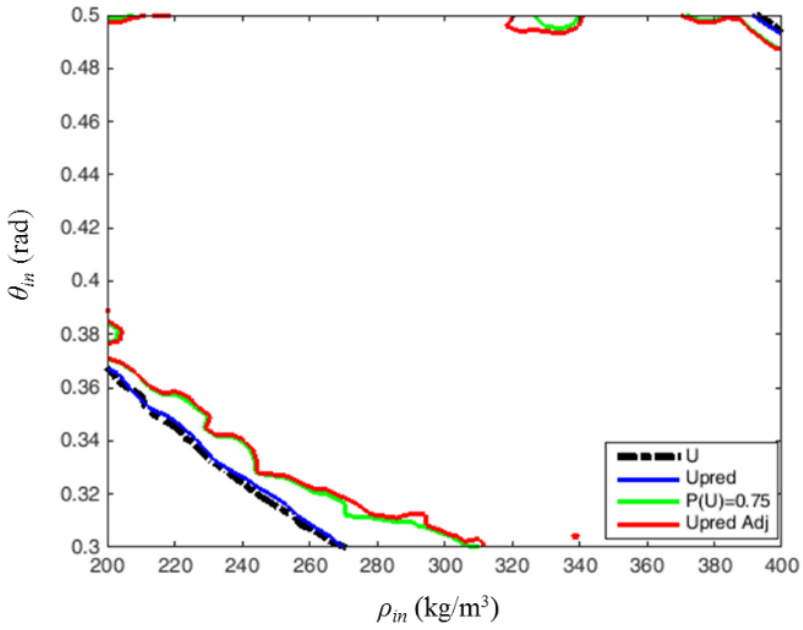

Figure 38: Feasible region for the roller compaction process. Contours of the feasible region identified by solving problem (65) are shown in black (*U*). Contours of the feasible region from the surrogate-based problem formulation (66) are shown in blue (*Upred*). The contour corresponding to *P(U)=0.75* is shown in green and the adjusted feasibility metric (*Upred Adj*) is shown in red.

In addition to guiding the selection of design space, the results of the steady state feasibility analysis can be used to suggest potential strategies to enhance the robustness of the roller compaction process. For instance, improving the estimate of the inlet angle could broaden the feasible region with respect to the inlet density. Specifically, if the estimated inlet angle was

between 0.38 and 0.48 radians with a high degree of confidence, then it would be possible to ensure that the entire range of anticipated inlet densities could be tolerated by the process. The feasible region also seems to suggest that higher inlet densities (>400 kg/m^3) might perform well with respect to the quality attributes of the ribbon. Therefore it might be worthwhile to consider a formulation with higher bulk density if possible.

The roller compaction process is dynamic in nature, as indicated by problem (62). Therefore it is also of interest to examine the dynamic feasibility of the system with respect to the uncertain parameters. This can be accomplished using the dynamic surrogate-based feasibility formulation shown in equation (67). Note that in this case the uncertain parameters are not explicit functions of time, as we consider only the case where a step change in the uncertain parameters is introduced at *time=0* seconds. The system then responds over the time horizon $t \in [0, 28]$ as shown in Figure 37.

$$\begin{aligned} &\min_{\rho_{in},\theta_{in},t} U_{pred}\left(\rho_{in},\theta_{in},t\right) \\ &s.t.\ \theta_{in} \in \left[\theta_{in}^{\min},\theta_{in}^{\max}\right] \\ &\rho_{in} \in \left[\rho_{in}^{\min},\rho_{in}^{\max}\right] \end{aligned} \tag{67}$$

The results of dynamic feasibility analysis can be used for a number of purposes. The first is to determine whether the roller compaction process will return to feasible operation for a given combination of the inlet density and inlet angle and, if so, how long this will take. This can be accomplished by examining the feasibility trajectory for the process, as depicted in Figure 39. It should be noted that since the two quality attributes (ribbon thickness and ribbon density) have very different magnitudes the value of the constraint violations (*u* in equation (65)) are scaled to the upper bound for each attribute. This ensures that the solution is not biased towards conditions that prevent a violation of the ribbon density constraint. Based on the feasibility trajectories shown in Figure 39, it is clear that the surrogate model can accurately capture the dynamic response of the feasibility function to changes in the uncertain parameters. In addition, it can be

seen that the feasibility function tends to respond as a decaying exponential to changes in the uncertain parameters, with the process generally returning to feasible operation within 20 seconds. Therefore it would be anticipated that the feasible region at 5 seconds would be much smaller than that at 10 seconds, but that the feasible region at 20 seconds would not be significantly different from that observed at 25 seconds. In fact this is consistent with the observed feasible regions, which are shown in Figure 40. The feasible regions in Figure 40 can also be used to identify changes in the uncertain parameters from which the process can recover most quickly. These results indicate an inverse relationship between the two uncertain parameters. Combinations of low inlet density with high values of the inlet angle or high inlet density with low values of the inlet angle are easiest for the process to recover from quickly. This is significant because it implies that the true value of the inlet angle can significantly influence the size of the feasible region immediately following a change in inlet material quality.

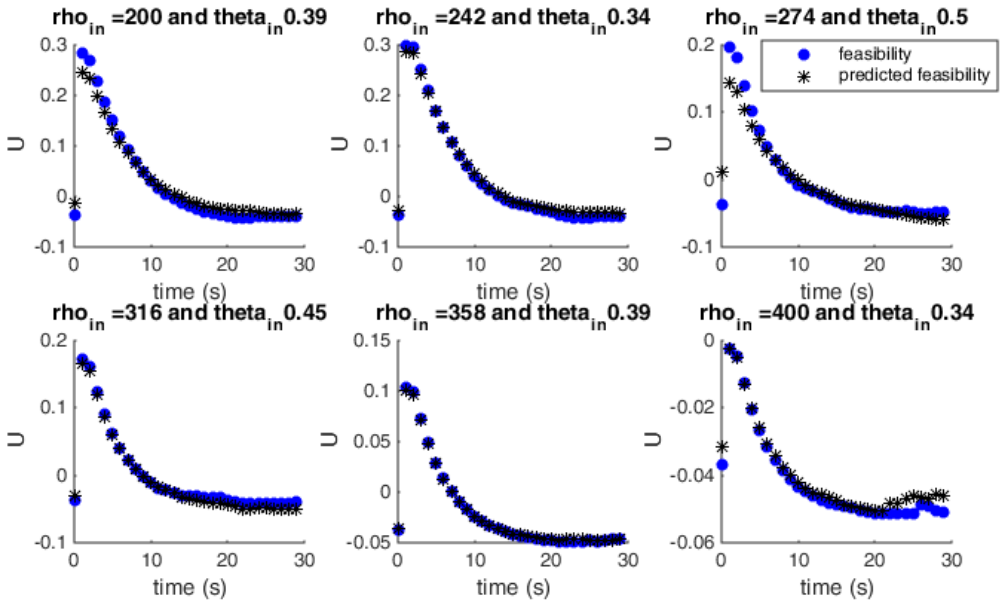

Figure 39: Feasibility trajectories for the dynamic roller compaction process at six different realizations of the uncertain parameters. Values of the inlet density (ρ_{in}) and inlet angle (θ_{in}) are given in the title. The feasibility trajectories obtained from the original process model are shown in blue. Those predicted by the surrogate function are shown in black.

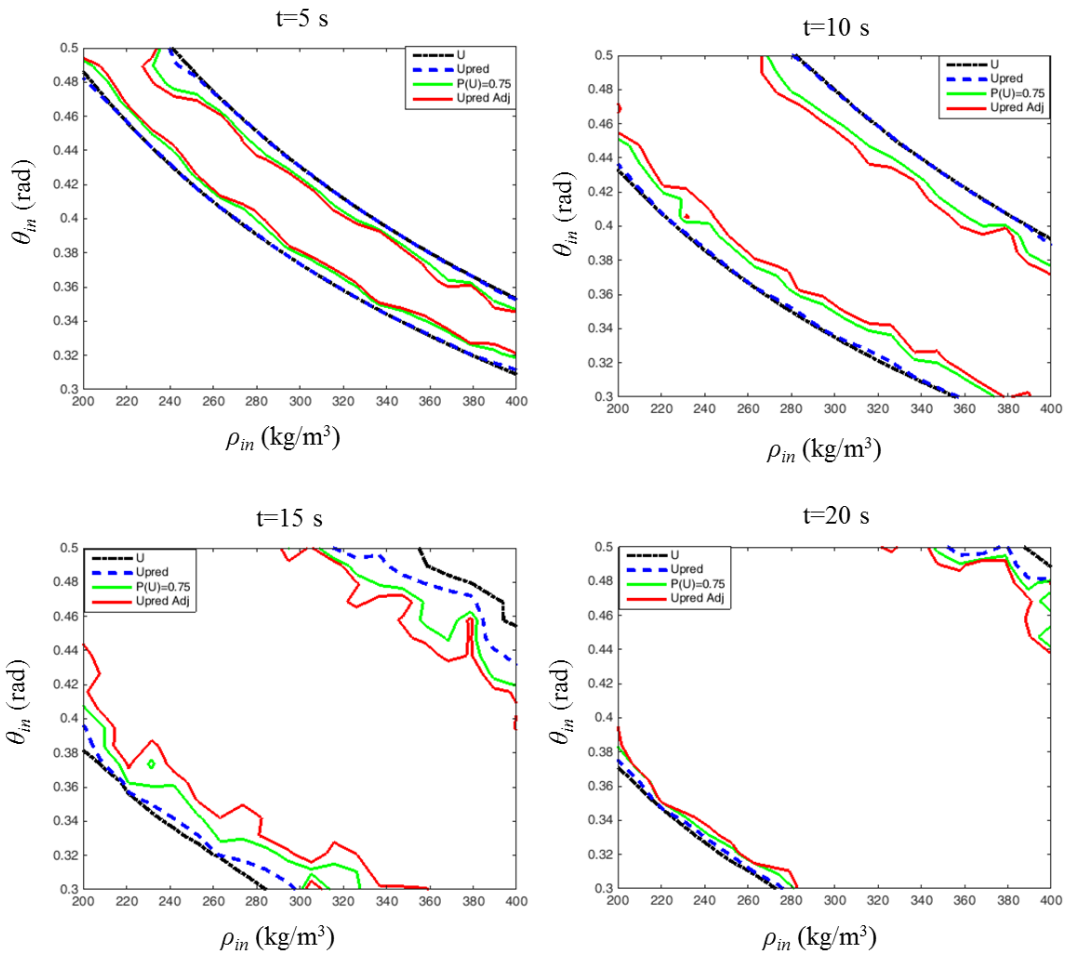

Figure 40: Feasible region for dynamic roller compaction process at various points along the time horizon, specifically *t=5,10,15* and *20* seconds. The feasible region obtained from the original process model (*U*) is compared with the feasible region obtained from the surrogate-based approach. The conservative feasibility metrics *P(U)* and the adjusted feasibility are also shown.

The surrogate-model developed for the feasibility function evaluates quickly relative to the original process model. The surrogate-based model can evaluate for a time horizon of $t \in [0, 28]$ seconds in less than 1 cpu-second while the original process model takes 39 cpu-seconds to evaluate. The computational time associated with the original process model would render it ineffective for use in model-predictive control applications. However, the reduced-order model could be used for such applications due to its efficiency. The surrogate-based model can also be used to rapidly generate many scenarios to evaluate the effect of the uncertain parameters on

process feasibility over time. It is therefore more appropriate for applications such as variance-based sensitivity analysis or constrained global optimization.

5.5. Conclusions and Areas for Future Work

In this chapter, the concept of feasibility analysis has been introduced. Feasibility analysis can be applied to pharmaceutical processes to aid in the selection of design space. Due to the complex and often dynamic nature of continuous pharmaceutical processes, traditional approaches to feasibility analysis may not be well-suited to these problems. Therefore a surrogate-based method for feasibility analysis has been developed. The method presented in this work is novel in that it introduces an initial model selection phase and incorporates an efficient local solution approach for the expected improvement function. Two feasibility metrics, the probabilistic and adjusted feasibility, have been introduced to ensure the surrogate-based feasible region is conservative relative to that from the original process model. This is particularly important when using a surrogate model to guide the selection of process design space. The proposed surrogate-based feasibility algorithm has also been extended to dynamic feasibility problems. The dynamic offers several advantages over traditional approaches to solving the dynamic feasibility analysis problem. It does not require explicit differentiation of the model equations with respect to the control variables and uncertain parameters. In addition, the surrogate model for the feasibility function evaluates rapidly and therefore has the potential to be used for applications requiring rapid simulation like model-predictive control.

The presented algorithms have been demonstrated to perform well for a series of test problems, including those with nonconvex and disjoint feasible regions, as well as feasible regions that change shape and size over time. They have also been used to evaluate the steady state and dynamic feasibility of a pharmaceutical roller compaction process. The results of feasibility analysis can be used to inform the selection of design space and suggest methods to enhance process robustness.

While the proposed methods perform well, they also have several drawbacks. The first of these is that, as is the case for surrogate-based methods generally, global convergence to a true optimal solution can be difficult to guarantee. (Biegler et al., 2014) In addition, the specific metamodel used in this work, kriging, is primarily an interpolating method. (Calder & Cressie, 2009) Therefore the surrogate model for the feasibility function cannot be used to extrapolate beyond the range of available design sites. There is still significant work to be done in the area of surrogate-based feasibility analysis, particularly for dynamic processes. For instance, the effect of implementing a variable time step in the dynamic surrogate-based feasibility algorithm should be considered. This could reduce the sampling requirements for this method and could ensure that the feasibility function is well-represented in areas where the process is changing rapidly with respect to time.

Table 12: Parameter definitions and values used in the roller compaction case study

Parameter	Symbol	Value/Range	Units
Roll radius	R	0.125	m
Roll width	W	0.05	m
Compression parameter	K	4.97	
Compression parameter	C_1	7.5×10^{-8}	$Pa/(kg/m^3)^{4.97}$
Compact surface area	A	0.01	m^2
Effective angle of friction	δ	0.7069	rad
Nip angle	α	0.173	rad
Angular position	θ	NA	rad
Hydraulic pressure (roll pressure)	Ph	0.9 – 1.1	MPa
Angular velocity of rolls	ω	4-6	rpm
Powder feed speed	u_{in}	$2.27 \times 10^{-4} – 4.27 \times 10^{-4}$	m/s
Inlet powder density	ρ_{in}	200 – 400	kg/m^3
Inlet angle	θ_{in}	0.3 – 0.5	rad
Ribbon thickness	h_0	$1.7 \times 10^{-3} – 1.9 \times 10^{-3}$	m
Ribbon density	ρ_{exit}	850 – 950	kg/m^3

Chapter 6

6. Flexibility analysis of pharmaceutical processes and design under uncertainty

6.1. Flexibility analysis and process design under uncertainty

A variety of terms can be used to describe the concept of process robustness, including flexibility, resiliency and operability. The first two are more common in process systems engineering and operations research, while the latter tends to be used more often in the controls literature.(I. E. Grossmann et al., 2014; Lima et al., 2010) This chapter will focus on process flexibility and its use in formulating problems of optimal process design under uncertainty. The concept of flexibility analysis is directly related to feasibility analysis in that both evaluate the ability of a process to operate in the presence of potential variability. They are distinguished by the fact that feasibility refers to the limits within which a process is operable, while flexibility quantifies the ability of a process to maintain feasible operation in the presence of inherent variability or external disturbances. (Biegler et al., 1997) Flexibility analysis can be applied to chemical process systems in order to evaluate robustness and has been employed extensively in the area of process design under uncertainty. (Floudas et al., 2001; Halemane, 1983; Mohideen et al., 1996b; Pistikopoulos & Ierapetritou, 1995) This field is concerned with the design of robust and cost effective chemical processes, objectives which are often competing as the most flexible design alternatives can be the most expensive. (Biegler et al., 1997; Pistikopoulos & Ierapetritou, 1995) Algorithms for process design under uncertainty are used in applications ranging from process synthesis (Mohideen et al., 1996a) and retrofit design (Bansal et al., 1998; Pistikopoulos & Ierapetritou, 1995) to supply-chain optimization (Guillén-Gosálbez & Grossmann, 2009; Tsiakis et al., 2001), and integrated design and scheduling (Subrahmanyam et al., 1994). However the use of these methods for the design of pharmaceutical processes has been relatively limited to supply-chain (Levis & Papageorgiou, 2004; Shah, 2004) and capacity planning (Linninger & Chakraborty, 2001) applications rather than manufacturing process design. In this chapter a

surrogate-based method for the evaluation of process flexibility will be introduced. This method is incorporated into a novel algorithm for the design of flexible processes under uncertainty. The proposed methods will be applied to a number of test problems, including a design problem for a continuous pharmaceutical manufacturing process.

6.1.1. The flexibility index problem

A rigorous mathematical formulation for process flexibility was introduced by Halemane and Grossmann(Halemane, 1983) in the 1980s. In this formulation, process robustness is quantified using the flexibility index. The flexibility index describes the extent to which a process is feasible over a range of potential deviations in process inputs, operating conditions and model parameters. The formulation of the flexibility index problem follows directly from the feasibility test problem introduced in Chapter 5. The flexibility test problem (68) can be used to determine if a particular process design d is feasible over a range of potential deviations in the uncertain parameters $\theta \epsilon T$. If $\chi(d) \leq 0$ in problem (68), then there exists a set of control actions that result in feasible process operation over the full range of deviations $\theta \epsilon T$. If not, the process is infeasible for some realizations of the uncertain parameters θ. (Biegler et al., 1997; I. E. Grossmann, Floudas, C.A., 1987; Swaney & Grossmann, 1985a, 1985b)

$$\chi(d) = \max_{\theta \in T} \min_{z} \max_{j \in J} \left\{ g_j(\mathrm{d}, \mathrm{z}, \theta) \right\} = \max_{\theta \in T} \Psi(d, \theta) \tag{68}$$

The flexibility index problem (69) follows directly from the flexibility test problem and can be used to quantify how flexible a particular process design is. In problem (69), the region $T(\delta)$ describes a range for the uncertain parameters that can vary in size depending on the flexibility of the process. The parameter δ is a non-negative scalar such that if $\delta < 1$, the corresponding feasible region is $T(\delta) \subset T$, if $\delta > 1$ the corresponding feasible region is $T(\delta) \supset T$ and if $\delta = 1$ the corresponding feasible region is $T(\delta)=T$, where T is the nominal range of the uncertain parameters. The feasible region $T(\delta)$ is defined in terms of the quantities $\Delta \theta^-$ and $\Delta \theta^+$, which represent positive and negative deviations from the nominal values of the uncertain parameters

respectively. (Biegler et al., 1997; I. E. Grossmann et al., 2014; Swaney & Grossmann, 1985a, 1985b)

$$
\begin{aligned}
&F = \max \delta \\
&s.t. \chi(d) = \max_{\theta \in T} \min_{z} \max_{j \in J} \left\{ g_j(\mathrm{d}, \mathrm{z}, \theta) \right\} \leq 0 \\
&T(\delta) = \{\theta \mid \theta^N - \delta \Delta\theta^- \leq \theta \leq \theta^N + \delta \Delta\theta^+ \} \\
&\delta \geq 0
\end{aligned}
\tag{69}
$$

The flexibility index F represents the largest deviation in the uncertain parameters that the process can tolerate in terms of the parameter deviations $\Delta\theta$. The corresponding region $T(F)$ represents the largest rectangle, or in the case of higher dimensional spaces hypercube, that can be inscribed within the multi-dimensional feasible region for the given design d. If $F>1$ the design is more flexible than it needs to be to handle the anticipated deviations, while if $F<1$ the design is only flexible enough to accommodate a fraction of these deviations. The process is exactly flexible enough to maintain feasible operation over the range of deviations from $\Delta\theta^-$ to $\Delta\theta^+$ when $F=1$. (Biegler et al., 1997; Swaney & Grossmann, 1985a, 1985b) The value of the uncertain parameters corresponding to the solution of the flexibility index problem is referred to as the critical parameter point θ^C. This is the set of conditions that limits the flexibility of the design. The critical point θ^C may occur at the extreme values of the uncertain parameter space T, in which case it is referred to as a vertex critical point. Otherwise it is referred to as a non-vertex critical point. The constraint(s) corresponding to θ^C are the critical constraints that limit flexibility. Relaxing these constraints can increase the size of the feasible region. (I. E. Grossmann et al., 2014) The flexibility test and flexibility index problems may also be subject to equality constraints $h_i(d,z,\theta)$, like those presented in the feasibility test problem in Chapter 5. These are not shown explicitly in the problem formulations in this chapter for simplicity.

Figure 41 depicts the concept of process flexibility for a case with two uncertain parameters, θ_1 and θ_2, and four constraints, one of which is nonlinear. For the processes shown, the flexibility index F is less than one, so the design cannot tolerate the full range of possible deviations in θ_1

and θ_2. In Figure 41a the critical point (θ^C) occurs at a vertex of the feasible region because it lies at a set of extreme values of the uncertain parameters corresponding to the upper bound of θ_1 and the lower bound of θ_2. In Figure 41b, θ^C is a non vertex critical point as it does not correspond to an upper or lower bound of the uncertain parameter θ_2. Solution strategies for the flexibility index problem vary depending on the nature of the critical point. For a vertex critical point the solution of problem (69) is a relatively straightforward enumeration procedure. This involves evaluating the flexibility index problem at each vertex of the feasible region and selecting the vertex solution corresponding to the greatest value of δ. For non vertex critical points the flexibility index problem is significantly more difficult to solve. Several approaches including the active set strategy (I. E. Grossmann & Floudas, 1987) and the split and bound strategy (Ostrovsky et al., 2000; Ostrovsky et al., 2002; Ostrovsky et al., 1994) have been developed to address the flexibility index problem for non vertex critical points.

The aforementioned methods of solving the flexibility index problem do have certain limitations. They require explicit, closed-form expressions for all of the process constraints. In addition, certain assumptions regarding the convexity of the feasible region must be satisfied. (Floudas, 2000; Floudas et al., 2001) Finally, for problems with large numbers of constraints and non vertex critical points the solution of the flexibility index problem can become computationally expensive. (Biegler & Grossmann, 2004)

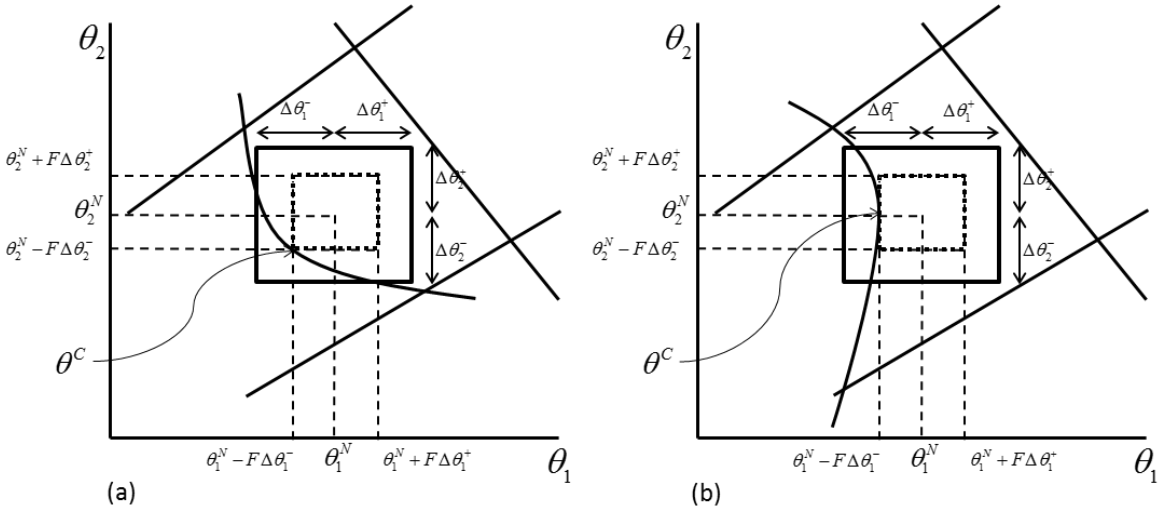

Figure 41: Geometrical representation of the concept of flexibility and the flexibility index. a) depicts a process design with a vertex critical point. b) depicts a process design with a non vertex critical point. Both a) and b) depict processes where the original design is not sufficiently flexible to tolerate the full range of possible deviations in θ_1 and θ_2.

6.1.2. Stochastic flexibility analysis

Stochastic flexibility analysis, introduced by Pistikopoulos and Mazzuchi (Pistikopoulos & Mazzuchi, 1990) and Straub and Grossmann (Straub & Grossmann, 1990), describes the extension of flexibility analysis to uncertainties described by probability distributions rather than fixed parameter ranges. The Stochastic Flexibility index, SF, describes the probability that a particular design *d* will achieve feasible operation given process uncertainty described by the joint probability distribution $j(\theta)$. The distributions for each uncertain parameter may be independent, but do not have to be. Upper and lower bounds on the uncertain parameters can be selected based on the standard deviation for the corresponding distributions. These so-called sigma bounds limit sampling to a high probability region of $j(\theta)$ (e.g. within +/- 3σ of the mean for a normal distribution) and create a closed feasible region, which simplifies solution of the stochastic flexibility problem. (Straub & Grossmann, 1993)

The stochastic flexibility analysis problem can be formulated as shown in equation (70), where the uncertain parameters θ are defined by the joint probability distribution function $j(\theta)$ and

$\Psi(d,\theta)\leq 0$ refers to a specific realization of the uncertain parameters being feasible. In order to solve problem (70), the integral $\int_{\theta:\Psi(d,\theta)\leq 0} j(\theta)d\theta$ must be evaluated over the feasible region in the space of the uncertain variables. This is illustrated conceptually in Figure 42, which shows the contours of a normal joint distribution function and the projection of a convex feasible region into the space of the uncertain parameters.

$$SF(d) = \int_{\theta:\Psi(d,\theta)\leq 0} j(\theta)d\theta \tag{70}$$

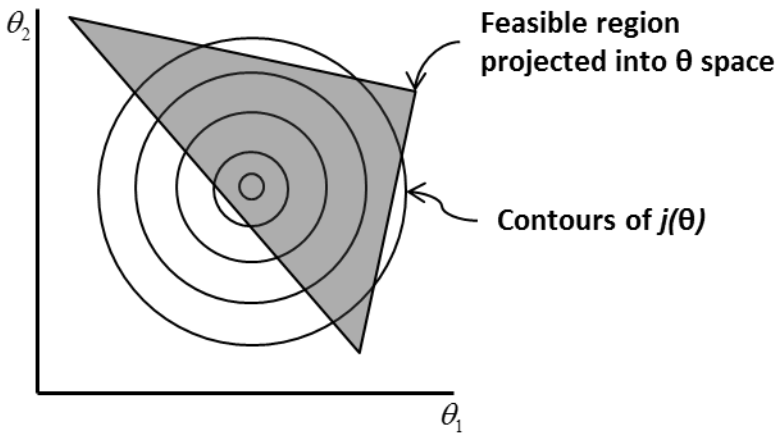

Figure 42: Geometrical representation of the concept of stochastic flexibility. The stochastic flexibility index *SF(d)* is obtained by integration over the feasible region projected into θ space.

The multiple integration in problem (70) makes it challenging to evaluate the stochastic flexibility index. Several solution strategies have been developed to solve problems defined by linear constraints, including the bounding procedure introduced by Pistikopoulos and Mazzuchi (Pistikopoulos & Mazzuchi, 1990) and the inequality reduction scheme proposed by Straub and Grossmann (I. E. Grossmann & Straub, 1996). These methods, while effective, have several limitations. They can only be applied to linear, and therefore convex, problems with relatively few constraints. (Straub & Grossmann, 1993) The bounding approach is appropriate only for uncertain parameters described by normal distributions, and requires *a priori* specification of the

set of active constraints. (Bansal et al., 2000; Pistikopoulos & Mazzuchi, 1990) The inequality reduction approach cannot be readily extended problems with design variables and therefore cannot be used for optimal design under uncertainty problems. (Straub & Grossmann, 1993) Several methods have since been proposed to evaluate stochastic flexibility in nonlinear and non convex systems. (Pistikopoulos & Ierapetritou, 1995; Straub & Grossmann, 1993) These formulations still require the feasible region to satisfy 1-d convexity conditions with respect to the uncertain parameters and tend to be limited to problems with relatively few constraints. Given the difficulties associated with evaluating the stochastic flexibility index, many approaches to the problem of optimal process design under stochastic uncertainty do not directly include calculation of the stochastic flexibility index. (Pistikopoulos & Ierapetritou, 1995; Sahinidis, 2004)

Remarks on the relationship between the flexibility and stochastic flexibility indices

In developing a surrogate-based method for flexibility analysis, it will be useful to apply the formulation of the stochastic flexibility problem to the general case of process flexibility. In doing this it will be useful to consider the relationship between the flexibility index (*F*) and the stochastic flexibility index (*SF*). First, it is important to note that in stochastic flexibility analysis the shape and the size of the feasible region is the same as that considered in the non stochastic case. The primary difference between the flexibility index problem (69) and the stochastic flexibility problem (70) is the probability of realizing different combinations of uncertain parameters. In problem (69), all parameter variations within the range $T = \left\{ \theta \,|\, \theta^N - \Delta\theta^- \le \theta \le \theta^N + \Delta\theta^+ \right\}$ are equally likely whereas in problem (70) some parameter combinations are more likely to occur than others. This can affect the flexibility index because feasible combinations of the uncertain parameters may become more or less likely depending on the distribution $j(\theta)$. If the uncertain parameters are selected from a uniform distribution with bounds $[\theta^N - \Delta\theta^-,\ \theta^N + \Delta\theta^+]$ then the stochastic flexibility index (*SF*) and the flexibility index (*F*) should provide comparable estimates of process robustness. Their magnitudes will not be

identical, as the value of F depends on the selection of $\Delta\theta$. However the ranking of various process designs with respect to their flexibility are expected to be comparable. The only exception would be for the case where multiple designs are feasible within the entire uncertain parameter range T. While the flexibility index problem can explore parameter values beyond the range of the nominal parameter deviations $\Delta\theta^+$ and $\Delta\theta^-$, the stochastic flexibility index is limited to values between 0 and 1. In order to further distinguish between these designs using the stochastic flexibility index it would be necessary to extend the bounds of the distribution $j(\theta)$ to a point where, at most, one design is feasible everywhere.

6.1.3. Optimal process design under uncertainty

The flexibility and stochastic flexibility index problems provide a means of characterizing a particular process design with respect to robustness. In order to select an optimal process design, algorithms are needed for systematically evaluating the available design alternatives and selecting one that is preferred based on economic and/or robustness objectives.(Pistikopoulos & Ierapetritou, 1995) The general problem of process design under uncertainty can be formulated as a mixed integer linear or nonlinear program (MIP or MINLP), with integer variables used to represent discrete design decisions.(Biegler et al., 1997) For process synthesis problems, e.g. the design of an entire manufacturing plant, the number of discrete design options can be quite large.(I. E. Grossmann, 1996) The resulting combinatorial optimization problem may be computationally intractable if all possible realizations the integer variables must be enumerated. A variety of methods, many of them employing branch-and-bound algorithms, can be used to solve these types of optimization problems efficiently. (Adjiman et al., 1997; Quesada & Grossmann, 1992) Branch-and-bound methods provide a way to explore the space of available process designs without having to explicitly enumerate all possible alternatives. This is accomplished by dividing the problem into a series of sub regions, determining lower and upper bounds on the objective function within these sub regions, and systematically eliminating regions

that do not have the potential to contain an optimal solution. (Adjiman et al., 1997; Lee & Grossmann, 2001; Ryoo & Sahinidis, 1995) In this way, only a limited number of integer combinations are evaluated, allowing the efficient solution of large MIP and MINLP problems. Recent advances in the development of algorithms for integer programming problems have made it possible to solve large-scale deterministic optimization problems like those found in process design applications. (Floudas & Gounaris, 2009; Nemhauser & Wolsey, 1999)

For problems with stochastic uncertainties, techniques based on two-stage stochastic programming have been developed to solve the optimal design problem. (Pistikopoulos & Ierapetritou, 1995) As their name implies, two-stage stochastic programming methods divide the decision variables for the problem into two groups. The first-stage variables must be set prior to evaluating the problem with respect to the uncertain variables. The second-stage variables are realized after the uncertain variables. (Sahinidis, 2004) In the context of design under uncertainty, the design variables d correspond to first-stage variables while the control variables z are considered second-stage variables. Two-stage stochastic programs are solved in two stages. In the first stage the design variables are selected to minimize the first stage cost (or maximize profit). The second-stage cost cannot be optimized directly at the first stage because it will vary depending on the realization of the uncertain parameters. However, the expected value of the second stage cost may be incorporated into the first stage objective if it can be estimated. (Tsiakis et al., 2001) At the second stage, the control variables are optimized to minimize cost and ensure operational feasibility. Alternatively, the first stage can focus exclusively on cost reduction or profit maximization and the second stage can be used to ensure feasibility and controllability. (Pistikopoulos & Ierapetritou, 1995) For problems with a limited number of uncertain parameters, parametric programming can be used to obtain closed-form solutions for the optimal design problem. (Bansal et al., 2002; Sahinidis, 2004) These solutions are explicit functions of the uncertain parameters, and can be used to identify optimal control strategies for new realizations

of these parameters. (Dua et al., 2002; Dua & Pistikopoulos, 1998) However these methods can only be used for relatively small problems with respect to the uncertain variables.

One significant challenge associated with process design and synthesis problems is the cost associated with simulating the model in order to determine objective function values for specific design alternatives. Several authors have sought to address this issue through the use of reduced-order models. (Agarwal et al., 2008; Henao & Maravelias, 2010, 2011; Lang et al., 2011) A majority of this work has focused on the use of reduced-order models at the simulation level. For instance, reduced-order models based on PCA (Lang et al., 2011) or proper orthogonal decomposition (POD) (Agarwal et al., 2008) have been used to optimize industrial processes including power plants and pressure swing adsorption systems. These reduced-order models are incorporated into the simulation locally, at the unit operation level. (Biegler et al., 2014; M. Martin & Grossmann, 2011) The use of reduced-order models for global optimization of process operating conditions has also been demonstrated by several authors. (Agarwal & Biegler, 2013; F. Boukouvala & Ierapetritou, 2013) However these implementations have focused on continuous variables and therefore have not included process design under uncertainty applications. The use of surrogate models in algorithms for process synthesis has also been reported. (Davis & Ierapetritou, 2009; Henao & Maravelias, 2011) However the surrogate models in these algorithms are used primarily for optimization over continuous variables. For instance, Davis and Ierapetritou (Davis & Ierapetritou, 2009) developed an MINLP algorithm that uses a branch and bound approach to fathom the integer variables and black-box models to optimize relaxed NLP sub problems at each realization of the integer variables. In other words, surrogate models are used for optimization over the continuous variables after the integer variables have been realized. This method can be extended to the case where the surrogate models are functions of both continuous variables and relaxed integer variables. (Davis & Ierapetritou, 2008) A similar approach has been demonstrated for process synthesis applications by Henao and Maravelias (Henao & Maravelias, 2010, 2011). These authors present a process synthesis algorithm in which

surrogate models based on artificial neural networks ANN are used at the unit operation level. The problem is then formulated as a generalized disjunctive program so that commercially available MINLP solvers can be used to optimize the superstructure over the integer variables.

In the remainder of this chapter, a surrogate-based method for estimating the flexibility of a process design is presented. This method is demonstrated using a number of test problems and is then applied to the roller compaction case study presented in Chapter 5. The surrogate-based flexibility index is then incorporated into an algorithm for surrogate-based process design under uncertainty. This method is similar to the MINLP algorithm of Davis and Ierapetritou (Davis & Ierapetritou, 2008, 2009) in that a branch and bound approach is used to explore the space of available design alternatives. However the proposed algorithm differs from prior work in that integer design variables are incorporated directly into the surrogate model. This method is applied to linear and nonlinear test problems involving discrete design variables and continuous uncertainties. Finally the proposed method is used to select a robust process design for the continuous feeding and blending process introduced in Chapter 3.

6.2. A method for surrogate-based flexibility analysis

The stochastic flexibility index can be interpreted as the probability that a particular design d will achieve feasible operation given process uncertainty described by the joint probability distribution $j(\theta)$. Mathematically, this means that equation (70) can be interpreted as a conditional probability statement of the form $P(A|B)$, as shown in expression (71).

$$SF(d) = P\left(\Psi(d,\theta) \le 0 \mid j(\theta)\right) \tag{71}$$

The conditional probability $P(\Psi(d,\theta)\le 0|j(\theta))$ can be empirically estimated using the formula shown in equation (72), where N is a sufficiently large number of samples, and $\Psi(d,\theta_i)$ is the feasibility function value for a specific realization of the uncertain parameters θ_i. (Congdon, 2006) The flexibility index determined by equation (72) will be referred to as the estimated flexibility and is denoted $\hat{F}(d)$. The term estimated flexibility is preferred over stochastic

flexibility, as the proposed approach can be applied to uncertain parameters sampled from arbitrary distributions. This includes the uniform distribution within the region $T = \left\{\theta \mid \theta - \Delta\theta^{-} \le \theta \le \theta + \Delta\theta^{+}\right\}$, in which case the uncertainties are not stochastic in nature.

$$\hat{F}(d) = P\left(\Psi(d,\theta) \le 0 \mid j(\theta)\right) = \frac{\sum_{i \in N} y_i}{N}$$
where N is the total number of samples drawn from the distribution $j(\theta)$
y_i is a binary variable such that
$$y_i = \begin{cases} 1 \text{ if } \Psi(d,\theta_i) \le 0 \\ 0 \text{ otherwise} \end{cases} \quad (72)$$

The number of samples *N* can be calculated based on the anticipated variance in the process response and the desired confidence level for the solution as shown in equation (73). This formula has been previously reported in the literature as a means of determining the number of simulations required to estimate objective function values in simulation-based stochastic optimization problems. (You et al., 2009) In equation (73), H represents the desired confidence level, *z* is the z-score corresponding to this confidence interval and *S(n)* is the variance in the process response. Since a kriging metamodel is used to approximate the feasibility function, an estimate of the maximum process variance can be obtained directly from the metamodel. This can be used to estimate the response variance term *S(n)*.

$$N = \left[\frac{z_{\alpha/2} S(n)}{H}\right]^2$$
where $H = 0.95$
$z_{\alpha/2} = 1.96$
$S(n) =$ response variance (73)

While the approach described in expression (72) would not be reasonable for an expensive process model, the surrogate feasibility model is computationally cheap to evaluate. For the surrogate feasibility models developed in Chapter 5, a large number of function evaluations can be obtained in a matter of cpu-seconds.

Since the variance for a kriging predictor is known, it is possible to use this information to obtain upper and lower bounds on the estimated flexibility. These can be calculated as shown in equation (74), where s_u is the variance of the kriging predictor.

$$\hat{F}^U(d) = \frac{\sum_{i \in N} y_i}{N}$$
$$y_i = \begin{cases} 1 \text{ if } \Psi(d, \theta_i) - \sqrt{s_u} \le 0 \\ 0 \text{ otherwise} \end{cases}$$
$$\hat{F}^L(d) = \frac{\sum_{i \in N} y_i}{N} \qquad (74)$$
$$y_i = \begin{cases} 1 \text{ if } \Psi(d, \theta_i) + \sqrt{s_u} \le 0 \\ 0 \text{ otherwise} \end{cases}$$

It should be noted that if a design is infeasible then $\hat{F}(d)$ will be 0. Therefore if a sufficiently large (based on equation (73)) number of samples is used to estimate the flexibility index and it is found that $\hat{F}(d) = 0$, it can be concluded that process design *d* is infeasible. This is true only if the surrogate feasibility model is sufficiently accurate to distinguish between feasible and infeasible regions of the uncertain space. This should be established as part of the surrogate-based feasibility algorithm described in Chapter 5.

Case studies

The proposed method for surrogate-based flexibility analysis is applied to the two test problems presented in Chapter 5, namely the nonlinear, nonconvex example problem and the Branin function. The estimated flexibility index is determined for uncertain parameters sampled from a uniform distribution and from a normal distribution for each example problem. Upper and lower bounds on the estimated flexibility are also determined using equation (74). The number of samples (*N*) used to estimate $\hat{F}(d)$ in all case studies is 10,000. This is significantly more than the minimum recommended number of samples obtained by evaluating equation (73), which is on the order of *N=100*. A larger number of samples is used for the two test problems because the

corresponding number of evaluations of the original process model were already available from the model validation at the surrogate-based feasibility stage. Thus the flexibility index estimated using equation (72) could be compared with that obtained from the original process model. Increasing the number of function evaluations from 100 to 10,000 changes the estimated flexibility index by less than two percent for the example problems, so equation (73) provides a good estimate of the required number of samples. In this case it is inexpensive (less than 5 cpu-s) to compute 10,000 evaluations of the reduced-order model so there is little detriment to considering a larger than required number of samples.

The results of the surrogate-based flexibility analysis are shown in Table 13. For both example problems the surrogate-based results compare favorably with those obtained using the original process model. The estimated flexibilities are found to agree, indicating that the degree of prediction error in the surrogate model is not significantly affecting the estimate of process flexibility.

Table 13: Surrogate-based flexibility analysis for the three example problems given in Chapter 5, including the steady state roller compaction process. The flexibility estimated using the original process model in equation (72) is compared with that obtained by using the reduced-order model for the two case studies and these are found to agree well. All flexibility metrics are estimated using 10,000 model evaluations.

Problem	UNIFORM CASE (Not Stochastic)				NORMAL DISTRIBUTION CASE (Stochastic uncertainty)			
	F(d) – original model	F(d) – surrogate model	FL(d)	FU(d)	F(d) – original model	F(d) – surrogate model	FL(d)	FU(d)
Nonlinear Nonconvex example	0.3648	0.3645	0.3314	0.4004	0.7094	0.7137	0.6718	0.7528
Branin function	0.0843	0.0840	0.0322	0.1320	0.0059	0.0059	0.0013	0.0195
Steady state Roller Compaction		0.9398	0.7207	0.9602		0.9999	0.9880	1

For the nonlinear, nonconvex example problem, the flexibility for a uniform distribution is lower than that obtained when the uncertain parameters are sampled from a normal distribution. This is consistent with Figure 43a, which shows the feasible region projected into the space of the uncertain parameters along with the contours of $j(x_1,x_2)$ sampled from a normal distribution. In this case it can be seen that the highest probability areas of $j(x_1,x_2)$ tend to fall within the feasible region. By sampling from a normal distribution, uncertain parameter realizations within the feasible region are more likely to occur than when sampling from a uniform distribution. This enhances the flexibility of the process for the stochastic case. For the Branin function, the process is less flexible in the stochastic case. This is consistent with Figure 43b, which shows that when sampling from a normal distribution two of the three disjoint feasible regions are very unlikely to be realized. In addition, a majority of the region encompassed by the normal distribution $j(x_1,x_2)$ falls outside of the feasible region for the Branin function.

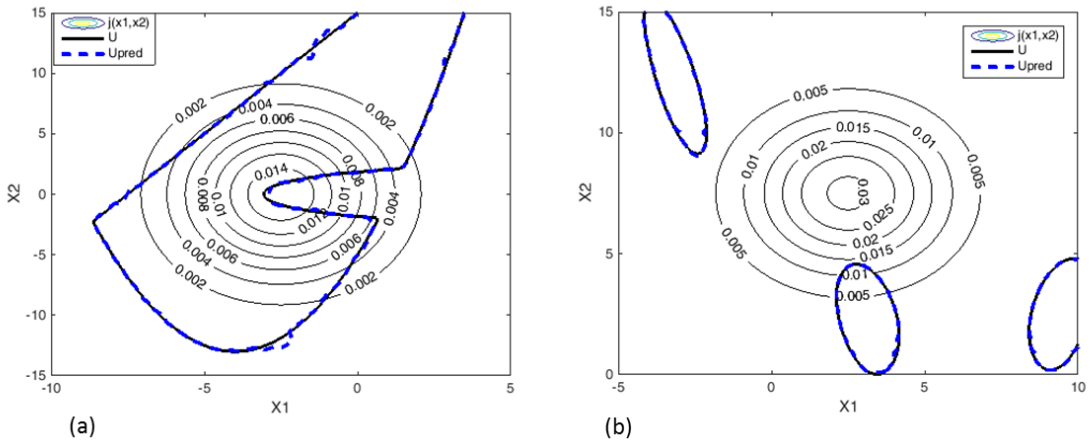

Figure 43: Feasible region (*U*) and surrogate-based feasible region (*Upred*) projected into the uncertain parameter space for a)nonlinear, non convex example problem and b)Branin function. The contours of the uncertain parameter distribution $j(x_1,x_2)$ when sampled from a normal distribution are also shown.

6.2.1. Surrogate-based flexibility analysis of a roller compaction process

The estimated flexibility approach has been applied to the steady state roller compaction example presented in Chapter 5. For this process, the flexibility is significantly increased when the uncertain parameters (ρ_{in} and θ_{in}) are described by a normal distribution rather than a uniform distribution. In fact the process is likely to be feasible everywhere within the joint probability distribution for the uncertain parameters $j(\rho_{in},\theta_{in})$, as indicated in Figure 44. It is also notable that the bounds on the flexibility index are much tighter in the normal distribution case. This is due to lower prediction variance within the feasible region. The prediction variance within the lower left-hand corner of the uncertain parameter space is high, because very few samples are collected within this region during the surrogate-based feasibility analysis. This contributes to higher uncertainty in the estimation of the flexibility index when this portion of the uncertain space is considered, as in the uniform distribution case.

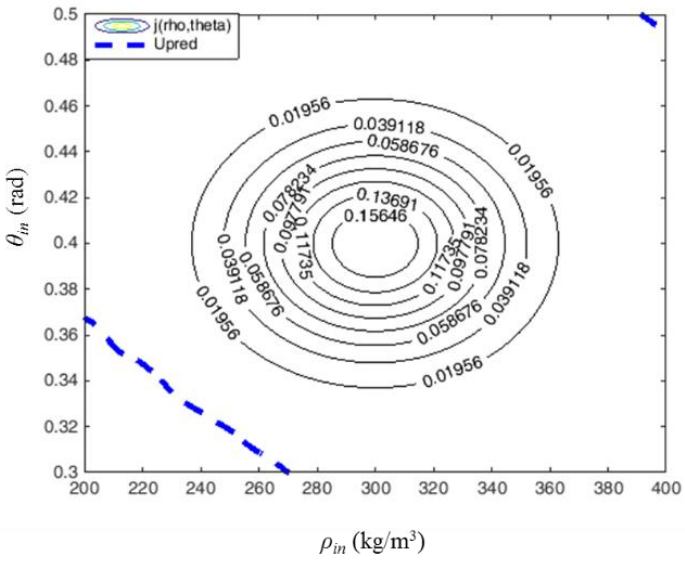

Figure 44: The surrogate-based feasible region (*Upred*) projected into the uncertain parameter space for the steady-state roller compaction problem. The contours of the uncertain parameter distribution $j(\rho_{in},\theta_{in})$ when sampled from a normal distribution are also shown.

6.3. An algorithm for surrogate-based process design under uncertainty

Having demonstrated that the flexibility index for a process can be estimated using a surrogate representation, it is now possible to incorporate the surrogate-based flexibility metric into an algorithm for robust process design. When developing an algorithm for optimal design under uncertainty there are several aspects to consider, including:

- The objective function that will be used to evaluate design optimality
- The metric that will be used to characterize robustness
- The method by which infeasible or suboptimal designs will be removed from consideration

In addition, for a surrogate-based approach it is important to balance the need for an accurate process representation with the desire for efficient sampling of the uncertain space.

In this work, designs are evaluated on both robustness and, for the pharmaceutical case study, cost. The process robustness is characterized using the estimated flexibility metric defined in section 6.2. Only uniform distributions for the uncertain parameters are considered in the case studies presented, but this approach can be readily extended to problems involving stochastic uncertainty defined by an arbitrary distribution. In order to minimize sampling requirements, an initial feasibility stage is included so that infeasible designs can be removed from consideration prior to surrogate modeling. Also, only samples that are integer feasible (i.e. that correspond to integer values for the design variables) are included in the surrogate model.

The proposed algorithm for surrogate-based optimal design under uncertainty is shown in Figure 45. The initial feasibility stage involves evaluating each process design at the nominal realization of the uncertain parameters θ^N. This is done using the original process model, not a surrogate representation. If a design is infeasible at the nominal operating conditions it is removed from the set of candidate process designs. This stage has the potential to become computationally expensive for process synthesis problems involving a large number of design alternatives. For

these problems, an initial surrogate model can be developed using a design of experiments that spans the space of available designs. The surrogate model can then be used to evaluate each design around the nominal conditions θ^N and infeasible designs can be eliminated. This latter approach should be used when the number of candidate signs is greater than *10k*, where *k* is the total number of variables in the synthesis problem (i.e. $k=|z|+|d|+|\theta|$ where $|s|$ indicates the cardinality of set *s*). *10k* is the recommended number of data points to build a preliminary surrogate model, so if *10k* is less than the number of possible design alternatives then a preliminary surrogate model will be less computationally expensive then evaluating each design at the nominal point θ^N.

After the initial feasibility phase, the set of candidate designs is iteratively evaluated and narrowed in the surrogate-based flexibility stage. First, preliminary surrogate models for the feasibility function and (optionally) the total cost are developed. Samples are taken from within the set of feasible integer variables, i.e. the candidate designs, and the range of the uncertain parameters $T = \left\{ \theta \,|\, \theta^N - \Delta\theta^- \le \theta \le \theta^N - \Delta\theta^+ \right\}$. In the case studies presented, a set of 9 design sites in the uncertain space is sampled for each feasible design. 9 design sites are selected because the fitting problem for the kriging parameters is underdetermined with fewer samples for the nonlinear case study. Surrogate model selection is carried out based on the minimum prediction variance criterion described in Chapter 5. Once surrogate models for the feasibility function and the total cost have been developed these are used to estimate the flexibility index and cost corresponding to each candidate design. Upper and lower bounds on cost and flexibility can also be obtained, as described in equation (74). In cases where both flexibility and cost criteria are considered, the number of samples used for the surrogate-based flexibility evaluation is set to twice the maximum value of *N* obtained by evaluating equation (73) for both objectives (cost and flexibility). In the next step, designs that are not potential optima are removed from the candidate set. Several criteria can be used to determine if a design is suboptimal. For instance, if the upper

bound on the estimated flexibility $\hat{F}^U$ is less than a user-specified requirement then a design can be eliminated due to inadequate robustness. If flexibility alone is used to select an optimal design, any design with $\hat{F}^U$ that is worse than $\hat{F}^L$ for the current optimal design can also be eliminated, as such a design does not have the potential to perform better than the current optimum. Alternatively, if the lower bound on the estimated cost for a particular design $\hat{C}^L$ is greater than the current upper bound within the set of candidate designs then this design can be eliminated, as it is guaranteed to be more expensive than the current optimum. The algorithm will terminate if the set of candidate designs contains only one entry after this step. If there are still several candidate designs, the expected improvement function introduced in Chapter 5 is maximized for each design. Note that it is important to remove suboptimal designs prior to the expected improvement stage to avoid additional sampling of the expensive process model for a design that is not a potential optimum. The corresponding realizations of design and uncertain variables are added to the set of design sites for the surrogate model. The surrogate model is updated and the algorithm returns to the flexibility and/or cost evaluation step. This continues until only one design is left in the candidate set, or until the upper and lower bounds on flexibility and/or cost are within a certain tolerance (ε_1) of one another. At this point the algorithm terminates. If only one design remains, then this is selected as the optimal design. If multiple designs satisfy the optimality criteria, the design corresponding to the lowest cost or the greatest flexibility can be selected. If two designs are equivalent, a secondary criterion may be introduced to select between them.

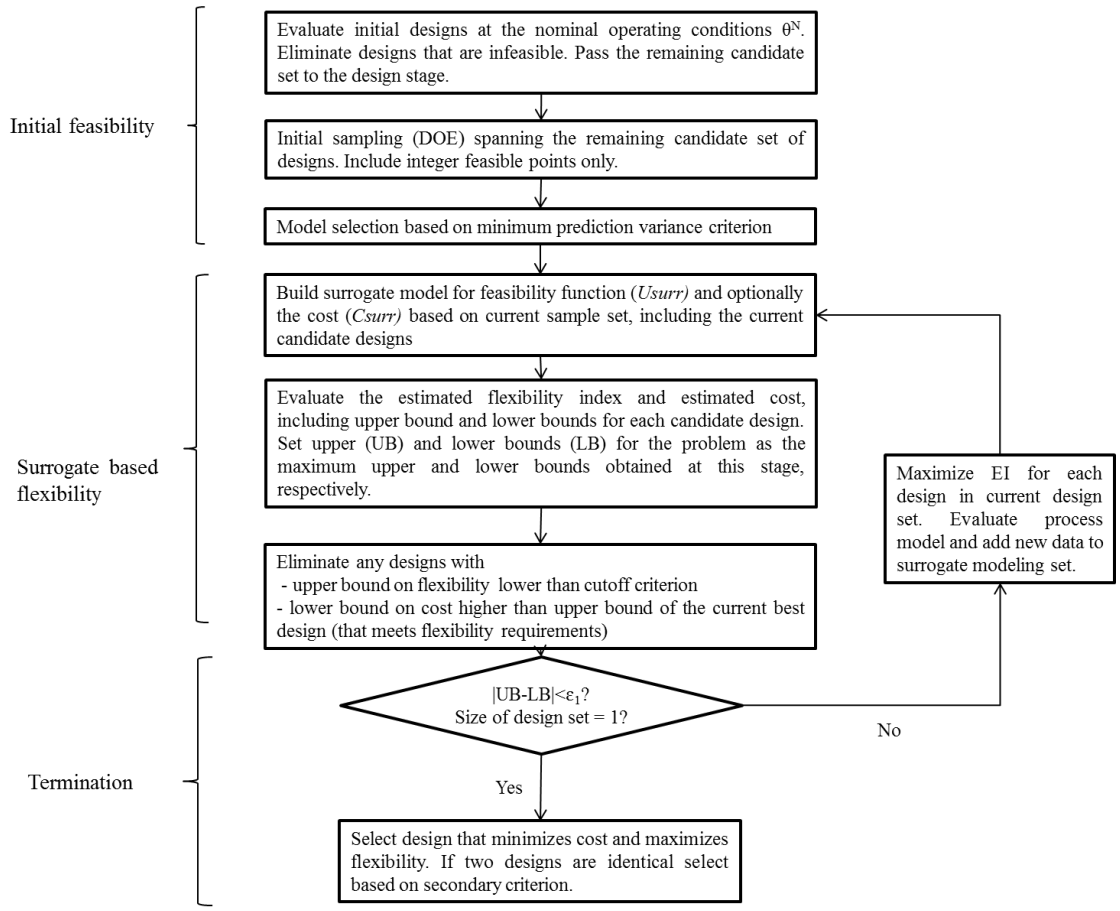

Figure 45: Surrogate-based algorithm for process design under uncertainty

Case studies

The proposed algorithm is tested on two design problems, one that is linear in the design and uncertain variables and a second that is nonlinear in both. The test problems are adapted from the literature. (Bansal et al., 2000) For simplicity, they do not include control variables. In each of these test problems the objective is to select the most flexible design, so cost criteria are not included. However the bounding procedure applied to the flexibility in these test problems can be applied to a cost objective for design problems like the one presented in section 6.4.

Linear process design problem

The first test problem is linear in both the design and uncertain variables. This problem is described by equation (75).

$$
\begin{aligned}
&g1: 0.5\theta_1 + 0.2\theta_2 + d_1 - 3.5d_2 \le 0 \\
&g2: -2.5\theta_1 + \theta_2 + 2d_2 - \frac{5}{3} \le 0 \\
&g3: 0.5\theta_1 - \theta_2 - d_1 + 0.5d_2 - 2 \le 0 \\
&0 \le \theta_1, \theta_2 \le 4 \\
&0 \le d_1, d_2 \le 5 \\
&\theta_1^N = 2 \\
&\theta_2^N = 2 \\
&d_1, d_2 \text{ are integer variables}
\end{aligned}
\tag{75}
$$

This process contains a total of 36 possible design alternatives. Of these, only a subset are feasible at the nominal operating conditions for the process, as shown in Figure 46. In Figure 46, the title of each chart indicates the values of d_1 and d_2 corresponding to the design. The contours of the feasibility function are shown in color. The contour corresponding to $\Psi(d,\theta)=0$ is shown in black. The nominal operating point $(\theta_1^N, \theta_2^N)=(2,2)$ is shown as a black dot.

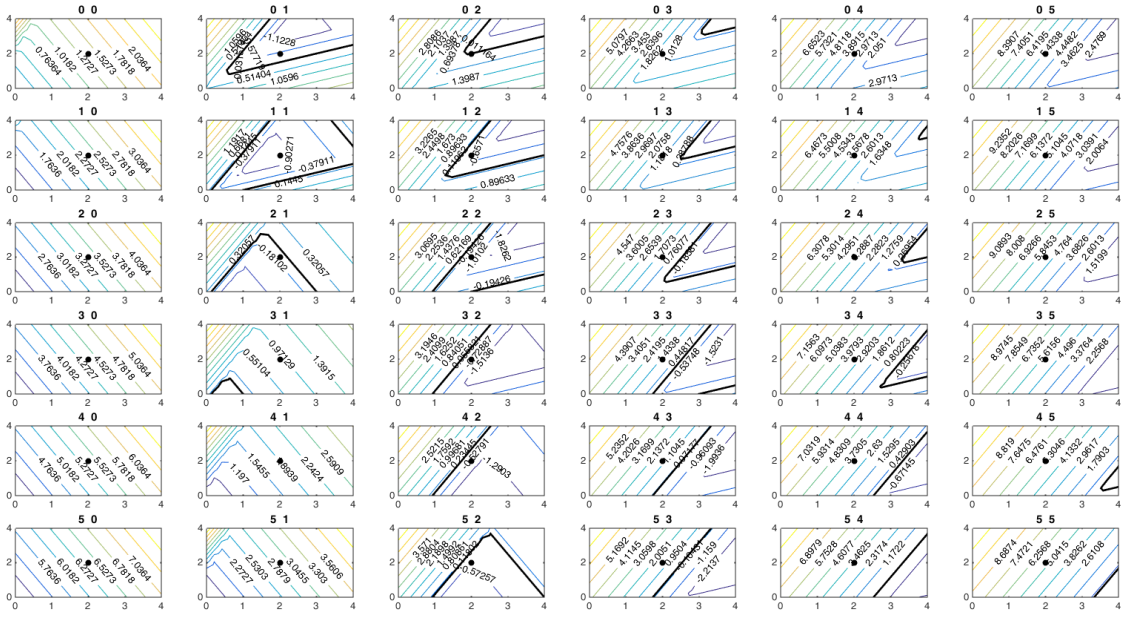

Figure 46: Contours of feasibility function for each of the designs in problem (75) For each design, the boundary of the feasible region is shown in black. The nominal operating conditions are indicated by a black dot.

Of the 36 designs shown in Figure 46, only 9 are feasible under the nominal operating conditions. These constitute the set of candidate designs. All candidate designs are successfully identified during the initial feasibility stage and are listed in Table 14.

Table 14: Set of feasible designs for the linear test problem shown in equation (75)

Feasible Designs d1, d2
0, 1
0, 2
1, 1
1, 2
2, 1
2, 2
3, 2
4, 2
5, 2

The initial DOE consists of 81 samples, 9 for each of the feasible designs. After the initial models are developed, the flexibility index for each design is estimated. This is accomplished using *N=10,000* samples (in θ-space) for each design. For all 9 feasible designs this requires only 24. 6 cpu-s. The values of $\hat{F}(d)$, $\hat{F}^L(d)$, and $\hat{F}^U(d)$ for each of the feasible designs is given in

Table 15. The lower bound on the flexibility is set by the largest $\hat{F}^L(d)$ among the candidate designs. This corresponds to the worst-case flexibility for the best current design. Any design that has a value of $\hat{F}^U(d)$ lower than this lower bound is removed from the candidate set, because it does not have the potential to be more flexible than the current best design. After this stage the set of 9 candidate designs is reduced to 8, as the design (0,2) is removed.

Table 15: Initial estimated flexibility indices for each nominally feasible design in the linear test problem (75), along with upper and lower bounds

Design d1, d2	$\hat{F}(d)$	$\hat{F}^U(d)$	$\hat{F}^L(d)$
0, 1	0.3990	0.6833	0.1791
0, 2	0.1719	0.3895	0.0328
1, 1	0.5892	0.8176	0.1275
1, 2	0.3331	0.5710	0.1464
2, 1	0.3186	0.7806	0.0222
2, 2	0.5010	0.6463	0.2775
3, 2	0.5660	0.6535	0.4276
4, 2	0.5660	0.6535	0.3348
5, 2	0.3648	0.6458	0.0541

During the surrogate-based flexibility stage the criterion for the difference between the upper and lower bounds was selected to be $\varepsilon_1=0.05$. A total of 65 iterations of the flexibility stage are completed, corresponding to 241 model evaluations. The final candidate set consists of the three designs shown in Table 16. All three have similar estimated flexibility indices, though the design (1,1) appears to be slightly more flexible than the others. A secondary criteria, such as cost, could be used to further select between designs.

Table 16: Final candidate set of designs for linear flexibility test problem shown in equation (75)

Design d1 d2	$\hat{F}(d)$	$\hat{F}^U(d)$	$\hat{F}^L(d)$
1 1	0.5936	0.6625	0.5006
3 2	0.5666	0.5777	0.5538
4 2	0.5661	0.5769	0.5559

The contours of the feasible region for the final set of designs are shown in Figure 47. It can be seen that the surrogate model provides an accurate estimate of the feasible region for the final set of candidate designs. In addition, it is clear that the three designs have comparable feasible regions, with designs (3,2) and (4,2) being nearly identical.

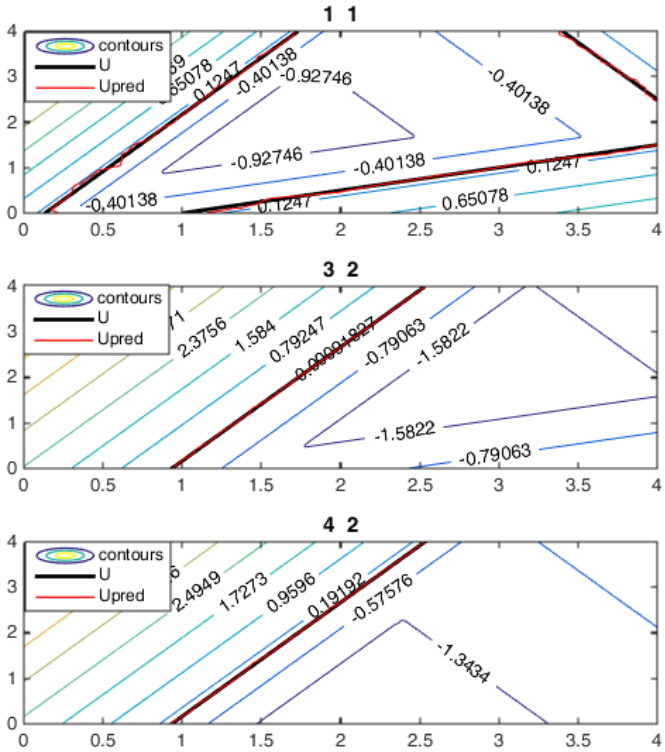

Figure 47: Contours of the feasibility function for the final set of candidate designs for the linear test problem (75). This includes the true contours of the feasible region, the true boundary of the feasible region (shown in black) and the contours of the feasible region predicted by the surrogate model (shown in red).

Nonlinear design problem

The proposed algorithm is also demonstrated for test problem (76), which is nonlinear in the design and uncertain variables.

$$
\begin{aligned}
&g1: -0.5\theta_1 + 0.5\theta_2 + d_1 - 3.5d_2 \le 0\\
&g2: -\frac{4\theta_1}{d_1} - \theta_2 + \frac{1}{d_1} + 2d_2 \le 0\\
&g3: 0.5\theta_1 - \theta_2 - d_1 + 0.5d_2 + d_2^2 \le 0\\
&0 \le \theta_1, \theta_2 \le 4\\
&0 \le d_1, d_2 \le 5\\
&\theta_1^N = 2\\
&\theta_2^N = 2\\
&d_1, d_2 \text{ are integer variables}
\end{aligned}
\qquad (76)
$$

The set of possible designs for this problem also consists of 36 combinations, which are shown in Figure 48.

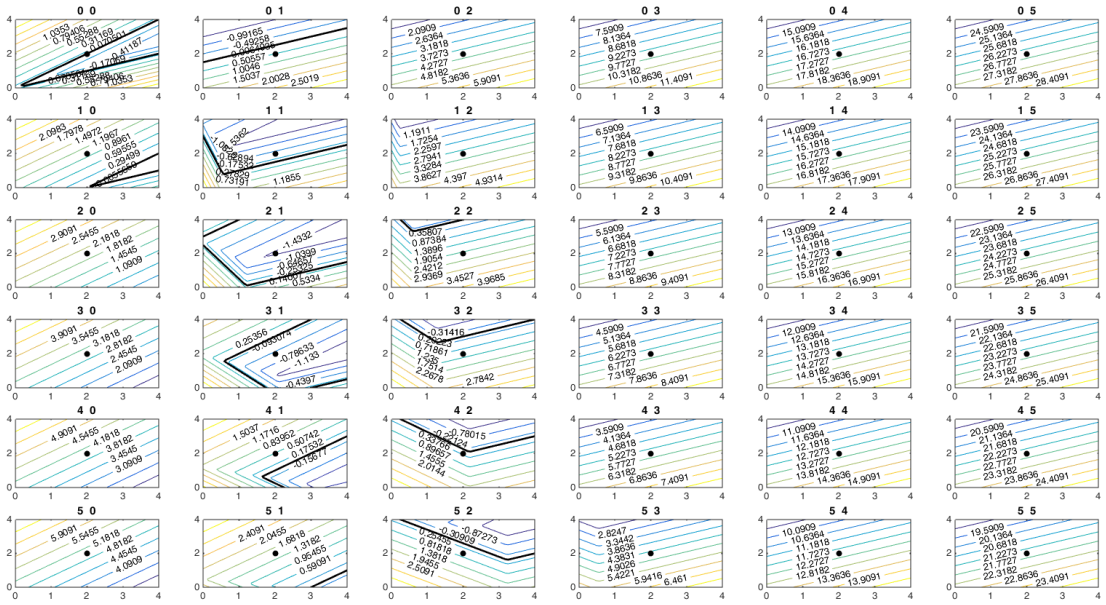

Figure 48: Contours of feasibility function for each of the designs in problem (76). For each design, the boundary of the feasible region is shown in black. The nominal operating conditions are indicated by a black dot.

Of the 36 designs, only four are feasible at the nominal operating conditions. These are correctly identified during the initial feasibility stage and are shown in Table 17. The initial model is built using a set of 36 samples, 9 per candidate design, as for the linear case study. Preliminary evaluation of the estimated flexibility index for the feasible designs is conducted using 10,000 samples, which evaluate in 11.5 cpu-s. The preliminary evaluation of the flexibility, as well as upper and lower bounds, is given in *Table 17*. The current best lower bound is 0.4803. All designs have a better upper bound than the current best lower bound, so they are all candidate designs for the surrogate-based flexibility stage.

Table 17: Initial estimated flexibility indices for each nominally feasible design in the nonlinear test problem (76), along with upper and lower bounds

Design d1 d2	$\hat{F}(d)$	$\hat{F}^U(d)$	$\hat{F}^L(d)$
0 0	0.2550	0.6170	0.0074
1 1	0.5788	0.7210	0.4389
2 1	0.7250	0.8756	0.4803
3 1	0.5931	0.8092	0.3229

Using the criterion $\varepsilon_1=0.05$, the surrogate based flexibility stage converges in 30 iterations, after collecting a total of 87 samples of the original process model. The final candidate set consists of a single design (2,1), which has an estimated flexibility of 0.724 and upper and lower bounds of 0.783 and 0.660 respectively. The feasible region corresponding to the most flexible design is shown in Figure 49. It can be seen that the surrogate feasibility model accurately represents the feasible region for this design.

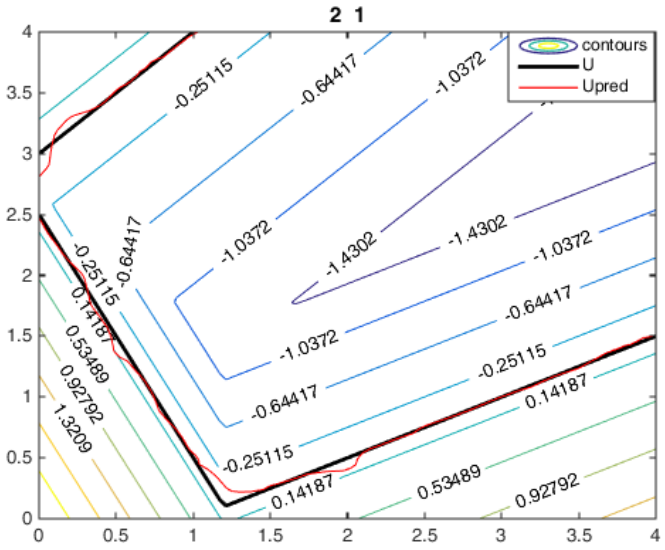

Figure 49: Contours of the feasibility function for the most flexible design for the nonlinear test problem (76). This includes the true contours of the feasible region, the true boundary of the feasible region (shown in black) and the contours of the feasible region predicted by the surrogate model (shown in red).

6.4. Design of a pharmaceutical blending process under uncertainty

The proposed method for surrogate-based design under uncertainty has been applied to a continuous feeding and blending process. The process model used in this application is the same model demonstrated in Chapter 3. The design variables considered are the feed hopper volume V and the blender length L. The uncertain parameters considered include the noise in the feed rate and the powder properties, specifically the compressibility. The agitator rotation rate in the blender is a control variable that can be used to mitigate the effect of feed rate fluctuations. The feed hopper volume affects the refill frequency (RF), which is measured in the number of times the feeder must be refilled per hour. It also influences the magnitude of each refill, which in turn affects the amount that the powder at the base of the feed hopper densifies during refill. The extent to which the powder densifies is also affected by the compressibility of the powder. These two factors contribute to disturbances in feed rate during refill of the feed hopper. The blender length and agitator rotation rate influence mixing performance, which affects the relative standard deviation (RSD) at the blender exit. The feasibility is limited by constraints on the blender capacity, the total flow rate through the system, the active ingredient (API) composition of the blend and the relative standard deviation (RSD) of the blend. The objective is to identify a flexible process design at the lowest possible cost. The costs considered include purchased equipment costs and the marginal labor cost associated with operating the feeder at different refill frequencies. Material costs are not included, as the process is operated at a constant flow rate set point.

The surrogate-based process design problem for this application can be formulated as in equation (77). The total cost is calculated using the cost factors and equations presented in Table 18. Note that the marginal labor cost associated with feeder refills is based on the refill frequency. The scaling factor of 0.05 is selected to ensure that labor costs are on the same order of magnitude as capital costs. A more extensive study of labor costs for continuous pharmaceutical manufacturing processes is required to more accurately estimate this factor. (Schaber et al., 2011)

$$\min_{L,V,rpm} C^{total}$$
$$s.t.\ \hat{F}(d) \geq \varepsilon \tag{77}$$

Table 18: Cost factors for continuous feeding and blending case study

Item	Size/scale factor	Cost equation ($)	Reference
Ribbon mixer	Volume, V, ft^3	$C_p^1 = 1,700V^{0.60}$	(F. Boukouvala & Ierapetritou, 2013)
Screw feeders	Vol. flow rate, ft^3/hr	$C_p^2 = 760S^{0.22}$	(F. Boukouvala & Ierapetritou, 2013)
Feed Hoppers	Volume,V, ft^3	$C_p^3 = 450V^{0.46}$	(F. Boukouvala & Ierapetritou, 2013)
Labor	Refill frequency (function of feed hopper size)	$C_{labor} = 160,000 \times RF \times 0.05$	(Schaber et al., 2011)
Total cost	Purchased equipment + annual marginal labor	$C^{total} = C_p^1 + C_p^2 + C_p^3 + C_{labor}$	--

The ranges of the design, control and uncertain parameters considered in this case study are shown in Table 19. The form of the feasibility constraints is shown in equation (78) and the corresponding limits on the design and operating variables are shown in Table 20.

Table 19: Summary of design, control and uncertain parameters for continuous blending case study

Variable	Type	Nominal value	Range or Levels
Feed hopper volume (m^3)	Design	--	0.0667, 0.1, 0.2
Blender Length (m)	Design	--	0.15, 0.3, 0.45
Powder densification during refill (%)	Uncertain	30	15-45
Feed rate variability (% of set point)	Uncertain	3	1-5
Blender blade speed (rpm)	Control	180	100 - 250

Operating constraints

$$FR^L \le FR \le FR^U$$

$$BF^L \le BF \le BF^U$$

Quality constraints

$$RSD \le RSD^U$$

$$C_{API}^L \le C_{API} \le C_{API}^U \quad (78)$$

Table 20: constraints on operating parameters and quality attributes for continuous blending case study

Variable	Symbol	Type	Target	Lower bound	Upper bound
Net flow rate (kg/h)	FR	Operating	40	35	45
Mixer fill level (% working volume)	BF	Operating	--	10	50
API Concentration (w/w)	C_{API}	Quality	0.3	0.27	0.33
Relative standard deviation (RSD)	RSD	Quality	--	--	0.6

The 9 possible designs for this problem, along with their nominal feasibility values, are shown in Table 21. Note that designs corresponding to a blender length of 0.15 meters are infeasible at the nominal operating conditions. For these designs the capacity constraints on the blender are exceeded. Therefore they are excluded from the set of candidate designs for the surrogate based flexibility phase.

Table 21: Designs and nominal feasibility values for the continuous blending case study

Feeder volume (m3)	Blender length (m)	Nominal Feasibility
0.2	0.15	0.0031
0.2	0.3	-0.0007
0.2	0.45	-0.0003
0.1	0.15	0.0011
0.1	0.3	-0.0006
0.1	0.45	-0.0003
0.0667	0.15	0.0007
0.0667	0.3	-0.0006
0.0667	0.45	-0.0003

Next, a surrogate model for the feasibility function and a surrogate model for the cost are developed. Upper and lower bounds on the cost can be determined based on the kriging variance of the cost model, just as they are in equation (74) for the flexibility index. The initial design consists of 54 samples, 9 per candidate design. The surrogate model is used to estimate the flexibility index and cost for each of the candidate designs. For this case study, a surrogate-based minimization problem is embedded in the flexibility calculation to manipulate the control variable (blender rotation rate) in such a way that constraint violations are minimized. Based on the prediction variance for the cost model, which is greater than that for the feasibility model, a minimum of 125 samples is required to estimate the cost. To be conservative, 500 samples are used to estimate the flexibility and cost. The resulting flexibility indices and costs, along with lower and upper bounds, are shown in Table 22.

Table 22: Initial estimated flexibility indices and costs for the set of feasible candidate designs

Feeder volume (m3)	Blender length (m)	$\hat{F}(d)$	$\hat{F}^L(d)$	$\hat{F}^U(d)$	$\hat{C}(d)$ ($x1000)	$\hat{C}^L(d)$	$\hat{C}^U(d)$
0.2	0.3	0.0016	0.0100	0.9700	14.8674	14.8243	14.9104
0.2	0.45	0.4464	0.0200	0.9800	14.9549	14.9119	14.9979
0.1	0.3	0.5228	0.0400	1.0000	21.9546	21.9175	21.9917
0.1	0.45	0.9963	0.0200	0.9800	22.0409	22.0038	22.0781
0.0667	0.3	0.9969	0.0300	0.9900	29.5154	29.4739	29.5568
0.0667	0.45	0.9981	0.0100	0.9700	29.6012	29.5598	29.6426

Initially the bounds for the estimated flexibility are wide, so it is not possible to eliminate any designs based on the preliminary flexibility estimates. During the surrogate-based flexibility stage, designs with a lower bound on estimated flexibility of less than 0.2 are eliminated. In addition, any design with cost lower bound ($\hat{C}^L$) greater than the upper bound ($\hat{C}^U$) of the best design that meets the flexibility criteria is eliminated as it cannot be optimal with respect to cost.

After collecting an additional 89 samples during the surrogate-based flexibility phase, it is possible to identify a subset of designs that meet the flexibility requirements. These are shown in Table 23 along with their corresponding cost and flexibility estimates. The most flexible design

corresponds to a more expensive alternative, but there is a less than one percent cost difference between the two designs. The two most flexible design alternatives correspond to the smallest feed hopper and therefore the most frequent feeder refills. In this case, the more frequent refills result in smaller disturbances during refill and therefore reduce the probability of exceeding the specifications on RSD or API concentration of the blend. The most flexible design corresponds to a larger blender, as this has a broader residence time distribution (RTD) which can dampen feed rate fluctuations more effectively.

Table 23: Final set of candidate designs and corresponding flexibilities and costs

	Feeder volume (m3)	Blender length (m)	$\hat{F}(d)$	$\hat{F}^L(d)$	$\hat{F}^U(d)$	$\hat{C}(d)$ ($x1000)	$\hat{C}^L(d)$	$\hat{C}^U(d)$
Design1	0.0667	0.3	0.2361	0.2277	0.2463	29.5182	29.4940	29.5424
Design2	0.0667	0.45	0.2821	0.2703	0.2944	29.6000	29.5720	29.6279

The feasible operating regions for both of the candidate designs are shown in the Figure 50. The less expensive, but also less flexible design is show in Figure 50a while the more expensive, but also more flexible design is shown in Figure 50b. Based on observed feasible regions, it may be preferable to select Design 2, as it is defined by nearly linear constraints which will likely be more straightforward to enforce during process operation.

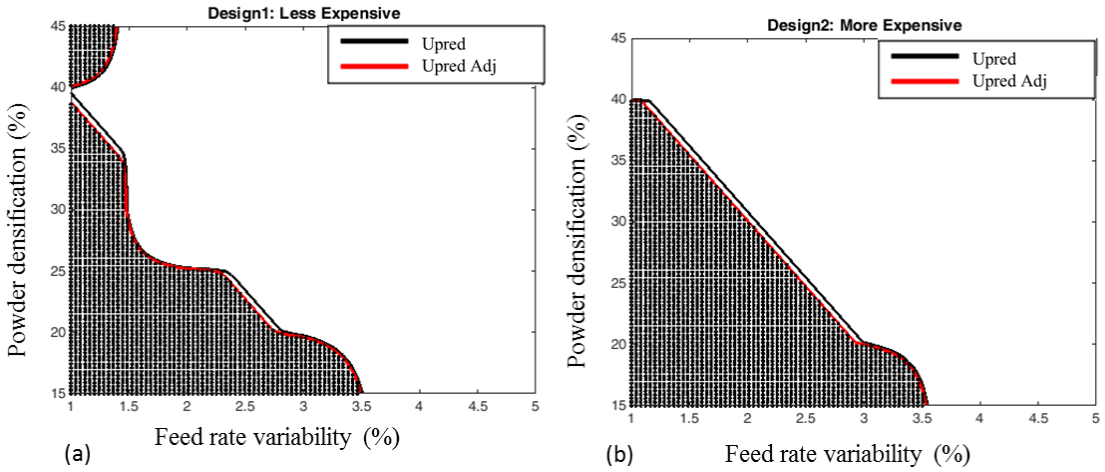

Figure 50: Design space for the two most flexible designs with respect to the uncertain parameters (feed rate variability, powder densification). The boundary of the predicted feasible region is shown in black, while the adjusted feasibility (see Chapter 5) is shown in red. a) indicates the less expensive design while b)indicates the more expensive but also more flexible design

6.5. Conclusions and future work

In this chapter, a novel, surrogate-based method for estimating process flexibility has been introduced. This method can be applied to both deterministic uncertainties and stochastic uncertainties characterized by arbitrary probability distributions. The proposed approach has been demonstrated on several test problems, including the steady state operation of a pharmaceutical roller compaction process. A surrogate-based algorithm for optimal design under uncertainty has been proposed based on the estimated flexibility index. This method differs from previous efforts at surrogate-based process synthesis in that the design variables are explicitly included in the surrogate model. The proposed algorithm has been demonstrated for both linear and nonlinear problems, and has been applied to the selection of design alternatives for a continuous pharmaceutical feeding and blending process.

A great deal of work remains to be done in the area of surrogate-based process design under uncertainty, particularly as it applies to pharmaceutical processes. For instance, more accurate data for the estimation of operating costs for continuous pharmaceutical processes are needed. In

addition, the proposed surrogate-based algorithm for design under uncertainty needs to be tested for larger design problems involving hundreds of design alternatives. These problems can be used to evaluate the initial feasibility approach based on surrogate modeling, which is described in section 6.3 but has not been tested in this work. In addition, the proposed method can be expanded to more structured synthesis problems (e.g. those modeled using a state-task-network)(Henao & Maravelias, 2010, 2011). This would require an additional branching and bounding step at the superstructure level during the surrogate-based flexibility stage.

Chapter 7

7. Conclusions and future perspectives

The primary objective of this dissertation has been to leverage concepts from process systems engineering (PSE) to help achieve the aim of Quality-by-Design in pharmaceutical process development. Specifically this has included the use of process modeling, especially reduced-order modeling, as well as sensitivity analysis, feasibility and flexibility analysis and process design under uncertainty. Each of these tools can contribute to enhanced process understanding and ultimately to the design of robust and well-understood pharmaceutical manufacturing processes. Modeling is integral to the implementation of the process analysis and optimization as a whole. In a process development context, integrated flowsheet models can supplement experimental studies during process scale-up and can be used evaluate design alternatives and control strategies prior to implementing these on-scale. (K. V. Gernaey & Gani, 2010; Singh et al., 2012; Troup & Georgakis, 2013) Within these integrated models, reduced-order modeling can bridge the gap between high fidelity simulations and the semi-empirical unit operation models.(F. Boukouvala, Gao, et al., 2013; Rogers & Ierapetritou, 2014) Once integrated process models have been developed, they can be used for process analysis. Sensitivity analysis can be conducted to identify potential critical process parameters, as well as manipulated variables for quality control strategies.(F. Boukouvala et al., 2012; Rogers et al., 2014) This has implications for quality risk assessment, as well as for establishing process design space. Feasibility analysis is also useful in determining design space. The results of feasibility analysis can be used to identify the edges of failure for a process with respect to product quality.(F. Boukouvala et al., 2010a; MacGregor & Bruwer, 2008) Knowing the shape and size of the feasible operating region is useful when establishing process design space.(Garcia et al., 2008) Finally, process flexibility analysis can be used to quantify process robustness.(Halemane, 1983) In conjunction with algorithms for optimal

design under uncertainty, this information can be used to design processes that are robust and cost effective.(Pistikopoulos & Ierapetritou, 1995)

In order to implement these PSE concepts, it has been necessary to develop and implement mathematical tools for process modeling, analysis and optimization. This has been challenging due to the complex nature of the solids-based processes considered in this dissertation. Particulate processes are known to be difficult to model for a number of reasons.(Muir Wood, 2008) Lack of understanding related to physical properties and flow behavior of pharmaceutical powders exacerbates this issue for pharmaceutical processes.(P. McKenzie et al., 2006; Muzzio et al., 2002) The modeling techniques that can most accurately represent these particulate processes (e.g. DEM), also tend to be computationally intensive.(Ketterhagen et al., 2009) As such they may not be suitable for many applications, including integrated process modeling, sensitivity analysis and optimization. Even methods that are well-suited to flowsheet modeling, like population balance (PBM) and residence time distribution (RTD) models, require experimental data (Gao, Vanarase, et al., 2011) or information from DEM(Wassgren et al., 2011) to ensure accuracy.

It is also difficult to apply commercially available algorithms for process optimization to pharmaceutical process models. This is due to the complex and often dynamic nature of flowsheet simulations for solids-based processes. (F. Boukouvala, Chaudhury, et al., 2013; F. Boukouvala et al., 2012) Therefore computationally efficient methods for optimization of complex process models must be developed.(F. Boukouvala & Ierapetritou, 2013; Caballero & Grossmann, 2008; Henao & Maravelias, 2011; Müller et al., 2013) Developing such methods presents its own challenges, including the need to develop efficient surrogate models, to balance sampling requirements with model accuracy (Jones, 2001; Jones et al., 1998) and the incorporation of branch and bound methods for problems involving both discrete and continuous variables.(Davis & Ierapetritou, 2009; Müller et al., 2013)

In this work, efforts have been made to address many of these challenges. A method for dynamic reduced-order modeling of high dimensional systems has been introduced, which can be used to incorporate information from DEM simulations into flowsheet models at low computational cost. This method has been applied to model a continuous blender, and the resulting flowsheet simulation demonstrates reasonable agreement with experimental systems. A surrogate-based algorithm for feasibility analysis has also been developed. This method has been shown to generate accurate representations of the feasible region for complex and dynamic processes. It is also computationally efficient relative to surrogate-based methods presented in prior work.(F. Boukouvala & M. G. Ierapetritou, 2012) The use of surrogate models to estimate flexibility indices for processes with both deterministic and stochastic uncertainties has also been presented in this thesis, for what is believed to be the first time. Surrogate-based flexibility analysis has been incorporated into an algorithm for process design under uncertainty. This algorithm is unique in the area of surrogate-based process synthesis in that it employs reduced-order modeling for both discrete and continuous variables.(Davis & Ierapetritou, 2009; Henao & Maravelias, 2011)

The role of PSE concepts in Quality-by-Design based pharmaceutical development has been demonstrated with case studies throughout this dissertation. Sensitivity analysis has been used to identify potential critical process parameters and control strategies for a continuous direct compaction process. Feasibility analysis has been applied to steady state and dynamic roller compaction processes to guide the selection of design space for this unit operation. Flexibility analysis has been demonstrated for a roller compaction process and also an integrated feeding and blending system. In the context of the roller compaction process the flexibility index has been used to evaluate the effect of different types of uncertainties (stochastic vs. deterministic) on process robustness. For the feeding and blending system, feasibility and flexibility analysis have been combined with a surrogate-based algorithm for process design under uncertainty to aid in the selection of a robust process design.

This dissertation has endeavored to develop methods that will facilitate the increased use of process systems engineering tools in the pharmaceutical industry. However there are many areas for future work that can improve and enhance these tools. For pharmaceutical process modeling, significant effort should be directed to the development and experimental validation of more predictive integrated models. Specifically, better models are needed to account for the influence of material properties on process performance in pharmaceutical flowsheeting applications. These models should consider not only raw material properties, but also the effect of mechanical forces like shear on the properties of pharmaceutical powders. In addition, further study is needed in the area of PCA-based reduced-order modeling to further optimize the selection of equipment discretization and sampling frequency for DEM applications. Residence time distribution models that are more predictive, particularly with respect to micro mixing behavior, are needed for continuous blending systems. Ideally these models should be able to incorporate the influence of material properties on parameters in the RTD model. In addition, residence time distributions for other unit operations, especially loss-in-weight feeders, hoppers and feed frames are needed to more accurately capture the dynamics of integrated pharmaceutical manufacturing processes. Finally, more extensive validation of flowsheet simulations against data collected from continuous manufacturing systems is needed. While integrated process models are not expected to provide exact quantitative agreement with experimental systems, it is important to demonstrate that they can capture general trends with respect to process performance and dynamics.(Chatterjee, 2008)

There is also a great deal of additional development needed in the area of efficient optimization strategies for complex processes involving black-box objectives and constraints. The most significant outstanding issue in this area is the need for rigorous proofs of global convergence for surrogate-based methods. This can be addressed in part through the use of algorithms that combine a surrogate-based global search with a trust region local search.(Agarwal & Biegler, 2013; F. Boukouvala & Ierapetritou, 2014) In the area of surrogate-based feasibility analysis,

these methods need to be demonstrated for case studies involving large numbers of uncertain parameters. Such case studies may necessitate the development of more efficient sampling criteria for model improvement. For dynamic surrogate-based feasibility analysis, sampling strategies with variable time steps should be evaluated to determine if these can improve the dynamic representation of the feasibility function. Further development of strategies for surrogate-based process synthesis is also needed. Existing strategies take advantage of surrogate models at the local optimization level, but this does not afford significant computational savings with respect to optimizing the process design. Surrogate-based algorithms for the solution of mixed-integer problems (MIP, MINLP) that can accommodate black-box functions of both continuous and discrete variables are needed in order to extend the capabilities of existing surrogate-based process synthesis algorithms.

Finally, the development of commercially available software to implement the various modeling and analysis tools described in this dissertation could result in more widespread adoption of PSE concepts in industry. The recent development of solids-processing libraries for flowsheet modeling environments like gPROMS™ and ASPEN Plus® are promising developments in this area. However more extensive model libraries are needed for pharmaceutical operations. In addition, user-friendly software that can implement sensitivity and feasibility analysis as well as efficient and robust optimization would greatly enhance the capacity for these methods to play a role in pharmaceutical process development.

Acknowledgement of Prior Works

All of the work presented in this dissertation represents original research by the author. However some of the concepts presented in Chapters 2-4 have previously appeared in published works, for which the author of this dissertation is also the first author. The text in this dissertation is not taken directly from these publications, but key concepts and figures in Chapters 2-4 may have appeared in prior works. The author of this dissertation is also first author on these prior works, which include:

Rogers, A., Hashemi, A., & Ierapetritou, M. G. (2013). Modeling of Particulate Processes for the Continuous Manufacture of Solid-Based Pharmaceutical Dosage Forms. Processes, 1, 67-127.

Rogers, A., & Ierapetritou, M. G. (2014). Discrete element reduced-order modeling of dynamic particulate systems. Aiche Journal, 60, 3184-3194.

Rogers, A., Inamdar, C., & Ierapetritou, M. G. (2014). An Integrated Approach to Simulation of Pharmaceutical Processes for Solid Drug Manufacture. Industrial & Engineering Chemistry Research, 53, 5128-5147.

Rogers, A., Ierapetritou, M. (Submitted) Challenges and Opportunities in Modeling Pharmaceutical Manufacturing Processes. Computers & Chemical Engineering. Special Issue – FOCAPD 2014. Submitted Novembr, 2014

Bibliography

Adjiman, C. S., Androulakis, I. P., & Floudas, C. A. (1997). Global optimization of MINLP problems in process synthesis and design. Computers & Chemical Engineering, 21, S445–S450.

Agarwal, A., & Biegler, L. T. (2013). A trust-region framework for constrained optimization using reduced order modeling. Optim Eng, 14, 3-35.

Agarwal, A., Biegler, L. T., & Zitney, S. E. (2008). Simulation and Optimization of Pressure Swing Adsorption Systems Using Reduced-Order Modeling. Ind. Eng. Chem. Res., 48, 2327-2343.

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. Statistics Surveys, 4, 40-79.

Bahri, P. A., Bandoni, J. A., & Romagnoli, J. A. (1997). Integrated flexibility and controllability analysis in design of chemical processes. Aiche Journal, 43, 997-1015.

Balci, O. (2010). Golden rules of verification, validation, testing, and certification of modeling and simulation applications. SCS M&S Magazine.

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

Banerjee, I., Pal, S., Maiti, S. (2010). Computationally efficient black-box modeling for feasibility analysis. Computers and Chemical Engineering, 34, 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, S817-S820.

Bansal, V., Perkins, J. D., & Pistikopoulos, E. N. (2000). Flexibility analysis and design of linear systems by parametric programming. Aiche Journal, 46, 335-354.

Bansal, V., Perkins, J. D., & Pistikopoulos, E. N. (2002). Flexibility analysis and design using a parametric programming framework. Aiche Journal, 48, 2851-2868.

Barasso, D., & Ramachandran, R. (2012). A comparison of model order reduction techniques for a four-dimensional population balance model describing multi-component wet granulation process. Chemical Engineering Science, 80, 380-392.

Basheer, I. A., & Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. J Microbiol Methods, 43, 3-31.

Basu, P., Joglekar, G., Rai, S., Suresh, P., & Vernon, J. (2008). Analysis of Manufacturing Costs in Pharmaceutical Companies. Journal of Pharmaceutical Innovation, 30-40.

Benyahia, B., Lakerveld, R., & Barton, P. I. (2012). A Plant-Wide Dynamic Model of a Continuous Pharmaceutical Process. Industrial & Engineering Chemistry Research, 51, 15393-15412.

Berthiaux, H., Marikh, K., & Gatumel, C. (2008). Continuous mixing of powder mixtures with pharmaceutical process constraints. Chemical Engineering and Processing, 47, 2315-2322.

Biegler, L. T. (2007). An overview of simultaneous strategies for dynamic optimization. Chemical Engineering and Processing, 46, 1043-1053.

Biegler, L. T., & Grossmann, I. E. (2004). Retrospective on optimization. Computers & Chemical Engineering, 28, 1169-1192.

Biegler, L. T., Grossmann, I. E., & Westerberg, A. W. (1997). Systematic Methods of Chemical Process Design. In. Upper Saddle River, NJ: Prentice Hall.

Biegler, L. T., Lang, Y. D., & Lin, W. (2014). Multi-scale optimization for process systems engineering. Computers & Chemical Engineering, 60, 17-30.

Bindhumadhavan, G., Seville, J. P. K., Adams, M. J., Greenwood, R. W., & Fitzpatrick, S. (2005). Roll compaction of a pharmaceutical excipient: Experimental validation of rolling theory for granular solids. Chemical Engineering Science, 60, 3891 – 3897.

Blower, S. M., & Dowlatabadi, H. (1994). Sensitivity and Uncertainty Analysis of Complex-Models of Disease Transmission - an Hiv Model, as an Example. International Statistical Review, 62, 229-243.

Boukouvala, F., Chaudhury, A., Sen, M., Zhou, R. J., Mioduszewski, L., Ierapetritou, M. G., & Ramachandran, R. (2013). Computer-Aided Flowsheet Simulation of a Pharmaceutical Tablet Manufacturing Process Incorporating Wet Granulation. Journal of Pharmaceutical Innovation, 8, 11-27.

Boukouvala, F., Dubey, A., Vanarase, A., Ramachandran, R., Muzzio, F.J., Ierapetritou, M. (2012). Computational Approaches for Studying the Granular Dynamics of Continuous Blending Processes, 2 – Population Balance and Data-Based Methods. Macromolecular Materials and Engineering, 297, 9-19.

Boukouvala, F., Gao, Y. J., Muzzio, F., & Ierapetritou, M. G. (2013). Reduced-order discrete element method modeling. Chemical Engineering Science, 95, 12-26.

Boukouvala, F., & Ierapetritou, M. (2012). Simulation-Based Derivative-Free Optimization for Computationally Expensive Function. In AIChE Annual Meeting. Pittsburgh, PA: AIChE.

Boukouvala, F., & Ierapetritou, M. G. (2012). Feasibility analysis of black-box processes using an adaptive sampling Kriging-based method. Computers & Chemical Engineering, 36, 358-368.

Boukouvala, F., & Ierapetritou, M. G. (2013). Surrogate-Based Optimization of Expensive Flowsheet Modeling for Continuous Pharmaceutical Manufacturing. Journal of Pharmaceutical Innovation, 8, 131-145.

Boukouvala, F., & Ierapetritou, M. G. (2014). Derivative-free optimization for expensive constrained problems using a novel expected improvement objective function. Aiche Journal, n/a-n/a.

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

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

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

Boukouvala, F., Niotis, V., Ramachandran, R., Muzzio, F. J., & Ierapetritou, M. G. (2012). An integrated approach for dynamic flowsheet modeling and sensitivity analysis of a continuous tablet manufacturing process. Computers & Chemical Engineering, 42, 30-47.

Box, G. E. P., & Wilson, K. B. (1951). On the Experimental Attainment of Optimum Conditions. Journal of the Royal Statistical Society Series B (Methodological), 13, 1-35.

Bratley, P., & Fox, B. L. (1988). Algorithm 659 Implementing Sobol's Quasirandom Sequence Generato. ACM Transactions on Mathematical Software, 14, 88-100.

Buchholz, S. (2010). Future manufacturing approaches in the chemical and pharmaceutical industry. Chemical Engineering and Processing, 49, 993-995.

Burnham, A. J., MacGregor, J. F., & Viveros, R. (1999). Latent variable multivariate regression modeling. Chemometrics and Intelligent Laboratory Systems, 48, 167-180.

Caballero, J. A., & Grossmann, I. E. (2008). An Algorithm for the Use of Surrogate Models in Modular Flowsheet Optimization. Aiche Journal, 54, 2633-2650.

Calder, C. A., & Cressie, N. (2009). Kriging and Variogram Models. In International Encyclopedia of Human Geography (pp. 49-55): Elsevier Science.

Chatterjee, S. (2008). Overview of Models Used in Design Space Determination: A Regulatory Perspective. In AIChE Annual Meeting. Philadelphia, PA: AIChE.

Chen, W., Chang, S. Y., Kiang, S., Marchut, A., Lyngberg, O., Wang, J., Rao, V., Desai, D., Stamato, H., & Early, W. (2010). Modeling of pan coating processes: Prediction of tablet content uniformity and determination of critical process parameters. J Pharm Sci, 99, 3213-3225.

Congdon, P. (2006). Bayesian Statistical Modeling (2 ed.). Chichester, West Sussex, England: John Wiley & Sons, Ltd.

Corredor, C. C., Bu, D., & Both, D. (2011). Comparison of near infrared and microwave resonance sensors for at-line moisture determination in powders and tablets. Anal Chim Acta, 696, 84-93.

Cryer, S. A., & Scherer, P. N. (2003). Observations and process parameter sensitivities in fluid-bed granulation. Aiche Journal, 49, 2802-2809.

Cukier, R. I., Fortuin, C. M., Shuler, K. E., Petschek, A. G., & Schaibly, J. H. (1973). Study Of Sensitivity Of Coupled Reaction Systems To Uncertainties In Rate Coefficients 1. Theory. Journal Of Chemical Physics, 59, 3873-3878.

Cukier, R. I., Levine, H. B., & Shuler, K. E. (1978). Nonlinear Sensitivity Analysis of Multiparameter Model Systems. Journal of Computational Physics, 26, 1-42.

Cukier, R. I., Schaibly, J. H., & Shuler, K. E. (1975). Study Of Sensitivity Of Coupled Reaction Systems To Uncertainties In Rate Coefficients .3. Analysis Of Approximations. Journal Of Chemical Physics, 63, 1140-1149.

Danckwerts, P. V. (1953). Continuous flow systems: Distribution of residence times. Chemical Engineering Science, 2, 1-13.

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

Davis, E., & Ierapetritou, M. (2008). A Kriging-Based Approach to MINLP Containing Black-Box Models and Nois. Ind. Eng. Chem. Res., 47, 6101-6125.

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

Dec, R. T., Zavaliangos, A., & Cunningham, J. C. (2003). Comparison of various modeling methods for analysis of powder compaction in roller press. Powder Technology, 130, 265-271.

Degerman, M., Westerberg, K., & Nilsson, B. (2009). A Model-Based Approach to Determine the Design Space of Preparative Chromatography. Chemical Engineering & Technology, 32, 1195-1202.

Dimitriadis, V. D., & Pistikopoulos, E. N. (1995). Flexibility Analysis of Dynamic-Systems. Industrial & Engineering Chemistry Research, 34, 4451-4462.

Dimov, I., & Georgieva, R. (2010). Monte Carlo algorithms for evaluating Sobol' sensitivity indices. Mathematics and Computers in Simulation, 81, 506-514.

Dua, V., Bozinis, N. A., & Pistikopoulos, E. N. (2002). A multiparametric programming approach for mixed-integer quadratic engineering problems. Computers & Chemical Engineering, 26, 715-733.

Dua, V., & Pistikopoulos, E. N. (1998). Optimization Techniques for Process Synthesis and Material Design Under Uncertainty. Chemical Engineering Research & Design, 76, 408-416.

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

Dubey, A., Vanarase, A. U., & Muzzio, F. J. (2012). Impact of Process Parameters on Critical Performance Attributes of a Continuous Blender—A DEM-Based Study. Aiche Journal, 58, 3676-3684.

Duchesne, C., & Macgregor, J. F. (2004). Establishing multivariate specification regions for incoming materials. Journal of Quality Technology, 36, 78-94.

Dyment, J., Reimers, C., & Beck, R. (2013). Overview of Solids Modeling for Chemical Processes. In. Burlington, MA: Aspen Technology Inc.

Engisch, W. E., & Muzzio, F. J. (2010). Hopper Refill of Loss-in-Weight Feeding Equipment. In AIChE Annual Meeting. Salt Lake City, UT.

Engisch, W. E., & Muzzio, F. J. (2012). Method for characterization of loss-in-weight feeder equipment. Powder Technology, 228, 395-403.

Faldu, B., Sharma, A., Sharma, A., & Chauhan, C. S. (2012). Roller Compaction: Imperative Process for Tablet Manufacturing: A Review. International Journal of Pharmaceutical Research and Development, 4, 40-47.

FDA. (1995). Guidance for industry: Immediate release solid oral dosage forms scale-up and postapproval changes: Chemistry, manufacturing, and controls, in vitro dissolution testing, and in vivo bioequivalence documentation. In F. a. D. A. CDER (Ed.). USA.

FDA. (1997a). Guidance for industry: Modified release solid oral dosage forms scale-up and postapproval changes: Chemistry, manufacturing, and controls, in vitro dissolution testing, and in vivo bioequivalence documentation. In F. a. D. A. CDER (Ed.). USA.

FDA. (1997b). Guidance for industry: Nonsterile semisolid dosage forms scale-up and postapproval changes: chemistry, manufacturing, and controls, in vitro dissolution testing, and in vivo bioequivalence documentation. In F. a. D. A. CDER (Ed.). USA.

Floudas, C. A. (2000). Nonconvex Optimization and Its Applications. In C. A. Floudas (Ed.), Deterministic Global Optimization: Theory, Methods and Applications (Vol. 37). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Floudas, C. A., & Gounaris, C. E. (2009). A review of recent advances in global optimization. Journal of Global Optimization, 45, 3-38.

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

Gabrielsson, J., Lindberg, N.-O., & Lundstedt, T. (2012). Multivariate methods in pharmaceutical applications. Journal of Chemometrics, 16, 141-160.

Gantt, J. A., & Gatzke, E. P. (2006). A stochastic technique for multidimensional granulation modeling. Aiche Journal, 52, 3067-3077.

Gao, Y. J., Ierapetritou, M., & Muzzio, F. (2012). Periodic Section Modeling of Convective Continuous Powder Mixing Processes. Aiche Journal, 58, 69-78.

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

Gao, Y. J., Muzzio, F. J., & Ierapetritou, M. G. (2012). Optimizing continuous powder mixing processes using periodic section modeling. Chemical Engineering Science, 80, 70-80.

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

Garcia-Munoz, S. (2014). Two novel methods to analyze the combined effect of multiple raw-materials and processing conditions on the product's final attributes: JRPLS and TPLS. Chemometrics and Intelligent Laboratory Systems, 133, 49-62.

García-Muñoz, S. (2009). Establishing multivariate specifications for incoming materials using data from multiple scales. Chemometrics and Intelligent Laboratory Systems, 98, 51-57.

García-Muñoz, S., & Polizzi, M. (2012). WSPLS — A new approach towards mixture modeling and accelerated product development. Chemometrics and Intelligent Laboratory Systems, 15, 116-121.
Garcia, T., Cook, G., & Nozal, R. (2008). PQLI key topics - criticality, design space, and control strategy. 3, 2.
Gavi, E., & Reynolds, G. (2014). System model of a tablet manufacturing process. Computers & Chemical Engineering, 71, 130-140.
Gentis, N. D., & Betz, G. (2012). Compressibility of binary powder formulations: investigation and evaluation with compaction equations. J Pharm Sci, 101, 777-793.
Gernaey, K. V., Cervera-Padrell, A. E., & Woodley, J. M. (2012). A perspective on PSE in pharmaceutical process development and innovation. Computers & Chemical Engineering, 42, 15-29.
Gernaey, K. V., Cervera-Padrell, C., Woodley, J.M. (2012). A perspective on PSE in pharmaceutical process development and innovation. Computers and Chemical Engineering, 42, 15-29.
Gernaey, K. V., & Gani, R. (2010). A model-based systems approach to pharmaceutical product-process design and analysis. Chemical Engineering Science, 65, 5757-5769.
Getaz, D., Butte, A., & Morbidelli, M. (2013). Model-based design space determination of peptide chromatographic purification processes. J Chromatogr A, 1284, 80-87.
Glen, G., & Isaacs, K. (2012). Estimating Sobol sensitivity indices using correlations. Environmental Modelling & Software, 37, 157-166.
Gremaud, P. A., Matthews, J. V., & Schaeffer, D. G. (2006). On the computation of steady hopper flows III: Model comparisons. Journal of Computational Physics, 219, 443-454.
Grossmann, I. E. (1996). Mixed-Integer Optimization Techniques for Algorithmic Process Synthesis. Advances in Chemical Engineering, 23, 171-246.
Grossmann, I. E., Calfa, B., & Garcia-Herreros, P. (2014). Evolution of Concepts and Models for Quantifying Resiliency and Flexibility of Chemical Processes. Computers & Chemical Engineering.
Grossmann, I. E., & Floudas, C. A. (1987). Active Constraint Strategy for Flexibility Analysis in Chemical Processes. Computers & Chemical Engineering, 11, 675.
Grossmann, I. E., Floudas, C.A. (1987). Active constraint strategy for flexibility analysis in chemical processes. Computers & Chemical Engineering, 11, 675-693.
Grossmann, I. E., & Straub, D. A. (1996). Recent developments in the evaluation and optimization of flexible chemical processes. In G. V. Reklaitis, A. K. Sunol, D. W. T. Rippin & O. Hortacsu (Eds.), Batch Processing Systems Engineering Fundamentals and Applications for Chemical Engineering (Vol. 143, pp. 495-516). Berlin Heidelberg: Springer in collaboration with NATO Scientific Affairs Division.
Guillén-Gosálbez, G., & Grossmann, I. E. (2009). Optimal design and planning of sustainable chemical supply chains under uncertainty. Aiche Journal, 55, 99-121.
Halemane, K. P. G., I.E. (1983). Optimal Process Design under Uncertainty. Aiche Journal, 49, 425.
Haware, R. V., Tho, I., & Baure-Brandl, A. (2009). Multivariate analysis of relationships betwee material properties, process parameters and tablet tensile strength for a-lactose monohydrates. European Journal of Pharmaceutics and Biopharmaceutics, 73, 424-431.
Heckel, R. W. (1961a). An analysis of powder compaction phenomena. Trans. Metall. Soc.AIME, 221, 1001-1008.
Heckel, R. W. (1961b). Density pressure relationships in powder compaction. Trans. Metall. Soc. AIME, 671-675.
Helton, J. C., & Davis, F. J. (2003). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliability Engineering & System Safety, 81, 23-69.

Helton, J. C., Davis, F. J., & Johnson, J. D. (2005). A comparison of uncertainty and sensitivity analysis results obtained with random and Latin hypercube sampling. Reliability Engineering & System Safety, 89, 305-330.

Henao, C. A., & Maravelias, C. T. (2010). Surrogate-Based Process Synthesis. Computer Aided Chemical Engineering, 28, 1129-1134.

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

Holt, B. R., & Morari, M. (1985a). Design of Resilient Processing Plants V-The Effect of Deadtime on Dynamic Resilience. Chemical Engineering Science, 40, 1229.

Holt, B. R., & Morari, M. (1985b). Design of Resilient Processing Plants VI-The Effect of Right Half Plane Zeros on Dynamic Resilience. Chemical Engineering Science, 40, 59.

Homma, T., & Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. Reliability Engineering & System Safety, 52, 1-17.

Hsu, S. H., Reklaitis, G. V., & Venkatasubramanian, V. (2010a). Modeling and Control of Roller Compaction for Pharmaceutical Manufacturing Part II: Control System Design. Journal of Pharmaceutical Innovation, 5, 24-36.

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

ICH. (2005a). ICH Q9 Quality Risk Management.

ICH. (2005b). ICH Q10 Pharmaceutical Quality Systems.

ICH. (2009). Guidance for Industry Q8(R2) Pharmaceutical Development.

Jackson, S., Sinka, I. C., & Cocks, A. C. (2007). The effect of suction during die fill on a rotary tablet press. Eur J Pharm Biopharm, 65, 253-256.

Jansen, M. J. W. (1999). Analysis of variance designs for model output. Computer Physics Communications, 117, 35-43.

Järvinen, M. A., Paaso, J., Paavola, M., Leivisk, K., Juuti, M., Muzzio, F., & Järvinen, K. (2012). Continuous direct tablet compression: effects of impeller rotation rate, total feed rate and drug content on the tablet properties and drug release. Drug Development and Industrial Pharmacy, 1-7.

Järvinen, M. A., Paaso, J., Paavola, M., Leivisk, K., Juuti, M., Muzzio, F., Järvinen, K. (2012). Continuous direct tablet compression: effects of impellerrotation rate, total feed rate and drug content on the tablet properties and drug release. Drug Development and Industrial Pharmacy, 1-7.

Jia, Z. Y., Davis, E., Muzzio, F. J., & Ierapetritou, M. G. (2009). Predictive Modeling for Pharmaceutical Processes Using Kriging and Response Surface. Journal of Pharmaceutical Innovation, 4, 174-186.

Joe, S., & Kuo, F. Y. (2003). Remark on algorithm 659: Implementing Sobol's quasirandom sequence generator. ACM Transactions on Mathematical Software, 29, 49-57.

Johanson, J. R. (1965). A rolling theory for granular solids. ASME Journal of Applied Mechanics, E32, 842-848.

Jones, D. R. (2001). A taxonomy of global optimization methods based on response surfaces. Journal of Global Optimization, 21, 345-383.

Jones, D. R., Schonlau, M., & Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. Journal of Global Optimization, 13, 455-492.

Kawakita, K., Hattori, I., & Kishigami, M. (1977). Characteristic constants in kawakita's powder compression equation. Journal of Powder Bulk Solids Technology, 1, 3-8.

Kawakita, K., & Lüdde, K. H. (1971). Some considerations on powder compression equations. Powder Technology, 4, 61-68.

Ketterhagen, W. R., Ende, M. T. A., & Hancock, B. C. (2009). Process Modeling in the Pharmaceutical Industry using the Discrete Element Method. J Pharm Sci, 98, 442-470.

Kleijnen, J. P. C. (2009). Kriging metamodeling in simulation: A review. European Journal of Operational Research, 192, 707-716.

Kleinebudde, P. (2004). Roll compaction/dry granulation: pharmaceutical applications. European Journal of Pharmaceutics and Biopharmaceutics, 58, 317-326.

Kremer, D. M., & Hancock, B. C. (2006). Process simulation in the pharmaceutical industry: A review of some basic physical models. J Pharm Sci, 95, 517-529.

Krige, D. (1951). A statistical approach to some mine valuations and allied problems at the Witwatersrand. University of Witwatersrand, Johannesburg.

Kroonenberg, P. M. (2008). Applied Multiway Data Analysis. Hoboken, NJ: John Wiley & Sons, Inc.

Kuentz, M., Lunenberger, H. (2000). A new model for the hardness of a compacted particle systems, applied to tablets of pharmaceutical polymers. Powder Technology, 111, 145-153.

Kuentz, M. L., H. (2000). A new model for the hardness of a compacted particle systems, applied to tablets of pharmaceutical polymers. Powder Technology, 111, 145-153.

Lakerveld, R., Benyahia, B., Braatz, R. D., & Barton, P. I. (2013). Model-Based Design of a Plant-Wide Control Strategy for a Continuous Pharmaceutical Plant. Aiche Journal, 59, 3671-3685.

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. Energy & Fuels, 23, 1695-1706.

Lang, Y. D., Zitney, S. E., & Biegler, L. T. (2011). Optimization of IGCC processes with reduced order CFD models. Computers & Chemical Engineering, 35, 1705-1717.

Lecompte, T., Doremus, P., Thomas, G., & Perier-Camby, L. (2005). Dry granulation of organic powders—dependence of pressure 2D-distribution on different process parameters. Chemical Engineering Science, 60, 3933 – 3940.

Lee, S., & Grossmann, I. E. (2001). A global optimization algorithm for nonconvex generalized disjunctive programming and applications to process systems. Computers & Chemical Engineering, 25, 1675–1697.

Leopore, J., & Spavins, J. (2008). PQLI design space. Journal of Pharmaceutical Innovation, 3, 79-87.

Levis, A. A., & Papageorgiou, L. G. (2004). A hierarchical solution approach for multi-site capacity planning under uncertainty in the pharmaceutical industry. Computers & Chemical Engineering, 28, 707-725.

Li, G., Rosenthal, C., & Rabitz, H. (2001). High Dimensional Model Representations. Journal of Physical Chemistry A, 105.

Li, G., Wang, S.-W., & Rabitz, H. High Dimensional Model Representations (HDMR): Concepts and Applications.

Li, G. Y., Rabitz, H., Hu, J. S., Chen, Z., & Ju, Y. G. (2008). Regularized random-sampling high dimensional model representation (RS-HDMR). Journal of Mathematical Chemistry, 43, 1207-1232.

Li, G. Y., Wang, S. W., & Rabitz, H. (2002). Practical approaches to construct RS-HDMR component functions. Journal of Physical Chemistry A, 106, 8721-8733.

Liberge, E., & Hamdouni, A. (2010). Reduced order modelling method via proper orthogonal decomposition (POD) for flow around an oscillating cylinder. Journal of Fluids and Structures, 26, 292-311.

Lieu, T., Farhat, C., & Lesoinne, A. (2006). Reduced-order fluid/structure modeling of a complete aircraft configuration. Computer Methods in Applied Mechanics and Engineering, 195, 5730-5742.

Lima, F. V., Jia, Z., Lerapetritou, M., & Georgakis, C. (2010). Similarities and Differences Between the Concepts of Operability and Flexibility: The Steady-State Case. Aiche Journal, 56, 702-716.

Linninger, A. A., & Chakraborty, A. (2001). Pharmaceutical waste management under uncertainty. Computers & Chemical Engineering, 25, 675-681.

López-Negrete de la Fuente, R., García-Muñoz, S., & Biegler, L. T. (2010). An efficient nonlinear programming strategy for PCA models with incomplete data sets. Journal of Chemometrics, 24, 301-311.

Lophaven, S. N., Nielsen, H. B., & Søndergaard, J. (2002). DACE: A Matlab Kriging Toolbox. In Informatics and Mathematical Modeling. Denmark: Technical University of Denmark.

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

Marikh, K., Berthiaux, H., Mizonov, V., Barantseva, E., & Ponomarev, D. (2006). Flow Analysis and Markov Chain Modelling to Quantify the Agitation Effect in a Continuous Powder Mixer. Chemical Engineering Research & Design, 84, 1059-1074.

Martin, J. D. (2009). Computational Improvements to Estimating Kriging Metamodel Parameters. J. Mech. Des., 131, 084501.

Martin, M., & Grossmann, I. E. (2011). Energy optimization of hydrogen production from lignocellulosic biomass. Computers & Chemical Engineering, 35, 1798-1806.

Martínez, L., Peinado, A., Liesum, L., & Betz, G. (2013). Use of near-infrared spectroscopy to quantify drug content on a continuous blending process: Influence of mass flow and rotation speed variations. European Journal of Pharmaceutics.

McKenzie, P., Kiang, S., Tom, J., Rubin, E., & Futran, M. (2006). Can Pharmaceutical Process Development Become High Tech? Aiche Journal, 52.

McKenzie, P. K., S. Tom, J. Rubin, E., Futran, M. (2006). Can Pharmaceutical Process Development Become High Tech? AIChE Journal, 52.

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. Powder Technology, 200, 105-116.

Metheron, G. (1963). Princinples of geostatistics. Economic Geology, 58.

Mohideen, M. J., Perkins, J., & Pistikopoulos, E. N. (1996a). Optimal synthesis and design of dynamic systems under uncertainty. Computers & Chemical Engineering, 20, S895-S900.

Mohideen, M. J., Perkins, J. D., & Pistikopoulos, E. N. (1996b). Optimal Design of Dynamic Systems under Uncertainty. Aiche Journal, 42, 2251-2272.

Morari, M. (1983). Flexibility and resiliency of process systems. Computers & Chemical Engineering, 7, 423-437.

Morin, G., & Briens, L. (2013). The Effect of Lubricants on Powder Flowability for Pharmaceutical Application. AAPS PharmSciTech, 14, 1158-1168.

Mortier, S. T., De Beer, T., Gernaey, K. V., Vercruysse, J., Fonteyne, M., Remon, J. P., Vervaet, C., & Nopens, I. (2012). Mechanistic modelling of the drying behaviour of single pharmaceutical granules. Eur J Pharm Biopharm, 80, 682-689.

Muir Wood, D. (2008). Modelling granular materials: discontinuum-continuum. In J. F. Chen, J. Y. Ooi & J. G. Teng (Eds.), Structures and Granular Solids: From Scientific Principles to Engineering Applications. Leiden, The Netherlands: CRC Press/ Balkema.

Müller, J., Shoemaker, C. A., & Piche, R. (2013). SO-MI: A surrogate model algorithm for computationally expensive nonlinear mixed-integer black-box global optimization problems. Computers & Operations Research, 40, 1383-1400.

Muzzio, F. J., Shinbrot, T., & Glasser, B. J. (2002). Powder technology in the pharmaceutical industry: the need to catch up fast. Powder Technology, 124, 1-7.

Nemhauser, G., & Wolsey, L. (1999). Integer and Combinatorial Optimization. New York, NY: John Wiley & Sons, Inc.

Olsson, A. M. J., & Sandberg, G. E. (2002). Latin hypercube sampling for stochastic finite element analysis. Journal of Engineering Mechanics-Asce, 128, 121-125.

Ostrovsky, G. M., Achenie, L. E. K., & Wang, Y. (2000). A New Algorithm for Computing Process Flexibility. Ind. Eng. Chem. Res., 39, 2368-2377.

Ostrovsky, G. M., Achenie, L. E. K., Wang, Y., & Volin, Y. M. (2002). A Unique Approach for Solving Sub-Problems in Flexibility Analysis. Chemical Engineering Communications, 189, 125-149.

Ostrovsky, G. M., Volin, Y. M., Barit, E. I., & Senyavin, M. M. (1994). Flexibility analysis and optimization of chemical plants with uncertain parameters. Computers & Chemical Engineering, 18, 755-767.

Pernenkil, L., & Cooney, C. L. (2006). A review on the continuous blending of powders. Chemical Engineering Science, 61, 720-742.

Peterson, J. J. (2008). A Bayesian approach to the ICH Q8 definition of design space. J Biopharm Stat, 18, 959-975.

Pistikopoulos, E. N., & Ierapetritou, M. G. (1995). Novel-Approach for Optimal Process Design under Uncertainty. Computers & Chemical Engineering, 19, 1089-1110.

Pistikopoulos, E. N., & Mazzuchi, T. A. (1990). A Novel Flexibility Analysis Approach for Processes with Stochastic Parameters. Computers and Chemical Engineering, 14, 991-1000.

Plumb, K. (2005). Continuous processing in the pharmaceutical industry - Changing the mind set. Chemical Engineering Research & Design, 83, 730-738.

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

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

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

PSE. (2012). Model Developer Guide Release v3.6.

Quesada, I., & Grossmann, I. E. (1992). An LP/NLP based branch and bound algorithm for convex MINLP optimization problems. Computers & Chemical Engineering, 16, 937-947.

Ramachandran, R., Arjunan, J., Chaudhury, A., & Ierapetritou, M. G. (2011). Model-Based Control-Loop Performance of a Continuous Direct Compaction Process. Journal of Pharmaceutical Innovation, 6, 249-263.

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

Reinhardt, U. E. (2001). Perspectives on the pharmaceutical industry. Health Aff (Millwood), 20, 136-149.

Reynolds, G., Ingale, R., Roberts, R., Kothari, S., & Gururagan, B. (2010). Practical application of roller compaction process modeling. Computers and Chemical Engineering, 34, 1049-1057.

Rios, L., & Sahinidis, N. (2012). Derivative-free optimization: a review of algorithms and comparison of software implementations. Journal of Global Optimization, 56, 1247-1293.

Rogers, A., Hashemi, A., & Ierapetritou, M. G. (2013). Modeling of Particulate Processes for the Continuous Manufacture of Solid-Based Pharmaceutical Dosage Forms. Processes, 1, 67-127.

Rogers, A., & Ierapetritou, M. G. (2014). Discrete element reduced-order modeling of dynamic particulate systems. Aiche Journal, 60, 3184-3194.

Rogers, A., Inamdar, C., & Ierapetritou, M. G. (2014). An Integrated Approach to Simulation of Pharmaceutical Processes for Solid Drug Manufacture. Industrial & Engineering Chemistry Research, 53, 5128-5147.

Roweis, S. (1998). EM algorithms for PCA and SPCA (Vol. 10).

Ryoo, H. S., & Sahinidis, N. V. (1995). Global optimization of nonconvex NLPs and MINLPs with applications in process design. Computers & Chemical Engineering, 19, 551-566.

Saboo, A. K., & Morari, M. (1984). Design of resilient processing plants—IV: Some new results on heat exchanger network synthesis. Chemical Engineering Science, 39, 579-592.

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

Saltelli, A. (2002). Making best use of model valuations to compute sensitivity indices. Computer Physics Communications, 145, 280-297.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., & Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Computer Physics Communications, 181, 259-270.

Saltelli, A., & Bolado, R. (1998). An alternative way to compute Fourier amplitude sensitivity test (FAST). Computational Statistics & Data Analysis, 26, 445-460.

Saltelli, A., Chan, K., & Scott, E. M. (2000). Sensivivity Analysis. Chichester: John Wiley & Sons Ltd.

Saltelli, A., Tarantola, S., & Campolongo, F. (2000). Sensitivity analysis as an ingredient of modeling. Statistical Science, 15, 377-395.

Saltelli, A., Tarantola, S., & Chan, K. P. S. (1999). A quantitative model-independent method for global sensitivity analysis of model output. Technometrics, 41, 39-56.

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

Sasena, M. J., Papalambros, P., & Goovaerts, P. (2002). Exploration of metamodeling sampling criteria for constrained global optimization. Eng Optim, 34, 263-278.

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Analytical Chemistry, 36, 1627-1639.

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. Industrial & Engineering Chemistry Research, 50, 10083-10092.

Schaibly, J. H., & Shuler, K. E. (1973). Study Of Sensitivity Of Coupled Reaction Systems To Uncertainties In Rate Coefficients. 2. Applications Journal Of Chemical Physics, 59, 3879-3888.

Seifert, T., Sievers, S., Bramsiepe, C., & Schembecker, G. (2012). Small scale, modular and continuous: A new approach in plant design. Chemical Engineering and Processing, 52, 140-150.

Sen, M., Chaudhury, A., Singh, R., John, J., & Ramachandran, R. (2013). Multi-scale flowsheet simulation of an integrated continuous purification-downstream pharmaceutical manufacturing process. Int J Pharm, 445, 29-38.

Sen, M., Dubey, A., Singh, R., & Ramachandran, R. (2013). Mathematical Development and Comparison of a Hybrid PBM-DEM Description of a Continuous Powder Mixing Process. Journal of Powder Technology, 2013, 1-11.

Sen, M., Rogers, A., Singh, R., Chaudhury, A., John, J., Ierapetritou, M. G., & Ramachandran, R. (2013). Flowsheet optimization of an integrated continuous purification-processing pharmaceutical manufacturing operation. Chemical Engineering Science, 102, 56-66.

Sen, M., Singh, R., Vanarase, A., John, J., & Ramachandran, R. (2012). Multi-dimensional population balance modeling and experimental validation of continuous powder mixing processes. Chemical Engineering Science, 80, 349-360.

Shah, N. (2004). Pharmaceutical supply chains: key issues and strategies for optimisation. Computers & Chemical Engineering, 28, 929-941.

Shvartsman, S. Y., Theodoropoulos, C., Rico-MartInez, R., Kevrekidis, I. G., Titi, E. S., & Mountziaris, T. J. (2000). Order reduction for nonlinear dynamic models of distributed reacting systems. Journal of Process Control, 10, 177-184.

Sin, G., Gernaey, K. V., & Eliasson Lantz, A. (2009). Good modelling practice (GMoP) for PAT applications: Propagation of input uncertainty and sensitivity analysis. Biotechnology Progress, 25, 1043–1053.

Singh, R., Gernaey, K. V., & Gani, R. (2010). ICAS-PAT: A software for design, analysis and validation of PAT systems. Computers & Chemical Engineering, 34, 1108-1136.

Singh, R., Ierapetritou, M., & Ramachandran, R. (2012). An engineering study on the enhanced control and operation of continuous manufacturing of pharmaceutical tablets via roller compaction. Int J Pharm, 438, 307-326.

Singh, R., Ierapetritou, M., & Ramachandran, R. (2013). System-wide hybrid MPC-PID control of a continous pharmaceutical tablet manufacturing process via direct compaction. European Journal of Pharmaceutics, 85, 1164-1182.

Skogestad, S., & Morari, M. (1987). Design of Resilient Processing Plants IX-The Effect of Model Uncertainty on Dynamic Resilience. Chemical Engineering Science, 42, 1765.

Sobol, I. M. (1993). Sensitivity analysis for nonlinear mathematical models. Mathematical Modelling & Computational Experiment, 1, 407-414.

Straub, D. A., & Grossmann, I. E. (1990). Integrated Stochastic Metric of Flexibility for Systems with Discrete State and Continuous Parameter Uncertainties. Computers & Chemical Engineering, 14, 967-985.

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

Subrahmanyam, S., Peknyt, J. F., & Reklaitis, G. V. (1994). Design of Batch Chemical-Plants under Market Uncertainty. Industrial & Engineering Chemistry Research, 33, 2688-2701.

Suresh, P., & Basu, P. K. (2008). Improving Pharmaceutical Product Development and Manufacturing: Impact on Cost of Drug Development and Cost of Goods Sold of Pharmaceuticals. Journal of Pharmaceutical Innovation, 3, 175–187.

Swaney, R. E., & Grossmann, I. E. (1985a). An Index for Operational Flexibility in Chemical Process Design Part I: Formulation and Theory. Aiche Journal, 36.

Swaney, R. E., & Grossmann, I. E. (1985b). An Index for Operational Flexibility in Chemical Process Design}Part II: Computational Algorithms. Aiche Journal, 31.

Timm, N. H. (2002). Applied Multivariate Analysis. New York, NY: Springer-Verlag.

Tolsma, J. E., Clabaugh, J. A., & Barton, P. I. (2002). Symbolic Incorporation of External Procedures into Process Modeling Environments. Ind. Eng. Chem. Res., 41, 3687 - 3876.

Tremblay, D., & Peers, Z. (2014). Jump Start: Aspen Custom Modeler V8. In. Burlington, MA: Aspen Technology Inc.

Troup, G. M., & Georgakis, C. (2013). Process systems engineering tools in the pharmaceutical industry. Computers & Chemical Engineering, 51, 157-171.

Tsiakis, P., Shah, N., & Pantelides, C. C. (2001). Design of Multi-echelon Supply Chain Networks under Demand Uncertainty. Ind. Eng. Chem. Res., 40, 3585-3604.

Vanarase, A. U., Alcala, M., Rozo, J. I. J., Muzzio, F. J., & Romanach, R. J. (2010). Real-time monitoring of drug concentration in a continuous powder mixing process using NIR spectroscopy. Chemical Engineering Science, 65, 5728-5733.

Vanarase, A. U., & Muzzio, F. J. (2011). Effect of operating conditions and design parameters in a continuous powder mixer. Powder Technology, 208, 26-36.

Vassiliadis, V. S., Sargent, R. W. H., & Pantelides, C. C. (1994a). Solution of a Class of Multistage Dynamic Optimization Problems .1. Problems without Path Constraints. Industrial & Engineering Chemistry Research, 33, 2111-2122.

Vassiliadis, V. S., Sargent, R. W. H., & Pantelides, C. C. (1994b). Solution of a Class of Multistage Dynamic Optimization Problems .2. Problems with Path Constraints. Industrial & Engineering Chemistry Research, 33, 2123-2133.

Vervaet, C., & Remon, J. P. (2005). Continuous granulation in the pharmaceutical industry. Chemical Engineering Science, 60, 3949-3957.

Vikhansky, A., & Kraft, M. (2006). Two methods for sensitivity analysis of coagulation processes in population balances by a Monte Carlo method. Chemical Engineering Science, 61, 4966-4972.

Wang, J., Wen, H., & Desai, D. (2010). Lubrication in tablet formulations. Eur J Pharm Biopharm, 75, 1-15.

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

Weir, G. J. (2004). A mathematical model for dilating, non-cohesive granular flows in steep-walled hoppers. Chemical Engineering Science, 59, 149-161.

Westerberg, K., Broberg-Hansen, E., Sejergaard, L., & Nilsson, B. (2013). Model-Based Risk Analysis of Coupled Process Steps. Biotechnol Bioeng.

Westerhuis, J. A., Kourti, T., & MacGregor, J. F. (1998). Analysis of multiblock and hierarchical PCA and PLS models. Journal of Chemometrics, 12, 301-321.

Wold, S., Sjostrom, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems, 58, 109-130.

Wu, C. Y., Ruddy, O. M., Bentham, A. C., Hancock, B. C., Best, S. M., & Elliott, J. A. (2005). Modelling the mechanical behaviour of pharmaceutical powders during compaction. Powder Technology, 152, 107-117.

Yang, S., & Evans, J. R. G. (2007). Metering and dispensing of powder; the quest for new solid freeforming techniques. Powder Technology, 178, 56-72.

You, F., Wassick, J. M., & Grossmann, I. E. (2009). Risk Management for a Global Supply Chain Planning under Uncertainty: Models and Algorithms. Aiche Journal, 55, 931-946.

Yu, L. X. (2008). Pharmaceutical quality by design: product and process development, understanding, and control. Pharm Res, 25, 781-791.

Yu, S., Gururajan, B., Reynolds, G., Roberts, R., Adams, M. J., & Wu, C. Y. (2012). A comparative study of roll compaction of free-flowing and cohesive pharmaceutical powders. Int J Pharm, 428, 39-47.

Zhang, X., Lionberger, R. A., Davit, B. M., & Yu, L. X. (2011). Utility of physiologically based absorption modeling in implementing Quality by Design in drug development. AAPS J, 13, 59-71.

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

Ziehn, T., & Tomlin, A. S. (2008). Global sensitivity analysis of a 3D street canyon model – part I: the development of high dimensional model representations. Atmospheric Environment, 42, 1857–1873.

Ziehn, T., & Tomlin, A. S. (2009). GUI–HDMR – A software tool for global sensitivity analysis of complex models. Environmental Modelling & Software, 24, 775-785.