ROBUST PROCESS MONITORING FOR CONTINUOUS PHARMACEUTICAL MANUFACTURING

by

Mariana Moreno

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy
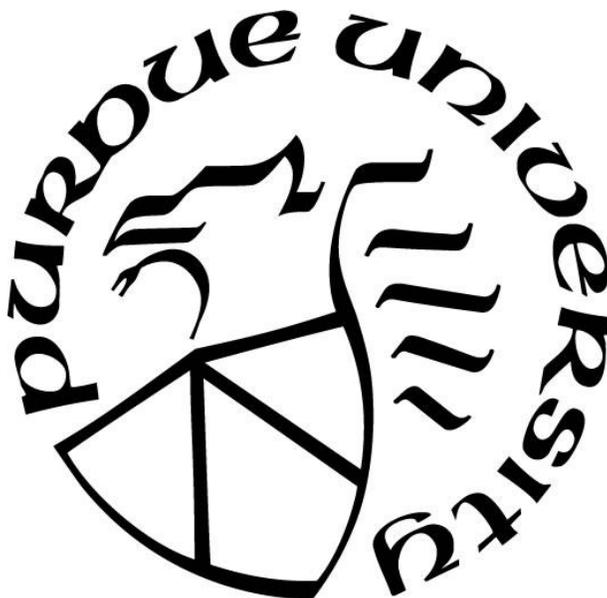
Davidson School of Chemical Engineering

West Lafayette, Indiana

December 2018

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Gintaras V. Reklaitis, Co-Chair

    Davidson School of Chemical Engineering

Dr. Zoltan K. Nagy, Co-Chair

    Davidson School of Chemical Engineering

Dr. Michael Harris

    Davidson School of Chemical Engineering

Dr. Carl Wassgren

    School of Mechanical Engineering

Dr. Marcial Gonzalez

    School of Mechanical Engineering

Approved by:

    Dr. John Morgan

        Head of the Graduate Program

A mis padres, sin ustedes esto no hubiera sido posible

ACKNOWLEDGMENTS

I want to express my appreciation for the assistance given by my undergraduate students, especially: Benjamin Rentz, Cody Leach, Alexander R. Milaszewski, Sarah M. Mathena, Alex Hein, and Isaac Mendoza. I hope that they learned from me, as much as I learned from them.

My special thanks are extended to my friends. First, the friends I made during this phase of my life: Dr. Martha E. Hay, Dr. Jennifer S. Laster, and Dr. Paulami Majumdar. Overcoming the stress and difficulties of a graduate degree would not have been possible without their friendship. Second, but not least, my longtime friends Miroslava Mendoza, Diego F. Velasquez, and Martha A. Arellano who have always been there for me for emotional support.

I am particularly grateful to my parents, Sofía Ricaño Rivas and Mario A. Moreno Martínez, for their support and counsel. It is only fair also to acknowledge my grandmother, Irma Martinez Grajales, who has been a second mother to me, and my sisters Irma S. Moreno Ricaño and Karla S. Ricano, who believed in me. I could not have been blessed with a better family than this one. Lastly, I would like to thank my fiancée Christopher F. Kulesza, for his relentless encouragement and filling my days with joy during my graduate studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Author: Moreno, Mariana. PhD
Institution: Purdue University
Degree Received: December 2018
Title: Robust Process Monitoring for Continuous Pharmaceutical Manufacturing
Major Professor: Gintaras V. Reklaitis & Zoltan K. Nagy

Robust process monitoring in real-time is a challenge for Continuous Pharmaceutical Manufacturing. Sensors and models have been developed to help to make process monitoring more robust, but they still need to be integrated in real-time to produce reliable estimates of the true state of the process. Dealing with random and gross errors in the process measurements in a systematic way is a potential solution. In this work, we present such a systematic framework, which for a given sensor network and measurement uncertainties will predict the most likely state of the process. As a result, real-time process decisions, whether for process control, exceptional events management or process optimization can be based on the most reliable estimate of the process state.

Data reconciliation (DR) and gross error detection (GED) have been developed to accomplish robust process monitoring. DR and GED mitigate the effects of random measurement errors and non-random sensor malfunctions. This methodology has been used for decades in other industries (i.e., Oil and Gas), but it has yet to be applied to the Pharmaceutical Industry. Steady-state data reconciliation (SSDR) is the simplest forms of DR but offers the benefits of short computational times. However, it requires the sensor network to be redundant (i.e., the number of measurements has to be greater than the degrees of freedom).

In this dissertation, the SSDR framework is defined and implemented it in two different continuous tableting lines: direct compression and dry granulation. The results for two pilot plant

scales via continuous direct compression tableting line are reported in this work. The two pilot plants had different equipment and sensor configurations. The results for the dry granulation continuous tableting line studies were also reported on a pilot-plant scale in an end-to-end operation. New measurements for the dry granulation continuous tableting line are also proposed in this work.

A comparison is made for the model-based DR approach (SSDR-M) and the purely data-driven approach (SSDR-D) based on the use of principal component constructions. If the process is linear or mildly nonlinear, SSDR-M and SSDR-D give comparable results for the variables estimation and GED. The reconciled measurement values generate using SSDR-M satisfy the model equations and can be used together with the model to estimate unmeasured variables. However, in the presence of nonlinearities, the SSDR-M and SSDR-D will differ. SSDR successfully estimates the real state of the process in the presence of gross errors, as long as steady-state is maintained and the redundancy requirement is met. Gross errors are also detected whether using SSDR-M or SSDR-D.

CHAPTER 1. INTRODUCTION

Product quality is one of the main concerns in the Pharmaceutical Industry [1]. Continuous manufacturing has the potential of improving product quality while also offering a more robust and flexible manufacturing configuration. Significant progress has been made in this area over the past couple of years, by adopting the Quality-by-Design (QbD) paradigm in pharmaceutical manufacturing [2]. The development of Process Analytical Technology (PAT) tools has been crucial to ensuring adequate process monitoring and potentially improving control of the Critical Quality Attributes (CQA's) and Critical Process Parameters (CPP's) in real-time. Models have also been developed to increase process understanding. Further steps have to be taken to integrate this information and make real-time decisions.

This research is focused on downstream manufacturing of solid oral dosage (e.g., tablets, capsules) which is the most common form of drug product. This project will be limited to continuous tableting processes, specifically direct compression and dry granulation. The methodologies applied to real-time management need to be developed to have an effective continuous process. Real-time management considers five components: process monitoring, process control, exceptional events management, strategies to deal with out of specification material, and real-time process optimization [3].

Process monitoring requires effective deployment of sensors and typically can be improved through the use of models, which can be at steady-state or dynamic. If the state of the process is known with sufficient accuracy, a control scheme can be designed to maintain the CQA's or CPP's on target. During manufacturing, exceptional events can occur in which the control system cannot correct, and thus they have to be detected, diagnosed and mitigated in real-time. Strategies to track any out of specification material also have to be developed so that a decision to reject the non-

conforming material can be made. Finally, all these components make it possible to carry out effective real-time optimization, under which the quality of the product can be maximized.

Process monitoring is the foundation of the real-time process management. Use of the right measurements together with a process model, in an integrated approach, can give the most reliable estimate of the state of the process. The goal of this work is to make process monitoring more robust through the use of data reconciliation and gross error detection. To achieve this, the development of appropriate online sensors and process models are needed. With reconciled values of the process variable known QbC can be based on the most reliable estimate of the process state.

1.1   Project Motivation

Regulatory agencies have encouraged the development of sensors schemes and a systematic approach for the estimation of the real process state. This project aims to improve the product quality, in the continuous tableting processes, through robust process monitoring in real-time. Steady-state data reconciliation framework was proposed as a solution, due to its short computational time and its ability to reconcile the process measurements using relatively simple process models. This approach directly takes into account the measurement uncertainties and uses well-known statistical tests for gross error detection.

This project was partially supported by the Food and Drug Administration (FDA), and it was part of the Engineering Research Center for Structured Organic Particulate Systems (C-SOPS). C-SOPS is an organization sponsored by the National Science Foundation, and it focuses on the scientific contributions for the design, scaling, control, and optimization of the processes involved in the manufactured of structured organic particulate systems.

.

1.2   Research Objectives

The objective of this dissertation is to develop a model-based framework that will predict the most likely state of the process. The framework is based on the using a given sensor network configuration, with known measurement uncertainties. The development of sensor schemes and the integration of models are necessary for the estimation of the process state in real-time. The objectives of this research work are outlined as follow:

•   Develop sensors for the CQA's in the variants of the continuous tableting line using direct compression and dry granulation.

•   Establish the model-based framework that can deal with random and gross errors in the sensors.

•   Test the framework on the two instances of the continuous tableting line implemented in the Purdue pilot plant

1.3   Dissertation Overview

This dissertation is organized as follow:

•   Chapter 2 describes the steady-state data reconciliation framework for a Direct Continuous Tableting Line. The data reconciliation problem is formulated as an optimization problem, and suitable statistical tests are used for gross error detection. It also reports on the performance of the framework on several cases scenarios.

•   Chapter 3 shows the implementation of the steady-state data reconciliation framework, presented in Chapter 2, on the direct compression continuous tableting line operating at Rutgers University. The performance of the framework was evaluated through at-line measurements.

•   Chapter 4 demonstrates the implementation of the steady-state data reconciliation framework in the dry-granulation line, which is a more complex process than direct compression. It also

investigates performance on an alternative instance of direct compression with uses an alternative sensor configuration, compared to that studied in Chapter 3.

• Chapter 5 represents the study of new PAT tools for particle-size measurement which provides an alternative means of increasing measurement redundancy for purposes of data reconciliation on the dry granulation tableting line.

• Chapter 6 introduces data reconciliation as a solution for robust process monitoring in continuous tableting processes. This chapter also presents the data reconciliation concept and how it supports the implementation of QbC.

• Chapter 7 summarizes the findings of this thesis and suggests future work to advance and improve process monitoring in continuous tablet manufacturing.

1.4 References

[1]   U. S. D. of H. and H. S. FDA, "Guidance for Industry PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance," no. September, p. 16, 2004.

[2]   S. L. Lee, T. F. O. Connor, X. Yang, C. N. Cruz, L. X. Yu, and J. Woodcock, "Modernizing Pharmaceutical Manufacturing : from Batch to Continuous Production," no. 1, pp. 1–25, 2015.

[3]   M. Ierapetritou, F. Muzzio, and G. V. Reklaitis, "Perspectives on the Continuous Manufacturing of Powder-Based Pharmaceutical Processes," *AIChE J.*, vol. 62, no. 6, pp. 1846–1862, 2016.

CHAPTER 2. STEADY-STATE DATA RECONCILIATION FRAMEWORK FOR A CONTINUOUS TABLETING LINE

2.1 Introduction

Maintaining critical quality attributes (CQA's) of the drug product within specified bounds is a significant concern in pharmaceutical manufacturing [1]. In traditional batch production, the CQA's are typically monitored by statistical sampling of the finished dosage form and rejection of the batch if sampling indicates deviations from CQA's specifications, leading to significant waste and increased cost [2]. Continuous manufacturing has been pursued by the industry and encouraged by regulators to overcome these limitations [3]. Effective continuous manufacturing is typically implemented using a real-time process management strategy, which encompasses multiple components. The first foundational component consists of real-time measurement and monitoring of the system including the use of Process Analytical Technology (PAT) tools. The second involves a robust process control system. The third involves the detection, diagnosis, and mitigation of exceptional events. Finally, the fourth component requires real-time procedures for tracking and isolating noncompliant product [2]. This paper focuses on the first component.

Establishing an effective sensor network for monitoring CQA's is a major step towards improving plant operations. However, measurements of CQA's or related process variables can fluctuate over time due to random errors or to the occurrence of gross errors [4], [5]. Random errors cannot be predicted with certainty; they can only be characterized by probability distributions [6]. This type of the error can be caused by a variety of factors such as analytical errors or normal flow fluctuations, which cannot be entirely eliminated. By contrast, gross errors are caused by non-random events, such as a bias due to miscalibration, sensor degradation or complete sensor malfunction [7]. Reliable process monitoring in real-time is still a challenge in

continuous pharmaceutical manufacturing. Dealing with random and gross errors of the process measurements in a systematic way is a potential solution. The objective of this paper is to demonstrate a model-based framework, which for a given sensor network configuration and set of measurement uncertainties, will predict the most likely state of the process. Moreover, the capability of the methodology for the detection of gross errors in the process (GED), through statistical tests will be shown. Although data reconciliation (DR) has been widely used in other industries (i.e., oil and gas), to the author's knowledge, it has not been applied to continuous pharmaceutical processes.

Alternatively, purely data-driven methods based on multivariate statistical models can also be used for state estimation [7], [8]. The model-based data reconciliation methods (SSDR-M) and the multivariate approaches generate comparable results under the assumption that the relationships between the measured variables are linear and certain additional assumptions hold true (Section 2.2.2) [9], [10], [11]. However, differences in reconciled values can occur if the process model involves nonlinear relationships between measured variables. SSDR-M also has the advantage that it can provide estimates of unmeasured variables through the model equations. In this paper, we report on the real-time application of both methodologies for DR and GED to a direct compression line, which involves the online use of load cells and near-infrared (NIR) spectroscopy-based sensors.

2.2   Theory

2.2.1   Data reconciliation

The purpose of data reconciliation is to obtain the best estimate of the process measurements in the presence of random errors in the measurements [10]. Data reconciliation can be posed as an optimization problem in which the objective function (see Eq.2.1), and the process

model (Eq. 2.2 and 2.3) are functions of the process variables. The process model usually consists of mechanistic relations (e.g. mass balances, energy balances), which are dependent on both measured and unmeasured variables as well as parameters. The key requirement is that the estimate of the process variables must satisfy the process model while optimizing a likelihood function. The mathematical formulation of an SSDR problem is given as below:

$$\min_{x,y} J = (\mathbf{x}^+ - \mathbf{x})^{\mathrm{T}} \mathbf{Q}^{-1} (\mathbf{x}^+ - \mathbf{x}) \tag{2.1}$$

$$\text{s.t.} \quad \mathbf{h}(\mathbf{x}, \mathbf{y}, \boldsymbol{\theta}) = 0 \tag{2.2}$$

$$\mathbf{g}(\mathbf{x}, \mathbf{y}, \boldsymbol{\theta}) \leq 0 \tag{2.3}$$

Where: $\mathbf{x}^+ \in \Re^n$ is a vector of measurements, $\mathbf{x} \in \Re^n$ is a vector of reconciled values of the n measurements, $\mathbf{y} \in \Re^m$ is a vector of the m unmeasured process variables, $\mathbf{Q}$ is the covariance matrix, $\Re^{n \times n}$, $\boldsymbol{\theta} \in \Re^p$ is a vector of p process model parameters, $\mathbf{h} \in \Re^k$ is a set of equations which describe the steady-state behavior of the process, and $\mathbf{g} \in \Re^q$ is a set of inequality constraints, n is the number of reconciled variables, m the number of unmeasured variables and k the number of relations or equations.

The measured variables are assumed to be contaminated with random errors that are normally distributed with zero mean and known variances/covariances. Note that the objective function only contains terms corresponding to the measured variables. The unmeasured variables can be calculated from known variables and the process model, providing that the unmeasured variables are observable. Tests for observability are available in the literature [9], [12]. In general, DR requires that the process under study have positive degrees of freedom, that is, $n+m-k > 0$. Otherwise, there may be no feasible solution to Eq. 2.2. Therefore, there must be measurement redundancy in the process, requiring the deployment of sufficient PAT measurement points.

The process model used in DR can either describe the steady-state behavior of the process as shown in equation 2.2 or its dynamic behavior, in which case the process model will consist of a system of algebraic and differential equations. Dynamic models can represent the departures of the process variables from the steady-state and thus allow dynamic DR to be performed during transitions from steady-state, but the computational time for solving the dynamic DR problem will be much higher [13]. Other techniques for dealing with noisy measurements which do not require steady-state requirement include the family of Kalman filters. Although the classical Kalman filter (KF) does not have the steady-state requirement, it is applicable to linear and unconstrained systems. If the process is nonlinear and with hard constraints, modifications have been proposed, such as the Extended Kalman Filter (EKF) [14].

Moreover, dynamic data reconciliation approaches are also available and these do allow capturing of process dynamics [15]. Moving horizon estimation (MHE) is another technique used for treating noisy dynamic processes which is found to perform better than EKF usually [16]. However, the primary limitation of these approaches is the computational time required for their use in real-time, especially for processes with fast dynamics.

The advantage of steady-state models is that they can be solved in computational time that allows the effective real-time use of the results; however, they are only valid during steady-state periods, hence, require that the observed variations be around a steady-state. Of course, steady state should be understood to mean that variations in process variables are within an allowable range determine via appropriate statistical tests. Consistent with the steady-state assumption, each measurement in the vector $\mathbf{x}$ is assumed to be made synchronously, that is, recorded at the same time point.

In general, steady-state process models (Eq. 2.2) consist of material balances, energy balances, design and other empirical relationships as well as equations relating process variables to material physical and chemical properties. While component material balances are linear in the component flows, the energy balances are typically bilinear while the physical property, equipment, performance are generally nonlinear. Consequently, although in the linear case, the solution of the DR problem can be expressed analytically and computed using projection matrix constructions [17], in the presence of nonlinear model elements, the DR problem must be solved numerically using nonlinear programming algorithms [10], [13]. In the continuous direct compression tableting line example discussed in subsequent sections of this paper, the SSDR model involves linear component material balances, energy balances are not required, and the only nonlinearities are introduced through property measurements, specifically those of stream composition.

The SSDR problem is typically solved using a two-step approach; first, for given values of the model parameters, the measured variables are reconciled. In the second step which usually is executed less frequently than the first step, the model parameters are (re-) estimated [18] using measurements from a sufficiently large number of sample points. The case studies reported in this paper only discuss the first step.

2.2.2   DR using Multivariate statistical methods

Multivariate Statistical Process Control (MSPC) methods involve the use of Principal Component Analysis (PCA) or Partial-Least Squares (PLS) to monitor measured variables in a reduced space and detect process upsets through statistical tests [19]. PCA models represents the system through a subset of the principal components (PC's) or latent variables by capturing most of the high-dimensional data ($\mathbf{X}$) variance [20], [21]. PCA model can be expressed in scores ($\mathbf{T}$) and loadings ($\mathbf{P}$). The scores are the projections of each observation onto a subspace. The loadings

are the eigenvectors associated with the eigenvalues of the covariance matrix. There are additional forms of multivariate models, such as PLS, which represents $\mathbf{X}$ using a reduced number of latent variables (LV) while also representing the space of response matrix ($\mathbf{Y}$) [22], [23]. In other words, PLS uses latent variables to capture the variation in $\mathbf{X}$, and that predicted in the $\mathbf{Y}$ [22], [24].

PCA is widely used for both monitoring and abnormal events detection. For the training of the PCA model, it is necessary first to obtain a sufficient number of instances of the measured variables, as training data. Then the loading matrix $\mathbf{P}$ is computed, as shown in Eqs. 2.4 and 2.5: the direct eigenvector-eigenvalue calculations or the Nonlinear Iterative Partial Least Squares (NIPALS) algorithm can be used for this purpose.

$$\mathbf{X} = \mathbf{TP}^{\mathrm{T}} + \mathbf{E} \tag{2.4}$$

$$\mathbf{T} = \mathbf{XP} \tag{2.5}$$

Where: $\mathbf{X} \in \mathfrak{R}^{Nxn}$ is the matrix of observed variables, $\mathbf{T} \in \mathfrak{R}^{Nxd}$ is the principal components or scores matrix, $\mathbf{P} \in \mathfrak{R}^{dxn}$ is the loading matrix, $\mathbf{E} \in \mathfrak{R}^{Nxn}$ is the random error matrix for $\mathbf{X}$, n is the observed values, N is the number of observations and d is the degrees of freedom or the number of PC's where $d \ll n$

While PCA models are most commonly used for monitoring, it is less known [10] that such models can also be used to compute reconciled values of the measured variables $\mathbf{X_m}$. As shown, Eq. 2.6. can be used to compute the updated scores from the new values of the measured variables and the loadings matrix. Then, the new estimated scores and loading matrix can be used to compute the reconciled values of the measured variables as shown in Eq. 2.7. These estimated values will be equivalent to the reconciled values obtained using the model-based approach, if certain conditions are met [10], [11], [25]. The scores and loadings can also be used in the statistical tests for GED (see section 2.2.3.2) for the reconciled values.

$$\mathbf{T_m} = \mathbf{X_m}\mathbf{P} \tag{2.6}$$

$$\widehat{\mathbf{X}} = \mathbf{T_m}\mathbf{P}^{\mathrm{T}} \tag{2.7}$$

Where: $\mathbf{X_m} \in \mathfrak{R}^{Nxn}$ is the matrix of the new measured variables, $\mathbf{T_m} \in \mathfrak{R}^{Nxd}$ is the new principal components or scores matrix used for the new measured variables, $\widehat{\mathbf{X}} \in \mathfrak{R}^{Nxn}$ is the matrix of the reconciled values for the measured variables through PCA, n is the observed values, N is the number of observations and d is the degrees of freedom or the number of PC's, where $d \ll n$

There are some key similarities and differences between the model and data-driven based approaches for reconciliation and monitoring. In both cases, the covariance matrix must be obtained from historical data, that is from a training set of measured variable observations while the process is at steady-state(s). If it is assumed that an $(n\text{-}k)$ set of principal components is used, the measurement errors are independent and identically distributed and the process model is linear, then the DR and PCA based approaches are also comparable, and the later method can be used for SSDR-D. Moreover, as shown in [10] the constraint matrix can also be estimated through PCA and will be equivalent to the DR model except for a rotational transformation of the data matrix. In [10], the authors further propose a framework for using DR and PCA in a complementary fashion and provide seven examples which employ simulated scenarios for a water distribution system.

There are differences that arise in how these approaches deal with gross errors and how the unmeasured variables are estimated. In SSDR-M, once the measurement with which the gross error is associated is identified, it is removed from the reconciliation process, the remaining measured variables are reconciled again, and the measurement with gross error is treated as an unmeasured variable. Since the process model already includes a mechanistic or empirical equation relating the

measured variables to the unmeasured variable, the new set of reconciled values can be used to generate an estimate of the measurement that is in gross error.

In the SSDR-D case, the prediction of the measurement that is in gross error usually is treated as a missing variable. There are three approaches for handling missing variables in PCA that are used in industrial practice [26]: single component projection derived from the NIPALS algorithm, projection to the model plane and conditional mean replacement, with the last one, generally regarded to be superior. By using this method, it is possible to replace the missing variable (i.e., measurement with gross error) by its conditional mean given by the other observed variables [26]. The conditional mean is estimated with the observed variables (with no gross error), using a standard score estimation routine. In common practice, the limit for application of these methods is to have no more than 20% of the variables missing; while in complex industrial problems it can go up to 30-60% [27]. Of course, PCA cannot estimate unmeasured variables when there is no training data.

The goal of this work is to demonstrate a systematic framework for dealing with measurement errors in order to estimate the "true' state of the process. We apply both approaches as applicable, establish their limitations and demonstrate their performance in simple cases arising from the continuous tableting line via direct compression.

2.2.3   Error Detection

2.2.3.1   Model-based alternative

The DR methodology, whether process model-based or data-driven, generates reconciled process variable values, under the assumption that errors in the measured variables are random. If a non-random, so-called gross, error arises in a measurement, then that measurement will bias the reconciliation outcome. Thus, if such a gross error occurs then it must be identified, the faulty

measurement discarded and the remaining measurements reconciled, as noted in the previous section. Detection of gross errors can readily be performed through the use of appropriate statistical tests, which are similar to outlier detection schemes employed in regression development.

There is a considerable literature on various tests for gross error detection (GED) proposed for use in DR. In this work, we employ two well-known tests: the global test and the measurement test. The global test (GT) serves to detect if there is an appreciable measurement error in the process [6], [5]. It can be shown that the maximum likelihood function (Eq. 2.1) evaluated at the reconciled values follows a Chi-squared ($\chi^2_{1-\alpha}$) distribution. The $\chi^2_{1-\alpha}$ is a function of the linearly independent equations (I) and the selected level of significance α (e.g., 5%), Thus, if the value of the objective function is greater than the Chi-squared test criterion, then it is likely that a gross error has occurred.

However, the global test does not pinpoint which measurement is at fault. A second test, called the measurement test, can help to indicate where the gross error is located. It uses the magnitude of the variable corrections to identify those that are "large" relative to standard deviation and thus suspected. Improved results can also be obtained by using Nonlinear programming (NLP) methods via the Modified Iterative Measurement Test (MIMT) algorithm [28]. The framework, shown in Figure 2.1, summarizes the steps which are followed in the reconciliation process used in this study.

The GED test can be applied to DR applications involving either linear or nonlinear process models. However, since the measurement test is defined in terms of the coefficient matrix of the linear model, in the nonlinear case this requires some modification. Specifically, this matrix must be replaced by a linearization of the process model constructed at the nominal steady-state values of the process variables. Eq. 2.8 represents the model constraints in linear form.

$$\mathbf{Ax} - \mathbf{c} = 0 \tag{2.8}$$

Where: $\mathbf{A} \in \Re^{kxn}$ is the linear constraint matrix, $\mathbf{x} \in \Re^n$ is a vector of reconciled values, $\mathbf{c} \in \Re^n$ is a vector of known coefficients.

The details of measurement test are summarized in the following relationships:

$$\mathbf{a_i} = (\mathbf{x_i^+} - \mathbf{x_i}) \tag{2.9}$$

$$\mathbf{V} = \mathbf{QA^T(AQA^T)^{-1}AQ} \tag{2.10}$$

$$\mathbf{mt_i} = \frac{|\mathbf{a_i}|}{\mathbf{V_{ii}}^{\frac{1}{2}}} \tag{2.11}$$

$$\beta = 1 - (1 - \alpha)^{1/n} \tag{2.12}$$

Where: $\mathbf{x^+} \in \Re^n$ is a vector of measurements, $\mathbf{x} \in \Re^n$ is a vector of reconciled values, $\mathbf{a} \in \Re^n$ is a vector of errors, $\mathbf{Q}$ is the covariance matrix, $\Re^{n \times n}$, $\mathbf{A} \in \Re^{kxn}$ is the matrix of coefficients of the linear/linearized process model, $\mathbf{mt} \in \Re^n$ is a vector of the measurement test values, i is the index, and n is the number of measurements, α is the level of significance (e.g., 5%), and β is the modified level of significance since it is a function of α and the number of measurements.

It can be shown that the difference vector ($\mathbf{a_i}$) follows a multivariate normal distribution with an expected value of zero and covariance given by Eq. 2.9 to Eq. 2.12. Furthermore, the measurement test value $\mathbf{mt_i}$ follows a normal distribution with mean 0 and standard deviation of 1 [28]. Thus, if the $\mathbf{mt_i}$ is larger than the upper critical value ($z_{\alpha/2}$), most likely that measurement $\mathbf{x_i}$ involves a gross error. Another type of critical value ($z_{1-\beta/2}$), can also be considered which involves the number of measurements and the probability of error in each measurement. If there is a gross error and the measurement test cannot detect it, it is necessary to take the highest measurement test value and suggest it as a gross error. However, it is important to note that this decision is effectively a heuristic.

Finally, it should be acknowledged that process faults can arise from causes other than sensor failures. Such faults will give rise to significant departures of process variables outside of the range of variations that will be accepted by the steady-state test and, thus such, exceptional events will require active intervention in the process. In general, they will lead to the production of material outside of quality limits. The gross error detection methodology outlined above is based on the assumption that the process is at steady state, that such exceptional events have not occurred and thus that the gross errors that arise lie with the measurements. There is an extensive literature on process fault detection and diagnosis which discusses these issues to which we direct the interested reader (see for example [29]).

2.2.3.2   Data-driven alternative

In the case of the SSDR-D approach, there is a two-criteria statistical test in common use to check if there are gross errors: The Square Prediction Error (SPE) and the Hotelling's T^2. The SPE, (see Eq. 2.13), considers the model and measurement mismatch; it relates the orthogonal distance between the model plane and a given observation test [30]. The Hotelling's T^2, Eq. 2.14, describes the overall distance from the model and the observation [31], [22]. The confidence intervals for these tests can be found in the literature [26], [28].

$$SPE = \sum_{i=1}^{n} (\mathbf{x}_i^+ - \mathbf{x}_i)^2 \tag{2.13}$$

$$T^2 = \mathbf{t}_d^T \mathbf{S}^{-1} \mathbf{t}_d \tag{2.14}$$

Where: $\mathbf{x}^+ \in \Re^n$ is a vector of measurements, $\mathbf{x} \in \Re^n$ is a vector of reconciled values, $\mathbf{t} \in \Re^n$ is the vector of scores for d degrees of freedom and $\mathbf{S}$ the variance matrix for the scores generated in the training data, $\Re^{n \times n}$.

In order to identify the location of the gross error, it is necessary to check the difference between the reconciled and measured variables, that is the contribution of each variable to the SPE and the Hotelling's T^2 criteria. Eq. 2.15 shows the power contributions ($\boldsymbol{spe_i}$) of each variable [32]. Eq. 2.16 represents the contributions done to the Hotelling's T^2.

$$\boldsymbol{spe_i} = (\boldsymbol{x_i^+} - \boldsymbol{x_i})^2 \tag{2.15}$$

$$T_d^2 = \sum_{i=1}^{d} \frac{\boldsymbol{t_i^2}}{\boldsymbol{s_{t_i}^2}} \tag{2.16}$$

2.3 SSDR frameworks

In this section, we summarize the logic flow for the reconciliation and error detection procedures, which were discussed separately in the previous sections. Both the model-based and the data-driven frameworks incorporate repeated cycles of error detection to accommodate the situation in which multiple gross errors may arise.

2.3.1 SSDR-M framework

The first two steps of the logic flow summarized below deal with defining the sensor network, process model and confirming observability of the unmeasured variables in the process model. The remaining steps 3-8 are repeatedly executed at each successive measurement time.

SSDR-M Logic

- Step 1. Formulate the DR problem: Define the sensor network structure and determine the variance/covariance of the sensor measurements. Establish the steady-state process model equations and the set of measured and unmeasured variables. Test the unmeasured variables for observability.

- Step 2. <u>Redundancy Check</u>: Confirm that the number of measurements (No. M) is greater than the system degrees of freedom (DOF) and go to step 3. If the number of measurements is equal or less than the system degrees of freedom, DR cannot be performed, and additional sensors need to be added to the sensor network

- Step 3. Test that the system is at steady-state and if so record all of the sensor measurements at a given time. It is possible to check for steady-state by applying statistical tests. In the literature there are different ways to detect steady-state reported [33]. One approach is to perform a linear regression over a data window and a t-test; if the slope is different from zero, then there is not steady state [34]. It is also possible to do an F-test or R-test to determine if the process is at steady-state [35]. A simpler way to determine steady–state is the use of statistical process control chart (SPC) moving average chart with a threshold of 3σ [25] [24]. If there is missing data (e.g., fouling, disconnection), treat the variable as an unmeasured variable.

- Step 4. Solve the reconciliation problem to obtain reconciled values of measured variables.

- Step 5. Execute the global test. Please note that the $\chi^2{}_{1-\alpha}$ is calculated based on the confidence interval and the number of independent equations. If the global test fails, proceed to step 6. If the global test is passed, go to step 8.

- Step 6: Compute the measurement test value. If the process model is nonlinear, use the linearization at the steady-state operating point to estimate matrix **A** in Eq. 2.8.

- Step 7. If one or more of the measurement test values is greater than the critical value (**mt**$_i$ $> z_{\alpha/2}$), the measurement test failed. If the measurement test passed and the global tests did not, then select the measurement test that has the highest value and remove said measurement from the objective function. However, it should be noted that this decision

criterion which is commonly used does constitute a heuristic. The measurement that was removed from the objective function will be considered an unmeasured variable; therefore, it has to be added to the said set. Go back to step 2.

- Step 8. Compute the unmeasured variables using the reconciled measurement vector. If the model involves parameters, it may be necessary to re-estimate them using the reconciled values of the variables. Go to step 3 and repeat the process with at the next measurement time.
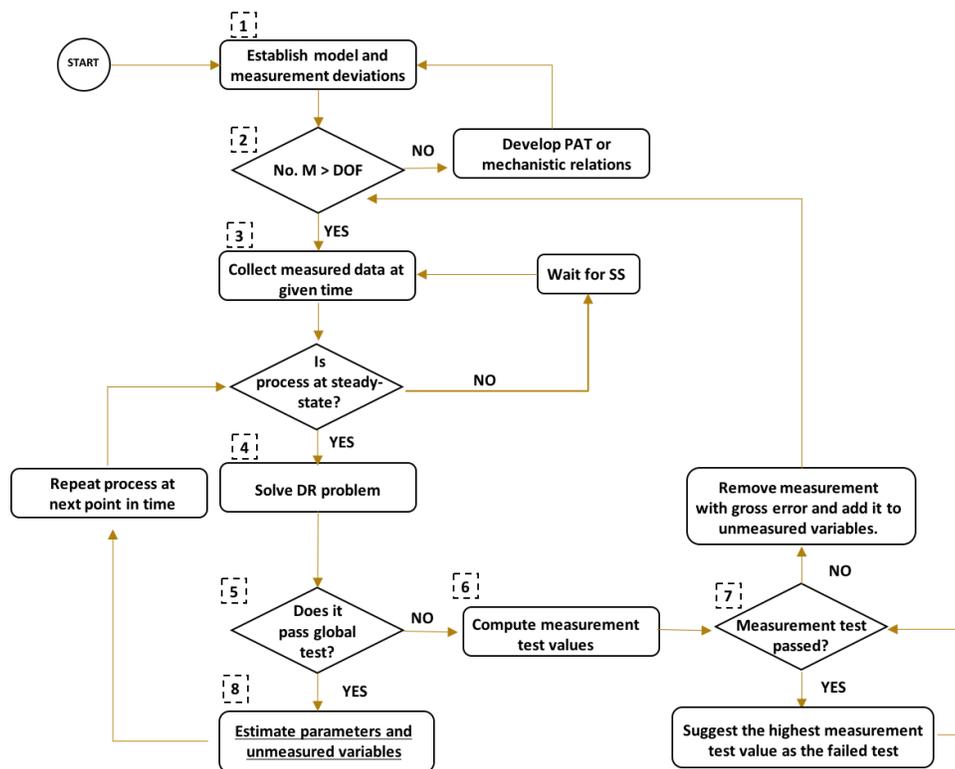
Figure 2.1. SSDR-M Logic Flow.

2.3.2 SSDR-D framework

In the multivariate statistical methods-based alternative, the process model is replaced by a PCA model that is constructed from historical data collected from the process at normal steady

state operating conditions (NOC). Once the PCA model is constructed and a subset of principal components selected, the process of computing reconciled measured, and corresponding estimates of unmeasured values is repeated with each new measurement set, followed by the application of selected statistical tests, such as the SPE, to detect gross errors. Figure 2.2 shows the SSDR-D framework flow.

SSDR-D logic flow

- Step1. Model Building: Determine the error-covariance matrix and select historical data at NOC, without outliers. Train the PCA model and evaluate the number of principal components. Cross-validate the model and validate it with a new data set. If the system has unmeasured variables, add a PLS model or mechanistic relations to estimate them using $\hat{\mathbf{X}}$.

- Step 2. Check that the process remains at steady-state [34]. Given the measured data at a given point in time, use the model to monitor the process; it is necessary to reconstruct the observations by using the $\mathbf{T_m}$ and $\mathbf{P}$ loadings as in Eq. 2.7.

- Step 3. Choose a confidence interval (i.e., 99%) and detect outliers in the process using equation 2.13, which represent the SPE statistical tests [36]. SPE follows a χ^2 distribution, similar to the GT. If any of these tests values are larger than the confidence interval, most likely there is a gross error, go to step 4. If not, proceed to step 5.

- Step 4. Evaluate the power contributions per variable and select the highest one; most likely, that is the measurement that has the gross error. Treat the measurement with gross error as missing data and go to step 1.

- Step 5. Accept the reconstructed values from equation 2.7 and estimate the unmeasured variables through a PLS model or appropriate mechanistic relations established in step 1.

- Step 6. Repeat algorithm at the next point in time.
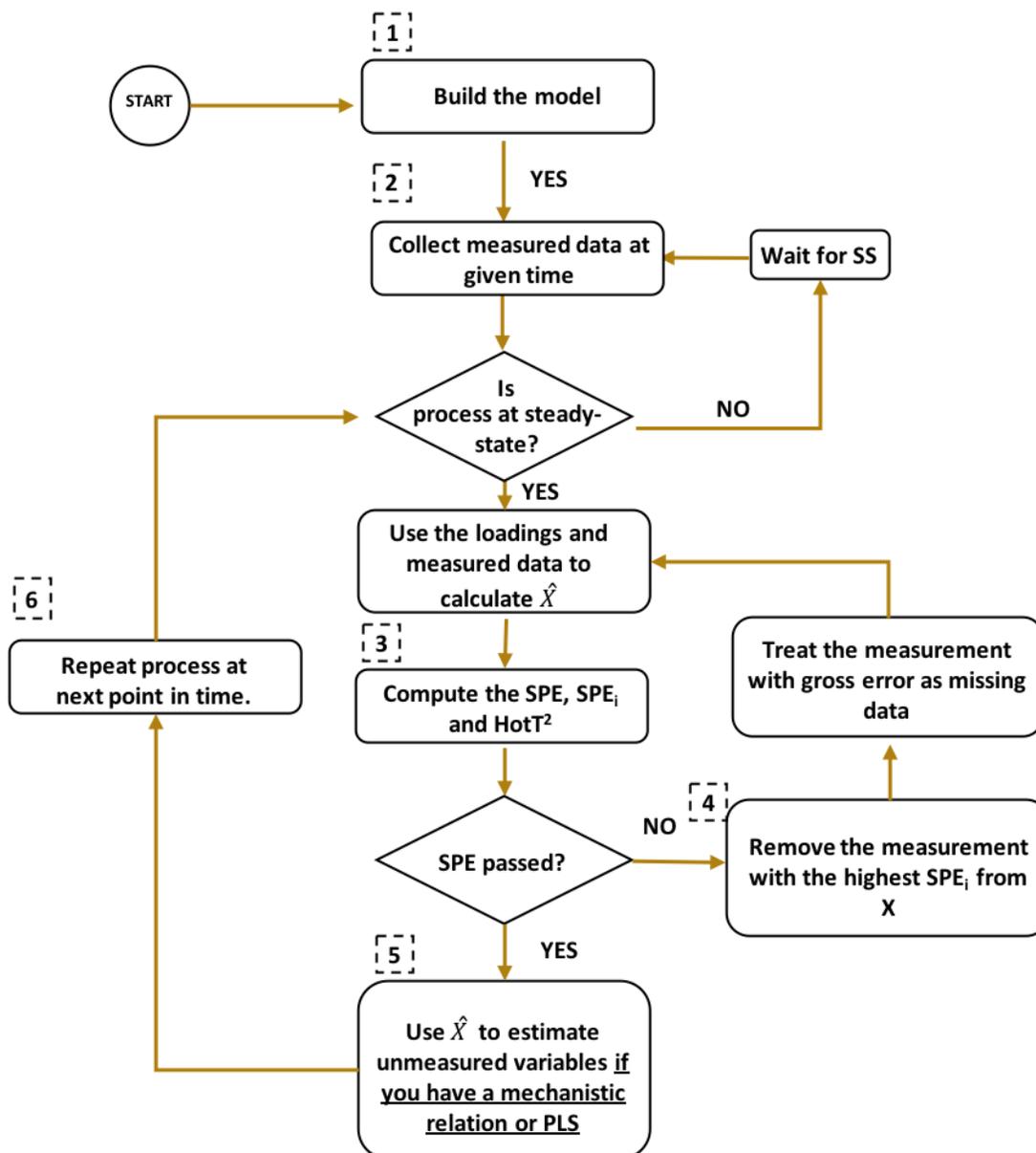
Figure 2.2. SSDR-D Logic Flow.

2.4 Application Studies

Solid dosage tablets, the dominant dosage form for oral drug delivery, can be produced via continuous direct compression (DC). In this process, the active pharmaceutical ingredient (API) and the excipients are fed using loss-in-weight feeders, continuously blended and conveyed into the tablet press to produce the tablet [37]. The main CQA's of tablets produced in a direct

compression line are tablet composition, tablet hardness, tablet weight and physical dimensions as well as tablet dissolution characteristics. In-process controls may include CPP's and CQA's such as mass flows of the feed materials and powder blend, blend composition and tablet production rate.

2.4.1 Experimental System

For illustrative purposes, two cases will be presented that focus on the powder feeding and mixing components of the process. Figure 2.3 shows the block diagram of the continuous tableting pilot plant under study and highlights the subsystem we studied. Example 1 (E1) consists of two feeders: feeder 1 provides the API flow (F_{API}) and feeder 2 the excipient flow (F_{exc}), followed by the blender. Example 2 (E2) consists of the E1 system followed a third feeder (F_{lube}) and another blender. The main CQA's and CPP's in this subsystem are: powder blend composition after each blender and the powder flows.
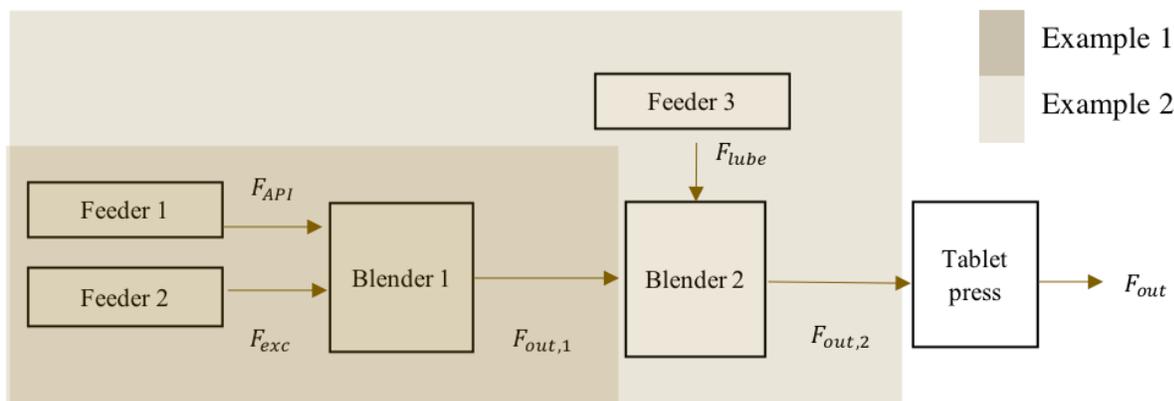
Figure 2.3. Block Diagram of the Direct Compression Tableting Line.
Feeder and Blender Subsystems.

2.4.2 Materials and Equipment

Feeder 1 and feeder 2 are Schenck AccuRate AP-300 feeders, which transport acetaminophen (APAP obtained from Mallinckrodt Inc., Raleigh, North Carolina) and Avicel

Microcrystalline Cellulose pH 102 (MCC 102, acquired from FMC BioPolymer Corporation in Philadelphia, Pennsylvania). Feeder 3 is a PUREFEED DP-4 Disc Feeder that introduces Magnesium Stearate (MgSt, from Spectrum Chemical Manufacturing Corporation in Gardena, CA) for Example 2 Case I and Silicon Dioxide (Cab-O-Sil Untreated Fumed Silica, obtained from Cabot) for Example 2 Case II. The feeders use load cells to measure the weight-loss, have a scan rate of 16 million counts per minute and use a window time of ten seconds to average said measurement. The feeder has an internal proprietary proportional-integral (PI) feedback controller, which based on changes in the weight-loss measurement, sends a signal of 4-20 mA, to the auger motor to increase or decrease the speed. The two blenders are Gericke GCM 250 mixers, which are operated at 200 rpm in the experiments reported herein.

2.4.3   PAT tools and Sensor Calibration

In the case of E1, the APAP composition ($x_{API,1}$) at the outlet of Blender 1 is measured using a CDI -256-1.7T1 NIR spectrometer and the total flow with a Mettler Toledo Balance. The Mettler Toledo Balance arrangement is not a common PAT tool, but it is used for these experiments for convenience in demonstrating the SSDR framework. The CDI NIR collects eight spectra of the flowing blend stream every second with an integration time of 4 ms, using wavelengths from 904 nm to 1687 nm [38]. The CDI spectra are related to APAP composition through a calibration function, which must be developed specifically for the materials processed. Calibration should ideally be done at production flowrate; although smaller flowrates are acceptable. The calibration was executed at a flow rate of 10 kg/h, and the Solvias Turbido NIR probe was positioned to be perpendicular to the powder flow. The combination of measurement and calibration function can be viewed as a chemometric model. The API composition measurement at the exit of the first blender ($x_{API,1}$) employs a chemometric model that uses a PLS

model (CDI-NIR sensor 1) with two principal components. The spectra pretreated using the Savitzy-Golay filter, first derivative and Standard Normal Variate (SNV), over a 5%-17% (wt.) APAP composition range and MCC200 as the excipient, following a method similar to that employed by Vanarase A. et al. [39]. For purposes of E2, the Mettler Toledo balance was moved to the exit of the second blender. The API composition measurement at the exit of the second blender ($x_{API,2}$) employs a chemometric model that uses a PLS model (CDI-NIR sensor 2) with three principal components. The spectra pretreated using the Savitzy-Golay filter, first derivative, and SNV, over a 5%-15% (wt.) APAP composition range, 0.2% Silicon Dioxide (SiO_2) and the rest of the mixture was MCC200. Figure 2.4 shows the equipment and the sequence used in the pilot plant for each example.
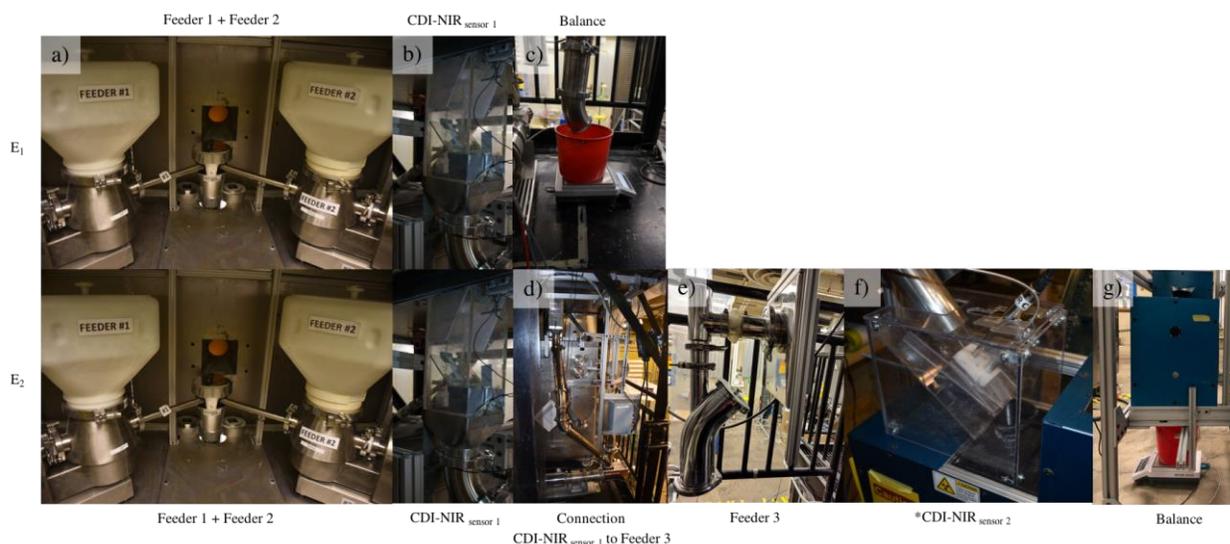
Figure 2.4. Feeder and blender PAT tools location according to E1 and E2.

The API and excipient feeders load cells (a), CDI-NIR probe at the exit of the first blender (b), the Mettler Toledo balance (c), the connection between the first to the second blender (d), the lubricant feeder load-cell (e), CDI-NIR probe at the exit of the second blender (f), and the Mettler Toledo balance at the exit of the second blender (g). *Only applicable for E2, case 2.

2.4.4   Estimation of Reconciled Values

2.4.4.1   SSDR model (SSDR-M)

At steady state, the feeder and blender subsystem can be represented through simple material balance relations, consisting of Eq. 2.17 and Eq. 2.18. It is assumed there is no material loss in the process. This assumption is only used for building the steady-state model. Material loss can be detected by data reconciliation. Of course, all flows have to be non-negative, and the composition in the range of 0 to 100 (%wt.).

$$F_{API} + F_{exc} - F_{out,1} = 0 \qquad\qquad (2.17)$$

$$\frac{x_{API,1}\ F_{out,1}}{100} - F_{API} = 0 \qquad\qquad (2.18)$$

Where: F_{API} is the reconciled value of the API flow (kg/h), F_{exc} is the reconciled value of the excipient flow (kg/h), $x_{API,1}$ is the reconciled value for the API composition at the exit of the first blender (%wt) and $F_{out,1}$ is the reconciled value of the outlet flow of the first blender (kg/h). In this simple example, there are four measured variables and two degrees of freedom.

In the E2 example, the lubricant flow and the composition at the exit of the second blender are additional variables. Eq. 2.19 represents Eq. 2.17 and Eq. 2.18 rearranged, while Eq. 2.20 to Eq. 2.22 represent the additional constraints.   Note that in E1, all variables are measured; while in E2, F$_{out,1}$ becomes an unmeasured variable. In addition, in E2, the composition measurements at the exit of the second blender are also unmeasured variables. Potentially, another NIR sensor could be added to monitor the composition of this stream. Table 1 lists the Q of the measurements, where Q is the known covariance matrix based on the relative standard deviation of the measurements. In the case of E2, there are five measured variables, three unmeasured variables and three degrees of freedom (see Table 1). Usually, **Q** for SSDR-M is determined from historical data. In these case

studies to facilitate comparison, the **Q** used in SSDR-M (shown in Table 2.1) is computed from the SSDR-D training data.

$$\frac{x_{API,1}\,(F_{API} + F_{exc})}{100} - F_{API} = 0 \tag{2.19}$$

$$F_{API} + F_{exc} + F_{lube} - F_{out,2} = 0 \tag{2.20}$$

$$\frac{x_{API,2}\,F_{out,2}}{100} - F_{API} = 0 \tag{2.21}$$

$$\frac{x_{lube}\,F_{out,2}}{100} - F_{lube} = 0 \tag{2.22}$$

Where: F_{API} is the reconciled value of the API flow (kg/h), $F_{exc,1}$ is the reconciled value of the excipient flow (kg/h), F_{lube} is the reconciled value of the lubricant flow (kg/h), $x_{API,2}$ is the estimated value for the API composition at the exit of the second blender (%wt), x_{lube} is the estimated value of the lubricant composition at the exit of the second blender (%wt) and $F_{out,2}$ is the reconciled value of the outlet flow of the second blender (kg/h).

Table 2.1. Normal operating conditions and covariance of the feeder blender subsystem

Variable	Equipment	Set-Point	Q (%)			Type of variable		
			E1	E2, Case I	E2, Case II	E1	E2, Case I	E2, Case II
F_{API} (kg/h)	Load cells	1	2	7	1	Measured	Measured	Measured
F_{exc} (kg/h)	Load cells	9	2	7	2	Measured	Measured	Measured
$x_{API,1}{}^{*}$ (% wt)	CDI NIR	10	6	6	4	Measured	Measured	Measured
$F_{out,1}{}^{*}$ (kg/h)	Load cells	10	8	--	--	Measured	Unmeasured	Unmeasured
F_{lube} (kg/h)	Load cells	0.05	--	15	15	NA	Measured	Measured
$x_{API,2}{}^{*}$ (% wt)	CDI NIR	9.9	--	--	6	NA	Unmeasured	Measured
$x_{lube}{}^{*}$ (% wt)	CDI NIR	0.5	--	--	--	NA	Unmeasured	Unmeasured
$F_{out,2}{}^{*}$ (kg/h)	Load cells	10.05	--	8	8	NA	Measured	Measured

* Average using a window time of 10 seconds. **Q** was determined using the training data for the multivariate model

For purposes of solving the maximum likelihood optimization problem, the feeder flows are bounded from 0 kg/h to 50 kg/h, the two API composition from 0% to 100%, the lubricant composition 0% to 2%, and the total flows from 0 kg/h to 100 kg/h. The problem was initialized at the nominal operating condition shown in Table 2.1. The solution of the reconciliation problem is obtained using the "fmincon" MATLAB function' which is an implementation of an "interior-point" optimization method. The fmincon parameter options selected were: tolerance function =1e-8, step tolerance =1e-8, relative maximum constraint violation =1e-9, maximum function evaluations=1e9, and maximum iteration count = 1e6. For the measurement test, the nonlinear composition constraints are linearized using Taylor series expansion evaluated at the set point.

2.4.4.2 SSDR Multivariate models (SSDR-D)

The PCA model for E1 is developed using the four measured variables: the API flow, the excipient flow, the API composition and the total flow. The PCA model, using normal operating data with values ranging from 1 kg/h to 1.7 kg/h for feeder 1, 5.7 kg/h to 9 kg/h for feeder 2, 10% to 15% API composition, and 6.7kg/h to 10 kg/h for $F_{out,1}$. Two principal components were sufficient to represent the process, the PCA model has a total explained variance (R^2) of 95% and after the cross-validation a prediction power (Q^2) of 91%.

For purposes of case E2 case I, the training data is for the same conditions as for E1, with the additional data for feeder 3 ranging from 50 g/h and 100 g/h; and MgSt composition from 0.5% to 1% (%wt.). Using three principal components, the model has an R^2 of 94% and after the cross-validation a Q^2 of 89%. For E2 case II, the system is the same, but we add a measurement: the composition of acetaminophen at the exit of the second blender. In this example, feeder 3 is operated from 0 g/h and 25 g/h; in the E2 case II, silicon dioxide (SiO_2) composition ranged from 0% to 0.2% (%wt.). Using three principal components, the model has an R^2 of 96.5% and after the

cross-validation a Q^2 of 95.5%. The SSDR-D models were established using *phi*, a MATLAB toolbox which uses the NIPALS algorithm to deal with missing variables [40].

2.5 Results and Discussion

2.5.1 E1 Results

Figure 2.5 shows the reconciliation results obtained using the SSDR-M framework as implemented in real time. On the left-hand block of plots, four traces are shown: the blue curves are the actual measurements, the green curves the corresponding reconciled values using the four measurements (SSDR-M all M), the magenta curves reconciled values using three measurements (SSDR-M no GE), where the API composition was deleted, and the red lines the gross error test limits. Initially, the reconciliation proceeds with no gross errors detected. However, at around 100s and again from 200 to 328s large gross errors are detected. The first gross error detection is due to a temporary fouling of NIR sensor, after which the process resumes acceptable operation. However, the second gross error which occurred during the time period 200s to 320s powder was persistent and was observed to be due to a significant accumulation of powder on the NIR sensor which required a compressed air burst for removal of the fouling material.

On the right-hand set of plots in Figure 2.5, the bottom figure shows the progress of the global test and the top three the measurement test results over time. Note that the measurement test failure occurs with the composition measurement. Therefore, the composition variable was removed from the reconciled values and added to the unmeasured variables for the period of times during which it failed (grey area). Reconciliation was executed again, and the global test passed from 200s to 328s. If the global test had failed again, it would not be possible to reconcile because a lack of redundancy.

At around 100s-200s and between 600s and 820s there are small but frequent gross errors detected. In this example, these errors are caused by fluctuation in the composition measurement; the measurement test value for the composition (right-hand image in Figure 2.5) proves it. Depending on the process limits, these might or might not have a significant effect. In this case, these fluctuations are believed to be due to fouling of the CDI probe at the exit of the first blender; a second measurement in another location might be needed for redundancy.
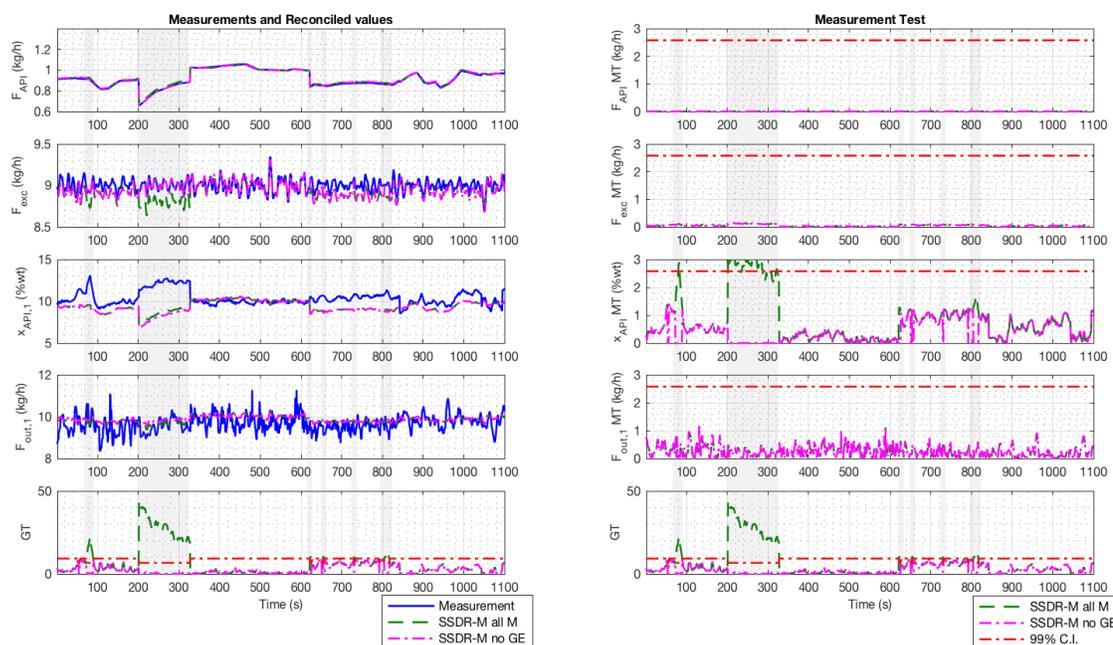
Figure 2.5. E1 SSDR-M framework results.

Reconciled values and measurements (left), global test and measurement test (right) in real-time. SSDR-M all M was reconciling with all measurements, SSDR-M no GE removes the gross error (grey area), the limit used was 99% confidence interval (C.I.).

As can be seen in Figure 2.6, the reconciled values obtained by the two approaches, (when all four measurements are used, and gross errors are absent) are quite comparable. The global test also gives comparable results to the SPE, since both tests follow a chi-square distribution using a 99% interval. In the SSDR-D framework, the power contribution (absolute value) at time 200s is

0.02 (%wt) for the API composition, while for the others it is below 0.015 (kg/h). This information

is comparable to the measurement test results. Both approaches indicate that the error is likely to

lie in the composition measurement.
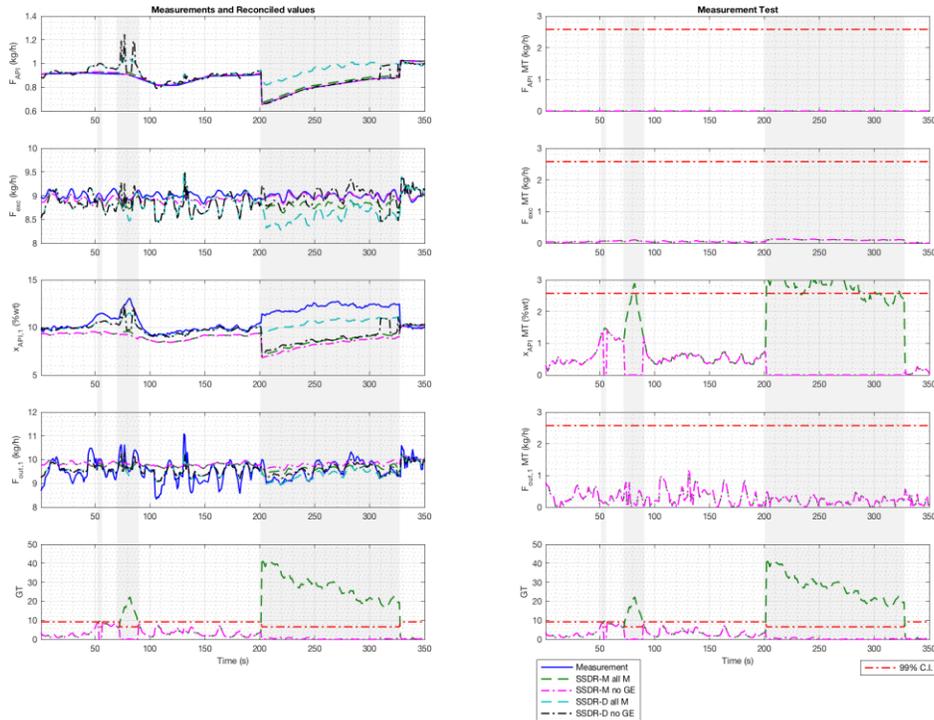
Figure 2.6. SSDR-D (PCA) and the SSDR-M framework results.

SSDR-M all M is reconciling with all measurements; SSDR-M no GE removes the gross error
when it triggers (grey area); SSDR-D all M considering all measurement and SSDR-D no GE
removes the gross error (grey area).

In order to evaluate SSDR-M and SSDR-D results, we use equation 2.1, the value of the

likelihood function to compare the results (see Table 2.2). The first column shows the points of

time for evaluation, which were randomly selected. In both cases, reconciliation was possible

because there are at least three measurements active in the algorithm. If there was a gross error, it

was located in the composition. If we compare the values of Eq. 2.1 for SSDR-D and SSDR-M,

there is not one method that consistently gives a lower value of the likelihood function than the

other. However, the material balance constraints (Eq. 2.17 and Eq. 2.18) evaluated at the

reconciled values obtained from SSDR-D, show deviations from zero, indication that the reconciled values obtained via SSDR-D do not satisfy the mass balances and thus the lower values of Eq. 2.1 arise simply because these reconciled values are infeasible. In the case of the reconciled values obtained via SSDR-M, the constraints (Eq. 2.17 and Eq. 2.18) have a maximum constraint tolerance of 1e-9 set in fmincon.

Table 2.2. Objective function evaluation for E1 SSDR-M and SSDR-D

Sample time (s)	Measurements in SSDR-M	Measurements in SSDR-D	SSDR-M Eq. 2.1	SSDR-D Eq. 2.1	SSDR-D Eq. 2.17	SSDR-D Eq. 2.18
3	4	4	1.93	8.10	0.18	0.04
30	4	4	3.30	6.54	0.18	0.03
38	4	4	2.10	1.82	0.18	0.03
92	4	4	5.56	6.70	0.18	0.03
141	4	4	3.46	2.82	0.18	0.03
271	3	3	0.02	0.06	0.18	0.02
280	3	3	0.01	0.57	0.18	0.02
286	3	3	0.01	0.66	0.18	0.02
304	3	3	0.01	0.56	0.18	0.02
337	4	4	0.07	2.03	0.19	0.03

Under the SSDR-M framework, the measurement with a gross error was discarded when the global test failed. This corresponded to the composition measurement, and so it was moved to the unmeasured variable set and estimated using re-reconciled values of the flows and the component balance equation. In the SSDR-D case, when the error was detected, the PCA model treated that measurement as a missing measurement and computed its value using the NIPALS algorithm. The error detection activates when the SPE tests fail, and the power contribution shows there is an error in the composition. Figure 2.6 also shows the SSDR-M and the SSDR-D results when gross errors arise. It is important to note that, under the SSDR-D approach, the SPE results shown are only of the input variables (SPE_X). The reconciled values from SSDR-M and SSDR-D

begin to differ only in the presence of the gross error for the variables where the gross error is located; while the API flow and excipient flow are still similar. Note that if an additional sensor were to fail (i.e., fouling, probe disconnection, etc.), it would no longer be possible to generate reconciled values because there would be no measurement redundancy in the system. Under these conditions, it is in general not likely that the variable values as measured would satisfy all of the equations of the process model exactly.

2.5.2 E2 Case I Results

The results for a run of E2 are shown in Figure 2.7. In this run, there are no gross errors (using 99% confidence interval), and thus the two frameworks give very similar reconciliation results since the additional underlying relationships between the measured variables are linear. However, there are differences in the treatment of the unmeasured variables. In the SSDR-M approach under normal operating conditions, the model equations will generate values of the unmeasured variables from the reconciled measured variable values. However, in the SSDR-D approach the training data consists only of measured values; thus, the corresponding PCA model cannot reflect information associated with the unmeasured variable. Therefore, to estimate the unmeasured variable, the PCA model must be augmented with equations (in this case a composition balance) in order to compute the unmeasured composition using the reconciled values of the measured variables. Of course, one can also add another composition measurement at the exit of the second blender. However, we can only estimate the output flow of the first blender through a mass balance (Eq. 2.17).
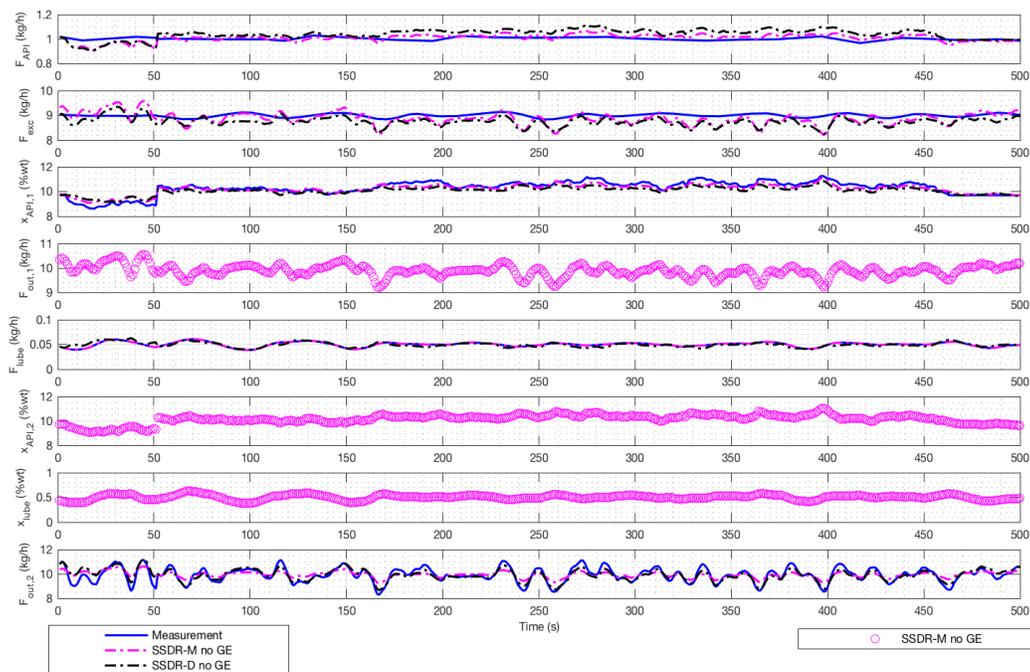
Figure 2.7. E2 case I framework results.

SSDR-M and SSDR-D (PCA) reconciled values, unmeasured variables are marked as 'o'.

2.5.3 E2 Case II Results

In E2 Case I, we are able to see how SSDR-M estimates unmeasured variables. However, we did not compare the performance between SSDR-M and SSDR-D. Since the data reported based on actual measurements (non-simulated data), the "true" values of the process variables are not known – they are only known within the limits of measurement errors. The objective of E2 Case II is to show the importance of redundancy of measurements in comparing SSDR-M and SSDR-D results. For E2 Case II we added a composition measurement to increase redundancy. After analysing the data from the pilot plant operation, we selected a period of time, where there were no gross errors, therefore, reconciled and measurement values were closed to each other. In the period of time from 11s to 110s, we inserted a bias of +3%wt in the second composition measurement offline (simulated scenario).

In Figure 2.8, are shown the reconciled values from SSDR-M and SSDR-D, with and without gross errors. Aside from the number of measurements and their location, the main difference between E1 and E2 is the way SSDR-D manages gross errors. In E1 if we have one gross error it represented 25% of missing data; in E2 one gross error represents approximately 17% of missing data (satisfying the heuristic rule). SSDR-D all m denotes the results of reconciling with all measurements using PCA; in SSDR-D no GE when the gross error is triggered (grey area), the measurement with gross error is removed and is treated as a missing variable. The NIPALS algorithm is then used to estimate the value of that missing variable.
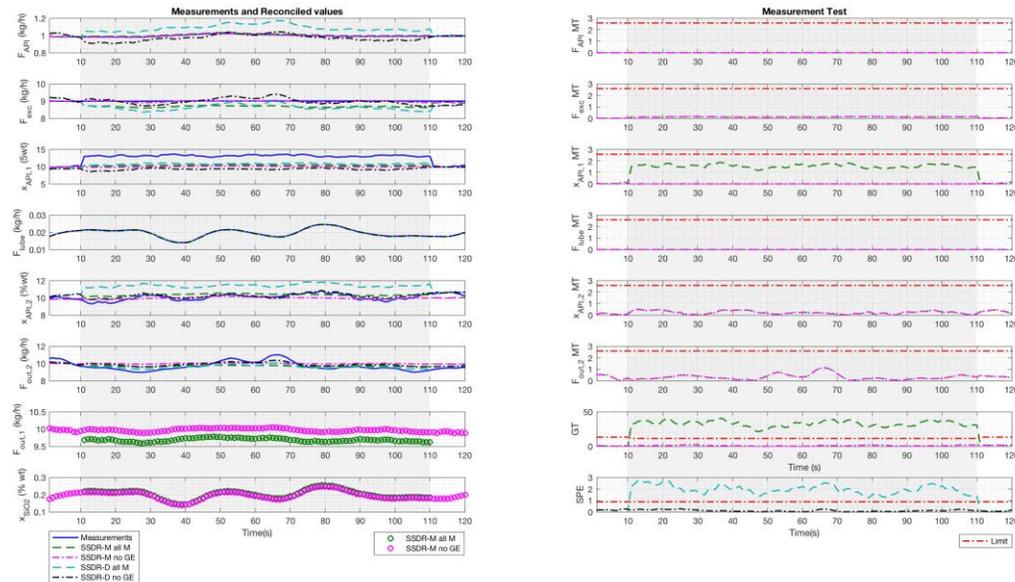
Figure 2.8. E2 case 2 SSDR-D (PCA) and the SSDR-M framework results.

SSDR-M all m indicates the results of reconciliation with all measurements; SSDR-M no GE removes the gross error when it triggers (grey area); SSDR-D all m considers all measurement and SSDR-D no GE treats gross errors as missing measurements; unmeasured variables are marked as 'o'.

In the left-hand of Figure 2.8, the reconciled values and measurements are shown. SSDR-D results show larger variations in the API flow and excipient flow than SSDR-M; while for the other variables, the SSDR-M and SSDR-D reconciled values are similar. At time 10s, the global

test and the SPE chart detect the injected error. The measurement test values suggest that the second API composition is likely at fault; this result is in accordance with the bias added in the measurement. This bias distorted the unmeasured variables estimates, especially the total flow estimated at the exit of the first blender. Nevertheless, the constraint values are not zero (see Table 2.3) for the SSDR-D solution, indicating some infeasibility. For SSDR-M the maximum constraint tolerance is 1e-9. Moreover, Eq. 2.18 and Eq. 2.22 give an estimate for unmeasured variables, such as the flow at the exit of the first blender and the lubricant composition.

Table 2.3. Objective function evaluation for E2 Case 2 SSDR-M and SSDR-D

Sample time (s)	Measurements in SSDR-M	Measurements in SSDR-D	SSDR-M Eq. 1	SSDR-D Eq. 1	SSDR-D Eq. 19	SSDR-D Eq. 20	SSDR-D Eq. 21
5	6	6	0.31	11.50	-0.06	0.11	0.02
6	6	6	0.19	8.11	-0.06	0.12	0.03
13	5	5	1.19	59.11	-0.05	0.10	0.07
34	5	5	0.82	8.81	-0.04	0.07	0.03
39	5	5	0.67	12.42	-0.05	0.04	0.01
46	5	5	0.16	9.36	-0.05	0.10	0.03
53	5	5	0.62	5.35	-0.07	0.13	0.02
84	5	5	1.43	3.18	-0.03	0.14	0.07
92	5	5	0.44	17.55	-0.04	0.09	0.05
96	5	5	0.56	22.29	-0.04	0.08	0.04
99	5	5	0.57	19.50	-0.04	0.08	0.04
114	6	6	1.56	2.21	-0.04	0.09	0.03

Normally, in order to compare the two different approaches, SSDR-M and SSDR-D, we would use the actual value of the variables to compute the RMSE and residuals. In a simulated scenario, this would be the measurement before adding noise. In the E2 case II we use actual measurements rather than simulations. Thus, in order to perform the comparison, we treat the measurement (before inserting the bias) as the point of comparison and treat it as the "actual value". Table 2.4 represent the RMSE between the "actual value" and the reconciliation results. In

this way, we can evaluate the performance of models in the presence of gross errors. In four variables out of six, SSDR-M gives the result closest to the measurement. For the composition at the exit of the second blender SSDR-D is slightly better. Figure 2.9 shows the average absolute residual per model, based on equation 2.23, which uses the real measurement value (before inserting the bias) to evaluate performance. The lower the residual, the better the performance. In most of the variables, dealing with gross errors by treating them as unmeasured variables, improves the estimated result. The flow measurement is the only exception. It should be noted that as expected, whether using the SSDR-M or SSDR-D, the variability, which results if gross errors are not eliminated, is very high.

Table 2.4. Performance of SSDR-M and SSDR-D Case 2

Variable	NOC	RMSE SSDR-M no GE	RMSE SSDR-D no GE
F_{API} (kg/h)	1.00	7.61E-04	3.86E-02
F_{exc} (kg/h)	9.00	4.45E-02	1.74E-01
$x_{API,1}{}^{*}$ (% wt)	10.15	2.94E-01	9.49E-01
F_{lube} (kg/h)	0.02	1.39E-05	4.14E-05
$x_{API,2}{}^{*}$ (% wt)	10.06	3.20E-01	2.31E-01
$F_{out,2}{}^{*}$ (kg/h)	9.66	5.80E-01	3.96E-01

$$Residual = \frac{100*|x_i - x^+_{i,r}|}{x^+_{i,r}} \tag{2.23}$$

Where: $x^+ \in \Re^n$ is a vector of measurements, $x \in \Re^n$ is a vector of reconciled values, r subscript stands for the real value.
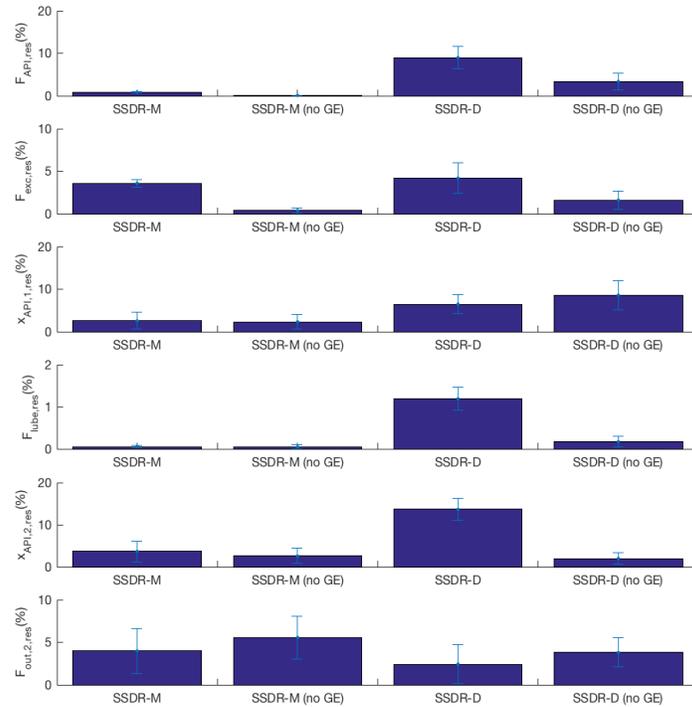
Figure 2.9. Average performance of SSDR-M and SSDR-D in the presence of gross errors (GE).

2.6 Conclusions

Data reconciliation and gross errors detection in real-time are important for the effective operation of a production line given that the online sensors invariably have significant measurement error. The SSDR-M framework proves to be an efficient way to determine the most likely value of process variables and to identify gross errors during operation. Providing the sensor network has measurement redundancy, the reconciled values are preferred to the raw measurements for use in effective process monitoring and control.

In the simple case studies reported here, the distortion in the reconciled variables resulting from the presence of gross errors is readily observed. However, the ability to effectively predict process variables in the presence of gross errors is dependent on measurement redundancy in the sensor network. This, of course, imposes equipment and operating cost in the process. For the

direct compression continuous tableting line, it is imperative that additional PAT tools be introduced for the remaining CQA's and CPP's [41]. For instance, on-line measurement of tablet hardness and weight or the deployment of PAT tools for powder flow. In continuing work in our team, these are being investigated.

As confirmed in this work, an SSDR-D approach which uses multivariate statistical models can generate reconciliation results equivalent to those obtained via SSDR-M under appropriate conditions: specifically, if the relationships between the measured variables are linear and n-k principal components are used. The reconciled values will differ, when nonlinearities arise, such as when the tablet property measurements used in the model associated with the tablet press, are added to the reconciliation problem. In terms of gross error detection for the SSDR-D approach, the SPE criterion gives results comparable to the global test used in SSDR-M, since by definition both instances follow a Chi-Squared distribution. The SSDR-M measurement test gives comparable results to the power contribution used in SSDR-D in these case studies. However, at this point, we have not confirmed that they are strictly speaking statistically equivalent. The two approaches also differ in the manner of treatment of unmeasured variables. In the SSDR-M approach, unmeasured variables simply are accommodated as dependent variables in the process model. In the SSDR-D approach, unmeasured variables can only be estimated if the PCA model is supplemented with appropriate mechanistic equations relating the unmeasured variables to the reconciled measured variables.

2.7 References

[1]   U. S. D. of H. and H. S. FDA, "Guidance for Industry PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance," no. September, p. 16, 2004.

[2] M. Ierapetritou, F. Muzzio, and G. V. Reklaitis, "Perspectives on the Continuous Manufacturing of Powder-Based Pharmaceutical Processes," *AIChE J.*, vol. 62, no. 6, pp. 1846–1862, 2016.

[3] S. L. Lee, T. F. O'Connor, X. Yang, C. N. Cruz, S. Chatterjee, R. D. Madurawe, C. M. V Moore, L. X. Yu, and J. Woodcock, "Modernizing Pharmaceutical Manufacturing: from Batch to Continuous Production," *J. Pharm. Innov.*, vol. 10, no. 3, pp. 191–199, Sep. 2015.

[4] Q. Su, M. Moreno, A. Giridhar, G. V Reklaitis, and Z. K. Nagy, "A Systematic Framework for Process Control Design and Risk Analysis in Continuous Pharmaceutical Solid-Dosage Manufacturing," *J. Pharm. Innov.*, vol. 12, no. 4, pp. 327–346, Oct. 2017.

[5] C. Knopf, *Introduction to Data Reconciliation and Gross Error Detection, in Modeling, Analysis and Optimization of Process and Energy Systems*. Hoboken, NJ, USA: John Wiley & Sons, 2011.

[6] S. Narasimhan and C. Jordache, *Data reconciliation and Gross Error detection*. Houston, TX.: Gulf Publishing Company, 2000.

[7] M. Alhaj-Dibo, D. Maquin, and J. Ragot, "Data reconciliation: A robust approach using a contaminated distribution," *Control Eng. Pract.*, vol. 16, no. 2, pp. 159–170, 2008.

[8] H. Tong, "Studies in data reconciliation using principal component analysis," McMaster University, 1995.

[9] O. Cencic and R. Fruhwirth, "A general framework for data reconciliation-Part I: Linear constraints," *Comput. Chem. Eng.*, vol. 75, pp. 196–208, 2015.

[10] S. Narasimhan and N. Bhatt, "Deconstructing principal component analysis using a data reconciliation perspective," *Comput. Chem. Eng.*, vol. 77, pp. 74–84, 2015.

[11] S. Narasimhan and S. L. Shah, "Model identification and error covariance matrix estimation from noisy data using PCA," *Control Eng. Pract.*, vol. 16, no. 1, pp. 146–155, Jan. 2008.

[12] C. Benqlilou, "Data Reconciliation as a Framework for Chemical Processes Optimization and Control," pp. 1–211, 2004.

[13] M. Diehl, H. G. Bock, J. P. Schlöder, R. Findeisen, Z. Nagy, and F. Allgöwer, "Real-time optimization and nonlinear model predictive control of processes governed by differential-algebraic equations," *J. Process Control*, vol. 12, pp. 577–585, 2002.

[14] J. Liu, Q. Su, M. Moreno, C. Laird, Z. Nagy, and G. Reklaitis, "Robust state estimation of feeding–blending systems in continuous pharmaceutical manufacturing," *Chem. Eng. Res. Des.*, vol. 134, pp. 140–153, Jun. 2018.

[15] S. Bai, J. Thibault, and D. D. McLean, "Dynamic data reconciliation: Alternative to Kalman filter," *J. Process Control*, vol. 16, no. 5, pp. 485–498, Jun. 2006.

[16] E. L. Haseltine and J. B. Rawlings, "Critical Evaluation of Extended Kalman Filtering and Moving-Horizon Estimation," *Ind. Eng. Chem. Res.*, vol. 44, no. 8, pp. 2451–2460, 2005.

[17] C. M. Crowe, Y. A. G. Campos, and A. Hrymak, "Reconciliation of process flow rates by matrix projection. Part I: Linear case," *AIChE J.*, vol. 29, no. 6, pp. 881–888, Nov. 1983.

[18] N. Arora and L. T. Biegler, "Redescending estimators for data reconciliation and parameter estimation," *Comput. Chem. Eng.*, vol. 25, no. 11, pp. 1585–1599, 2001.

[19] A. Almaya, L. De Belder, R. Meyer, K. Nagapudi, H.-R. H. Lin, I. Leavesley, J. Jayanth, G. Bajwa, J. DiNunzio, A. Tantuccio, D. Blackwood, and A. Abebe, "Control Strategies for Drug Product Continuous Direct Compression—State of Control, Product Collection Strategies, and Startup/Shutdown Operations for the Production of Clinical Trial Materials and Commercial Products," *J. Pharm. Sci.*, vol. 106, no. 4, pp. 930–943, Apr. 2017.

[20] J. E. Jackson, *A User's Guide to Principal Components*. Wiley, 2005.

[21] H. Tong and C. M. Crowe, "Detection of gross erros in data reconciliation by principal component analysis," *AIChE J.*, vol. 41, no. 7, pp. 1712–1722, 1995.

[22] J. MacGregor and A. Cinar, "Monitoring, fault diagnosis, fault-tolerant control and optimization: Data driven methods," *Comput. Chem. Eng.*, vol. 47, pp. 111–120, 2012.

[23] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemom. Intell. Lab. Syst.*, vol. 58, no. 2, pp. 109–130, Oct. 2001.

[24] T. Kourti and J. F. J. F. MacGregor, "Process analysis, monitoring and diagnosis, using multivariate projection methods," *Chemom. Intell. Lab. Syst.*, vol. 28, no. 1, pp. 3–21, Apr. 1995.

[25] S. A. Imtiaz, S. L. Shah, and S. Narasimhan, "Missing Data Treatment Using Iterative PCA and Data Reconciliation," *IFAC Proc. Vol.*, vol. 37, no. 9, pp. 197–202, Jul. 2004.

[26] P. R. C. Nelson, P. A. Taylor, and J. F. MacGregor, "Missing data methods in PCA and PLS: Score calculations with incomplete observations," *Chemom. Intell. Lab. Syst.*, vol. 35, no. 1, pp. 45–65, Nov. 1996.

[27] A. Folch-Fortuny, F. Arteaga, and A. Ferrer, "PCA model building with missing data: New proposals and a comparative study," *Chemom. Intell. Lab. Syst.*, vol. 146, pp. 77–88, 2015.

[28] I.-W. Kim, M. S. Kang, S. Park, and T. F. Edgar, "Robust data reconciliation and gross error detection: The modified MIMT using NLP," *Comput. Chem. Eng.*, vol. 21, no. 7, pp. 775–782, Mar. 1997.

[29] A. Gupta, A. Giridhar, V. Venkatasubramanian, and G. V. Reklaitis, "Intelligent alarm management applied to continuous pharmaceutical tablet manufacturing: An integrated approach," *Ind. Eng. Chem. Res.*, vol. 52, no. 35, pp. 12357–12368, 2013.

[30] J. F. MacGregor, H. Yu, S. García Muñoz, and J. Flores-Cerrillo, "Data-based latent variable methods for process analysis, monitoring and control," *Comput. Chem. Eng.*, vol. 29, no. 6, pp. 1217–1223, May 2005.

[31] H. Hotelling, "The Generalization of Student's Ratio," *Ann. Math. Stat.*, vol. 2, no. 3, pp. 360–378, 1931.

[32] P. Miller, R. E. Swanson, and C. E. Heckler, "Contribution plots: a missing link in multivariate quality control," *Appl. Math. Comput. Sci.*, vol. Vol. 8, no, pp. 775–792, 1998.

[33] G. A. C. Le Roux, B. F. Santoro, F. F. Sotelo, M. Teissier, and X. Joulia, "Improving steady-state identification," *Comput. Aided Chem. Eng.*, vol. 25, pp. 459–464, 2008.

[34] M. J. Bagajewicz, *Smart Process Plants: Software and Hardware Solutions for Accurate Data and Profitable Operations: Data Reconciliation, Gross Error Detection, and Instrumentation Upgrade*, 1st ed. McGraw-Hill Education, 2010.

[35] S. Cao and R. R. Rhinehart, "An efficient method for on-line identification of steady state," *J. Process Control*, vol. 5, no. 6, pp. 363–374, 1995.

[36] P. Nomikos and J. F. MacGregor, "Multivariate SPC Charts for Monitoring Batch Processes," *Technometrics*, vol. 37, no. 1, pp. 41–59, 1995.

[37] D. M. Parikh, *Handbook of Pharmaceutical Granulation Technology, Third Edition*. CRC Press, 2016.

[38] J. Austin, A. Gupta, R. McDonnell, G. V. Reklaitis, and M. T. Harris, "A novel microwave sensor to determine particulate blend composition on-line," *Anal. Chim. Acta*, vol. 819, pp. 82–93, 2014.

[39] A. U. Vanarase, M. Alcalà, J. I. Jerez Rozo, F. J. Muzzio, and R. J. Romañach, "Real-time monitoring of drug concentration in a continuous powder mixing process using NIR spectroscopy," *Chem. Eng. Sci.*, vol. 65, no. 21, pp. 5728–5733, 2010.

[40] S. García-Muñoz, "Phi MATLAB Toolbox." Personal Communication, 2015.

[41] S. Ganesh, R. Troscinski, N. Schmall, J. Lim, Z. Nagy, and G. Reklaitis, "Application of X-Ray Sensors for In-line and Noninvasive Monitoring of Mass Flow Rate in Continuous Tablet Manufacturing," *J. Pharm. Sci.*, Sep. 2017.

[42] A. U. Vanarase and F. J. Muzzio, "Effect of operating conditions and design parameters in a continuous powder mixer," *Powder Technol.*, vol. 208, no. 1, pp. 26–36, 2011.

CHAPTER 3. STEADY-STATE DATA RECONCILIATION FOR A CONTINUOUS TABLETING LINE VIA DIRECT COMPRESSION: A CASE STUDY

3.1 Introduction

Reliable process monitoring is one of the important technologies that support the management of product quality [1] in pharmaceutical manufacturing. The use of quality-by-design and the development of Process Analytical Technology (PAT) tools have been highly encouraged by the FDA and regulatory agencies outside of the US [2]. However, although PAT tools are quite powerful, it is important to recognize that every measurement, especially those carried out in real time, is subject to errors. Thus it is necessary to take appropriate steps to correct these measurements to obtain the best possible estimate of the real state of the process [3] on which to base manufacturing decisions. The objective of this paper is to show how data reconciliation (DR) can be carried out in real time for this purpose and to demonstrate its application to a direct compression continuous tableting line operating at steady state. In this work it is shown how to integrate both at-line and offline measurements to improve the DR process.

There are two types of errors that are encountered in the operation of a process sensor network, namely, random errors and gross errors (e.g., sensor failures, miscalibration). One of the well-established methods to deal with random sensor errors is data reconciliation [4]. This methodology corrects the information provided by the sensors through the use of a process model, which captures the relationship between the measured variables [5]. Since gross, that is non-random, errors will generally distort the results of DR [6], it is necessary to identify and remove gross errors before one can obtain reliable reconciled values of the measured (and unmeasured) process variables. Gross error detection can be readily carried out using well known statistical tests,

which in essence are outlier detection methods. Once the measurement set has been cleared of gross errors and the measurement values have been rectified, the rectified values can be used to take appropriate process decisions, such as process control actions [7] [8].

In [9], the authors establish a steady-state data reconciliation (SSDR) framework and demonstrate its use in two process applications. The framework employs both a data-driven model (SSDR-D) and an equation-based model (SSDR-M) approach. The SSDR-D approach uses a multivariate model, in which the process is represented through principal components (PC's). The SSDR-M approach formalizes the reconciliation process as an optimization problem (see Eq. 3.1) in which a likelihood-based objective function is minimized subject to the constraints imposed on the measured and unmeasured process variables by the process model (see Eq. 3.2 to Eq. 3.3). The process model consists of the relevant relationships between the process variables, including material and energy balances, property relationships, etc. Both approaches can reconcile measurements provided that there are redundant measurements available. Moreover, SSDR-M can also generate estimates of unmeasured variables provided they are observable [10]. Since both approaches are only applicable to process measurements obtained at steady-state, this condition must be assessed through suitable tests, a number of which are reported in the literature [11]. It is important to note that it is possible to reconcile at non-steady state; however, in that case, the process model must represent the process dynamics [12]. The solution of the dynamic data reconciliation problem generally requires longer the computational time: the steady-state approach offers shorter computational time, which permits more frequent execution of reconciliation in real-time.

$$\min_{\mathbf{x},\mathbf{z}} J = (\mathbf{x}^+ - \mathbf{x})^{\mathrm{T}} \mathbf{Q}^{-1} \left(\mathbf{x}^+ - \mathbf{x} \right) \tag{3.1}$$

$$\text{s.t.} \quad \mathbf{h}\left(\mathbf{x},\mathbf{y},q\right) = 0 \tag{3.2}$$

$$\mathbf{g}\left(\mathbf{x},\mathbf{y},q\right) \pounds 0 \tag{3.3}$$

In [13], it was shown that PCA-based and process model-based approaches yield comparable results if the process model is linear, and the selected principal components are selected to represent the process adequately. In feeder-blender subsystem studies reported in [14], results obtained using SSDR-D based on PCA models and SSDR-M using material balances were indeed shown to be comparable. However, without knowing the number of equations in the process model (or equivalently the degree of freedom of the model), the selection of the number of principal components has to be based on heuristics. A poor selection of PC's can lead to the inaccurate representation of the process and thus infeasible values of the reconciled variables. It is also demonstrated that the two approaches yield reconciled values which will be distorted in the presence of gross errors and thus that such errors must be excluded. Once the measurements identified to be in error are removed and the remaining measurements reconciled, the variables associated with gross errors are treated as unmeasured in SSDR-M, while in SSDR-D they are treated as missing variables [15]. Since reducing the number of measured variables affects the redundancy of the process, it is important that the sensor network has sufficient redundancy to allow reconciliation to be carried out even in the presence of gross errors.

3.2   Equipment and Experimental Set-up

The experiments were executed using the pilot plant at Rutgers University, a schematic of which is shown in Figure 3.1. Three K-Tron feeders were used in this case study, each carrying a different component. For specific runs, the lactose flow (F_{exc}) and APAP flow (F_{API}) pass through a Quadro 197-S Comil, which has a 1000 microns screen and operates at 3045 rpm. This blend

and the Magnesium Stearate flow (F_{lube}) are conveyed into the Glatt GC70 blender. The blender is connected via a chute to the feed frame of a Fette 1200i tablet press, which has 20 stations producing tablets with a diameter (D) of 10 mm [16]. The runs included online, at-line and offline sensors to monitor the CQA'S (see Section 3.3).

The process uses measurements which are provided as integral parts of the individual unit operations as well as measurements made via external sensors. For instance, the K-Tron feeders have load cells to calculate the flow rate of each component by difference with sampling times of up to one second. Nevertheless, to avoid problems with the K-Tron software, the sampling time of the feeders was set to 3s. The blenders provide blender rpm, while, the Fette 1200i tablet press measurements includes: the feed speed (FS); the turret speed (TS); production rate; the fill depth; the thickness at pre-compression force (t_{PCF}); pre-compression force (PCF); the thickness at main compression force (t_{MCF}); main compression force (MCF); the standard deviation of the MCF. The MCF will change according to the t_{MCF}. The sampling time of the measurements of the tablet press is 0.5 s. The material is introduced into the press via a feed frame which operates at a set rpm. The external measurements are provided by suitably placed microNIR and Raman probes (see Section 3.3).
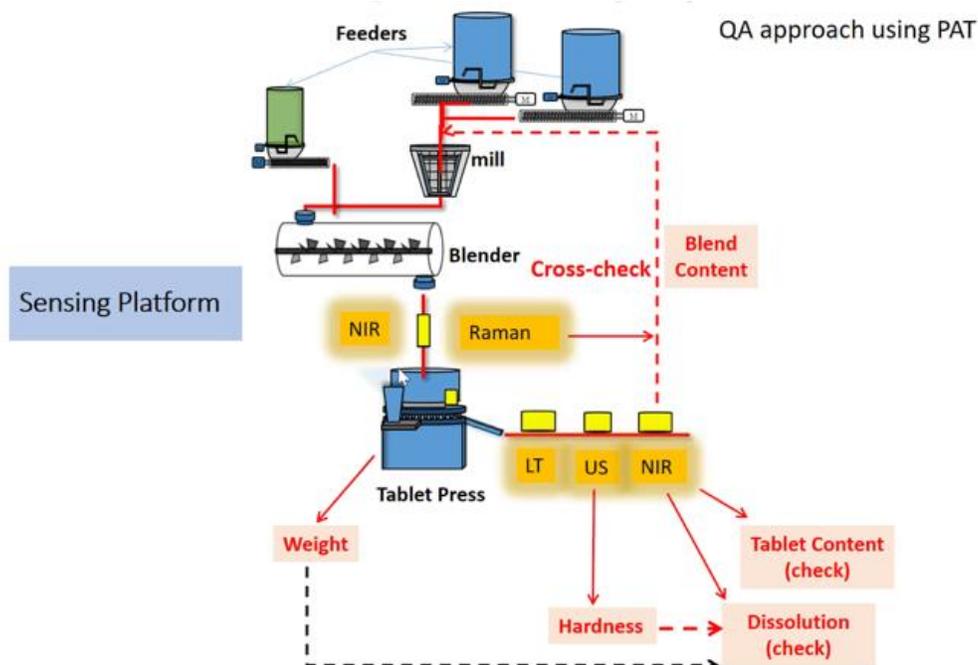
Figure 3.1. Rutgers Pilot Plant Diagram

3.3 Materials and Methods

3.3.1 Materials

The materials processed in the plant were semi-fine Acetaminophen (Mallinckrodt Inc. Raleigh, NC) which constitutes the API, Lactose Monohydrate (Foremost, farms, Rothschild, Wisconsin) is the excipient (Exc.) and Magnesium Stearate N.F. (Tyco Healthcare/Mallinckrodt, St. Louis, Missouri, USA.) is the lubricant. The formulations used are: 10% APAP, 89% to 89.5% Lactose, and 0.5% to 1% MgSt.

3.3.2 RAMAN and NIR method for powder flow

For this study, the Viavi microNIR probe and the Kaiser Raman were used for measurement of composition at the exit of the blender. These PAT tools were calibrated using data presented in [17]. For these calibration runs, samples of 1.3 kg were prepared using a V-blender and at different weight composition percentages: from 1%wt. to 15%wt. APAP, 0.5%wt. to 1%wt.

Magnesium Stearate, and 89% wt. to 97% wt. Lactose. The mixtures were introduced through the same chute that was used in the end-to-end continuous operation of the pilot plant. The Viavi microNIR and the Raman spectrometer Kaiser probe were incorporated in this chute and placed alongside each other. Each blend was used twice, and 200g were required inside the chute for the feed-frame to operate properly.

The Viavi microNIR settings were: spectral range from 908 nm to 1676 nm, a resolution of 6 nm, the integration time of 40 ms, the number of scans of 25 and a sampling time of 4s. In the case of the Raman spectrometer Kaiser probe, 3mm PhAT with a 12.5 focal length, the settings were: laser wavelength of 785nm line from an Invictus NIR diode operated at 400 mW, the number of scans of one, exposure time of 2s, overhead time of 3s, and sampling time of 5s. The collected spectra were used to create two Partial-Least-Squares (PLS) models in "Phi" MATLAB toolbox [18]. The spectra for the RAMAN spectrometer and the NIR were pretreated using the Savitzy-Golay filter, first derivative and Standard Normal Variate (SNV). The PLS model for the Viavi microNIR consists of two principal components (PC's) and has an explained variance (R^2) 93.4% and after cross-validation a prediction power (Q^2) of 93.2%. The model representing the Raman spectra has two PC'S and an R^2 98% and a Q^2 of 97.6%. In the case of the inline experiments, the scans of both spectrometers were set so they were synchronized. The sampling time from these sensors was approximately 5s; for the Raman measurement, the sampling time was fixed, while for the NIR it was between 4s-5s.

3.3.3   Tablet weight, thickness, density and tensile strength

Since real-time measurements of tablet weight, thickness, density and tensile strength are not available, it is necessary to develop suitable surrogates which can allow real-time monitoring of these properties. Accordingly, for each given set of processing conditions (see Section 3.4 for

details) a sample of six tablets was collected at set times during the continuous runs. The hardness, thickness, and weight of these tablets were then measured at-line. In particular, the tensile strength of the tablets was measured with Dr. Schleuniger, Pharmatron, model 6D, and estimated as it was in [16]. The thickness was measured with an Absolute digimatic digital caliper ($\pm$ 0.01 mm). Finally, the weight of the tablets was measured with an Adventurer Ohaus precision balance ($\pm$ 0.001g). The bulk density was estimated from these measurements, and the true density was calculated from the powder properties [16].

Since the weight and tensile strength were at-line measurements ($W_{At-line}$ and $T_{At-line}$, respectively), a model is required to predict said variables during the periods of time when direct measurements were not available. Therefore, for the inline runs, a PLS model was created to relate the tablet press measurements to the weight and tensile strength of the tablets. The inputs of the model were the setpoint composition of APAP in a range of 1%wt. to 12% wt.; the FS in a range of 20 rpm to 60 rpm; the TS in a range of 10 rpm to 50 rpm; the production in a range of 15 ktablets/h -60 ktablets/h; the fill depth in a range of 14mm -15mm; the PCF in a range of 0.5 kN to 3.5kN; the thickness at MCF (t_{MCF}) in a range of 5.7 mm to 7.2 mm; and the MCF in a range of 7kN to 24 kN. Four PC'S are used to represent the process and an R^2 77% and a Q^2 76.6%.

The PLS models serve to give estimates of the real-time values of the tablet tensile strength and weight since these are not available in real time. To improve these approximate values, the estimated tablet weight, and tensile strength real-time measurement, W_m and T_m respectively, are corrected through the use of at-line measurements. In particular, the differences between the at line measured values at a given point in time (t_i) and the PLS predictions (W_{PLS} for tablet weight and $T_{PLS,}$ for tensile strength) are used to compute the model bias (See Eq. 3.4 and Eq. 3.5). The bias (W_{Bias} for tablet weight and T_{Bias} for tensile strength) is used to estimate the real-time

measurements (See Eq. 3.6 and Eq. 3.7). The bias values are kept fixed until the next at-line

sampling data becomes available.

$$W_{Bias} = W_{PLS,t_i} - W_{At-line,t_i} \qquad (3.4)$$

$$T_{Bias} = T_{PLS,t_i} - T_{At-line,t_i} \qquad (3.5)$$

$$W_m = W_{PLS} + W_{Bias} \qquad (3.6)$$

$$T_m = T_{PLS} - T_{Bias} \qquad (3.7)$$

3.3.4   Tablets content uniformity offline

As in [17], the UV-Visible spectrophotometer, from Varian Inc., Santa Clara, CA, USA,

was used to measure the tablet content. The tablets were collected during the run, but the content

uniformity measurement was done offline. The tablets were dissolved using Phosphate Buffered

Saline (PBS) at pH 5.8. Nylon filters of 2µm were used to filter the solution, which was analyzed

at 245 nm (APAP peak location). The tablet content measurement was not used in Equation 3.1

because it is not a real-time value. However, the performance of the framework can be evaluated

by using this measurement and comparing it to the SSDR-M estimation (see section 3.5.1).

3.4   Experimental conditions and systems

3.4.1   Operating Conditions

A total of twenty-four conditions were run on the continuous tableting line. These

conditions were selected according to the Placket-Burman design, and each of them required 7 kg

of material. The design of experiments considered ten parameters: the total flow, the three

components composition, the use of the Comil, the blender speed, the chute height, the tablet press

feed speed, the fill depth of the cam and the target force in the tablet press. From these twenty-four

conditions, twenty-one were used for training data, and the rest was used to execute the SSDR framework.

The total flow through the line was set at values from 15 kg/h to 30 kg/h. For half of the runs, the lactose and APAP flow passed through the Comil set at 1000 μm, and the other half were operated without Comil. The composition range, for each component, was 2% wt. to 10%wt. APAP, 0.5% wt. to 1% wt. Magnesium Stearate, and 89% to 97% wt. Lactose. The blender, which mixed the three components, was operated between 100 rpm and 250 rpm. The mixture was fed to the chute, where the microNIR and Raman probes were located, and the powder level was kept constant during a run to assure steady-state. However, the powder level was varied at different heights over the range of 48cm to 81cm.

Under normal operating conditions (NOC) of the tablet press, the feed speed was selected between 15 rpm to 30 rpm. The fill depth was set between 14 mm and 15 mm. Finally, the MCF target was kept from 10kN to 20kN. The duration of each run depended on the selected flow rate and was approximately 14 to 28 min. Once the feeders reach steady-state, it was observed that the tablet press reached steady-state in 7 min to 10 min. The conditions used to demonstrate the SSDR framework are shown in Table 3.1.

Table 3.1. NOC used to show the SSDR framework

Label	F_{API} (kg/h)	F_{exc} (kg/h)	F_{lube} (kg/h)	Mill Speed (rpm)	Screen Size (μm)	Blender speed (rpm)	Min. Chute Height (cm)	Feed Frame (rpm)	Fill Cam (mm)	Force Target (kN)
19	1.5	13.425	0.075	3	1000	100	81	60	15	10
21	1.5	13.425	0.075	3	1000	250	81	20	14	20
7	1.5	13.350	0.150	NA	NA	250	48	60	14	10

Note. NA means not applicable.

3.4.2 Case Studies

Two cases were selected to demonstrate the SSDR framework. The first example (E1) involves the feeder and blender subsystem, which extends from the three feeders to the chute where the RAMAN and microNIR probes are positioned. The data from the upstream portions of runs 19 and 21 were used for E1. The second example (E2) extends the study to include the tablet press and the at-line measurements for tablet CQA's. The full set of data from runs 7, 19 and 21 were used for E2. Both examples only consider the steady-state operation of the pilot plant. Since the sampling time for each equipment (i.e., feeders flow and tablet press equipment) and PAT tools differed (i.e., composition measurement at the exit of the blender), the data was synchronized, and the same measured values were maintained until the measurement was updated at the next sampling time.

3.5   SSDR for the Continuous Tableting Line

3.5.1   SSDR-M Implementation

In [9], the SSDR framework was demonstrated with a feeder and blender system with three components, where the composition measurement was located at different points, and the total flow was measured. Case E1 is a similar case, except that the composition measurement is made at the same location using two different types of sensors and there is no blender outlet flow measurement. Equation 3.8 represents the mass balance in the feeder-blender subsystem for the API components for the powder blend. It is assumed there is no material loss. Equation 3.9 states that the compositions measurement from the Raman ($x_{API,1}$) and microNIR ($x_{API,2}$) probe should be equal since they are located in the same position. There are two linear-independent equations for this subsystem (derived from Eq. 8 and Eq. 9 through Taylor series expansion) and five

measurements, as a consequence, the system is redundant, and it is possible to reconcile. Since the composition measurements of the tablets leaving the line are known, they can be used as the actual composition if the tablet content is assumed to be equal to the powder blend composition (see Equation 3.10). The tablet content measurements are done offline; therefore, in real-time, this variable is an observable unmeasured variable.

$$x_{API,1}\left(F_{API} + F_{exc} + F_{lube}\right) - F_{API} = 0 \tag{3.8}$$

$$x_{API,1} - x_{API,2} = 0 \tag{3.9}$$

$$x_{API,1} - x_{API,Tab} = 0 \tag{3.10}$$

For E2, the measurements associated with the tablet press were included. Equation 3.11 represents the overall mass balance for the line. The inlet flow, which is the summation of the three components flow, must be equal to the tablet outlet flow. The tablets flow is calculated from the tablet weight (W), the TS, and the number of stations ($n_{\frac{tab}{rev}}$). In ideal steady-state, there should be no accumulation in the hopper over time. However, if there is a steady accumulation of material in the hopper, then that accumulation can be modeled as an additional flow (F_{ac}) associated with the tablet press. This pseudo-steady state model is commonly used in systems involving tanks to capture the effect of varying tank levels [19] [20].

$$F_{API} + F_{exc} + F_{lube} - \left(W*TS*n_{\frac{tab}{rev}} + F_{ac}\right) = 0 \tag{3.11}$$

The tablet press has additional valuable measurements such as the MCF and t_{MCF}. These can be included in the reconciliation through Equation 3.12, derived from the Kawakita model [21] [22] [23]. This equation represents the behavior of powder under pressure, where a represents the relative volume decrease at infinite pressure and $1/b$ represents the necessary pressure to get the

relative volume decrease *a/2*. These parameters are estimated through Monte Carlo simulations in MATLAB, using a formulation composed of 10% wt. APAP, and 0.5% wt. Magnesium Stearate and the rest was lactose. The information was gathered from the training data. Parameter *a* was estimated as 0.2744 (0.2645, 0.2844), while parameter *b* was 0.03758 (0.02732, 0.04785) MPa^{-1}. The R^2 for this model was 75%.

$$\frac{MCF}{\frac{t_{PCF} - t_{MCF}}{t_{PCF}}} - \frac{MCF}{a} - \frac{\pi D^2}{4ab} = 0 \tag{3.12}$$

The actual thickness (t) of the tablets is obtained as an at-line measurement while the measurement obtained in real time from the press is t_{MCF}. Under the conditions of these runs, the tablets underwent a relaxation which resulted in an increase of approximately 7% in the thickness from the t_{MCF} value. Accordingly, in the SSDR-M model, a thickness bias (t_{Bias}) that accounts for the difference between the tablet thickness and t_{MCF} is introduced. This thickness bias is normalized by the t_{MCF} value. In the data reconciliation, the initial t_{Bias} is set to 0.07 and kept at that value until the bias is corrected using the next at-line measurement. Equation 3.13 relates the at line tablet thickness to t_{MCF} as used in the SSDR-M framework.

$$t = \left(1 + t_{Bias}\right) t_{MCF} \tag{3.13}$$

To complete the process, model a relationship is needed to relate tablet properties to tensile strength. The percolation theory proposed by Kuentz and Leuenberger offers such a relationship as shown in Equation 3.14 [16] [24]. This equation, relating the hardness to the relative density (ρ) requires the estimation of two key parameters: the tensile strength at maximum density (σ_{max}) and the critical density at which the density start to show strength ($\rho_{c\sigma}$). These parameters were calculated from Montecarlo samples using MATLAB. The data used for this parameter estimation

was collected from the twenty-one conditions that were executed for data training described in section 3.4.1. The results obtained were σ_{max} = 4.376 MPa and $\rho_{c\sigma}$ = 0.825, with an R^2 of 87%.

$$T - S_{max}\left[1 - \left(\frac{1 - \left(\frac{4W}{\pi Dt r_t}\right)}{1 - r_{cs}}\right) e^{\left(\frac{4W}{\pi Dt r_t} - r_{cs}\right)}\right] = 0 \tag{3.14}$$

The process model for case E2 consists of Eq. 3.8 to Eq. 3.14. The system can be summarized in four independent linear equations, derived from Eq. 3.8, Eq. 3.9, Eq. 3.12, and Eq. 3.14 (Eq. 3.13 is implied) through Taylor series expansion, and nine measurements. The covariance of the measured variables was estimated from the training data, from the runs made at 10% wt. APAP and 15 kg/h. Table 3.2 shows the covariance matrix for E1 and E2. The measurements were used as initial values for the reconciled variables in the maximum likelihood optimization problem. The reconciled variables were also bounded within an operational range. In SSDR-M, the ranges used for each variable were defined as follow: the APAP flow from 0 kg/h to 7 kg/h; the lactose flow from 0 kg/h to 35 kg/h; the lubricant feeder from 0 kg/h to 0.5 kg/h, the two API composition range from 0% to 100%; the t_{MCF} range was determined from 5 mm to the t_{PCF}; the MCF from 1kN to 30 kN; the weight 600mg to 900mg; the tensile strength from 0.3 MPa to 5 MPa; the accumulation flow from -30 kg/h to 20 kg/h. It is important to note that the TS measurement was treated as a parameter due to its low variability.

For cases E1 and E2, The SSDR-M nonlinear problem is solved in MATLAB using the "fmincon" MATLAB function' which is an implementation of an "interior-point" optimization method. The fmincon parameter options selected were: tolerance function = 1e-8, step tolerance =1e-8, relative maximum constraint violation = 1e-9, maximum function evaluations = 1e9, and maximum iteration count = 1e6. The nonlinear composition constraints are linearized using Taylor

series expansion evaluated at the set point, and this version was used for the measurement test evaluation.

Table 3.2. Covariance matrix

Variable	Equipment	Q (%)		Type of variable	
		E1	E2	E1	E2
F_{API} (kg/h)	Load cells	4.23	4.23	Measured	Measured
F_{exc} (kg/h)	Load cells	1	1	Measured	Measured
F_{lube} (kg/h)	Load cells	30	30	Measured	Measured
$x_{API,1}$ *(% wt.)	Raman	6	6	Measured	Measured
$x_{API,2}$*(% wt.)	NIR	8	8	Measured	Measured
t_{MCF} (mm)	Tablet Press	--	2	Unmeasured	Measured
MCF (kN)	Tablet Press	--	Tablet Press	Unmeasured	Measured
W (mg)	Soft Sensor	--	2	Unmeasured	Measured
Ten (MPa)	Soft Sensor	--	10	Unmeasured	Measured

3.5.2 SSDR-D Implementation

For SSDR-D, PCA was used to represent the process. In the case of E1, the model is trained based on five measured variables: the API flow, the excipient flow, the lubricant flow, API composition by two different instruments. The PCA model, using normal operating data with values ranging from 0.05 kg/h to 3.3 kg/h for the APAP flow, 13.2 kg/h to 29.3 kg/h for the lactose flow, 0.01 kg/h to 0.51 kg/h for the lubricant flow, 2% to 15% API composition. E1 can be represented using three principal components (PC's), with an R^2 of 96.3% and a Q^2 of 93.7%.

Two PCA models were built for case E2 using nine measured variables. The training data, used for these models is the same historical data that was used to estimate the covariance matrix for the SSDR-M framework. The main difference between the SSDR-D models is the number of PC's selected to represent the process. The PC's, for the first model, were selected based on the

magnitudes of the eigenvalues, which had to be close to one. Based on this heuristic rule, four PC's were selected: the resulting model has an R^2 of 92.8% and a Q^2 of 91.1%. In the second model, the selection of PC's is based on the number of measurements and the number of equations [25]. The system has nine measured variables and two equations that related them. Six PC's were selected to represent the system; this decision was based on the heuristics suggested in [13] and cross validation. The second SSDR-D model has an R^2 of 98.9% and a Q^2 of 98.7%, and the PC's selection was based on the eigenvalues.

The training data, used for the two PCA models, was centered and scaled to build the models. it also covered the following range for each variable: 0.3 kg/h to 3.3 kg/h APAP flow; 13.2 kg/h to 29.3 kg/h lactose flow; 0.012 kg/h to 0.5 kg/h lubricant flow; 2% to 15% API composition; 5.9 mm to 8 mm for the thickness; 1.2 kN to 26.1 kN for main compression force; 672 mg to 877mg tablet weight; the tablet tensile strength from 0.6 MPa to 2.7 MPa. The models were built using the "phi" MATLAB toolbox as a '. mat' and implemented in MATLAB for SSDR-D and GED.

3.6   Results and Discussion

3.6.1   SSDR-M and SSDR-D reconciled values for the feeder and blender system

The first results focus on the E1 system, where the feeder, comil and blender were the only unit operations considered. For most of the run, Experiment 19 has no gross errors detected. Experiment 21 has a gross error in the Raman composition measurement since this measurement is not available during a time-period. Here the absolute error is estimated using the variable value one second before the bias correction for the tablet weight, tensile strength, and thickness takes place. In this way, at-line measurements can be used as the true values of the variables and a point of comparison to evaluate the performance of the measurements, SSDR-M and SSSDR-D. For the

composition measurements, the composition measurement of the tablets offline was used to estimate the error. For experiment 19 the composition measurement offline was 12.4 %wt. (± 0.08) and for experiment 21 the composition was 10.3 %wt. (± 0.10).

Table 3.3 shows relevant results at a given point in time: the measurements; the reconciled values through SSDR-M for measured and unmeasured variables, plus the parameters estimated; the reconciled values through SSDR-D for measured variables; if the measurements are being active; the error estimation for the variables with at-line or offline measurements. If the measurement is active, it will have a value of one, while if it is classified as having a gross error, then it will have a value of zero in the "active" column.

Table 3.3. Feeder and Blender Data Reconciliation Performance

Exp	Var	NOC			Active		Absolute Error		
		M	SSDR-M	SSDR-D	SSDR-M	SSDR-D	M	SSDR-M	SSDR-D
19	F_{API} (kg/h)	1.6	1.6	1.5	1.0	1.0	--	--	--
	F_{EXC} (kg/h)	13.4	13.4	13.8	1.0	1.0	--	--	--
	F_{lube} (kg/h)	0.1	0.1	0.1	1.0	1.0	--	--	--
	$x_{AP1,1}$ (%wt)	11.3	10.6	11.2	1.0	1.0	1.13	1.85	1.22
	$x_{AP1,2}$ (%wt)	9.9	10.6	10.3	1.0	1.0	2.46	1.85	2.10
21	F_{API} (kg/h)	1.6	1.6	1.2	1.0	1.0	--	--	--
	F_{EXC} (kg/h)	13.4	13.4	15.2	1.0	1.0	--	--	--
	F_{lube} (kg/h)	0.1	0.1	0.1	1.0	1.0	--	--	--
	$x_{API,1}$ (%wt)	0.0	10.5	9.4	0.0	0.0	10.30	0.15	0.89
	$x_{API,2}$ (%wt)	10.2	10.5	8.9	1.0	1.0	0.09	0.15	1.45

As shown in Table 3.3 for experiment 19, the raw Raman composition measurement is the closest value to the offline composition measurement. SSDR-D reconciled value for the API composition, given by the Raman measurement, is the next nearest value to the actual API composition. However, both the NIR direct measurement and the SSDR-D API composition reconciled value, based on the NIR measurement, are far from the actual value. The SSDR-M API

composition, given by the Raman and NIR, have the same value and therefore the same error, which is distributed among variables. Of course, the SSDR-M results have the merits of satisfying the mass balances for the process.

However, it is important to note the NIR sampling time was increased to match the Raman sampling time. Moreover, if the Raman measurement is missed, as in experiment 21, then there might not be a reliable state estimation of the system without reconciliation.

Table 3.3 shows how SSDR-M and SSDR-D detect the gross error in the Raman measurement. The system is redundant since four measurements are still active. The NIR direct measurement is the closest to the real API composition. However, SSDR-M provides a more robust solution because it detects the error in real-time and estimates the powder blend composition. SSDR-M also complies with the mass balance of the system.

3.6.2   SSDR-M and SSDR-D for end-to-end operation

In this section, the E2 system, which includes the tablet press, is analyzed. For the evaluation of the SSDR performance in the tablet press, the at-line measurements were used for tablet weight and tensile strength. The accumulation flow (F_{ac}) and thickness are unmeasured variables that are estimated using Eq. 3.11 and Eq. 3.13, which are part of the model used in SSDR-M. Of course, they can also be used to estimate the unmeasured variables from the results of the SSDR-D reconciliation. Recall that in general estimation of these unmeasured variables is not possible in the SSDR-D framework unless these equations are used.

Table 3.4 shows the performance in the end-to-end run. The selected confidence interval was 99%, giving an initial $\chi^2 = 13.28$. In experiment 19, the error evaluations for the API composition differ from those reported in section 3.6.1 as a consequence of differences in the process model used. SSDR-D 6PC detects a gross error in the MCF, while the other alternatives (SSDR-M and SSDR-D 4PC) do not. In theory, SSDR-M and SSDR-D 6PC should give the same or very similar results if the process model is linear or mildly nonlinear. This statement holds true for the E1 system, but the E2 system is highly nonlinear because it includes the relationships associated with the tablet press operation.

Table 3.4. Performance End-to-End Operation

No	Var	NOC				Active			Absolute Error			
		M	SSDR-M	SSDR-D 6PC	SSDR-D 4PC	SSDR-M	SSDR-D 6PC	SSDR-D 4PC	M	SSDR-M	SSDR-D 6PC	SSDR-D 4PC
19	F_{API} (kg/h)	1.56	1.59	1.61	1.51	1.00	1.00	1.00	--	--	--	--
	F_{EXC} (kg/h)	13.42	13.40	12.21	13.88	1.00	1.00	1.00	--	--	--	--
	F_{lube} (kg/h)	0.08	0.07	0.08	0.06	1.00	1.00	1.00	--	--	--	--
	$x_{API,\,1}$ (%wt)	11.27	10.55	10.88	11.25	1.00	1.00	1.00	1.13	1.85	1.52	1.15
	$x_{API,2}$ (%wt)	9.94	10.55	9.97	10.33	1.00	1.00	1.00	2.46	1.85	2.43	2.07
	t_{MCF} (mm)	7.20	6.90	6.98	7.04	1.00	1.00	1.00	--	--	--	--
	MCF (kN)	8.70	7.11	8.20	6.57	1.00	0.00	1.00	--	--	--	--
	W (mg)	738.12	764.65	760.92	761.35	1.00	1.00	1.00	49.46	22.93	26.66	26.23
	T (MPa)	0.80	0.78	1.02	0.88	1.00	1.00	1.00	0.07	0.05	0.29	0.15
	t (mm)*	7.70	7.38	7.47	7.53	--	--	--	4E-03	0.32	0.23	0.17
	F_{ac} (kg/h) *	1.77	1.30	0.21	1.76	--	--	--	--	--	--	--
	a	--	0.26	--	--	--	--	--	--	--	--	--
	1/b (MPa)	--	36.60	--	--	--	--	--	--	--	--	--

SSDR-M shows the best performance for the tablet weight and tensile strength estimation. The calculated thickness, based on the thickness at main compression force and bias estimation, is the closest value to the thickness at-line measurement value, followed by the SSDR-D 4 PC thickness. The accumulation flow is estimated between 1.3 kg/h to 1.8 kg/h for all cases except SSDR-D 4PC, where the accumulated flow is approximately 0.2 kg/h. This constitutes a significant difference between SSDR-D 6 PC and SSDR-D 4 PC reconciliation. SSDR-D 6 PC result for the lactose flow is also significantly different than the other methods, underestimating the lactose flow by more than 1kg/h.

Experiment 7 is the third run used to analyze the E2 results (see Table 3.5). Here SSDR-D 6PC reconciliation values give the best estimation for the real tablet weight and tensile strength, followed by SSDR-M results. SSDR-M gives the best estimation for the unmeasured thickness. The accumulation flow is calculated between 2.89 kg/h to 2.98 kg/h for all cases except SSDR-D 4PC, where the accumulated flow is approximately 8.62 kg/h. This discrepancy again shows the impact of the selection of principal components. More information for real-time implementation can be found in Appendix B.

Table 3.5. Performance End-to-End Effects on Unmeasured Variables

No	Var	NOC				Active			Absolute Error			
		M	SSDR-M	SSDR-D 6PC	SSDR-D 4PC	SSDR-M	SSDR-D 6PC	SSDR-D 4PC	M	SSDR-M	SSDR-D 6PC	SSDR-D 4PC
7	F_{API} (kg/h)	1.58	1.70	1.48	2.06	1.00	1.00	1.00	--	--	--	--
	F_{EXC} (kg/h)	13.32	13.26	13.28	18.79	1.00	1.00	1.00	--	--	--	--
	F_{lube} (kg/h)	0.21	0.20	0.21	0.11	1.00	1.00	1.00	--	--	--	--
	$x_{API,1}$ (%wt)	12.70	11.22	12.61	12.48	1.00	1.00	1.00	--	--	--	--
	$x_{API,2}$ (%wt)	11.72	11.22	11.79	11.56	1.00	1.00	1.00	--	--	--	--
	t_{MCF} (mm)	6.50	6.52	6.55	6.67	1.00	1.00	1.00	--	--	--	--
	MCF (kN)	8.90	8.90	8.65	9.14	1.00	1.00	1.00	--	--	--	--
	W (mg)	727.52	725.22	716.69	734.45	1.00	1.00	1.00	9.00	6.70	1.83	15.93
	T (MPa)	0.88	0.88	0.95	1.15	1.00	1.00	1.00	0.045	0.042	0.028	0.23
	t (mm)*	6.96	6.98	7.01	7.14	--	--	--	0.025	3E-03	0.032	0.16
	F_{ac} (kg/h)*	2.89	2.98	2.93	8.62	--	--	--	--	--	--	--
	a	--	0.28	--	--	--	--	--	--	--	--	--
	1/b (MPa)	--	22.92	--	--	--	--	--	--	--	--	--

If the API composition is considered one variable instead of two, then the average error of the NIR and Raman measurement can be selected as an indicator of prediction performance. For experiment 19, the best performance was given by SSDR 4 PC for API composition; for tablet thickness, the thickness at MCF with bias was the best option; and SSDR-M gave the best estimate for all other variables. For experiment 7, SSDR-D 6PC yields the lowest error for tablet weight and tensile strength; however, SSDR-D 4PC gives the worst estimate, and this affects the value of unmeasured variables significantly. The estimate for the lactose flow is also significantly different from the estimates provided by the other approaches (approx. 6kg/h). In this run, SSDR-M also yields the best estimate of the tablet thickness. The SSDR-M performance may have been better if the initial parameter estimation for a and 1/b have been done for Magnesium Stearate 1% instead of 0.5% wt. However, the data required for estimating these parameters were not available.

3.6.3   Loss of redundancy

During experiment 21, the tablet press is not operating at steady-state, and as a consequence, it was not possible to reconcile the variables for the entire line. The TS was in constant change during this period. However, this run is informative as it shows how the loss of redundancy affects the estimate of the real-state of the process. The SSDR-M, SSDR-D 6PC, and SSDR-D 4PC results are shown in

Table 3.6. Here SSDR-M has six measurements active, while the minimum of redundancy is seven for this system. The gross errors were identified at the Raman measurement, the MCF, and the tablet weight. As a consequence, they were classified as unmeasured variables. SSDR-D has seven measurements active, yet, generally having more than 20% measurements missing is expected to affect the reconciliation results negatively. Moreover, a gross error is identified in the NIR measurement, whereas the sensor seems to be working satisfactorily. SSDR-D 4PC is the only one of the alternatives not to lose redundancy. However, using an inadequate number of PC's in the model was shown in section 3.6.2 to be a cause for concern in the reliability of the estimates of the unmeasured variables.

Table 3.6. Loss of redundancy

No	Var	NOC				Active			Absolute Error			
		M	SSDR-M	SSDR-D 6PC	SSDR-D 4PC	SSDR-M	SSDR-D 6PC	SSDR-D 4PC	M	SSDR-M	SSDR-D 6PC	SSDR-D 4PC
21	F_{API} (kg/h)	1.59	1.58	1.52	1.20	1.00	1.00	1.00	--	--	--	--
	F_{EXC} (kg/h)	13.43	13.43	13.77	17.99	1.00	1.00	1.00	--	--	--	--
	F_{lube} (kg/h)	0.13	0.13	0.13	0.08	1.00	1.00	1.00	--	--	--	--
	$x_{API,1}$ (%wt)	0.00	10.45	6.29	6.38	0.00	0.00	1.00	10.30	0.15	4.01	3.92
	$x_{API,2}$ (%wt)	10.21	10.45	6.30	6.52	1.00	0.00	1.00	0.09	0.15	4.00	3.78
	t_{MCF} (mm)	5.80	6.21	5.89	5.88	1.00	1.00	1.00	--	--	--	--
	MCF (kN)	18.20	30.00	19.15	19.15	0.00	1.00	1.00	--	--	--	--
	W (mg)	714.48	722.13	701.63	713.35	0.00	1.00	1.00	23.42	15.77	36.27	24.55
	T (MPa)	1.79	1.79	1.78	1.81	1.00	1.00	1.00	1.38	1.38	1.39	1.36
	t (mm)*	6.20	6.64	6.30	6.29	--	--	--	0.30	0.14	0.20	0.21
	F_{ac} (kg/h) *	0.58	0.42	1.11	4.72	--	--	--	--	--	--	--
	a	--	0.28	--	--	--	--	--	--	--	--	--
	1/b (MPa)	--	20.90	--	--	--	--	--	--	--	--	--

On the bases of prediction error, the SSDR-M approach yields the best estimate for all variables except for tensile strength, while the SSDR-D 4PC approach yields the smallest error value for the tensile strength. However, if there is no redundancy, it cannot be assured that the process model will be satisfied. For instance, the MCF reconciled value is at the upper bound of this variable (see

Table 3.6).

Table 3.7 shows the values of the constraints (Eq. 3.8-3.9 and Eq. 3.11-3.14), using the raw measurements and the reconciliation results. Please note that Eq. 3.13 is implicit (the thickness was substituted) into the other equations. For evaluating equation 3.12 and 3.14, the parameters estimated in SSDR-M were used.

Table 3.7. Constraint evaluation

Method	No. M	Eq. 3.8	Eq. 3.9	Eq. 3.11	Eq. 3.12	Eq. 3.14
Measurement	9	-1.6	-10.2	1.7	-12.5	1.4
SSDR-M	6	0.0	0.0	1.7	2.1E-14	-1.1E-02
SSDR-D 6PC	7	-0.6	-1.2E-02	1.7	-10.7	0.5
SSDR-D 4PC	9	-0.21	-0.22	-6.85	-11.31	1.01

Note that equation 3.1 measures the difference between the measured and reconciled variable values weighted by the covariance matrix and will take on the value zero when evaluated at the measurement point. However, its value at when computed at the reconciled values generated via the alternative approaches can serve as a comparison criterion.

SSDR-M no longer passed the global tests, and the number of measurements was less than the degrees of freedom. As a consequence, the solution is distorted (i.e., the MCF is inaccurate) and it does not satisfy all the constraints. The overall mass balance represented in Eq. 3.11 does not equate to zero. Equation 3.14, which represents the tensile strength of the tablet, is also outside the limits established in fmincon. In the case of the SSDR-D models, none of them comply with the constraints. SSDR-D 6 PC is also non-redundant because more than 20% of the measurements are missing. SSDR-D 4 PC has all measurement active, but it does not identify the RAMAN

measurement as a gross error, even though the measurement is missing. As a consequence, its solution is distorted. For SSDR-D 4PC, Eq. 3.11 solution is significantly different from the other methods. It is possible that this model is not accurate for identification of gross error, highlighting the importance of the PC's selection. For experiment 21, none of the methods can give an accurate solution due to the loss of steady-state.

3.7 Conclusions

Based on these case studies, the SSDR framework does give reasonable estimates of the state of the continuous tableting line state. Using the at-line and offline measurements of the pilot plant, the framework was shown to be able to deal with random and gross errors. In section 3.6.1, the focus was on E1 and how it was possible to used the offline composition value as a measure of performance evaluation. For the second E1 experiment, both the SSDR-D and SSDR-M do detect the gross error arising with the Raman composition measurement. On average, SSDR-M gives the best estimation.

In section 3.6.2, the offline and at-line measurements were used to evaluate the performance of the framework in case studies involving the end-to-end process. The at line measurement are shown to be useful for estimating the prediction error in DR and most important as a means of correcting the bias of soft sensors. For the first E2 experiment, SSDR-M provided the best estimates for the tablet weight and tensile strength, while the SSDR-D 4 PC model, on average, offered the best estimate for the composition measurement. For the second E2 experiment, SSDR-D 6 PC gave the best estimate for tablet weight and tensile.

The steady-state process model must be used to estimate unmeasured variables, such as the tablet thickness and the flow accumulated in the hopper. While this occurs automatically in the

SSDR-M approach, in the SSDR-D alternatives the relevant model equations must be added to the PCA model to compute the unmeasured variables.

Since real-time date for hopper accumulation is not available, it is not possible to verify the predictions of the accumulated flow in the hopper: the results show a wide range of values between 0 kg/h to 3 kg/h. Three out of four approaches provide similar estimates. The SSDR-D results vary according to the number of PC's selected: the difference between the accumulated flows estimated by SSDR-D 6 PC and 4PC is as large as 5 kg/h. This reaffirms the importance of proper selection of PC's without overfitting or underestimating the training data. As expected, the SSDR-M approach offers the most reliable method for estimating the unmeasured variables. Gross error detection is also a point of difference. SSDR-M GED results differ from SSDR-D GED results due to the effects of process model nonlinearities. As a consequence, the SSDR-GED and the SPE detect errors at different points in time. The results of Section 3.6.3 shows that if the at-line measurements are considered, SSDR-M gives the best estimates. Overall, the SSDR-M approach seems to be the most robust: it effectively distributes the measurement error and yields reliable estimates of the unmeasured variables. The SSDR-D 6PC can offer acceptably good estimates of the process state, but the selection of PC's must be made carefully, and the PC-based model must be complemented with equations for estimating unmeasured variables. Moreover, the loss of redundancy can distort the final solution for other variables.

Improvements could be made to the continuous tableting line sensor network to increase redundancy and prediction reliability. For instance, the installation of a quantitative level sensor would be useful to improve the estimate of the accumulation in the tablet press hopper. The use of an ultrasound sensor to improve the real-time estimate of the tablet tensile strength would likewise be desirable. The use of dynamic models would permit extension of the framework to dynamic

data reconciliation under which the constraint of requiring the measurements to be at steady-state would not be necessary. Of course, dynamic data reconciliation would come at the cost of increased computational complexity, longer computation times and thus less frequent reconciliation.

3.8   References

[1]   M. Ierapetritou, F. Muzzio, and G. V. Reklaitis, "Perspectives on the Continuous Manufacturing of Powder-Based Pharmaceutical Processes," *AIChE J.*, vol. 62, no. 6, pp. 1846–1862, 2016.

[2]   U. S. D. of H. and H. S. FDA, "Guidance for Industry PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance," no. September, p. 16, 2004.

[3]   S. Narasimhan and C. Jordache, *Data reconciliation and Gross Error detection*. Houston, TX.: Gulf Publishing Company, 2000.

[4]   J. A. Romagnoli and S. M.C., *Data Processing and Reconciliation for Chemical Process Operations*. San Diego, CA, 2000.

[5]   C. Knopf, *Introduction to Data Reconciliation and Gross Error Detection, in Modeling, Analysis and Optimization of Process and Energy Systems*. Hoboken, NJ, USA: John Wiley & Sons, 2011.

[6]   I.-W. Kim, M. S. Kang, S. Park, and T. F. Edgar, "Robust data reconciliation and gross error detection: The modified MIMT using NLP," *Comput. Chem. Eng.*, vol. 21, no. 7, pp. 775–782, Mar. 1997.

[7]   Q. Su, M. Moreno, A. Giridhar, G. V Reklaitis, and Z. K. Nagy, "A Systematic Framework for Process Control Design and Risk Analysis in Continuous Pharmaceutical Solid-Dosage Manufacturing," *J. Pharm. Innov.*, Oct. 2017.

[8]   S. L. Lee, T. F. O'Connor, X. Yang, C. N. Cruz, S. Chatterjee, R. D. Madurawe, C. M. V Moore, L. X. Yu, J. Woodcock, T. F. O'Connor, X. Yang, C. N. Cruz, S. Chatterjee, R. D. Madurawe, C. M. V Moore, L. X. Yu, and J. Woodcock, "Modernizing Pharmaceutical

Manufacturing: from Batch to Continuous Production," *J. Pharm. Innov.*, vol. 10, no. 3, pp. 191–199, Sep. 2015.

[9]   M. Moreno, J. Liu, Q. Su, C. Leach, A. Giridhar, N. Yazdanpanah, T. O'Connor, Z. K. Nagy, and G. V. Reklaitis, "Steady-state Data Reconciliation Framework for a Direct Continuous Tableting Line," *J. Pharm. Innov.*, 2018.

[10]   C. Benqlilou, "Data Reconciliation as a Framework for Chemical Processes Optimization and Control," pp. 1–211, 2004.

[11]   G. A. C. Le Roux, B. F. Santoro, F. F. Sotelo, M. Teissier, and X. Joulia, "Improving steady-state identification," *Comput. Aided Chem. Eng.*, vol. 25, pp. 459–464, 2008.

[12]   J. Liu, Q. Su, M. Moreno, C. Laird, Z. Nagy, and G. Reklaitis, "Robust state estimation of feeding–blending systems in continuous pharmaceutical manufacturing," *Chem. Eng. Res. Des.*, vol. 134, pp. 140–153, Jun. 2018.

[13]   S. Narasimhan and N. Bhatt, "Deconstructing principal component analysis using a data reconciliation perspective," *Comput. Chem. Eng.*, vol. 77, pp. 74–84, 2015.

[14]   M. Moreno, A. Giridhar, G. V. Reklaitis, and Z. K. Nagy, "Nonlinear Steady-State Data Reconciliation for a Continuous Tableting Line," in *AIChE Annual Meeting*, 2015.

[15]   P. R. C. Nelson, P. A. Taylor, and J. F. MacGregor, "Missing data methods in PCA and PLS: Score calculations with incomplete observations," *Chemom. Intell. Lab. Syst.*, vol. 35, no. 1, pp. 45–65, Nov. 1996.

[16]   S. M. Razavi, G. Callegari, G. Drazer, and A. M. Cuitiño, "Toward predicting tensile strength of pharmaceutical tablets by ultrasound measurement in continuous manufacturing," *Int. J. Pharm.*, vol. 507, no. 1–2, pp. 83–89, Jun. 2016.

[17]   S. Panikar, J. Li, V. Rane, S. Gillam, G. Callegari, B. Kurtyka, S. Lee, and F. Muzzio, "Integrating Sensors for Monitoring Blend Content in a Pharmaceutical Continuous Manufacturing Plant," *J. Pharm. Innov.*, 2018.

[18]   S. García-Muñoz, "Phi MATLAB Toolbox." Personal Communication, 2015.

[19]   M. J. Bagajewicz and Q. Jiang, "Comparison of Steady and Integral Dynamic Data Reconciliation," *Comput. Chem. Eng.*, vol. 24, pp. 2367–2383, 2000.

[20]   M. J. Bagajewicz, *Smart Process Plants: Software and Hardware Solutions for Accurate Data and Profitable Operations: Data Reconciliation, Gross Error Detection, and Instrumentation Upgrade*, 1st ed. McGraw-Hill Education, 2010.

[21] Q. Su, Y. Bommireddy, M. Gonzalez, G. V. Reklaitis, and Z. K. Nagy, "Variation and Risk Analysis in Tablet Press Control for Continuous Manufacturing of Solid Dosage via Direct Compaction," in *Proceedings of the 13th International Symposium on Process Systems Engineering PSE 2018*, M. Eden, G. Towler, and M. Ierapetritou, Eds. San Diego, CA: Elsevier, 2018, p. 2912.

[22] R. Singh, K. V. Gernaey, and R. Gani, "ICAS-PAT: A software for design, analysis and validation of PAT systems," *Comput. Chem. Eng.*, vol. 34, no. 7, pp. 1108–1136, 2010.

[23] V. Mazel, V. Busignies, S. Duca, B. Leclerc, and P. Tchoreloff, "Original predictive approach to the compressibility of pharmaceutical powder mixtures based on the Kawakita equation," *Int. J. Pharm.*, vol. 410, no. 1, pp. 92–98, 2011.

[24] M. Kuentz and H. Leuenberger, "A new model for the hardness of a compacted particle system, applied to tablets of pharmaceutical polymers," *Powder Technol.*, vol. 111, no. 1–2, pp. 145–153, Aug. 2000.

[25] S. A. Imtiaz, S. L. Shah, and S. Narasimhan, "Missing Data Treatment Using Iterative PCA and Data Reconciliation," *IFAC Proc. Vol.*, vol. 37, no. 9, pp. 197–202, Jul. 2004

CHAPTER 4. NONLINEAR STEADY-STATE DATA RECONCILIATION FOR CONTINUOUS TABLETING PROCESSES

4.1   Introduction

In 2002, the concepts of quality-by-design (QbD) and Process Analytical Technology (PAT) were introduced by the FDA as important milestones in the path towards improved pharmaceutical manufacturing. The goal of the QbD initiative was to establish the importance of identifying the variables of the process connected to product quality, called critical quality attributes (CQAs) and critical quality parameters (CPP's) and of defining the domain of values of these variables within which the product was acceptable for release. The goal of the PAT initiative was to advance the use of PAT tools for monitoring the CQA's and CPP's online. By doing so, it was possible to both improve the understanding and enable the automatic control of the manufacturing process [1] [2]. With QbD and PAT realized, the next development towards advancing manufacturing in the pharmaceutical industry is real-time process management, which spans real-time process monitoring, reconciliation of process measurements, detection and mitigation of gross errors and process control. In this work, the focus is on data reconciliation and gross error detection. Although these methodologies have been employed for decades in other industries [3], their use in pharmaceutical manufacturing, particularly continuous manufacturing, has lagged.

Data reconciliation serves to provide estimates of the most likely state of the manufacturing system given known random errors in the process measurements and can be implemented in data-driven [4], or model-based [5] forms. In this paper, model-based steady-state data reconciliation (SSDR-M) was adopted and its use was studied in two configurations of continuous tableting processes: direct compression and dry granulation. Since the basic design and implementation

details of a data reconciliation framework have been presented in [6], the objective of this work is to investigate the application of SSDR to these nonlinear processes and demonstrate how estimates of the state of the process can be improved by effective management of process errors. The process model-based form of SSDR was selected because it can execute in short computational times, while also accommodating process nonlinearities [7].

SSDR-M is usually formulated as a form of the maximum likelihood optimization problem. The objective function is defined to minimize the difference between the raw measurements ($\mathbf{x}^+$) and their reconciled values ($\mathbf{x}$), weighted by the covariance matrix ($\mathbf{Q}$) of the measurements [8]. Equation 1.1 represents the objective function.

$$\min_{\mathbf{x},\mathbf{y}} J = (\mathbf{x}^+ - \mathbf{x})^{\mathrm{T}} \mathbf{Q}^{-1} (\mathbf{x}^+ - \mathbf{x}) \tag{4.1}$$

The $\mathbf{Q}$ matrix is estimated from a priori information on the measurement errors inherent to the measurement technologies employed. The problem variables are constrained to satisfy the process model, which consists of fundamental relationships such as material balance equations and mechanistic relationships (Eq. 1.2), as well as possibly inequality constraints (Eq. 1.3), such as bounds on the variables [9].

$$\text{s.t. } \mathbf{h}\left(\mathbf{x},\mathbf{y},q\right) = 0 \tag{4.2}$$

$$\mathbf{g}\left(\mathbf{x},\mathbf{y},q\right) \pounds 0 \tag{4.3}$$

Where: $\mathbf{x}^+ \in \Re^n$ is a vector of measurements, $\mathbf{x} \in \Re^n$ is a vector of reconciled values of the n measurements, $\mathbf{y} \in \Re^m$ is a vector of them unmeasured process variables, $\mathbf{Q}$ is the covariance matrix, $\Re^{n \times n}$, $\boldsymbol{\theta} \in \Re^p$ is a vector of p process model parameters, $\mathbf{h} \in \Re^k$ is a set of equations which describe the steady-state behavior of the process, and $\mathbf{g} \in \Re^q$ is a set of inequality

constraints, *n* is the number of reconciled variables, *m* the number of unmeasured variables and *k* the number of relations or equations.

In general, the equations of the process model are functions of the measured variables, the unmeasured variables, and model parameters [10]. In [11] and [12], the authors established that a system has to have redundant measurements for data reconciliation to be applicable. The system is considered redundant if there are more measurements than degrees of freedom. As long as this condition is met, the SSDR-M results are considered the best estimates of the actual values of the process variables. If the condition is not met, then the solution is distorted and the measurements are the only information available to estimate the process state.

SSDR, as its name implies, can only be applied at steady-state. In SSDR-M, the process variables are in two main categories: measured and unmeasured [12]. Moreover, the measurements can be classified as redundant or non-redundant. A redundant measurement can be corrected through the model, while a non-redundant measurement can only be accepted in its raw form. The unmeasured variables can be observable or non-observable [13]. If they are observable, they can be estimated through the SSDR-M process model. Non-observable variables are neither measured nor computable from the process model.

Before reconciling process measurements, gross error detection (GED) has to be applied. If the reconciliation is executed with gross errors present, the solution will be distorted. A number of error detection methods are available in the literature [12] [14]. In this work, the established methods of global test (GT) and measurement test [6] are used. The former test is solved through nonlinear programming (NLP). However, its statistical limit, the χ^2 distribution, is a function of the system number of linear-independent equations (I) and the selected confidence interval (e.g., 90% to 95%) [11] [15]. If the optimal value of Eq. 1 is greater than χ^2, then there is likely a gross

error in the system. To locate the gross error, the measurement test is used, which is based on the

critical test value (z_c) [6]. Equation 4.4 represents the model constraints in linear form.

$$\mathbf{Ax} - \mathbf{c} = 0 \tag{4.4}$$

Where: $A \in \mathfrak{R}^{kxn}$ is the linear constraint matrix, $x \in \mathfrak{R}^n$ is a vector of reconciled values,

$c \in \mathfrak{R}^n$ is a vector of known coefficients.

This transformation can be achieved through Taylor series expansion, and is needed for the

measurement test and the evaluation of I. If the measurement test fails, most likely the

measurement which corresponds to the failed test has a gross error, and thus it has to be rejected

from the reconciliation process and treated as an unmeasured variable. Equation 4.5 and Equation

4.6 establish how to estimate the measurement test.

$$\mathbf{a}_i = (\mathbf{x}_i^+ - \mathbf{x}_i) \tag{4.5}$$

$$\mathbf{mt}_i = \frac{\left|\mathbf{a}_i\right|}{\text{cov}\left(\mathbf{a}\right)} \tag{4.6}$$

Where: $\mathbf{x}^+ \in \mathfrak{R}^n$ is a vector of measurements, $\mathbf{x} \in \mathfrak{R}^n$ is a vector of reconciled values, $\mathbf{a} \in \mathfrak{R}^n$ is

a vector of errors, $\mathbf{Q}$ is the covariance matrix, $\mathfrak{R}^{n \times n}$, $\mathbf{A} \in \mathfrak{R}^{kxn}$ is the matrix of coefficients of the

linear/linearized process model, and $\mathbf{mt} \in \mathfrak{R}^n$ is a vector of the measurement test values;

The level of significance is defined as α, while β is the modified level of significance and

it is dependent on the number of active measurements (see Equation 4.7).

$$b = 1 - \left(1 - a\right)^{1/n} \tag{4.7}$$

Where: n is the number of measurements, α is the level of significance (e.g., 5%), and β is the

modified level of significance since it is a function of α and the number of measurements.

4.2   Experimental Systems

4.2.1   Tableting line via direct compression (DC)

The DC tableting line comprises two loss-in-weight (LIW) Schenck AccuRate AP-300 feeders. One feeder transports API at a mass flow rate (F_{API}) of 1 kg/h, while the second feeder conveys an excipient at mass flow rate (F_{exc}) of 9 kg/h in normal operating conditions. The flow rates are measured using load cells built into the feeders. The API and excipient are blended in a Gericke GCM 250 mixer, operated at 200 rpm. The content uniformity of this stream is measured in real-time using a Near Infrared (NIR) spectroscopy method. Subsequently, a Purefeed DP-4 Disc feeder adds a glidant at a mass flow rate (F_{gli}) of 20 g/h into the system. The final mixture is blended in a second Gericke GCM 250 mixer, which is also operated at 200 rpm.   The NIR spectroscopy method is used at the exit of the second blender to measure the content uniformity in real-time [16]. The total flow rate of this stream was measured in real-time by means of an X-ray sensor [17]. More details on the PAT tool development are provided in Section 4.3. Once the flow rate is measured by the X-ray sensor, the blended powder is transported to the tablet press.

A "Z" incline conveyor is used to transport the material from the last real-time PAT tool to a Natoli BLP-16 Tablet Press. Round and flat tablets with a diameter of 8 mm are produced in the 16-station tablet press using B-type tooling. The in-die tablet thickness was limited to 4.2 mm and was set manually at the beginning of the experimental campaign. The dosing position (DS) was set at 15 mm, feed frame rotation speed (FS) set-point was 50 rpm, and the turret speed (TS) was set at 36 rpm. Other measurements provided by the tablet press are the main compression force (MCF), pre-compression force (PCF), the ejection force (EJ), temperature, and humidity. It is important to note that the press is a research unit, and thickness at MCF and PCF cannot be adjusted during operation.

4.2.2 Tableting line via dry granulation (DG)

In the dry granulation line, the API and excipient feeders and blender used in the DC tableting line are operated at the same conditions. The mixture content uniformity is measured by a NIR probe, and the blend is fed into an Alexanderwerk WP120x40 roller compactor. In the compactor, the powder blend is conveyed by the feed screw to two counter-rotating rolls. The feed screw speed (N_S) was set at 32 rpm. The rolls of this equipment have a diameter (D_R) of 12 cm, and their width (W) is of 4 cm and were operated at roll speed (N_R) of 6.5 rpm. The powder is gripped by the rolls and pulled into the nip area of the compacting rolls using the vacuum deaeration system before being compacted into a ribbon. The hydraulic pressure (P_h) was set at 45 bar, and the roll gap (S) control was disabled. The roller compacted ribbon was presented to the NIR probe to measure its density. A horizontal conveyor was used to feed the ribbon into the granulator. The granulator has two hammer mills in series and is operated at 50 rpm. The hammer mill of the roller compactor has two screens. An upper screen size of 1.25 mm, while the lower screen size of 1.0 mm is used.

No glidants or lubricants were added to the granules. The granule flow rate was measured by the X-ray sensor. Depending on the case scenario (refer section 4.4.2), the granules flow was transported, by the "Z" incline conveyor, into the Natoli BLP-16 Tablet Press. For the DG runs, flat tablets were also produced, with an 8mm and a 4.5 mm thickness. For the DG tableting line, the DS, FS, and TS were not varied; rather they were set at 10.7 mm, 29.7 rpm, and 41.5 rpm, respectively. The MCF, PCF, and EJ were measured but not controlled.
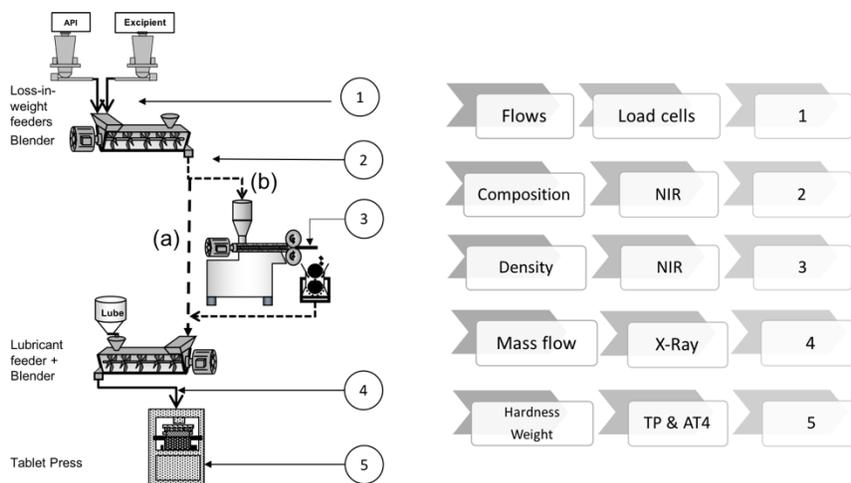
Figure 4.1. DC and DG continuous tableting line

4.3 Material and Methods

4.3.1 Materials

The active pharmaceutical ingredient (API), used for both tableting lines, was acetaminophen (APAP) (Mallinckrodt Inc., Raleigh, NC). Microcrystalline Cellulose (FMC BioPolymer Corporation, Philadelphia, PA) grades PH 200 and PH 102 (MCC 102 and MCC 200, respectively), were used as the excipient. The grade of MCC used depended on the tableting line configuration; MCC200 was used for direct compression (DC), while MCC102 was used in dry granulation (DG) experiments. Silicon dioxide (SiO_2), also known as Cab-O-Sil Untreated Fumed Silica (Cabot, Tuscola, Illinois), was employed as glidant only in the DC line, whereas in the DG line there was no lubricant or glidant used.

4.3.2 Content uniformity and ribbon density by NIR method

In the DC tableting line, the APAP composition was measured at the exit of the first blender ($x_{API,B1}$) and the exit of the second blender ($x_{API,B2}$) using CDI -256-1.7T1 NIR spectrometers, one located at each position. The CDI NIR collects eight spectra of the flowing blend stream every second with an integration time of 4 ms, using wavelengths from 904 nm to 1687 nm [18]. Both

spectrometers used Solvias Turbido NIR probes, which were positioned perpendicular to the powder flow. Two Partial-Least-Squares (PLS) models, one per position, with two principal components, were employed to relate the spectra and the APAP composition [19]. The spectra were treated with a Savitzy-Golay filter, first derivative, and SNV [16]. For $x_{API,B1}$, the calibration was done from 5% to 17% wt. APAP composition; it was established at an average flow of 10 kg/h with the rest of the mixture being MCC 200 and no lubricant or glidant added. For $x_{API,B2}$, the calibration ranged from 5% to 15% wt. APAP composition, with 0.2% SiO_2 and the rest being MCC 200; this calibration was also established at an average flow of 10 kg/h. The PLS models were developed using "Phi," a latent variable toolbox in MATLAB [20].

In the DG tableting line, one of the CDI -256-1.7T1 NIR spectrometers was used to measure $x_{API,B1}$. A PLS model with two PC's relates the spectra to the APAP composition. The calibration was done spanning a range of 5% to 15% wt. APAP composition, at an average flow of 10 kg/h with the rest of the mixture being MCC 102 and no lubricant or glidant. There was no APAP composition measurement used at the exit of the second blender. Instead, the second CDI -256-1.7T1 NIR spectrometer was used to measure the roller compacted ribbon density (ρ_{RC}) [21]. The physical arrangement of the roller compactor was modified to accommodate the NIR sensor. To ensure robust sampling from the NIR sensor, a 3D printed part was designed as a sensor holder to maintain the position and angle of the NIR sensor. A PLS model with two PC's was fit to relate the raw NIR spectra to the ribbon density in the range of 0.9 to 1.15 g/cc with a root mean squared error of 0.026 g/cc.

4.3.3   X-ray and balance flow rate measurements

In the DC tableting line studies, the X-ray was used to measure the flow at the exit of the second blender ($F_{out,B2}$), while the Mettler Toledo Balance was used for the tablet mass flow rate

measurement ($F_{out,TP}$). With the DG tableting line studies, there are two alternative configurations were investigated (see section 4.4). In the first configuration of the DG line, the tablet press was not used, and thus the Mettler balance was placed in the same spot as the X-ray sensor ($F_{out,B2,bal}$) to measure the granule flow. In the second configuration of the DG line, the balance was located at the exit of the tablet press to measure the tablet flow ($F_{out,TP}$).

4.3.4   Sotax AT4 tablet tester

The Sotax AT4 tablet tester measures the weight ($\pm$ 0.1 mg), thickness ($\pm$ 0.01 mm), diameter ($\pm$ 0.01 mm) and hardness ($\pm$ 1 N) of the tablets produced [22]. Of these measurements, the hardness test is the only one that is destructive. The tablets are transported pneumatically from the tablet press exit to the Sotax AT4 bowl, and from there to the measurement stations. These measurements are executed at-line with ten tablets sampled in intervals of approximately three minutes. However, the tester has a self-cleaning step between samples that adds delay time. Therefore, the sampling time was set to every six minutes so that the equipment would capture the tablets automatically during the run.

4.4   SSDR-M for the Continuous Tableting Processes

4.4.1   SSDR-M for the DC line end-to-end operation

The DC continuous tableting line was studied from the feeders to the tablet press running at steady-state. It is assumed there is no material loss or accumulation in the hopper level as in [7]. Figure 4.2 shows the block diagram for the DC tableting line in an end-to-end operation. This system is the first example (E1) employed to show the SSDR-M framework in this work.

E1
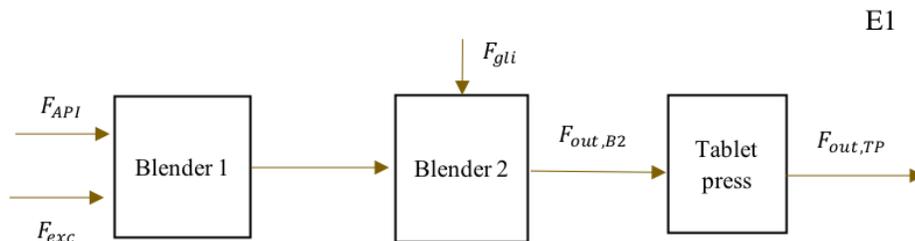
Figure 4.2. DC tableting line

In a typical data reconciliation problem, the variables that are usually measured are the flows. In this case, only some of the flows are measured, namely, the APAP flow, the excipient flow, the glidant flow, the flow at the exit of the second blender and the flow at the exit of the tablet press. For the other streams, other physical properties relating powder flow through mechanistic equations were used. The APAP composition was measured at the exits of the first and second blender. The MCF and the TS were measured in the tablet press. In total, this system employed nine real-time measurements, with a sampling time of one second each. For the flow rates and APAP compositions, a window of 10s was used as a moving average.

The APAP and excipient flows have to be equal to the flow at the exit of the first blender. Equation 4.8 represents the mass balance in the first blender, where the exit flow is unmeasured, but the API composition is measured.

$$x_{API,B1}\left(F_{API} + F_{exc}\right) - F_{API} = 0 \tag{4.8}$$

Once the glidant flow through the second blender is included this example becomes a three-component system. Equation 4.9 represents the mass balance in the second blender, where the summation of the components flows is equal to the flow at the exit of the second blender.

$$F_{API} + F_{exc} + F_{glid} - F_{out,B2} = 0 \tag{4.9}$$

Equation 4.10 considers the mass balance for the APAP component in the second blender. Equation 4.11 is used to estimate the glidant composition in the system.

$$x_{API,B2}F_{out,B2} - F_{API} = 0 \qquad (4.10)$$

$$x_{glid,B2}F_{out,B2} - F_{gli} = 0 \qquad (4.11)$$

The feeder and blender system was previously investigated in [6]. In this work, the study was expanded to include the tablet press. Equation 4.12 relates the MCF to the in-die relative density of the tablet.

$$MCF - \frac{\left(\left(\frac{4W}{\rho Dt\, r_t}\right) - r_c\right)\rho\frac{D^2}{4}}{b\left[\left(\frac{4W}{\rho Dt\, r_t}\right)(a-1) + r_c\right]} = 0 \qquad (4.12)$$

This equation, which is derived from the Kawakita model, represents the behavior of powder under compaction pressure [23]. Here parameter a represents the relative volume decrease, $1/b$ represents the pressure that is needed to get a volume decrease of $a/2$, and ρ_c is the critical relative density, and t is the tablet thickness. These parameters were estimated from prior design experiments, reported in [22], and are given by $a = 0.7911$, $1/b = 14.15$ MPa, and $\rho_c = 0.2499$. The model has a coefficient of determination (R^2) of 99.2%.

The flow at the exit of the second blender is equal to the tablet press inlet flow. The mass balance in the tablet press is represented in Equation 4.13.

$$F_{out,B2} - F_{out,TP} = 0 \qquad (4.13)$$

Equation 4.14 indicates that the tablet flow should be equal to the number of tablets ($n_{\frac{tab}{rev}}$) multiplied by turret speed and the tablet weight. The tablet weight is an unmeasured variable that

is observable through the SSDR-M model. The tablet hardness is also an observable variable in the system, through the use of equation 4.15 [24] [25].

$$F_{out,TP} - W \cdot TS \cdot \frac{n_{tab}}{rev} = 0 \tag{4.14}$$

$$H - \frac{\rho D \bar{t} s_{max}}{2} \left[1 - \left(\frac{1 - \left(\frac{4W}{\rho D t r_t} \right)}{1 - r_{cS}} \right) e^{\left(\frac{4W}{\rho D t r_t} - r_{cs} \right)} \right] = 0 \tag{4.15}$$

The parameter estimation for equation 4.15 was also reported in [22]. The tensile strength at zero porosity (σ_{max}) is 9.976 MPa, the powder true density (ρ_t) was 1.52 g/cc, the critical relative density at which the compact begins to exhibit strength ($\rho_{c\sigma}$) is 0.5875. The model R^2 is 98.9%. These parameters were estimated at the given composition of 10% APAP, 89.8% MCC 200 and 0.2% SiO$_2$. For $\bar{t}$ we measured a couple of tablets with the AT4 and took the average value as the parameter, since it had little or no variation (less than 0.6% RSD).

For this system, the initial number of independent equations is equal to five, and the initial number of measurements is nine. Table 4.1 shows the measurement set-point or normal operating conditions (NOC) employed in the DC tableting line. This table also shows the covariance used for each reconciled measurement and a summary of the variables classification from the SSDR-M perspective. The values were determined from a priori information of the DC tableting line NOC. Once the relative standard deviation (RSD) of each variable is evaluated at the selected NOC, the values were used for the covariance matrix. The confidence interval used was of 95%.

Table 4.1. NOC and covariance of the DC tableting line

Variable	Equipment	Set-Point	Q (%)	Type of variable
F_{API} (kg/h)	Load cells	1	1.6	Measured
F_{exc} (kg/h)	Load cells	9	0.3	Measured
$x_{API,B1}$ (% wt)	CDI NIR	10	7.1	Measured[1]
F_{gli} (g/h)	Load cells	20	15.0	Measured
$x_{API,B2}$ (% wt.)	CDI NIR	9.9	7.1	Measured[1]
x_{gli} (% wt.)	--	0.2	--	Unmeasured[1]
$F_{out,B2}$ (kg/h)	Load cells	10.02	7.0	Measured[1]
TS (rpm)	Tablet Press	36	1.4	Measured
MCF (kN)	Tablet Press	7.8	5.0	Measured[1]
$F_{out,TP}$ (kg/h)	Balance	10.02	7.0	Measured[1]
W (mg)	Sotax AT4	293	--	Unmeasured[1,2]
H (N)	Sotax AT4	370	--	Unmeasured[1,2]

Notes:

1) These variables do not have a set-point; therefore, the table shows NOC.

2) There are at-line measurements available for these variables

4.4.2 SSDR-M for the DG Line

4.4.2.1 SSDR-M from the feeders to the roller compactor

For the DG tableting line (Figure 4.3), the powder flow is not transferred from the exit of

the first blender, directly into the second blender as was the case with the DC tableting line. Instead,

the powder flow at the exit of the first blender feeds into the roller compactor, where the blend is

compressed by the rolls into a compacted ribbon. Equation 4.8 is the mass balance for the first

blender. Equation 4.16 assumes that all the powder, entering the roller compactor, is compacted to

a given density (ρ_{RC}). Therefore, at the point of compaction, the ribbon thickness is assumed to be

equal to the roll gap, and the ribbon width is the width of the rolls. The ribbon flow can be estimated

from its volume, density and roll speed.

$$F_{API} + F_{exc} - \rho r_{RC} N_R W D_R S = 0 \qquad (4.16)$$
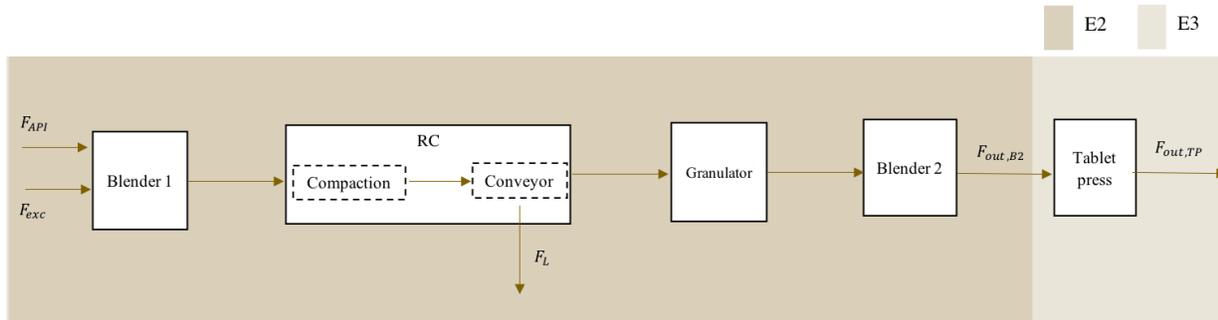
Figure 4.3. DG tableting line

A mechanistic model, based on Johanson's rolling theory [26], is used to represent the roller compaction process [16] [21]. Equations 4.17 through 4.19 relates the ribbon density to the hydraulic pressure (P_H), material properties and equipment geometry.

$$r_{RC} - r_0 P_0^{\frac{1}{K}} = 0 \tag{4.17}$$

and

$$1/2\, P_0 W D_R F = P_H A \tag{4.18}$$

where

$$F = \int_{q=0}^{q=a} \left[\frac{S/D}{\left(1 + S/D - \cos q\right)\cos q} \right]^K \cos q\, dq \tag{4.19}$$

Where: A is the compact surface area parameter, and K is an experimental constant related to the material compressibility.

After compaction, the CDI-NIR sensor is used to measure the density of the compacted ribbon. A breaker cuts the ribbon into smaller flakes, which are transported through a horizontal conveyor to feed the granulator. Due to this modified setup, there is some loss ribbon flakes and this loss (F_L) is treated as an unmeasured but observable variable. In the granulator itself, the loss of material is negligible, and granules are fed directly to the second blender. Equation 4.20 represents the mass balance in roller compactor to blender 2, considering the material loss.

$$\rho r_R N_R W D_R S - F_{out,B2} - F_L = 0 \tag{4.20}$$

In the second example (E2) the material flow goes from the feeders to blender 2 of the DG tableting line. The flow rate at the exit of the second blender is measured using the X-ray sensor, and thus a balance equation which makes the ($F_{out,B2,Bal}$) measurement redundant is added to the process model. Equation 4.21 states that the flow measurement from the X-ray should be equal to the flow measured by the balance. If this balance was not added to the model, the flow at the exit of the second blender would have been considered a non-redundant measurement. Initially, E2 has seven measurements and four DOR.

$$F_{out,B2} - F_{out,B2,Bal} = 0 \qquad (4.21)$$

Table 4.2 shows the set-point at NOC, and the covariance used for the reconciled variables. The covariance used for the measurements was determined from a priori information reported in the appendix. In the case of the roll gap, it was based on the variance of the ribbon thickness. The roll diameter, the roll width, and the roll pressure are considered parameters in this process, due to their low variability in the process.

4.4.2.2   SSDR-M for the DG tableting line, End-to-End

The third example (E3) involves the end-to-end DG tableting line. Equations 4.13 to 4.15 were used to represent the tablet press process. However, the parameters had to be re-estimated because the physical properties of the granule stream are not the same as the physical properties of the powder flow used in DC. Monte Carlo sampling and MATLAB are used to estimate the Kawakita parameters as $a = 1.0$, $b = 0.00205$ MPa^{-1}, and $\rho_c = 0.6452$. The model has a coefficient of determination (R^2) of 88.24%. The σ_{max} is 5.44 MPa, and the $\rho_{c\sigma}$ is 0.6743. The model R^2 is 81.01%. These parameters were estimated at the given composition of 10% APAP and 90% MCC 102.

For E3, four scenarios were studied for SSDR-M and compared the results to the measurement performance. The first scenario is the current SSDR-M solution. The second scenario considers the correction of the APAP composition measurement at the exit of the first blender. This scenario is possible if the raw spectra of the CDI-NIR are saved, and the PLS model is improved. The third scenario is with an improvement in the X-ray flow. A correction for this measurement, after the runs are executed, is not possible. However, the NOC of E2 were used to have an estimate on the flow correction to see its effects in the SSDR-M performance. The last scenario covers all measurements corrected for the SDDR-M solution. The initial number of measurements for this system is nine and six independent equations.

Table 4.2 has the set-point (SP) or NOC, and the covariance used for the reconciled variables in E3. The confidence interval used was of 95%.

Table 4.2. NOC and covariance of the DC tableting line

Variable	Equipment	SP	Q (%) E2	Q (%) E3	Type of variable E2	Type of variable E3
F_{API} (kg/h)	Load cells	1	2.0	1.60	Measured	Measured
F_{exc} (kg/h)	Load cells	9	0.3	0.30	Measured	Measured
$X_{API,B1}$ (% wt)	CDI NIR	10	7.0	7.10	Measured[1]	Measured[1]
S (mm)	RC	1.8	10.5	10.41	Measured[1]	Measured[1]
ρ_R (g/cc)	CDI NIR	1.06	2.0	1.73	Measured[1]	Measured[1]
$F_{out,B2}$ (kg/h)	Load cells	9	8.0	7.50	Measured[1]	Measured[1]
$F_{out,B2,bal}$ (kg/h)	Load cells	9	8.0	--	Measured[1]	NA
TS (rpm)	Tablet Press	29.7	--	1.40	NA	Measured
MCF (kN)	Tablet Press	11	--	5.00	NA	Measured[1]
$F_{out,TP}$ (kg/h)	Balance	9	--	5.00	NA	Measured[1]
W (mg)	Sotax AT4	318	--	3	NA	Unmeasured[1,2]
H (N)	Sotax AT4	230	--	8	NA	Unmeasured[1,2]

Notes:
1) These variables do not have a set-point; therefore, the table shows NOC.
2) There are at-line measurements available for these variables

4.5   Results and Discussion

4.5.1   SSDR-M for the DC line end-to-end operation

The SSDR-M results for E1 are reflected in Figure 4.4. In Figure 4.4a, the left-hand side plots show the measurements and reconciled values for the APAP flow, excipient flow and APAP composition at the exit of the first blender. The initial SSDR-M result is the reconciled solution obtained when all measurements are used. The final SSDR-M solution is the result obtained when the measurements that exhibited a gross error were deleted and the remaining measurements reconciled. The measurements in gross error were treated as unmeasured and estimated using the process model evaluated at the reconciled values. The right-hand plots indicate if the reconciled values passed or failed the measurement test. The value of one indicated that the measurement had no gross error and was considered in the reconciliation; while, if the value was zero, then that measurement had a gross error, was excluded from reconciliation and computed as an unmeasured variable using the model. In the trajectories shown in Figure 4.4a, no gross error was detected.

In Figure 4.4b the left-hand side plots show the measurements and reconciled values of the glidant flow, the APAP composition and flow at the exit of the second blender. In the Figure 4.4b right-hand side plots, the measurement tests results are shown. The APAP composition at the exit of the second blender had an intermittent gross error detected around 600s to 700s. The flow measurement at the exit of the second blender had a small bias, so a gross error was identified in that measurement across the run.
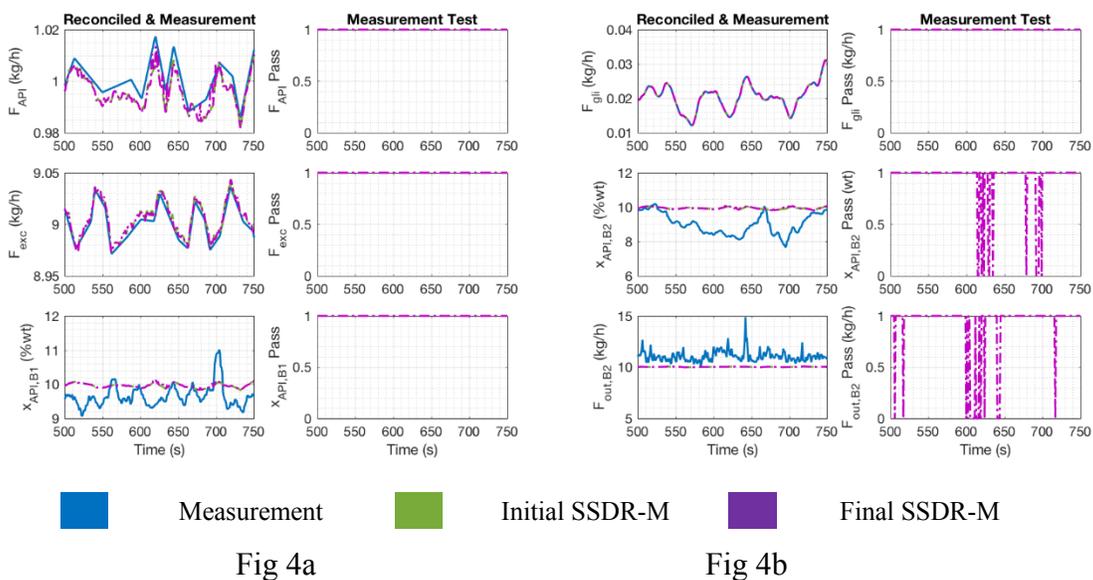
Figure 4.4. E1 DC reconciled and measurement values, no accumulation in the hopper: a) Feeder1, Feeder 2 to Blender1 b) Glidant Feeder to Blender 2

Figure 4.5 shows the measured and unmeasured variable associated with the tablet press. The left-hand plots in Figure 4.5a show the measurements and reconciled values for the turret speed, the main compression force, and the tablet flow. The right-hand portion of Figure 4.5a displays the measurement test status as a binary variable, where one meant the measurement did not have a gross error and zero meant that it did. The gross error was detected in the tablet flow, due to the disturbance caused by the manual replacement of the container used to collect tablets on Mettler Toledo balance.

Figure 4.5b shows the time series of the unmeasured variables over the run. For the tablet weight and hardness, there are at-line measurements available, and the sample average was used as the real value. To demonstrate the value of the SSDR-M model, the tablet flow computed from the model was used to estimate the weight and hardness at the same point in time at which the at-line measurements were recorded on the Sotax. The at-line measurements are highlighted in the red dot, filled with yellow. The weight and hardness equations were used to evaluate these

variables using the balance flow and marked it with the blue dot. The average error of estimation (AEE) was used as the performance indicator, which is a dimensional quantity[8]. The AEE between the measurement and the actual value is 0.0315 for the tablet weight and 0.182 for the tablet hardness. On the other hand, the AEE between the reconciled value and the actual value is 0.0242 for tablet weight and 0.0282 for tablet hardness. Showing that the AEE by using the SSDR-M.

Fig 5a   Fig 5b
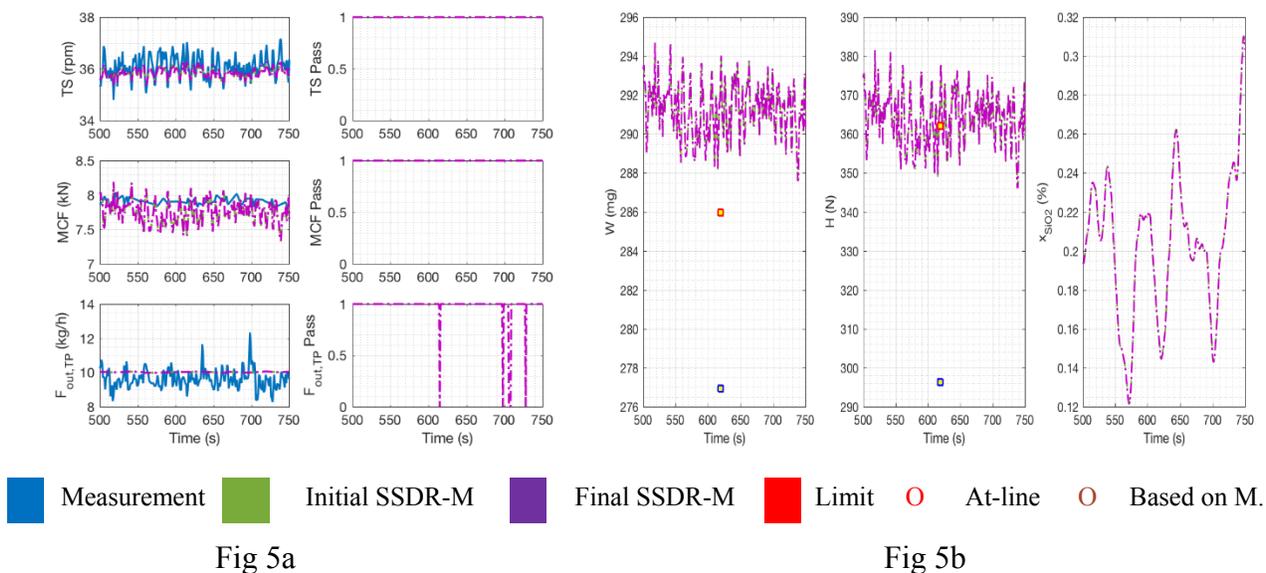
Figure 4.5. E1 DC data reconciliation results, no accumulation in the hopper: a) Tablet Press b) Unmeasured variables

The global test results for the DC tableting line over the time interval of the run are shown Figure 4.6. It is important to note that the selected limit for the global test, χ^2, is a function of the number of linear-independent equations (*I*). The number of linearly independent equations depends on the number of measured and unmeasured variables. If gross errors are detected, then depending on the measurement test results, variables that initially were considered measured variables and failed the measurement test would be classified as unmeasured. As a consequence, redundancy would be reduced, and the test statistic χ^2 changed, making the limit for the global

test stricter. The left-hand side shows the global test value with its limit (red line). The initial solution for SSDR-M is in green, and it is the result of the Equation 4.1 value when the reconciliation is based in all measurements, regardless if there are gross errors present. The purple line is the objective function value, once the gross errors are detected and treated as unmeasured variables. The right-hand side shows the binary variable result based on the final global test value (purple line) where one means pass and zero loss of redundancy. During this interval, the initial global test was violated at a different moment between 600s to 650s and 670s to 700s (see Figure 4.6 left-hand side). Nevertheless, there was still enough measurements to keep reconciling, and the final global test always passes since the global test binary variable was equal to one in this period (see the purple line in the left-hand side and its binary variable in the right-hand side).
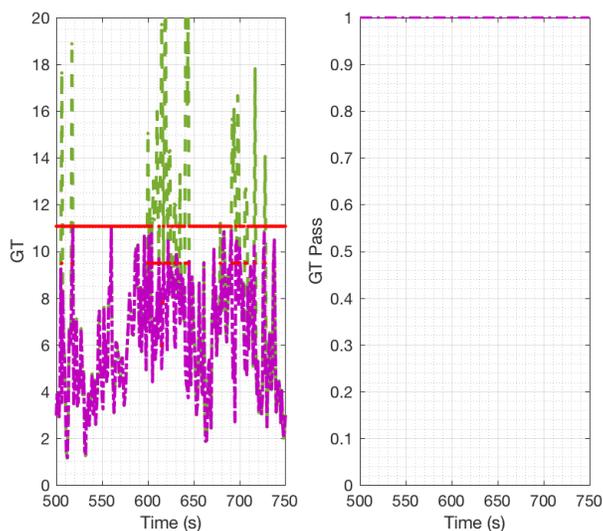
Figure 4.6. E1 DC tableting line global test

4.5.2 SSDR-M for the DG Line

4.5.2.1 E2, feeders to the exit of the second blender subsystem

Figure 4.7 shows the reconciliation results from the feeders to the roller compactor. The APAP flow, the excipient flow, and composition at the exit of the first blender are includes in Figure 4.7a. The flows do not involve measurement errors. However, the APAP composition at the exit of the first blender has a bias, and thus a measurement error is indicated across the run. However, the reconciled composition obtained when that variable is treated as unmeasured corrects that bias. Figure 4.7b shows the ribbon density and measured roll gap: these measurements have no measurement errors.

During the run, several ribbon samples were taken to compare the ribbon density and ribbon thickness to the reconciled results. It is important to note that these samples measurements were not executed at-line. Rather, the ribbon density was measured the next day, while the ribbon thickness was measured during the run with a caliper. The interval of time, in which the sample was taken, is known but not the exact time. Both offline measurements were plotted at random points (see red dot filled with yellow) through the experiment to provide a comparison of the variables actual value. It should be noted that the ribbon thickness has a high variability because it is not uniform across the width. One side of the ribbon is thicker than the other side; the ribbon has a trapezoidal shape instead of a rectangular one. Therefore, the assumption that the roll gap is equal to the ribbon thickness is only valid for the largest side of the roller compacted ribbon. As shown in this figure the roll gap reconciled value is closer to the average ribbon density of the run. The measurement test did not trigger because the standard deviation of the ribbon thickness used in the SSDR-M algorithm was sufficient. The ribbon density sample shows that the real-time measurement is accurate, although the SSDR-M result is slightly biased. However, independent at-line measurements would be desirable to make a proper comparison of AEE values.

The SSDR-M reconciled value for the roll gap is underestimated, but it is closer to the ribbon thickness average obtained with offline measurements. Since the ribbon in actuality has a trapezoidal shape, the assumption that the roll gap and ribbon thickness are equal at compaction is an approximation that needs improvement.
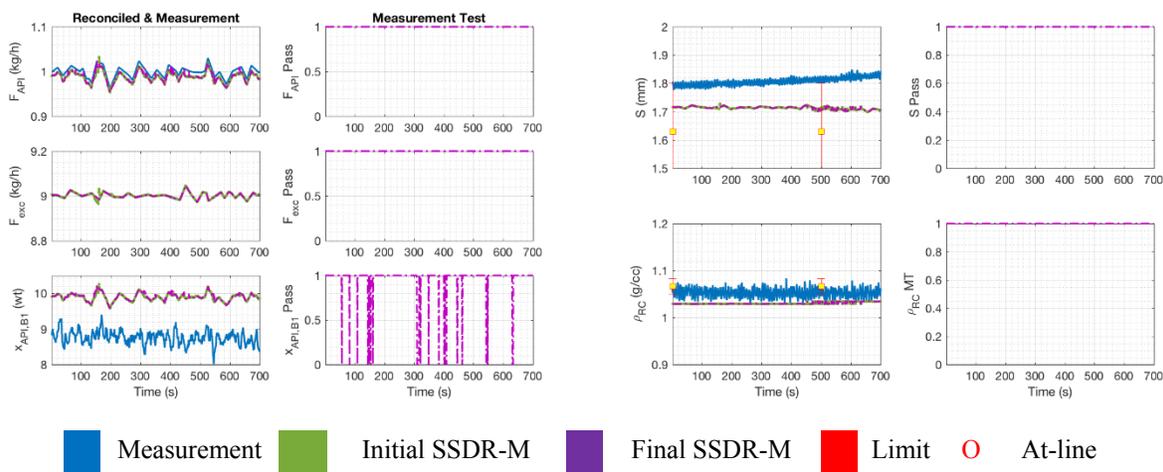
Fig 7a   Fig 7b

Figure 4.7. E2 DG reconciled and measurement values: a) Feeder1 and Feeder 2 to Blender1 b) RC

Figure 4.8a shows the measurements and reconciled values of the flow at the exit of the second blender. Note that the flow is measured both via the X-ray as well the Metter Toledo balance. After passage through the X-ray, the powder flow was collected in a container on the Mettler Toledo balance. Every time the container was changed, a disturbance was created, and a measurement error was detected. It is interesting to note that during the change of container, the X-ray sensor was also disturbed and thus for a few seconds, both sensors registered measurement errors and redundancy was lost. Figure 4.8b shows the global test results. A gross error was detected sporadically between 100s to 200s and 300s to 550s. The measurement test was used to identify the measurement with the gross error. Once the gross error was detected, that measurement was considered unmeasured. In this case, the gross error was caused by the bias in the composition

measurement at the exit of the first blender (see Figure 4.7a.) and the container change for the mass flow at the exit of the second blender (see Figure 4.8a). Figure 4.8b also shows that the global test binary variable was always one. Therefore, during this period of time, redundancy was not loss in the system.

Fig 8a   Fig 8b
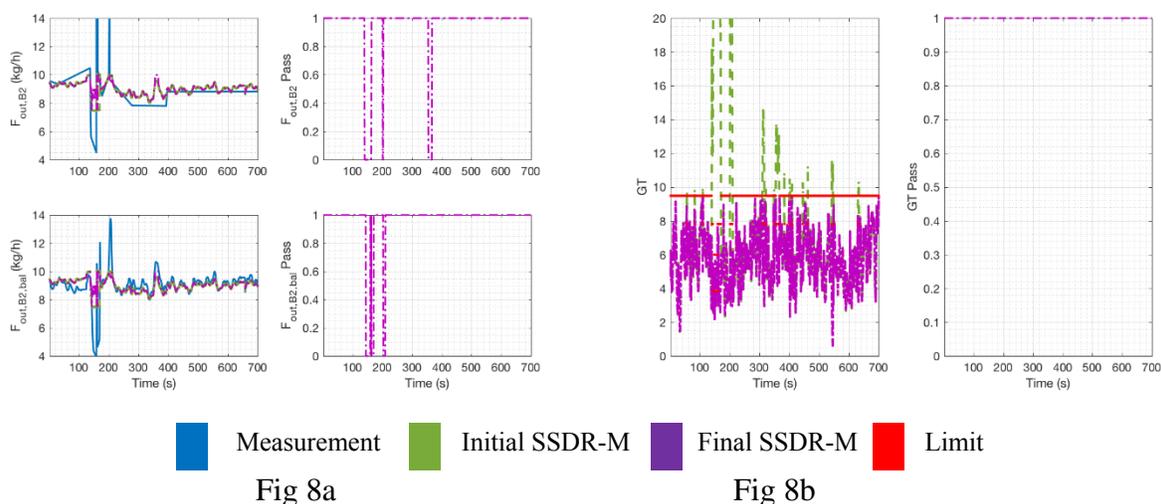
Figure 4.8. E2 DG reconciled and measurement values: a) Flow at the exit of blender 2 b) Global Test

4.5.2.2   E3, End-to-end operation

This section reports on the DG tableting line run conducted end-to-end including the tablet press. The results shown are for a period where the global test passed uninterrupted, but measurement errors were detected. In contrast to the previous run, the ribbon was not constrained from fluctuating distance from the CDI-NIR probe when the ribbon hit the breaker; as a result, there were some disturbances recorded in the ribbon density measurement. Figure 4.9 shows the plots for the six measurements and the reconciliation results from the feeders to the second blender in the DG tableting line. As shown in Figure 4.9a, a bias was detected for the APAP composition at the exit of the first blender, and thus a measurement error was detected across the run. Figure 4.9b indicates intermittent deviations in the ribbon density measurement caused by ribbon position

fluctuations, and thus a measurement error was detected at those points. The roll gap also shows a small drift over time; having a large deviation at 670s. One disadvantage of the measurement test is that it can be misidentified the location of the gross error because it is based on a linear solution. In this run, the flow rate at the exit of the second blender showed a measurement drift over time.
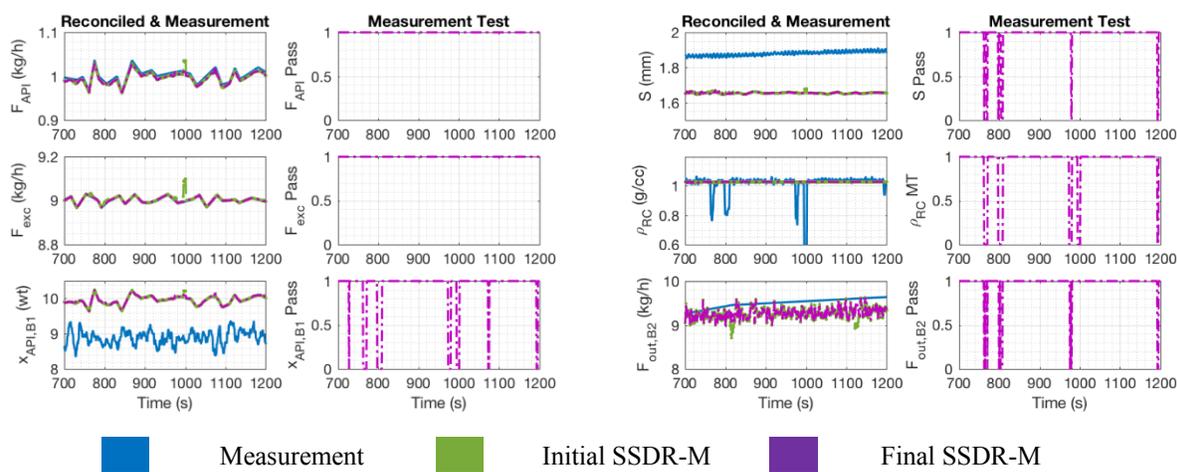
Fig 9a   Fig 9b

Figure 4.9. E3 DG reconciled and measurement values: a) Feeder 1 and Feeder 2 to Blender1 b) RC

Figure 4.10a shows the measurements and reconciled values for the three tablet press variables, turret speed, main compression force, and tablet mass flow rate. A measurement error was only detected in the tablet press flow, again caused by the disturbance to the balance caused by changing of the container or the by the removal of tablets separated for use by the Sotax AT4. Figure 4.10b includes the unmeasured variables (i.e., material loss in the RC, the tablet weight, and tablet hardness) of the system. The estimated material loss in the system ranges from 0.5 kg/h to 1 kg/h. For the tablet weight and hardness, at-line measurements were acquired. The red dot filled with yellow represents the at-line measurement from the Sotax equipment; these measurements were considered the real value for those variables. The blue dot is based on the

balance flow measurement (based on M), and Equation 4.14 and Equation 4.15 were used to estimate them.

Fig 10a   Fig 10b

Figure 4.10. E3 DG reconciled and measurement values: a) Tablet Press b) Unmeasured variables

Figure 4.11 is the global tests results for the end-to-end results. If a gross error was detected, that measurement was treated as unmeasured. Thus, the number of linear-independent equations changed and the limit for the global test too. This figure proves that during the selected period, the global test passed and the system was redundant, except at 979s-980s and 1193s-1195s. If redundancy is lost, the SSDR-M estimation is distorted, and the measurement will have to be taken as the process state instead of the reconciled values.

Measurement	**Initial SSDR-M**	**Final SSDR-M**

Figure 4.11. The global test for the end-to-end DG tableting line

Table 4.3 shows the mean and relative standard deviation (RSD) of: the measurements, the current SSDR-M solution, the SSDR-M solution with the a corrected APAP composition at the exit of the first blender, the SSDR-M solution with the corrected flow measurement at the exit of the second blender, and the ideal SSDR-M, with all measurements corrected (see section 4.4.2.2 for cases details). If RSD for the current measurement is compared with the RSD of the original SSDR-M results, it can be seen that the variability per variable was reduced. If the measurements are improved, such as correcting the PLS model for the APAP composition or reducing the bias of the flow rate, the RSD can be reduced even further, being the ideal SSDR-M the best-case scenario.

Table 4.3. Measured variable estimates after SSDR-M and GED

Variable	NOC Mean	RSD	SSDR-M Mean	RSD	Corrected $x_{API,B1}$ Mean	RSD	Corrected $F_{out,B2}$ Mean	RSD	Ideal SSDR-M Mean	RSD
F_{API} (kg/h)	1.0	1.3	1.0	1.25	1.0	1.25	1.0	1.25	1.0	1.25
F_{exc} (kg/h)	9.0	0.1	9.0	0.14	9.0	0.14	9.0	0.14	9.0	0.14
$x_{API,B1}$ (% wt.)	8.9	2.3	10.0	1.13	10.0	1.13	10.0	1.13	10.0	1.13
S (mm)	1.9	0.7	1.65	0.23	1.65	0.23	1.65	0.23	1.65	0.23
ρ_R (g/cc)	1.0	8.2	1.03	0.14	1.03	0.14	1.03	0.14	1.03	0.14

$F_{out,B2}$ (kg/h)	9.5	1.0	9.3	1.25	9.3	1.26	9.3	1.22	9.3	1.23
TS (rpm)	29.8	1.3	29.8	1.19	29.8	1.20	29.8	1.19	29.8	1.19
MCF (kN)	11.3	2.1	11.3	2.11	11.3	2.12	11.2	2.03	11.2	2.04
$F_{out,TP}$ (kg/h)	9.0	6.0	9.3	1.25	9.3	1.26	9.3	1.22	9.3	1.23

The sampling of tablets hardness and weight was done across the run, and AEE was used as the performance indicator. Table 4 shows the AEE for the tablet weight and tablet hardness, according to the case. For tablet weight and hardness, the SSDR-M results in all cases give the best results because the error is minimized. If the sensors are improved, then it is possible to reduce the error of the tablet weight, while the tablet hardness error is the same or slightly higher.

Table 4.4. AEE for SSDR-M results and measurements

Variable	M	SSDR-M	Corrected $x_{API,B1}$	Corrected $F_{out,B2}$	Ideal SSDR-M
W	2.41E-02	1.65E-02	1.65E-02	1.50E-02	1.50E-02
H	9.49E-02	6.16E-02	6.16E-02	6.81E-02	6.80E-02

4.6 Conclusion

The SSDR-M framework is shown to accomplish its two objectives: (i) improved accuracy of measured variables in the presence of random errors and (ii) detecting and compensating for measurement errors. In so doing, it improves the monitoring reliability by providing a framework for model-based validation of the measurements. SSDR-M also provides estimates of unmeasured variables through the mechanistic equations, such as tablet weight and hardness. The results of case E1 show the accurate detection of gross errors, and it has enough measurements that it never loses redundancy. The tablet weight and hardness estimation improve with the SSDR-M framework.

For the DG tableting line, the SSDR-M framework correctly detects the gross errors. For E2, the Mettler Toledo balance measurement was added as a flow measurement to have a redundant measurement. Further improvements on the ribbon modeling have to be made. For the end-to-end operation of the DG tableting line, described in E3, four cases of SSDR-M were studied and compared their reconciled values to the measurements. The RSD was significantly reduced if SSDR-M was used. The best scenario is if the measurements are corrected for biases (SSDR-M ideal). At-line measurements were employed as best estimates of the real values of these variables. In that case, SSDR-M ideal gives us the best estimation for tablet weight and hardness, since it minimizes the error. Further improvements can be made in the DG tableting line by incorporating additional PAT tools (i.e., level sensor in the roller compactor hopper, level sensor in the tablet

press hopper, composition measurement in the feed frame) as in [27]. In this way, the variable estimates can be improved, and the probability that redundancy will not be lost increases.

In conclusion, SSDR-M provides a reliable and robust process monitoring tool for CQA's. SSDR-M reconciles the variables related to the product quality or estimates them through mechanistic equations. Given the reliable estimate of the process variables, robust process control systems can be implemented to assure the quality of the product.

4.7   References

[1]   U. S. D. of H. and H. S. FDA, "Guidance for Industry PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance," no. September, p. 16, 2004.

[2]   S. L. Lee, T. F. O'Connor, X. Yang, C. N. Cruz, S. Chatterjee, R. D. Madurawe, C. M. V Moore, L. X. Yu, J. Woodcock, T. F. O'Connor, X. Yang, C. N. Cruz, S. Chatterjee, R. D. Madurawe, C. M. V Moore, L. X. Yu, and J. Woodcock, "Modernizing Pharmaceutical Manufacturing: from Batch to Continuous Production," *J. Pharm. Innov.*, vol. 10, no. 3, pp. 191–199, Sep. 2015.

[3]   C. Benqlilou, "Data Reconciliation as a Framework for Chemical Processes Optimization and Control," pp. 1–211, 2004.

[4]   M. Martinetz, A.-P. Karttunen, S. Sacher, P. Wahl, J. Ketolainen, J. G. Khinast, and O. Korhonen, "RTD-based Material Tracking in a Fully-Continuous Dry Granulation Tableting Line," *Int. J. Pharm.*, vol. 547, no. 1–2, pp. 469–479, 2018.

[5]   S. Y. Park, S. C. Galbraith, H. Liu, H. W. Lee, B. Cha, Z. Huang, T. O'Connor, S. Lee, and S. Yoon, "Prediction of critical quality attributes and optimization of continuous dry granulation process via flowsheet modeling and experimental validation," *Powder Technol.*, vol. 330, pp. 461–470, 2018.

[6]   M. Moreno, J. Liu, Q. Su, C. Leach, A. Giridhar, N. Yazdanpanah, T. O'Connor, Z. K. Nagy, and G. V. Reklaitis, "Steady-state Data Reconciliation Framework for a Direct Continuous Tableting Line," *J. Pharm. Innov.*, 2018.

[7]   M. Moreno, J. Li, B. Rentz, S. Panikar, Q. Su, G. Callegari, N. Yazdanpanah, S. Lee, F. Muzzio, Z. K. Nagy, and G. V. Reklaitis, "Case Study: Steady-State-Data Reconciliation for a Continuous Tableting Line via Direct Compression," *in prep.*, 2018.

[8]   S. Narasimhan and C. Jordache, *Data reconciliation and Gross Error detection*. Houston, TX.: Gulf Publishing Company, 2000.

[9]   O. Cencic and R. Fruhwirth, "A general framework for data reconciliation-Part I: Linear constraints," *Comput. Chem. Eng.*, vol. 75, pp. 196–208, 2015.

[10]   D. B. Özyurt and R. W. Pike, "Theory and practice of simultaneous data reconciliation and gross error detection for chemical processes," *Comput. Chem. Eng.*, vol. 28, no. 3, pp. 381–402, 2004.

[11]   C. Knopf, *Introduction to Data Reconciliation and Gross Error Detection, in Modeling, Analysis and Optimization of Process and Energy Systems*. Hoboken, NJ, USA: John Wiley & Sons, 2011.

[12]   J. A. Romagnoli and S. M.C., *Data Processing and Reconciliation for Chemical Process Operations*. San Diego, CA, 2000.

[13]   C. M. Crowe, "Observability and redundancy of process data for steady state reconciliation," *Chem. Eng. Sci.*, vol. 44, no. 12, pp. 2909–2917, 1989.

[14]   I.-W. Kim, M. S. Kang, S. Park, and T. F. Edgar, "Robust data reconciliation and gross error detection: The modified MIMT using NLP," *Comput. Chem. Eng.*, vol. 21, no. 7, pp. 775–782, Mar. 1997.

[15]   M. J. Bagajewicz, *Smart Process Plants: Software and Hardware Solutions for Accurate Data and Profitable Operations: Data Reconciliation, Gross Error Detection, and Instrumentation Upgrade*, 1st ed. McGraw-Hill Education, 2010.

[16]   A. U. Vanarase, M. Alcalà, J. I. Jerez Rozo, F. J. Muzzio, and R. J. Romañach, "Real-time monitoring of drug concentration in a continuous powder mixing process using NIR spectroscopy," *Chem. Eng. Sci.*, vol. 65, no. 21, pp. 5728–5733, 2010.

[17]   S. Ganesh, R. Troscinski, N. Schmall, J. Lim, Z. Nagy, and G. Reklaitis, "Application of X-Ray Sensors for In-line and Noninvasive Monitoring of Mass Flow Rate in Continuous Tablet Manufacturing," *J. Pharm. Sci.*, Sep. 2017.

[18]   J. Austin, A. Gupta, R. McDonnell, G. V. Reklaitis, and M. T. Harris, "A novel microwave sensor to determine particulate blend composition on-line," *Anal. Chim. Acta*, vol. 819, pp. 82–93, 2014.

[19]   A. U. Vanarase and F. J. Muzzio, "Effect of operating conditions and design parameters in a continuous powder mixer," *Powder Technol.*, vol. 208, no. 1, pp. 26–36, 2011.

[20]   S. García-Muñoz, "Phi MATLAB Toolbox." Personal Communication, 2015.

[21]   S. Ganesh, Y. Shah, M. Moreno, Z. K. Nagy, and G. V. Reklaitis, "Integrating process knowledge for ribbon density monitoring in dry granulation based continuous tablet manufacturing," *in prep.*, 2018.

[22]   Q. Su, Y. Bommireddy, M. Gonzalez, G. V. Reklaitis, and Z. K. Nagy, "Variation and Risk Analysis in Tablet Press Control for Continuous Manufacturing of Solid Dosage via Direct Compaction," in *Proceedings of the 13th International Symposium on Process Systems Engineering PSE 2018*, M. Eden, G. Towler, and M. Ierapetritou, Eds. San Diego, CA: Elsevier, 2018, p. 2912.

[23]   V. Mazel, V. Busignies, S. Duca, B. Leclerc, and P. Tchoreloff, "Original predictive approach to the compressibility of pharmaceutical powder mixtures based on the Kawakita equation," *Int. J. Pharm.*, vol. 410, no. 1, pp. 92–98, 2011.

[24]   M. Kuentz and H. Leuenberger, "A new model for the hardness of a compacted particle system, applied to tablets of pharmaceutical polymers," *Powder Technol.*, vol. 111, no. 1–2, pp. 145–153, Aug. 2000.

[25]   S. M. Razavi, G. Callegari, G. Drazer, and A. M. Cuitiño, "Toward predicting tensile strength of pharmaceutical tablets by ultrasound measurement in continuous manufacturing," *Int. J. Pharm.*, vol. 507, no. 1–2, pp. 83–89, Jun. 2016.

[26]   S.-H. Hsu, G. V. Reklaitis, and V. Venkatasubramanian, "Modeling and Control of Roller Compaction for Pharmaceutical Manufacturing. Part I: Process Dynamics and Control Framework," *J. Pharm. Innov.*, vol. 5, no. 1–2, pp. 14–23, 2010.

[27]   S. Ganesh, M. Moreno, M. Gonzalez, J. Liu, M. Gonzalez, Z. K. Nagy, and G. V. Reklaitis, "Sensor Network for Continuous Tablet Manufacturing," *Proc. 13th Int. Symp. Process Syst. Eng. PSE 2018*, 2018.

CHAPTER 5. PARTICLE-SIZE REAL-TIME MONITORING IN A CONTINUOUS TABLETING LINE VIA DRY GRANULATION

5.1 Introduction

The roller compactor is a key unit operation in dry granulation-based processes for the continuous manufacture of tablets. In this unit, the powders are fed into a hopper and then conveyed by the feed screw to two counter-rotating rolls. The powder blend is gripped by the rolls, pulled into the nip area of the compacting rolls and then compressed into a ribbon. Upon exiting the rolls, the ribbon is broken up into granules by a hammer mill [1]; generating a granular feed stream for the tablet press. The result of these transformations of the blend is to achieve a granular material that has flow and compaction properties that facilitate high-speed tableting.

The particle size distribution (PSD) of this granular stream directly affects the tablet compaction, dissolution, and other properties, which define the final quality of the tablet [2]. Thus, the PSD of the granules flow is among the Critical Quality Attributes (CQA's) of the process. According to the FDA, product quality can be improved if the CQA's and Critical Process Parameters (CPP's) are monitored and controlled [3] and thus Process Analytical Technology (PAT) tools are necessary for effective real-time processing decisions [4] [5]. In the case of roller compaction, measurement of the PSD in real time can facilitate obtaining the good quality estimate of the state of the process, which is essential for real-time process monitoring and process control of the continuous tableting line [6].

It is well known that ribbon density has a direct impact on the PSD of the resulting granules [7]. Hence, it can be considered to be an important intermediate CQA. There are alternative real-time sensors for measuring the ribbon density, including microwave (MW) and Near-Infrared (NIR)

based sensors [8]. Moreover, the ribbon density is known to be affected by the feed screw speed, the roll speed, and the roll pressure [9]. At constant blend composition, the roll pressure and feed screw speed are the main CPP's that affected the ribbon density and thus should be monitored in real-time. Since the ribbon density does affect the PSD, the latter can also be indirectly monitored by tracking the former. Indeed in [10] [11], a model-based approach has been reported in which this has been achieved in real-time.   However, to increase process robustness and sensor network redundancy, it is desirable to investigate PAT tools for directly monitoring PSD. Redundancy in measurements is particularly important for effective data reconciliation, and ultimately control [12].

One sensing technology that has the potential to measure the PSD in real-time is the Eyecon® which uses image capture and image analysis to determine PSD. The Eyecon® had been evaluated for determining particle size in the offline analysis [13] and was found to be effective in PSD measurement providing that the particles sample was presented in such a manner that the particles were adequately separated. This high-speed camera has also been evaluated for real-time monitoring for other granulation operations such as wet granulation [14]. For dry granulation, there have been at-line experiments reported which involve monitoring the ribbon density in real-time [8] [15], and relating ribbon density to the particle size distribution offline and at-line by using sieve analysis and the Eyecon®, respectively [15]. However, in [15], the Eyecon® was not integrated into the dry granulation line, and challenges for real-time measurements (i.e., identification of steady-state) were not addressed.

The primary objective of this study is to investigate the effect of the ribbon density on the granules particle size distribution (PSD) in real-time using the Eyecon® as an in-line PSD sensor. To that end, the first step is to experimentally determine the effects of variables of the roller

compactor on the particle size distribution using the Eyecon® in the off-line mode with validation conducted using other off-line PSD measurement techniques. The second step is to establish the relation between roller compactor ribbon density with the granule PSD measured using the Eyecon® in in-line mode while the dry granulation line was operated in continuous mode.

5.2 Experimental Conditions

5.2.1 Materials

The active pharmaceutical ingredient (API), used was acetaminophen (APAP) provided by Mallinckrodt Inc., Raleigh, North Carolina. Microcrystalline Cellulose pH 200 and pH 102 (MCC 200 and MCC 102, respectively), from FMC BioPolymer Corporation in Philadelphia, PA, was used as the excipient. The composition used for all experiments was 25% APAP and 75% MCC, with MCC200 used in the batch experiments described in section 5.2.2 and MCC102 in the continuous in-line experiments described in section 5.2.3.

5.2.2 Effects of pressure and milling speed in the PSD

The roll pressure (P) is the primary CPP affecting the ribbon density, which in turn affects the PSD. The PSD can potentially also be affected by other CPP's, such as the mill speed, the screen size or the distance between the mill rotor and the screen [10] [16] [17]. In the Alexanderwerk WP120 roller compactor used in this work, neither the screen position nor the screen size can be changed in real-time, hence cannot be used for real-time control purposes. The milling speed can be manipulated in real time and thus potentially could be used to control PSD. Indeed it is reported in [10], that the milling speed does affect the PSD. However [16] reports that the effect is not significant. These different assessments could well be due to the difference in

equipment characteristics and range of milling speeds (MS) investigated. To clarify this issue, the MS as an experimental variable was included in this work.

In the first phase of this study, we will investigate the effect of the MS and P on the PSD where PSD is determined using the Eyecon® off-line. For each condition shown in Table 1, we prepared a batch blend consisting of 500 g APAP and 1500 g of MCC 200. The material was transported into the roller compactor hopper at 8 kg/h. The roller compactor was operated at a constant feed screw speed of 26 rpm, a roll speed of 8 rpm and without roll gap control. The rolls of this equipment have a diameter of 12 cm, and their width is of 4 cm. The upper screen size of the mill is 1.25 mm, while the lower screen has a size of 1.0 mm. The roll pressure and mill speed set-point were varied, as shown in Table 5.1.

Table 5.1. Experimental Condition for PSD as a function of MS and P

MS (rpm)/P(bar)	25	35	45	60
35	x	x	x	x
45	x	x	x	x
90	x	x	x	x

5.2.3   Real-time monitoring of the PSD

For the real-time monitoring experiments, the production line from the feeders to the compactor was operated in continuous mode. Two LIW Schenck AccuRate AP-300 feeders were used with the first feeder transporting APAP flow at 2 kg/h, while the second one conveyed MCC102 at 6 kg/h. The flowrates of the feeders are determined by load cells measurements. The APAP and MCC102 are blended in a Gericke GCM 250 mixer, operated at 200 rpm. The roller compactor was operated at a constant feed screw speed of 26 rpm, a roll speed of 6 rpm, the MS was set at 50 rpm, there was no roll gap control, and the pressure was set at 25, 35, 45, and 60 bar. The Eyecon® was attached to the Y-connection, designed by Innopharma. A plastic bag was

fastened at the outlet of the roller compactor, after the Eyecon®, to collect granules for offline analysis. Once the process conditions (i.e., feeders flow, roll pressure, and feed screw speed had no fluctuations) were at steady-state, the granule sample was collected for 5-8 minutes.

Figure 5.1. Roller compactor process

5.3   Methods

5.3.1   Sample Preparation

The collected granules tested in an offline configuration was divided by a Retsch rotary divider into eight equal partitions. The rotary divider has a vibratory feeder that was set at an amplitude of 50. These partitions were placed into containers of 500 ml. These containers were either used in sieve analysis (see section 5.3.3.2), one of them was separated a second time, and the rest was stored. For those samples that were separated a second time by the Retsch rotary divider, the material was placed into eight containers of 175 ml. The samples in the small container

were either used for laser diffraction (see section 5.3.3.3) and Eyecon® offline (see section 5.3.3.1) measurements.

5.3.2   Ribbon Density Measurement Offline

The density of weighed samples was determined by the volumetric displacement of a free-flowing powder. The Micromeritics GeoPyc 1360 was used to measure the bulk density of the ribbons that were created via the roller compactor. Three ribbons were collected at each pressure (section 5.2.2) and measured two samples of approximately 1cm by 1cm per ribbon. The samples were weighed, and 1.7 grams to 2.0 grams of material was put into the GeoPyc. The data collected from these experiments were used to analyze whether or not the pressure of the roller compaction process affects the density of the product. The ribbon was collected from the experiments described in section 5.2.2.

5.3.3   Particle Size Measurement Offline

5.3.3.1   Eyecon® System Benchtop/At-line Configuration

For offline analysis of the PSD from the batch experiments, the Innopharma Technology Eyecon® was used in benchtop configuration. In this configuration, the Eyecon® is vertically mounted above a sliding tray. A sample of granules is placed on the sliding tray and moved through a sequence of twenty-five predetermined positions. The illumination parameters are relevant; they control a combination of blue, red and green LED's. Based on the image taken and the surface colors on the particle, a map of the surface height is built [13] [15].

Based on the image gradient, the Eyecon® fits an ellipse to each particle and determines a maximum diameter (D_{max}) and minimum diameter (D_{min}). From the D_{max} and D_{min} values, an average diameter (d) is calculated, and the particle represented as a ellipsoid with a constant density. The relevant information, namely, the camera setting (i.e., colors, camera gain, and other

parameters), the eccentricity, the diameters considered, the cumulative volumetric distribution and the numeric cumulative distribution are exported as a CSV file. For section 5.2.2, the software used by the Eyecon® was a 32-bit Windows XP with the particle sizer software. Nine samples, of at least 0.5 g, per condition, were presented sparsely in a pan to measure their PSD. For section 5.2.3, the Eyecon® system was upgraded from the software used in section 5.2.2 to a Windows 10 64-bit with the EyePass software. Three samples, of at least 0.5g, and three repetitions per condition were presented sparsely in a pan to measure their PSD.

5.3.3.2 Sieve Analysis

The sieve analysis is a mass-based particle size distribution measurement. The offline particle size distribution of the granules was measured using the W.S. Tyler Ro-Tap® Model E Test Sieve Shaker for 10 minutes on the fine analysis setting. The sieve set, manufactured by VWR, uses the mesh sizes: 90, 125 180, 250, 355, 500, 710, 1000, and 1400 μm. The sieve analysis was performed for the samples generated for both the batch and continuous experiments. Three samples were tested per condition, with each sample weighing at least 100g each.

5.3.3.3 Laser Diffraction

Laser diffraction is one of the most common techniques for particle size measurements. Particle size distributions are created when particulate material passes in front of a laser beam. The particles split the beam at angles and frequencies relative to the particle size and are recorded by a detector in the back of the unit [18]. Modern laser diffraction units use Mie's theory of light scattering to relate the scattered laser beams to the actual size of the particle. It assumes the shape can be described as a sphere. Laser diffraction units inherently produce volumetric particle size distributions [19].

Particle size distribution testing using laser diffraction was performed using a Malvern Mastersizer 3000 for the materials generated under the continuous operating conditions. The Malvern Mastersizer 3000 is a modern laser diffraction (LD) unit, able to produce particle size distributions ranging from 0.01 to 3500 µm. During testing, the dry method was used, the air pressure was set 2 bar, and the feed rate was set at 40%. Nine samples, from 0.5 g to 1 g, per condition, were tested: three small containers were selected per condition, and three samples were taken per each container (see section 5.3.1).

5.3.4   PSD Measurement in Real-time

5.3.4.1   Eyecon® System Integrated In-line on to the DG tableting line

In the real-time experiments, the PSD of the particles, which were produced while the process operated at steady state, was measured with the Innopharma Technology Eyecon®, using the integration shown in Figure 5.2. While the measurement can be based on particle count, the analysis software EyePass generates a transformation to volume-based PSD. The Eyecon® system has built-in specific settings for real-time measurements for particular operations, such as milling, twin-screw wet granulation, fluid bed granulation and coating. The settings for the milling process were set as followed: image storage interval 2s, integration time 30s, minimum reported size 25 μm, maximum reported size 1400 μm , maximum detection size 1000 μm, false agglomeration filter 0.25, edge contrast threshold 50, analysis block size 301, material presentation offset 0, no alarms active, camera gain 43%, red camera gain 25%, green camera gain 23%, blue camera gain 75%, red LED gain 60%, green LED gain 50%, and blue LED gain 100%.

To select the colors and camera gain, the standard set-up procedure provided by Innopharma Technology was followed. The sample of granules was well-distributed in the window, and then the camera was auto-calibrated (i.e., the position was focused, and the colors calibration

were selected based on the colors). The window was made from a special polycarbonate anti-static to avoid fouling. The real sampling time was 2.5s, and the equipment has a scanning rate of 10m/s. The funnel design was made to present the sample well dispersed for the image analysis. The design was provided by Innopharma. The files recorded in real-time runs consist of similar information to the files recorded in benchtop mode, except that the results, including images, are shown at every sampling time. Although the Eyecon PSD measurement can be particle count-based, the transformed cumulative volumetric distribution was used to compare the Eyecon® results to the offline methods. It should be noted that the real-time data can be directly recorded in the pilot plant Emmerson DeltaV Historian through the OPC UA server provided by Innopharma Technology.

Figure 5.2. Y-connection configuration for real-time granules PSD monitoring

5.3.4.2   Steady-State Determination

For purposes of evaluating real-time PSD measurement, the granule images must be captured when the flow rate and CPP's of the roller compactor were at steady-state. To assure that all of the data analyzed from the Eyecon® was a steady-state, a steady-state test was employed. In the literature, there are a number of different procedures reported to test time series data for the attainment of process steady-state [20] [21]. One approach is to perform a linear regression over a data window and then to use a t-test; if the slope is sufficiently different from zero, then the process

is not a steady state [22]. The F-test or R-test can also determine if the process is at steady-state [23]. Statistical process control chart (SPC) moving average chart with a threshold of 3σ are also in common use [14]. The Phillips-Perron test was used for the one-unit root, which is a type of R-test. For this test, the shape was used as well as number count and volume-based lengths: Dv10, Dv50, and Dv90. It was assumed that these seven quantities are sufficient to characterize the process dynamics, thus, to determine the stationarity of the process.

5.3.4.3   Granule or Crystal Size Underestimation Compensation

One of the challenges in in-line imaging-based particle size analysis is the tendency to underestimate the particle size due to the random spatial orientations of particles[20]. Thus, the most likely observed diameter is not the actual diameter of the particle. Moreover, for purposes of estimating the PSD, an assumption must be made about the particle shape. The authors in [20] used a crystal underestimation correction due to the orientation of the crystal. Since the measurement of the PSD presents similar challenges, the same correction is used in this work. An ellipsoid shape is assumed for the particles because this assumption is consistent with the method used in the Eyecon image processing software, EyePass. An approximate method to estimate the real size of the particles was developed. This method was based on the observed diameters, which are directly provided by the Eyecon® image analysis unit. The derivation of the equations for the correction can be found in the appendix. Similar underestimation trends have been predicted by the image-analysis based aspect ratio soft sensor development, which was recently published [20].
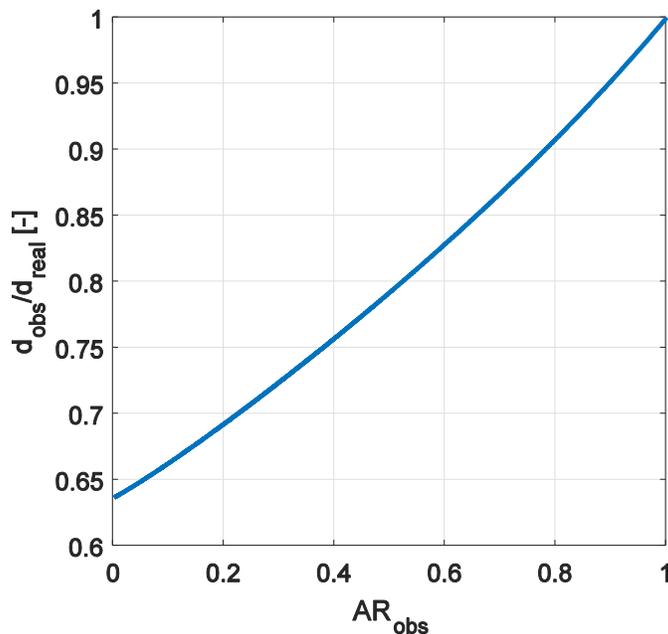
Figure 5.3. Mean particle diameter underestimation due to random spatial orientation of the ellipsoidal particles as a function of observed aspect ratio

5.4 Results and Discussion

5.4.1 Effects of pressure and milling speed in the PSD

Figure 5.4 shows that the relation between ribbon density, measured by the Geopyc, and pressure for the selected conditions. In general, in the roller compaction process, it is expected that the ribbon density is proportional to the applied pressure until a plateau effect reached. Fig 4 demonstrates that there is a linear relationship between ribbon density and pressure from 25 bar to 60 bar and a plateau has not yet been reached at 65 bar. The samples, which were acquired from the edge and the center of the ribbon, have a relative standard deviation (RSD) of 3.5%.
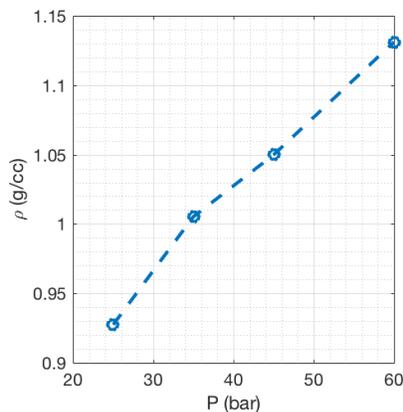
Figure 5.4. Ribbon density as a function of pressure.

Figure 5.5 shows the PSD as a function of milling speed and pressure. Figure 5.5a shows the cumulative mass distribution, which represents the number of particles per unit control volume that is less than a certain size [24] [25]. The cumulative mass distribution, given by the sieve analysis, shows a strong trend between PSD and pressure; as the pressure increases, the particle size increases. The relation is clear at dv_{50} and dv_{90}, while dv_{10} has the same trend, but some observations do not follow that trend (i.e., 25 bar and 45 rpm). It is evident that the milling speed has little or no effect on PSD in the selected normal operating conditions.

The Eyecon® in benchtop configuration was also used to measure the granules particle size distribution off-line. The PSD measurement from the Eyecon® is count-based, but for comparison purposes, the volume-based cumulative distribution was used. Figure 5.5b shows the Eyecon® results. The cumulative volume distribution generated by the Eyecon® shows the same trend as the sieve analysis; the peak in the fines increases, as the pressure decreases. Qualitative speaking the trend is the same, quantitative speaking the results are not comparable.
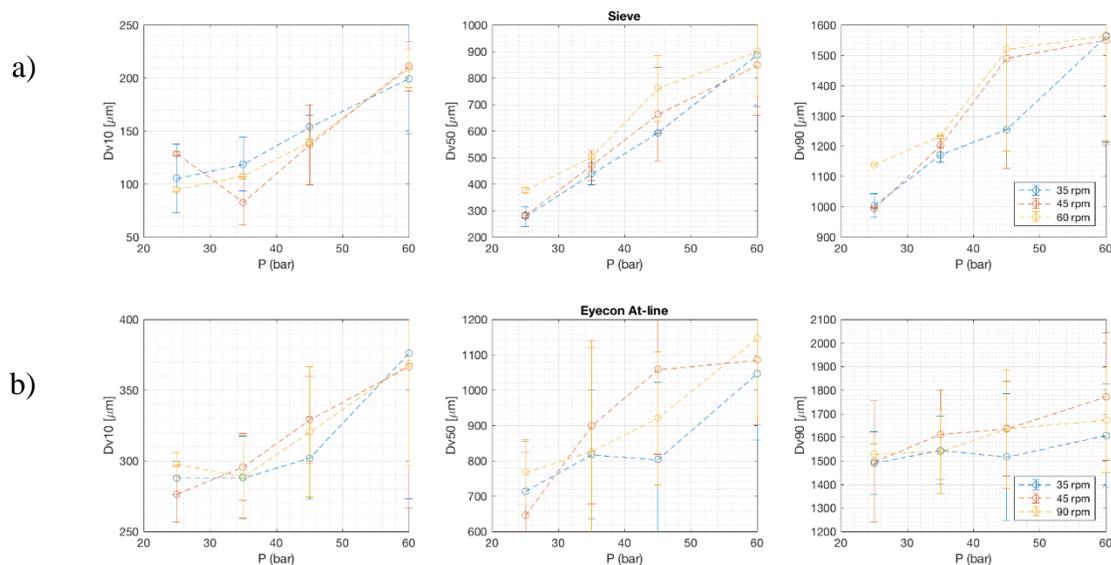
Figure 5.5. PSD as a function of pressure and milling: a) is the mass-based cumulative distribution results based on the sieve analysis results, b) is the volume-based cumulative distribution results based on the Eyecon in bench-top configuration results

5.4.2 PSD real-time monitoring

The Eyecon® can pose challenges In-line as a result of fouling on the viewing port window. The problem can be mitigated by using an anti-static window material such as polycarbonate. Figure 5.6 is an example of in-line measurements from a continuous run with the compactor pressure at 45 bar. The Phillips-Perron test employed to ensure that the data collected was at steady-state: the red squares indicate the time interval (i.e., 398s to 482s) identified by the Phillips-Perron test as being at steady-state. As can be seen from the plots, the APAP flow (F_{APAP}) has a higher variance than the MCC102 flow (F_{EXC}). The blender speed (BS), the feed screw speed (NS), the roll speed (NR), and the milling speed (MS) have an RSD below 1%. The BS and MS measurement have an offset, while the other parameters are overlapping with the set-point.
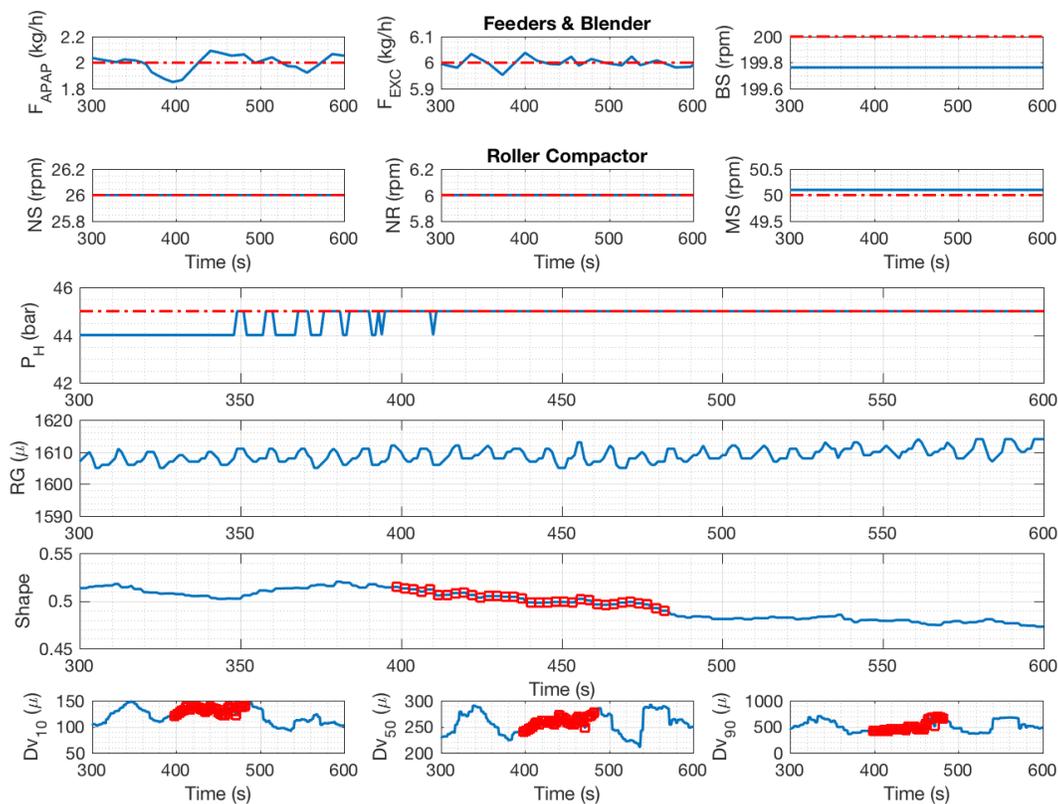
Figure 5.6. Feeder, blender, and roller compactor measurements

The pressure (P_H) and the roll gap (RG) also has an RSD below 1%, but as seen in Figure 5.6, at time 348s the pressure measurement drifted. Even though all the parameters have an RSD below 1%; the ribbon density has a higher variability because it is not uniform across the width of the ribbon, this measurement is not available in real-time for the results in Figure 5.6. This is due to the basic design of the compactor: cantilevered rolls with unconfined sides. Therefore, the PSD will also have a greater RSD.

Figure 5.7 shows the initial and final image of the granular flow taken during the steady-state period of Figure 5.6. Being an optically based measurement system, it only returns what it is seen through the lens. If the presentation of material to port window is limited then so will the information the EyePass software extracts. For instance, although the Eyecon® integration bracket

has a funnel to distribute the material toward the port window, there are particles falling in the background that are out of focus (circled in red). The out of focus particles are filtered out by EyePass. Moreover, there are small particles accumulating in the corners (circled in orange) and particles overlapping with each other (circled in yellow). The smaller accumulating particles in the corner will be detected over and over again unless new material is presented, showing the importance of avoiding fouling.

a)	b)
t = 398.67 s	t = 481.91 s
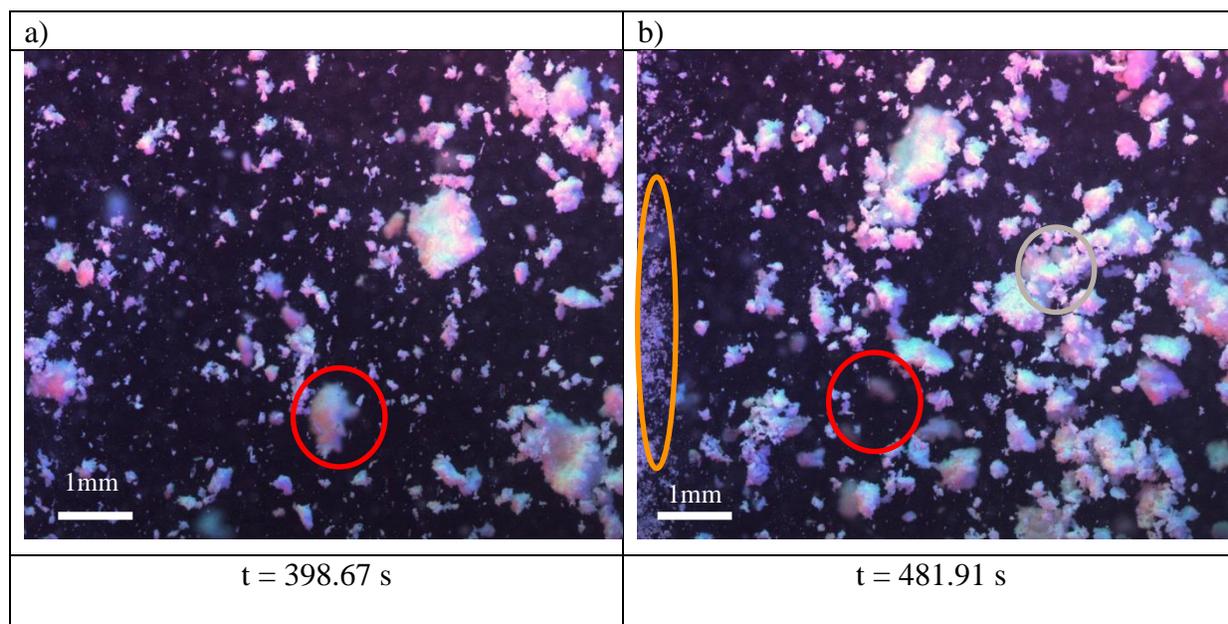
Figure 5.7. Eyecon at steady-state real-time images

For each steady-state run, a comparison was made between the real-time Eyecon® results with the offline methods.

Figure 5.8 shows the volume distribution for the results of the Eyecon® in real-time, the corrected measurement in real-time (see 3.4.3), the Eyecon® at-line, and the LD. This figure also shows the distribution from the sieve analysis, but in this case, it is mass distribution.

Figure 5.8a represented the complete distribution, while

Figure 5.8b is a close up of the peaks located in the fines region, and the

Figure 5.8c is a close up of the peaks located in the coarse region. In

Figure 5.8 it can be seen that the Eyecon® real-time distribution is bimodal. The peaks in the fines and coarse particle size regions are not as well defined as the peaks given by the offline and at-line methods. However, there is a consistent trend in the effects of pressure on particle size distribution. The peak located in the fines region decreases as the pressure increases, that is, it shifts towards the right. The peak in the coarse increases as pressure increase, that is, it shifts to the right. Although the correction to the Eyecon® results does shift the distribution in the correct direction, the correction is insufficient to overcome underestimation of the particle size.
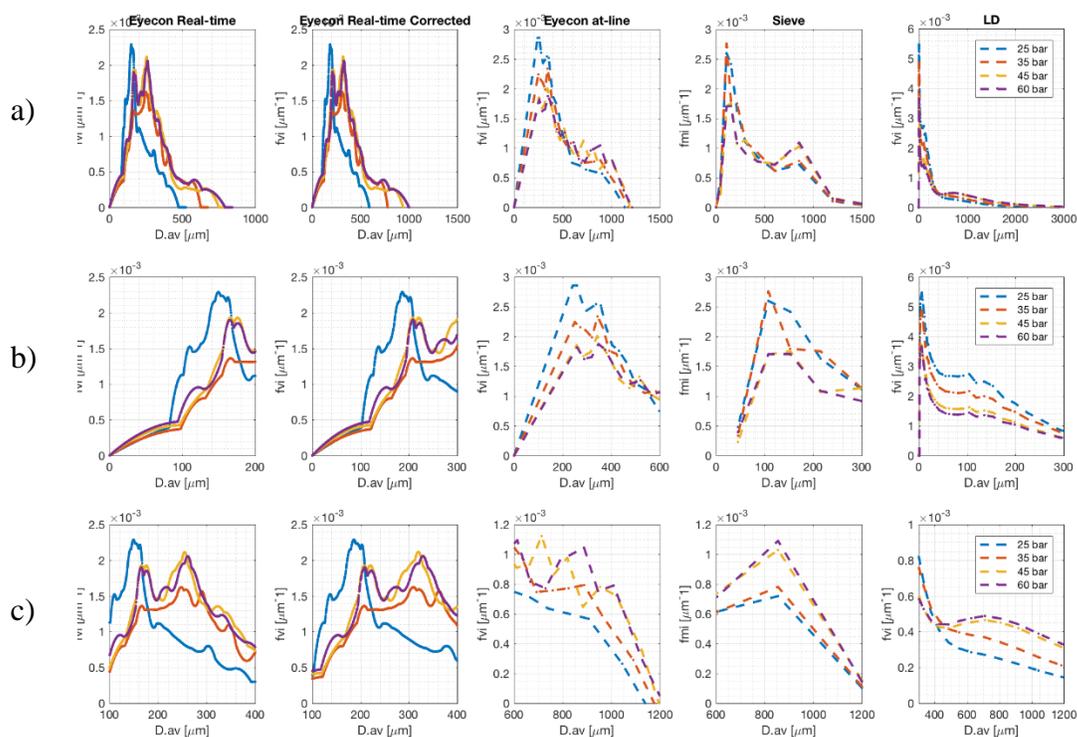
Figure 5.8. Volume and mass distribution for real-time and offline measurements

The Eyecon® at-line, the sieve analysis, and LD results do give a better definition of the bimodal distribution. The shift in the location of the peak in the fines is inversely proportional to the pressure, while the shift in the location of the peak in the coarse section is proportional to the pressure. The Eyecon® at-line, the fines peak is located around 250 µm, and the peak in the coarse region of particle size is located around 900 µm. For the sieve analysis, the peak in the peak in the fines is around 110 µm, while the peak in the coarse is estimated at 850 µm. The LD has three peaks. The first one is below 50 µm, which the other methods do not show because it is not covered in their range. The second peak follows the same trend as the other methods (the peak decreases and the pressure increases), and it is located approximately at 110 µm. For the LD, the peak in the coarse range is around 750 µm. While the methods differ in their quantitative results, the

qualitative trends are all consistent. Figure 5.9 represented the cumulative volume distribution generated by each method. The trend is the same for all method: the particle size increases as the pressure increases. The corrected estimation for the Eyecon® in real-time increases the PSD 60 µm to 120 µm, representing approximately 25% improvement.
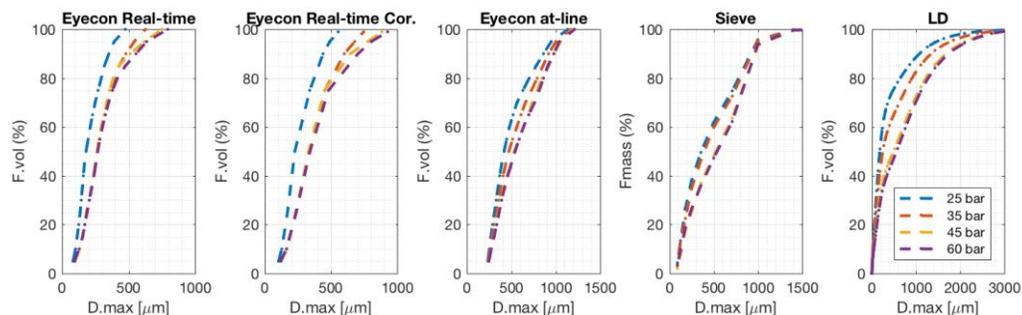
Figure 5.9. Volume and mass cumulative distribution

Figure 5.10 represents the cumulative volume distribution as a function of the applied pressure. Here a comparison between the estimation of the PSD, given by each method (e.g., LD, sieve, Eyecon® at-line), and the real-time in-line results can be made. All the methods show the monotonic relation that the PSD has with the pressure. Dv50 shows the best response for pressure, the other ones show a response, but it is not as strong as Dv50. The granules PSD start reaching the plateau effect at 45 bar, and at 60 bar it no longer increases. Since each method is based on a different property, the results do not match in all cases. However, all of them show that DV50 has a good response to pressure. Thus, this variable is an excellent candidate to control the PSD by manipulating the pressure.
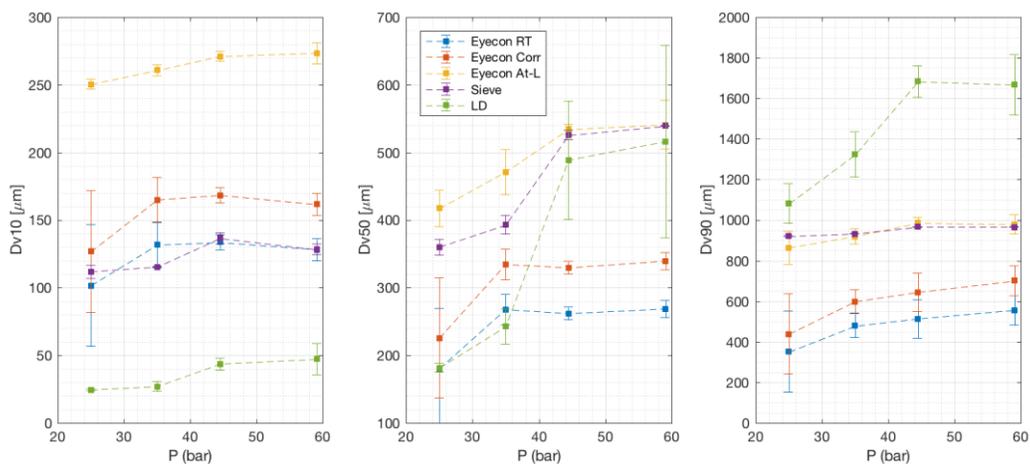
Figure 5.10. Volume cumulative distribution as a function of pressure and method used for PSD estimation

5.5   Conclusions

Based on the sieve analysis and Eyecon® in benchtop results, the MS had no significant effect on the granules PSD. This discrepancy with [10] might be due to a difference in the range tested and the difference in equipment. From the sieve, LD analysis, and Eyecon® at-line analysis, it is evident that the granules PSD is a bimodal distribution for all the conditions whether expressed in mass or volumetric terms. The peak in the coarse particle size region is higher as the pressure applied in the ribbon increases. The peak at the fines region decrease as the pressure increases. The mass and volumetric cumulative distribution also show that as the pressure increases, the particle size increases. The standard deviation is minimum at Dv10, but the trend representing PSD and pressure is more apparent at Dv50.

For in-line configuration, the Eyecon® shows the same trend. If the pressure is increased, the particle size increases. Fouling will skew the calculation of the camera. Moreover, one of the challenges presented with the Eyecon® in real-time is the underestimation of particle size. The correction proposed in this work does shift the PSD towards increasing size.

Nevertheless, for Dv50 and Dv90, the results given by the corrected PSD are still far from the offline methods results. The main reason is that the correction only can compensate for the random spatial orientations of particles, while the underestimation of the PSD by the Eyecon® in real-time is caused by additional phenomena, such as fouling. For example, the particles in the background are out of focus and thus might not be considered for the cumulative PSD. This is not an issue for the Eyecon® in bench-top configuration all the particles are at the same distance and same focus (i.e., the particles are well defined in the image). This last method overlaps with the sieve analysis results at Dv90, but it overestimates the value of Dv10.

Comparing offline methods with the Eyecon® is particularly difficult because each equipment measures the PSD based on a different property (i.e., count, mass, volume). The assumptions for each method are also different. For instance, LD assumes the particle is a sphere, while the Eyecon® assumes a fitted ellipse and then computes an equivalent ellipsoid. As a consequence, none of the methods PDS measurements overlap. The Eyecon has the advantage of real-time measurement; thus, the information retrieved from the Eyecon has the potential of being reconcile with the other dry granulation measurements. However, all the methods (in-line and offline) show the effect of pressure in the PSD and its bimodal distribution. For the real-time PSD measurement, the peaks are not as well defined in the distribution, especially the peak in the coarse particle range, which might be contributing to the particle underestimation. Nevertheless, all methods show Dv50 has a clear response with respect to pressure. If a model is developed, this variable can be integrated into the framework for data reconciliation. Once we can estimate this variable in real-time, we can design a control structure to set the PSD.

5.6 References

[1]   R. W. Miller and P. J. Sheskey, "Roller Compaction Technology for Pharmaceutical

Industry." Marcel Dekker Inc., 2003.

[2]   M. Juhnke and E. John, "Size reduction as integral element for development and manufacturing of engineered drug particles," *Chem. Eng. Technol.*, vol. 37, no. 5, pp. 757–764, 2014.

[3] U. S. D. of H. and H. S. FDA, "Guidance for Industry PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance," no. September, p. 16, 2004.

[4]   S. L. Lee, T. F. O'Connor, X. Yang, C. N. Cruz, S. Chatterjee, R. D. Madurawe, C. M. V Moore, L. X. Yu, J. Woodcock, T. F. O'Connor, X. Yang, C. N. Cruz, S. Chatterjee, R. D. Madurawe, C. M. V Moore, L. X. Yu, and J. Woodcock, "Modernizing Pharmaceutical Manufacturing: from Batch to Continuous Production," *J. Pharm. Innov.*, vol. 10, no. 3, pp. 191–199, Sep. 2015.

[5]   M. Ierapetritou, F. Muzzio, and G. V. Reklaitis, "Perspectives on the Continuous Manufacturing of Powder-Based Pharmaceutical Processes," *AIChE J.*, vol. 62, no. 6, pp. 1846–1862, 2016.

[6]   S. Ganesh, M. Moreno, M. Gonzalez, J. Liu, M. Gonzalez, Z. K. Nagy, and G. V. Reklaitis, "Sensor Network for Continuous Tablet Manufacturing," *Proc. 13th Int. Symp. Process Syst. Eng. PSE 2018*, 2018.

[7]   S. Yu, "Roll Compaction Of Pharmaceutical Excipients," University of Birmingham, 2012.

[8]   A. Gupta, J. Austin, S. Davis, M. Harris, and G. Reklaitis, "A novel microwave sensor for real-time online monitoring of roll compacts of pharmaceutical powders online - A comparative case study with NIR," *J. Pharm. Sci.*, vol. 104, no. 5, pp. 1787–1794, 2015.

[9]   S.-H. Hsu, G. V. Reklaitis, and V. Venkatasubramanian, "Modeling and Control of Roller Compaction for Pharmaceutical Manufacturing. Part I: Process Dynamics and Control Framework," *J. Pharm. Innov.*, vol. 5, no. 1–2, pp. 14–23, 2010.

[10]   S. Y. Park, S. C. Galbraith, H. Liu, H. W. Lee, B. Cha, Z. Huang, T. O'Connor, S. Lee, and S. Yoon, "Prediction of critical quality attributes and optimization of continuous dry granulation process via flowsheet modeling and experimental validation," *Powder Technol.*, vol. 330, pp. 461–470, 2018.

[11] D. Barrasso, S. Oka, A. Muliadi, J. D. Litster, C. Wassgren, and R. Ramachandran, "Population balance model validation and prediction of CQAs for Continuous milling

processes: Toward QbDin pharmaceutical drug product manufacturing," *J. Pharm. Innov.*, vol. 8, no. 3, pp. 147–162, 2013.

[12]   M. Moreno, J. Liu, Q. Su, C. Leach, A. Giridhar, N. Yazdanpanah, T. O'Connor, Z. K. Nagy, and G. V. Reklaitis, "Steady-state Data Reconciliation Framework for a Direct Continuous Tableting Line," *J. Pharm. Innov.*, 2018.

[13]   A. F. T. Silva, A. Burggraeve, Q. Denon, P. Van der Meeren, N. Sandler, T. Van Den Kerkhof, M. Hellings, C. Vervaet, J. P. Remon, J. A. Lopes, and T. De Beer, "Particle sizing measurements in pharmaceutical applications: Comparison of in-process methods versus off-line methods," *Eur. J. Pharm. Biopharm.*, vol. 85, no. 3 PART B, pp. 1006–1018, 2013.

[14]   A. S. El Hagrasy, P. Cruise, I. Jones, and J. D. Litster, "In-line size monitoring of a twin screw granulation process using high-speed imaging," *J. Pharm. Innov.*, vol. 8, no. 2, pp. 90–98, 2013.

[15]   M. A. P. McAuliffe, G. E. Omahony, C. A. Blackshields, J. A. Collins, D. P. Egan, L. Kiernan, E. Oneill, S. Lenihan, G. M. Walker, and A. M. Crean, "The use of PAT and off-line methods for monitoring of roller compacted ribbon and granule properties with a view to continuous processing," *Org. Process Res. Dev.*, vol. 19, no. 1, pp. 158–166, 2015.

[16]   M. T. A. Ende, S. K. Moses, A. J. Carella, R. A. Gadkari, T. W. Graul, A. L. Otano, and R. J. Timpano, "Improving the content uniformity of a low-dose tablet formulation through roller compaction optimization," *Pharm. Dev. Technol.*, vol. 12, no. 4, pp. 391–404, 2007.

[17]   R. McCann, "Investigating the Density Distribution of Roller Compacted Ribbons (Doctoral Dissertation)," *ProQuest Diss. Publ.*, 2014.

[18]   T. Allen, *Powder Sampling and Particle Size Determination*. Elsevier, 2003.

[19]   C. Bernhardt, "Particle Size Analysis," pp. 1–27, 1994.

[20] B. Szilagyi and Z. K. Nagy, "Aspect ratio distribution and chord length distribution driven modeling of crystallization of two-dimensional crystals for real-time model-based applications," *Cryst. Growth Desing*, vol. under revision, 2018.

[21]   G. A. C. Le Roux, B. F. Santoro, F. F. Sotelo, M. Teissier, and X. Joulia, "Improving steady-state identification," *Comput. Aided Chem. Eng.*, vol. 25, pp. 459–464, 2008.

[22]   M. J. Bagajewicz, *Smart Process Plants: Software and Hardware Solutions for Accurate*

Data and Profitable Operations: Data Reconciliation, Gross Error Detection, and Instrumentation Upgrade, 1st ed. McGraw-Hill Education, 2010.

[23] S. Cao and R. R. Rhinehart, "An efficient method for on-line identification of steady state," *J. Process Control*, vol. 5, no. 6, pp. 363–374, 1995.

[24] E. B. Litster, J. D., *The Science and Engineering of Granulation Processes*, vol. 53, no. 9. 2013.

[25] J. Litster, *Design and Processing of Particulate Products*. Cambridge: Cambridge University Press, 2016.

CHAPTER 6. ROBUST MONITORING FOR QUALITY-BY-CONTROL

6.1 Introduction

The variability of the product quality in the Pharmaceutical Industry is often caused by the poor understanding of the process and lack of robust process monitoring and control methods [1]. Continuous manufacturing has the potential to solve this problem, through the implementation of Quality-by-Design (QbD) [2]. Adequate process monitoring and control design are needed to accomplish this goal. By improving the real-time management of a continuous process, it is also possible to reduce the quantity of the out of specification material [3][4].

Process Analytical Technology (PAT) tools have been developed to monitor the variables related to the final product quality. Moreover, if the critical quality attributes (CQA's) and critical process parameters (CPP's) are measured or estimated, they can also be controlled. Thus, introducing the Quality-by-Control (QbC) concept as a subset of QbD. Control strategies can improve the quality of the product, but it has to be noted that they are not the final solution. The system can run into a fault that the control scheme cannot solve or the measurement can be faulty. A more robust monitoring scheme is needed to overcome these challenges.

In this work, data reconciliation is proposed as a method for robust process monitoring of the continuous tableting processes, specifically direct compression and dry granulation. This method has been implemented in other industries (i.e., Oil and Gas) for decades, but it has not been used for pharmaceutical processes. Data reconciliation can be posed as an optimization problem, in which the measurements of the process variables at a point in time together with the process model (i.e., in the simplest case, just mass balances) relating these variables are used as the basis for estimating adjustments to the measured values that will yield the most likely values of the

process variables. Different software packages had been developed to implement the data reconciliation framework in real-time [5]. Once the state of the process is estimated, it can be used for QbC.

My contributions made in continuous tableting manufacturing refer to process monitoring, while other authors pursued research in QbC, parameter estimation and dynamic modeling of the continuous tableting processes. The objective of this chapter is to show the different contributions made to the research of other authors. This chapter is organized as follow. Section 6.2 describes the continuous tableting processes (CTP). Section 6.3 focuses on the process monitoring of two CTP (i.e., direct compression and dry granulation) implemented in our pilot plant. Current QbD process monitoring is based on sensors, process models or multivariate models. Section 6.4 proposes data reconciliation as a method to achieve robust process monitoring, which can provide better process variable estimates for QbC. Section 6.5 describes the process control implemented in CTP; it extends from the description of the control framework to its implementation in the CTP and its challenges.

6.2 Continuous Tableting Manufacturing

The manufacturing processes in the pharmaceutical industry are designed for the production of active pharmaceutical ingredients (API) and drug products [7]. This project will be limited to drug product manufacturing, specifically the solid dosage tablet form which constitutes the dominant drug product form [4] [8].

There are three common manufacturing routes to produce tablets: direct compression, dry granulation and wet granulation [8]. Direct compression is the simplest method for drug manufacturing, but its use will depend on the flow and compaction characteristics of the powder

blend of which the table is composed. If the powder blend does not flow and compact readily, then direct compression is not an option, and a granulation method will be applied.

The main objective of granulation methods is to create larger granules which have improved powder flow characteristics and adequate compaction properties that allow a tablet of good mechanical integrity to be formed. These methods also reduce the tendency of the blend to segregate, improving the content uniformity, and narrowing particle size distributions. Wet and dry granulation methods are the most common granulation methods.

In wet granulation, the granules are created by combining the dry blended powder with a liquid through agitation. This operation is well developed but is not always useful for all materials. Some drugs are sensitive to moisture which can cause chemical stability problems or tendency to of the API to transform to a different polymorphic form [8]. The drying step that is required after wet granulation also can cause problems for compounds that are thermally unstable. As a consequence, dry granulation methods are more appropriate for this type of compound or drug [4] [8]. Among the dry granulation methods, roller compaction process is the most frequently used method. Roller compaction is inherently a continuous process, but it is operated intermittently because the unit operations up and downstream are traditionally operated in batch mode.

In the dry granulation line, there can be two or more feeders to convey the API and the excipients into a blender. After the material gets mixed, it is transported into the roller compactor (RC). In the roller compaction unit, the powders are fed into a hopper and then conveyed by the feed screw to two counter-rotating rolls. The powder is gripped by the rolls and pulled into the nip area of the compacting rolls before being compacted into a ribbon. The ribbons are then milled into granules by a rotary mill. The granules are conveyed into a mixer, where lubricant is added

through a low-in-weight feeder at low flow-rates. Finally, the granules are compressed into tablets [4].

6.3   Process Monitoring for Continuous Tableting Process

6.3.1   PAT tools for CTP

In this section, the developed PAT tools for the CTP (located at Purdue University) are described. The CQA's and CPP's are monitored through the sensors built-into the equipment or through externally mounted PAT tools. The sensor network and the PAT tools development are designed according to the structure and the needs of the CTP.

6.3.2   DC tableting line

In the direct compression process, the active pharmaceutical ingredient (API) and the excipients are fed to the process using individual feeders. The flowrates leaving the feeders are determined by the difference of the mass contained in the feeder which is measured over time via load cells on which the feeders are positioned. The feed streams are combined in a continuous blender, and then the blended powder is conveyed to the press to produce tablets. For the DC tableting line, the blend and tablet content uniformity and the tablet hardness are CQA's, while the flows are CPP's.

Near-infrared (NIR) spectroscopy is a non-destructive technique that can analyze samples without interfering in the process [9]. This technique can determine the chemical composition, density and moisture content [10] and in the studies carried out in this thesis is implemented to measure the composition of the blended material [11]. Its signal region extends from the visible spectral region (700 nm) to the beginning of the infrared region (2500 nm) [10]. However, it requires a calibration model to relate the reflected spectra to the powder blend composition. There

are two models capable of capturing this relationship for use online: Partial-least-Squares (PLS) and Artificial Neural Network (NN) combined with Principal Component Analysis (PCA) [1]. The powder blend API composition measurement can be located at different points of the CTP. In [12] it is used in the tablet press (TP) feed frame, while in [13] and [11] is measured at the exit of the blender. Figure 1 shows the points where the powder blend composition is measured in the pilot plant. The content uniformity measurement in this work is executed using an improved PLS model for the NIR.

There are other CPP's and CQA's sensors in the CTP tableting line. The total flowrate is a CPP's. The mass flow of the blended powder stream can be measured in real-time through the X-ray based sensor (SETXvue XP-300; Enurga Inc.) [14]. The tablet hardness is a CQA in the DC tableting line. Different measurement in the tablet press (i.e., dosing, main compression force, pre-compression force) can be related to this property. However, the tablet hardness measurement is a destructive test. Thus, it can only be executed at-line. There are different tablet testers available, one of them is the Sotax AT4 tablet tester (AT4). This equipment measures the weight, thickness, diameter, and hardness by diametrical compression of the tablets [15]. Of these measurements, the hardness test is the only destructive test. However, all of the AT4 measurements are at-line since the maximum sampling rate is slower than the tablet production rate. The tablet composition is also a CQA that can be measured at-line through NIR (transmittance). However, this last measurement is still under development for the CTP shown in Figure 6.1.

6.3.3 DG tableting line

In addition to the CQA's and CPP's of the DC tableting line, the DG tableting line has other properties related to the final quality of the product. The roller compacted ribbon density, and the particle size distributions (PSD) are part of the CQA's. These properties of the granules

produced by the roller compactor unit can be related to the final tablet properties. The ribbon density has an effect on the particle size distribution (PSD) of the granules [7] [8]. The PSD of the granules affects the flowability of the granules and is related to the final tablet hardness and dissolution [16]. Therefore, the ribbon density measurement is crucial to monitoring and controlling the product quality.

Two online methods can be used to measure the flow composition and ribbon density using a suitable calibration model: NIR spectroscopy and the microwave (MW) resonance sensor. The microwave resonance sensor will give similar information to the NIR, except that it can measure the density across the entire cross-section of the ribbon [9] whereas the NIR in reflectance mode does not penetrate the entire ribbon; rather it only measures the composition of a portion of the ribbon. However, the sampling time of the MW sensor is higher than the sampling time of the NIR.

The next unit operation, in the dry granulation tableting line, is a hammer mill. This operation is structured to be part of the roller compactor equipment but does function as a separate unit operation which takes the ribbon produced and breaks it into granules. Here the CQA is the particle size distribution (PSD) of the granule stream produced. This CQA can be measured offline using sieve analysis and laser diffraction, or at-line with the Innopharma Eyecon® Camera [17] [18]. So far, online measurements, with the Eyecon® Camera, have only been implemented in other continuous processes (e.g., wet granulation, high-shear granulator) [19] [20]. The application and testing of the Eyecon® Camera have yet to be done for dry granulation measurements in real-time. The investigation of the capability of this sensor to provide reliable online PSD information will be reported in Chapter 5 of this dissertation.
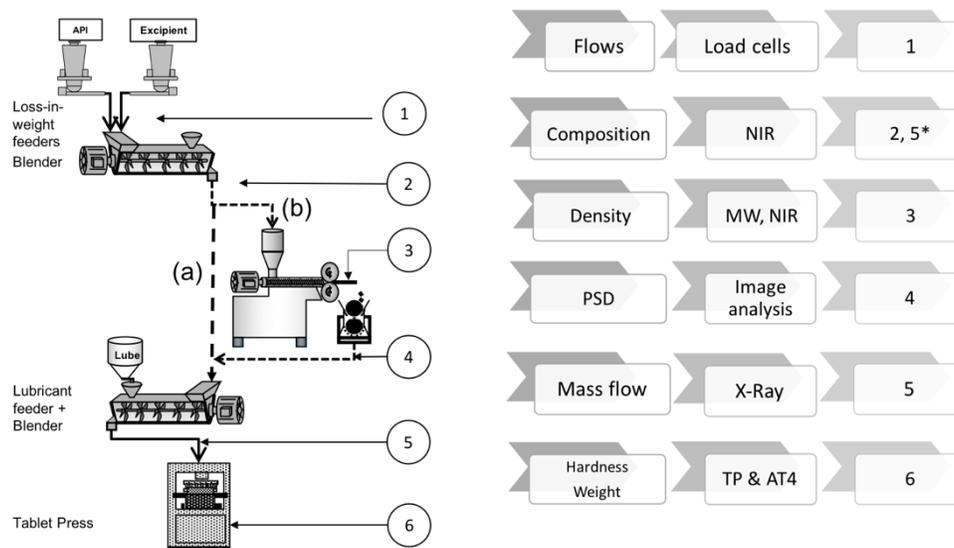
Figure 6.1. DC and DG continuous tableting line.

* Only applies to the DC tableting line.

6.3.4   Flowsheet modeling

The use of integrated flowsheet models has increased in the pharmaceutical industry and has been very beneficial to understanding variable interactions as well as the process dynamics of DC and DG tableting lines [7] [21] [22]. Steady-state and dynamic reduced order models have been developed to represent each unit in the CTP. In this section, the common models used in flowsheets for CTP are briefly reviewed.

6.3.5   DC tableting line

The three tableting processes share common unit operations. The unit operations of feeding and blending of the API and excipients flows are common to all three processes. These unit operations can be modeled in either at steady-state or in a dynamic form. The authors in [23] represent the feeder and blender system as a Plug-Flow Reactor (PFR). For the tablet press, there are also models available. For instance, the Kawakita model represents the behavior of powder

under pressure [24] (See Eq. 4.12). The tablet hardness can also be modeled through first order equations [25] [26] (see Eq. 4.15).

6.3.5.1   DG tableting line

In the literature, different models have been developed to represent the formation of the roller compacted ribbon as a function of the hydraulic pressure, roll speed, and feed screw speed. Through these models, the roller compacted ribbon density can be estimated. One of the models used to represent the roller compaction process is the model developed by Hsu, using the Johanson's rolling theory with dynamic mass balance [27].

The granular output of the hammer mill which is associated with the roller compactor can be predicted by a population balance model [28] [29]. The population balance equations include terms which model the breakage rate using a suitable semi-empirical breakage model. The parameters must be estimated from experimental data. The experiments involve running the mill at varying speed and measuring the particle size distribution of the resulting granules. There are other models, that can represent the milling process and have been applied to the dry granulation process, such as the Vogel and Peukert model [22].

$$\tag{6.1}$$

$$\tag{6.2}$$

In the previous equations, the variables are: f_{Mat} is the particle resistance against fracture; $W_{m,min}$ is the mass specific energy a particle can absorb without fracture; P_B represents the breakage probability; k is the integer number of successive impacts a particle receives,; w is the particle size before impacts; $W_{m,kin}$ represents the mass specific impact energy for each impact; B is the cumulative breakage function; v is the size of the fragment particle; v' is the fragment size for

additional fading of the particle size distribution; and q is a power-law exponent relevant to the particle impact velocity.

6.4   Robust Process Monitoring and Gross Error Detection

The robust operation of a continuous process requires a sufficient level and quality of online process measurements and an appropriate level of process control. As the pharmaceutical industry transitions from batch to continuous operation the questions of how many measurement points, what quality of online measurements are necessary or sufficient, and the level of sophistication of the process control systems required are very much under discussion. The focus of this chapter is to answer questions concerning the selection and impact of online measurements.

6.4.1   Steady-state data reconciliation and Gross Error Detection

A nonlinear steady-state data reconciliation model was developed for the DC and DG tableting line to address these questions. The data reconciliation model of such a process network consists of the relationship between all of the measured and relevant unmeasured variables of the process together with a description of the error distributions of these measurements [30] [31]. It is assumed that the number of measurements is larger than the degree of freedom of the network model, hence that there is redundancy.

Given a set of measurements, an optimization problem is posed which has as its objective function a likelihood type function (see equations 6.1, subject to equations 6.2 and 6.3). This function minimizes the difference between the measured values and their best estimates, weighted by the covariances of the measurement errors of the associated sensors, with the above-mentioned process model forming the constraint set [30].

$$\min_{x,y} J = (\mathbf{x}^+ - \mathbf{x})^\mathrm{T} \mathbf{Q}^{-1} \left(\mathbf{x}^+ - \mathbf{x} \right) \tag{6.1}$$

$$\text{s.t.} \quad \mathbf{h}\left(\mathbf{x},\mathbf{y},q\right) = 0 \tag{6.2}$$

$$\mathbf{g}\left(\mathbf{x},\mathbf{y},q\right) \pounds\, 0 \tag{6.3}$$

Where: $\mathbf{x}^+ \in \mathfrak{R}^n$ is a vector of measurements, $\mathbf{x} \in \mathfrak{R}^n$ is a vector of reconciled values of the n measurements, $\mathbf{y} \in \mathfrak{R}^m$ is a vector of the m unmeasured process variables, $\mathbf{Q}$ is the covariance matrix, $\mathfrak{R}^{n \times n}$, $\boldsymbol{\theta} \in \mathfrak{R}^p$ is a vector of p process model parameters, $\mathbf{h} \in \mathfrak{R}^k$ is a set of equations which describe the steady-state behavior of the process, and $\mathbf{g} \in \mathfrak{R}^q$ is a set of inequality constraints, *n* is the number of reconciled variables, *m* the number of unmeasured variables and *k* the number of relations or equations.

The solution of the maximum likelihood problem constitutes the best real-time estimate of the state of a processing system considering all of the measurements and their error characteristics. Such an estimate can serve as the basis for checking whether all process variables are within acceptable bounds, for identifying gross measurement errors, for control action as well as for process optimization steps.

Additionally, the reconciliation formulation provides a framework for investigating the impact of adding or eliminating measurements, of replacing instruments with given error characteristics with those with tighter variances or of introducing measurement redundancy [32]. Two tests can identify gross errors. The first one is the global test; if the objective function is greater than $\chi\,(k)^2_{1-\alpha}$, there is a gross error. The measurement test (equations 6.4 to 6.6) is another way to identify gross errors. Since the test is done per measurement, most likely any $\mathbf{mt}_i$ bigger than $z_{\text{crit, } (1-\alpha)}$ (statistical value) will be a gross error.

$$\mathbf{a}_i = (\mathbf{x}_i^+ - \mathbf{x}_i) \tag{6.4}$$

$$\mathbf{V} = \mathbf{QA}^\mathbf{T}\left(\mathbf{AQA}^\mathbf{T}\right)^{-1}\mathbf{AQ} \tag{6.5}$$

$$\mathbf{mt}_i = \frac{|\mathbf{a}_i|}{\mathbf{V}_{ii}^{\frac{1}{2}}} \tag{6.6}$$

Where: $\mathbf{x}^+ \in \Re^n$ is a vector of measurements, $\mathbf{x} \in \Re^n$ is a vector of reconciled values, $\mathbf{a} \in \Re^n$ is a vector of errors, $\mathbf{Q}$ is the covariance matrix, $\Re^{n \times n}$, $\mathbf{A} \in \Re^{kxn}$ is the matrix of coefficients of the linear/linearized process model, $\mathbf{mt} \in \Re^n$ is a vector of the measurement test values, i is the index, and n is the number of measurements, and α is the level of significance (e.g., 5%).

The data reconciliation problem has received much attention in the process systems engineering literature. However, most of the attention has been directed at applications with linear models, which most commonly consist of the component mass balances of the process [30]. Nonlinearities have principally been introduced through the addition of energy balances, which in their most common form involve equations that are bilinear in flow and composition or flow and temperature [30]. Additionally, it is generally assumed that measurement errors are normally distributed. In this work, the nonlinearities that arise because of the inclusion of online measurements of properties (i.e., density and variables such as screw speed, roll speed, etc.) are considered.

The model nonlinearities have implications for computation performance, which is important for the real-time application. Steady-state data reconciliation is used to reduce the computational time, but it is only applicable to the data collected when the process is at steady-state. Since the dynamic of the CTP is fast, steady-state data reconciliation was selected to be implemented in the DC and DG tableting line. Real-time tests are available to evaluate if the process is steady-state [33]. Different case studies were carried out on the CTP to illustrate the benefits of data reconciliation. In [34], a steady-state data reconciliation (SSDR) framework is proposed (see Chapter 2). Its application is shown in [35] for the DC tableting line (see Chapter

3), where the data provided by Rutgers University. In [36], SSDR was implemented for the DC and DG tableting line (see Chapter 4) at Purdue University. Improvements to the SSDR framework implementation can be made by increasing the number of measurements, as shown in [37], such as the PSD measurement in real-time (see Chapter 5). The PSD measurement in real-time is done through image analysis and provides potential information for data reconciliation. However, it can be challenging due to fouling during operation and PSD underestimation. This underestimation of the particle size can be caused by the particles being out of focus and the orientation of the particle. The former reason was addressed in Chapter 5; fellow co-authors contributed experimentally and modeling of the PSD underestimation correction.

6.4.2 Dynamic-state data reconciliation

Dynamic data reconciliation can represent the process more accurately, but its use does require an increase in the computational time. If the computational time to solve the problem can be reduced, then one could employ dynamic data reconciliation in real-time. Instead of only using a steady state model, as described by equations 6.2 and 6.3, the process model has to be modified to represent the process dynamics through differential equations 6.7 and 6.8.

$$f_x\left(\frac{dx}{dt}, x, y, q\right) = 0 \tag{6.7}$$

$$f_y\left(\frac{dy}{dt}, x, y, q\right) = 0 \tag{6.8}$$

Where: $\mathbf{x}^+ \in \mathfrak{R}^n$ is a vector of measurements, $\mathbf{x} \in \mathfrak{R}^n$ is a vector of reconciled values of the n measurements, $\mathbf{y} \in \mathfrak{R}^m$ is a vector of the m unmeasured process variables, $\boldsymbol{\theta} \in \mathfrak{R}^p$ is a vector of p process model parameters, n is the number of reconciled variables, m the number of unmeasured variables and k the number of relations or equations.

In [38], the lead author used dynamic data reconciliation to estimate the state of the feeder and blender subsystem. The approach is demonstrated using a simulation of the process rather than pilot plant runs, and it is only meant to investigate the application of dynamic data reconciliation online. Developing of measurements, and data collection was needed for this contribution. This information was provided to the lead author of [38]. This work was also applied experimentally in the tablet press operation, as stand-alone equipment (i.e., only the tablet press was operated) [15]. Further improvements can be made, and the next objective is to implement this framework in the CTP tableting line in an end-to-end operation.

6.5   Process Control for Continuous Tableting Process

QbC is defined as the improvement in the quality of the product attained through a suitable control design [39], and the use of property feedback control. For a control scheme to be well-design, it has to go through a systematic design process based on clear understanding of the process and a rigorous demonstration of how process control can improve process performance. Therefore, a systematic framework was proposed by [40]; the framework is described in section 6.5.1 and the feeder and blender model, proposed in [23], is used to show the benefits of the control structure in a subsystem of the CTP. The authors in [41] expanded the study of the framework for real-time implementation and executed it in the pilot plant, (see section 6.5.2). Although I was not the lead author, contributions to these deliverables were made at different levels: performance indicators design, initial simulation studies, measurement development and sensors connectivity to the process historian.

6.5.1   Systematic Framework for Process Control Design and Risk Analysis

This work introduces the importance of QbC and a new control framework for the design and analysis of Control Systems for Continuous Manufacturing Processes (CMP). The objective of this framework is to define the steps that must be taken to develop a control system design for a pharmaceutical process. The framework includes the classification of risk scenarios and control schemes according to their complexity level. It also proposes suitable performance indicators for the evaluation and comparison of the control system designs. This work also includes a case study, where it is shown how to apply the control framework to an existing subsystem (i.e., the feeder and blender unit operations) of the tableting process. This case study demonstrates the importance of applying Property Feedback Control (PFC) in the CQA's and CPP's to improve product quality.

For this work, the hierarchical three-layer control design was proposed by Prof. Zoltan Nagy (see 6.5.1.1). I conducted an initial study, in SIMULINK and OPTCON, for the feeder and blender control structures design. In this study, I also suggested two performance indicators for continuous pharmaceutical processes: Time-to-product (T2P) and Magnitude-to-Product (M2P), see section 6.5.1.3. For the publication, Dr. Qinglin Su improved the feeder-blender model and control structures through parameter estimation and state estimation. The lead author of [40] included other performance indicators suitable for the continuous pharmaceutical processes, such as Duration-to-Reject (D2R). He also added other indicators commonly used for control performance evaluation (i.e., ITAE). Dr. Qinglin Su also defined the risk scenarios that can be encounter in the continuous tableting line and mapped them (see 6.5.1.2). Finally, he integrated all these concepts into the systematic framework.

6.5.1.1   Hierarchical Three-Layer Control Design

Process control can be executed at three layers: layer 0, layer 1 and layer 2. The controllers at layer 0 (L0) are built-into the equipment by the manufacturer. For example, the loss-in-weight feeders have a built-in controller which adjusts the flow measurement to bring it to the set-point. The controllers at L0 only access the control variables of that specific unit operation, and the control designs are usually Proportional-Integral (PI) and Proportional-Integral-Derivative (PID), which are the most commonly used ones in the industry [42][43]. This type of layer does not have PAT tools to control the CQA's.

At layer 1 (L1) control, the PAT tools are used to measure CQA's and these values used by the control system. Thus, PFC can be applied to this layer. The common control schemes used in Layer L1 are ratio control, cascade control, feed-forward control, etc. These schemes can be implemented in a multiloop control design [42]. Controlling one or multiple unit operations is possible. Finally, the last layer of control is layer 2 (L2). At this layer, the common control schemes are Model Predictive Control (MPC), adaptive control, among others [44]. Here the controller offers control actions integrated across multiple unit operations, typically, so as to achieve plant-wide control.

6.5.1.2   Risk Mapping, Assessment, and Planning

The risk scenarios to which the control system must respond are ranked according to their severity and frequency: low risk is classified as R0, a risk at a medium level is R1, and high risk is R2 [40]. R0 disturbances can be measured and are of a short duration of time. R1 are medium risks scenarios with uncertainties in the sensor calibration or unmeasured disturbances. Finally, R2 risks are serious and rare problems, such as gross errors in the calibration or model-plant mismatch.

6.5.1.3 Control Performance Indicators

There are two indicators described in [40] that provide useful information to determine the performance of a control scheme. Figure 6.2 is an illustrative comparison between open loop and closed loop. For instance, if the system is being operated in open loop and there is an increase in the API flow, it is expected that the API content will also increase. In this example, the API content is the CQA, while the API flow is the manipulated variable. The first performance indicator is Time to Product (T2P). This indicator is illustrated in Figure 6.2a, and it consists of the length amount of time the disturbance takes to propagate from its source to the product, and it is detected as a deviation in one or more in the critical values of the product. If the system is operating in closed loop, the increase in the API content could be faster. Therefore, the amount of segregated or out of specification product can be decreased. The second performance indicator is Magnitude to Product (M2P). This indicator shows the maximum deviation in the CQA from the target value caused by the disturbance. Figure 6.2b is a graphical representation of the M2P concept.

a)   b)
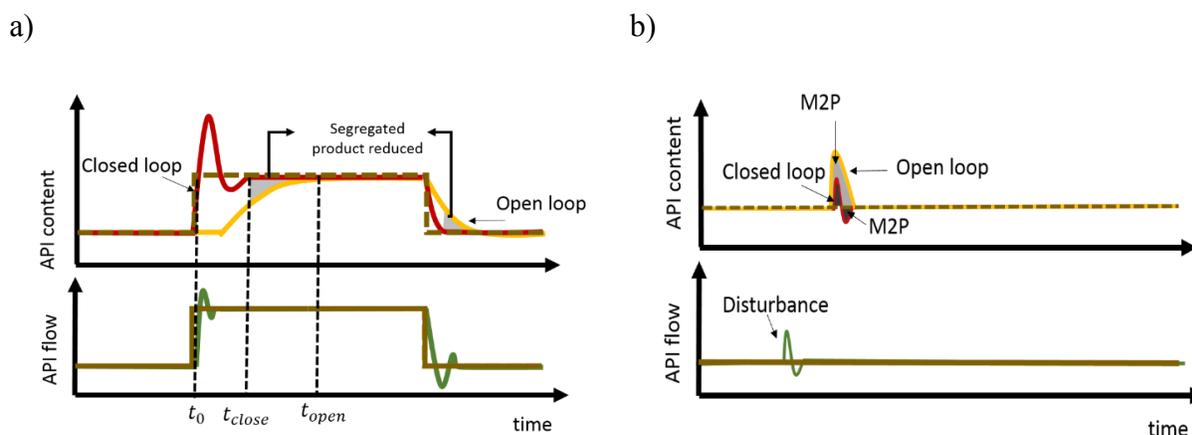
Figure 6.2. Performance Indicators: a) T2P definition, b) M2P definition.

Disturbances can affect the control variable, but operating in closed loop can reduce its effect. If the control variable is between the regulatory limits, the product can still be considered in-specification. Minimizing the disturbances effect on the control variables is important. Both indicators are relevant for the control analysis, T2P will define *when* the product is out of specification and M2P will show *if and by how much* it is out of specification.

6.5.2   Resilience and risk analysis of fault-tolerant process control design

The objective of the systematic framework, proposed in section 6.5.1, is to identify the potential risks associated with the control system design, the material property variations, and process uncertainties. In this way, QbD and process safety can be accomplished. In [40], the framework is applied to the feeding-blending subsystem in a simulated scenario. In [41], the systematic framework was demonstrated experimentally. This work included process operation failures and product quality variations in the feeding-blending system. The control performance evaluation showed the importance of QbC to assure product quality in the CTP. However, random and gross errors still pose a challenge for process control. Robust process monitoring has to be implemented to minimize the risk level (see section 6.4).

For this work, my contributions were made from the experimental point of view. Initial experiments to control the Acetaminophen (APAP) composition were conducted at a pilot plant scale using controllers at Layer 1 in Emerson DeltaV. The composition measurement needed improvement. Thus, I created a Partial Least Square (PLS) model to relate the CDI-NIR spectra to the APAP composition. The PLS model was developed using *phi*. MATLAB toolbox and two principal components were selected. The spectra were pretreated using the Savitzy-Golay filter, first derivative and Standard Normal Variate (SNV), and it was trained over a 5%-17% APAP composition range and MCC200 as the excipient. I also connected this information from the PAT

tools computer to the computer station, where the Emerson DeltaV Historian is installed (see Figure 6.3). The lead author of [41] executed the control structures in real-time, which includes control on the flowrate by including the Mettler Toledo Balance in the system. Moreover, he was able to implement the systematic framework in real-time.
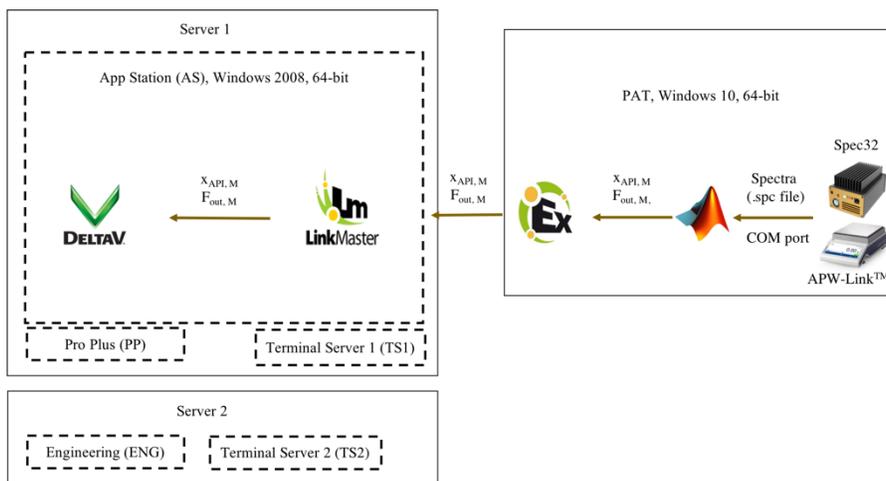

Figure 6.3. PAT tools station connected to the historian in DeltaV.

6.6   Conclusions

Robust process monitoring is the first step that has to be taken to improve product quality in real-time. This part of the real-time management scheme is crucial for QbC to be implemented in CTP. Robust process monitoring can be facilitated through data reconciliation and gross error detection. However, the development of PAT tools and process models is required. Data reconciliation was used at steady-state only to reduce the computational time to allow it to be implemented it in real-time. The value of robust process monitoring and its implementation in CTP is demonstrated in this dissertation.

6.7 References

[1] A. Gupta, J. Austin, S. Davis, M. Harris, and G. Reklaitis, "A novel microwave sensor for real-time online monitoring of roll compacts of pharmaceutical powders online - A comparative case study with NIR," *J. Pharm. Sci.*, vol. 104, no. 5, pp. 1787–1794, 2015.

[2] U. S. D. of H. and H. S. FDA, "Guidance for Industry PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance," no. September, p. 16, 2004.

[3] M. Ierapetritou, F. Muzzio, and G. V. Reklaitis, "Perspectives on the Continuous Manufacturing of Powder-Based Pharmaceutical Processes," *AIChE J.*, vol. 62, no. 6, pp. 1846–1862, 2016.

[4] R. Ramachandran, J. Arjunan, A. Chaudhury, and M. G. Ierapetritou, "Model-based control-loop performance of a continuous direct compaction process," *J. Pharm. Innov.*, vol. 6, no. 4, pp. 249–263, 2011.

[5] L. D. R. Beal, D. Hill, R. A. Martin, and J. D. Hedengren, "GEKKO Optimization Suite," pp. 1–26, 2018.

[6] A. Gupta, A. Giridhar, V. Venkatasubramanian, and G. V. Reklaitis, "Intelligent alarm management applied to continuous pharmaceutical tablet manufacturing: An integrated approach," *Ind. Eng. Chem. Res.*, vol. 52, no. 35, pp. 12357–12368, 2013.

[7] A. Rogers, A. Hashemi, and M. Ierapetritou, "Modeling of Particulate Processes for the Continuous Manufacture of Solid-Based Pharmaceutical Dosage Forms," *Processes*, vol. 1, no. 2, pp. 67–127, 2013.

[8] R. W. Miller and P. J. Sheskey, "Roller Compaction Technology for Pharmaceutical Industry." Marcel Dekker Inc., 2003.

[9] J. Austin, A. Gupta, R. McDonnell, G. V. Reklaitis, and M. T. Harris, "A novel microwave sensor to determine particulate blend composition on-line," *Anal. Chim. Acta*, vol. 819, pp. 82–93, 2014.

[10] FOSS, "A guide to Near-Infrared Spectroscopic Analysis of Industrial Manufacturing Processes," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2013.

[11] A. U. Vanarase and F. J. Muzzio, "Effect of operating conditions and design parameters in a continuous powder mixer," *Powder Technol.*, vol. 208, no. 1, pp. 26–36, 2011.

[12] E. M. Hetrick, Z. Shi, L. E. Barnes, A. W. Garrett, R. G. Rupard, T. T. Kramer, T. M. Cooper, D. P. Myers, and B. C. Castle, "Development of Near Infrared Spectroscopy-based Process Monitoring Methodology for Pharmaceutical Continuous Manufacturing Using an Offline Calibration Approach," *Anal. Chem.*, vol. 89, no. 17, pp. 9175–9183, 2017.

[13] J. Austin and M. T. Harris, "In-situ monitoring of the bulk density and the moisture content of rapidly flowing particulates using a microwave resonance sensor," *IEEE Sens. J.*, vol. 14, no. 3, pp. 821–828, 2014.

[14] S. Ganesh, R. Troscinski, N. Schmall, J. Lim, Z. Nagy, and G. Reklaitis, "Application of X-Ray Sensors for In-line and Noninvasive Monitoring of Mass Flow Rate in Continuous Tablet Manufacturing," *J. Pharm. Sci.*, Sep. 2017.

[15] Q. Su, Y. Bommireddy, M. Gonzalez, G. V. Reklaitis, and Z. K. Nagy, "Variation and Risk Analysis in Tablet Press Control for Continuous Manufacturing of Solid Dosage via Direct Compaction," in *Proceedings of the 13th International Symposium on Process Systems Engineering PSE 2018*, M. Eden, G. Towler, and M. Ierapetritou, Eds. San Diego, CA: Elsevier, 2018, p. 2912.

[16] M. Juhnke and E. John, "Size reduction as integral element for development and manufacturing of engineered drug particles," *Chem. Eng. Technol.*, vol. 37, no. 5, pp. 757–764, 2014.

[17] A. F. T. Silva, A. Burggraeve, Q. Denon, P. Van der Meeren, N. Sandler, T. Van Den Kerkhof, M. Hellings, C. Vervaet, J. P. Remon, J. A. Lopes, and T. De Beer, "Particle sizing measurements in pharmaceutical applications: Comparison of in-process methods versus off-line methods," *Eur. J. Pharm. Biopharm.*, vol. 85, no. 3 PART B, pp. 1006–1018, 2013.

[18] M. A. P. McAuliffe, G. E. Omahony, C. A. Blackshields, J. A. Collins, D. P. Egan, L. Kiernan, E. Oneill, S. Lenihan, G. M. Walker, and A. M. Crean, "The use of PAT and off-line methods for monitoring of roller compacted ribbon and granule properties with a view to continuous processing," *Org. Process Res. Dev.*, vol. 19, no. 1, pp. 158–166, 2015.

[19] D. Treffer, P. R. Wahl, T. R. Hörmann, D. Markl, S. Schrank, I. Jones, P. Cruise, R. K. Mürb, G. Koscher, E. Roblegg, and J. G. Khinast, "In-line implementation of an image-based particle size measurement tool to monitor hot-melt extruded pellets," *Int. J. Pharm.*, vol. 466, no. 1–2, pp. 181–189, 2014.

[20] A. S. El Hagrasy, P. Cruise, I. Jones, and J. D. Litster, "In-line size monitoring of a twin screw granulation process using high-speed imaging," *J. Pharm. Innov.*, vol. 8, no. 2, pp. 90–98, 2013.

[21] A. J. Rogers, C. Inamdar, and M. G. Ierapetritou, "An integrated approach to simulation of pharmaceutical processes for solid drug manufacture," *Ind. Eng. Chem. Res.*, vol. 53, no. 13, pp. 5128–5147, 2014.

[22] S. Y. Park, S. C. Galbraith, H. Liu, H. W. Lee, B. Cha, Z. Huang, T. O'Connor, S. Lee, and S. Yoon, "Prediction of critical quality attributes and optimization of continuous dry granulation process via flowsheet modeling and experimental validation," *Powder Technol.*, vol. 330, pp. 461–470, 2018.

[23] P. M. Portillo, F. J. Muzzio, and M. G. Ierapetritou, "Using compartment modeling to investigate mixing behavior of a continuous mixer," *J. Pharm. Innov.*, vol. 3, no. 3, pp. 161–174, 2008.

[24] V. Mazel, V. Busignies, S. Duca, B. Leclerc, and P. Tchoreloff, "Original predictive approach to the compressibility of pharmaceutical powder mixtures based on the Kawakita equation," *Int. J. Pharm.*, vol. 410, no. 1, pp. 92–98, 2011.

[25] M. Kuentz and H. Leuenberger, "A new theoretical approach to tablet strength of a binary mixture consisting of a well and a poorly compactable substance," *Eur. J. Pharm. Biopharm.*, vol. 49, no. 2, pp. 151–159, 2000.

[26] S. M. Razavi, G. Callegari, G. Drazer, and A. M. Cuitiño, "Toward predicting tensile strength of pharmaceutical tablets by ultrasound measurement in continuous manufacturing," *Int. J. Pharm.*, vol. 507, no. 1–2, pp. 83–89, Jun. 2016.

[27] S.-H. Hsu, G. V. Reklaitis, and V. Venkatasubramanian, "Modeling and Control of Roller Compaction for Pharmaceutical Manufacturing. Part I: Process Dynamics and Control Framework," *J. Pharm. Innov.*, vol. 5, no. 1–2, pp. 14–23, 2010.

[28] J. Litster, *Design and Processing of Particulate Products*. Cambridge: Cambridge University Press, 2016.

[29] E. B. Litster, J. D., *The Science and Engineering of Granulation Processes*, vol. 53, no. 9. 2013.

[30] S. Narasimhan and C. Jordache, *Data reconciliation and Gross Error detection*. Houston, TX.: Gulf Publishing Company, 2000.

[31] J. A. Romagnoli and S. M.C., *Data Processing and Reconciliation for Chemical Process Operations*. San Diego, CA, 2000.

[32] O. Cencic and R. Fruhwirth, "A general framework for data reconciliation-Part I: Linear constraints," *Comput. Chem. Eng.*, vol. 75, pp. 196–208, 2015.

[33] L. L. Simon, "Continuous Manufacturing: Is the Process Mean Stationary?," *AIChE J.*, vol. 00, no. 00, pp. 1–12, 2018.

[34] M. Moreno, J. Liu, Q. Su, C. Leach, A. Giridhar, N. Yazdanpanah, T. O'Connor, Z. K. Nagy, and G. V. Reklaitis, "Steady-state Data Reconciliation Framework for a Direct Continuous Tableting Line," *J. Pharm. Innov.*, 2018.

[35] M. Moreno, J. Li, B. Rentz, S. Panikar, Q. Su, G. Callegari, N. Yazdanpanah, S. Lee, F. Muzzio, Z. K. Nagy, and G. V. Reklaitis, "Case Study: Steady-State-Data Reconciliation for a Continuous Tableting Line via Direct Compression," *in prep.*, 2018.

[36] M. Moreno, S. Ganesh, Y. Shah, M. Gonzalez, Z. K. Nagy, and G. V. Reklaitis, "Nonlinear Steady-State Data Reconciliation for Continuous Tableting Processes," *in prep.*, 2018.

[37] S. Ganesh, M. Moreno, M. Gonzalez, J. Liu, M. Gonzalez, Z. K. Nagy, and G. V. Reklaitis, "Sensor Network for Continuous Tablet Manufacturing," *Proc. 13th Int. Symp. Process Syst. Eng. PSE 2018*, 2018.

[38] J. Liu, Q. Su, M. Moreno, C. Laird, Z. Nagy, and G. Reklaitis, "Robust state estimation of feeding–blending systems in continuous pharmaceutical manufacturing," *Chem. Eng. Res. Des.*, vol. 134, pp. 140–153, Jun. 2018.

[39] E. Icten, *Manufacture of Individualized Dosing: Development and Control of a Dropwise Additive Manufacturing Process for Melt Based Pharmaceutical Products*. 2016.

[40] Q. Su, M. Moreno, A. Giridhar, G. V Reklaitis, and Z. K. Nagy, "A Systematic Framework for Process Control Design and Risk Analysis in Continuous Pharmaceutical Solid-Dosage Manufacturing," *J. Pharm. Innov.*, Oct. 2017.

[41] S. Qinglin, M. Moreno, S. Ganesh, G. V. Reklaitis, and Z. K. Nagy, "Resilience and risk analysis of fault-tolerant process control design in continuous pharmaceutical manufacturing," *MKO Process Saf. Int. Symp.*, 2018.

[42] T. Marlin, *Process Control*, 2nd ed. 2000.

[43] D. E. Seborg, T. F. Edgar, D. A. Mellinchamp, and F. J. Doyle III, *Process Dynamics, and Control*, 3rd ed. 2011.

[44] F. Allgöwer, R. Findeisen, and Z. K. Nagy, "Nonlinear model predictive control: From theory to application," *J. Chinese Inst. Chem. Eng.*, vol. 35, no. 3, pp. 299–315, 2004.

CHAPTER 7. ACCOMPLISHMENTS AND FUTURE DIRECTION

7.1 Accomplishments

Process monitoring is the foundation for real-time process management. The Steady-State Data Reconciliation (SSDR) framework, established in this dissertation, gives a systematic approach to make process monitoring more robust because it deals with random and sensors gross errors in real-time. Thus, improving the process monitoring for the Continuous Tableting Process (CTP) is the key focus of the research presented in this thesis.

Specifically, PAT tools were developed to improve the process monitoring of the continuous direct compression tableting line and dry granulation tableting line. It was also shown that it is possible to estimate the process state through a model-based approach (SSDR-M) or a data-driven approach (SSDR-D). Both approaches include statistical tests for outlier detection, and their estimation of the process give comparable results when the model is linear or mildly-nonlinear. Whereas SSDR-M can estimate unmeasured observable variables, SSDR-D can only estimate unmeasured variables if the appropriate model equations or empirical relations are added to the data driven model. These constraints are already included in SSDR-M. This thesis provides four deliverables for the continuous pharmaceutical manufacturing, which include process measurement development, the steps to execute data reconciliation to deal with errors and its implementation in the direct compression and dry granulation continuous tableting line (see Figure 7.1).

The first deliverable was the SSDR framework, which gives the best estimate of the state variables of the two continuous tableting line studied, as long as the sensor network is redundant (see Chapter 2). The framework can be part of a continuous improvement process under which the number of PAT tools in the system is increased or the process model is improved through more

rigorous mechanistic relations. In this way, further control decisions can be based on the most reliable state of the process, making process monitoring robust for continuous pharmaceutical processes, such as continuous tableting manufacturing via direct compression and dry granulation.
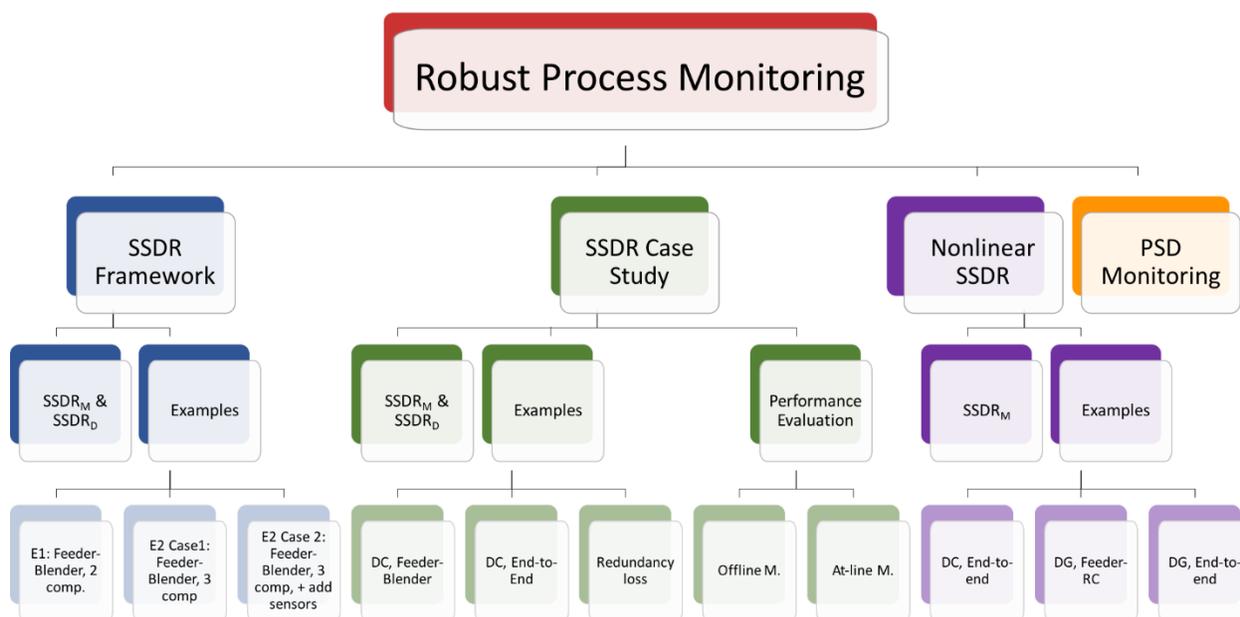

Figure 7.1. Dissertation deliverables for robust process monitoring for continuous tableting manufacturing

The second deliverable, is the SSDR case study, which shows the SSDR framework applied to the a continuous tableting line via direct compression (see Chapter 3). The nonlinear SSDR framework also uses the SSDR framework and it is considered the third deliverable; it is implemented in another continuous tableting line via direct compression and dry granulation (see Chapter 4). In this thesis, the development and implementation of more PAT tools is suggested for the particle size distribution of the granules produced in the dry granulation continuous tableting line (see Chapter 5). In this way, more measurements can be used for reconciliation. Finally, chapter 6 is not a deliverable but a summary of different contributions to other research topics related to process control. Each contribution was important for process control and the development of real-time process management for the pharmaceutical industry.

7.2   Future Directions

The framework proved to be able to deal with random and gross errors to make robust monitoring possible. However, it can always be improved by considering expanding the sensor network and improving the process models. Moreover, dynamic data reconciliation offers the potential to improve the estimation of the state of the process, by relaxing the steady-state requirement and considering transitions. Finally, process control has to be implemented over the reconciled values to improve product quality. In this section, the suggested improvements per area are explained.

7.2.1   Steady-State Data Reconciliation

SSDR has proven a useful method for process monitoring. However, this framework does not consider the bias correction of the models for soft sensors (e.g., composition measurement). In the literature [1], different methods for bias correction have been reported. Data reconciliation also does not consider the errors introduced by the process (e.g., ratholing and bridging in the hopper). These errors should be detected through the proper alarm design [2]; in which SSDR has to be integrated.

7.2.1.1   Sensor network

The sensor network shown for direct compression and dry granulation can be expanded by increasing the number of API composition measurements in the line. The MultiEye NIR is a suitable PAT tool to monitor the composition of the powder blender at multiple locations because it has four probes. This equipment has a wavelength range of 1500nm to 2200nm, and a sampling frequency of 25 kHz. It is important to note that the CDI-NIR has a different wavelength range (904 nm to 1781 nm). Thus, the MultiEye provides complementary information. Moreover, this equipment has OPC Unified Architecture (UA), meaning that it has its own server and the

capability to be connected into the tableting line sensor network (i.e., the data is sent in real-time to the process historian).

For the direct compression line, the composition measurement can be added at the feed frame of the tablet press, and at the exit of the second blender as a second composition measurement in the same location. For the dry granulation line, it can be added in the same location (i.e., tablet press feed frame and exit of the second blender); however, the calibration has to be done for granule flow. Other measurements can be developed through the MultiEye NIR, such as the ribbon density measurement to have more redundancy in the roller compactor process.

If done correctly, the increase of PAT tools can improve the data reconciliation results. Performance can be evaluated by comparing the data reconciliation results with the at-line and offline measurements of the critical quality attributes. Nevertheless, if the sensor is not well calibrated or is blocked (i.e., fouling) it can cause gross errors. If there is not a system to deal with said gross errors, the solution is distorted; thus, the importance of the SSDR framework. If the sensor is consistently unreliable, then it might not add benefits to the data reconciliation problem because the variable will always be considered unmeasured. Therefore, it will be important to develop mechanistic relations for those variables that cannot be measured in real-time.

7.2.1.2 Particle Size Distribution (PSD) modeling for the granules flow

As described in Chapter 5, it is possible to measure the cumulative volume granules PSD in real-time through the Eyecon camera. However, a model for the milling model needs to be included in the SSDR framework to reconcile the PSD values. If not, the measurement is non-redundant. Meaning that the measurement can only be evaluated and corrected by itself because there are not mechanistic equations that serve to correct the measurement. A model has been used to represent the effects of the ribbon density in the granules PSD using a ball mill [3]. Other models

represent the hammer mill in the roller compactor, but do not include the ribbon density or pressure effect in the PSD [4]. Therefore, a model that includes the effect of the ribbon density on the granules PSD has to be established.

7.2.2 Dynamic Data Reconciliation

If the implementation of dynamic data reconciliation is desired, then the development of dynamic models, for each unit operation, is required. The dynamic models for the feeder and blender unit operations have been developed and used in data reconciliation [5], but practical implementation is still needed. Dynamic data reconciliation for the tablet press also needs to be studied and implemented. The same challenges are faced for the dry granulation process. The roller compactor is a highly nonlinear process. Thus, this is a bigger challenge to execute in real-time with a low computational time. In the literature, different dynamic models are representing each unit operation [6]. Determination of normal operating conditions and parameter estimation are required to adapt said models to the continuous tableting processes presented in this work.

7.2.3 Robust Monitoring for Quality-by-Control

In Chapter 6, it has been shown that Quality-by-Control (QbC) implementation can improve product quality. Although, this chapter is not a deliverable of this dissertation, it shows the need of improving process monitoring for control purposes. The use of reconciled values instead of the direct measurement can improve the robustness of the system since the control decisions will be based on the true state of the process. Current work has shown control structures designed for each unit operation or an end-to-end simulation. However, a master control structure, for end-to-end operation in the continuous tableting process, has yet to be implemented with data reconciliation.

For instance, now that the PSD of the granules can be measured in real-time, this measurement can be integrated to the SSDR framework shown in Chapter 4 for the continuous tableting line via dry granulation. For the continuous dry granulation tableting line, the control structures in the literature are conducted using simulated scenarios. This is a good approach for process understanding, but the implementation is required in real-time. If a model is established for the milling process, the PSD distribution measurement can be reconciled and controlled. Thus, improving the quality of the product through robust process monitoring and control is possible. Marking the importance of measurements development, model development and a framework that considers both to estimate the most likely state of the process: data reconciliation.

7.3 References

[1] J. A. Romagnoli and S. M.C., *Data Processing and Reconciliation for Chemical Process Operations*. San Diego, CA, 2000.

[2] A. Gupta, "Intelligent Alarm System Management and Value of Information Analysis Applied to Continuous Pharmaceutical Manufacturing," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2013.

[3] D. Barrasso, S. Oka, A. Muliadi, J. D. Litster, C. Wassgren, and R. Ramachandran, "Population balance model validation and prediction of CQAs for Continuous milling processes: Toward QbD in pharmaceutical drug product manufacturing," *J. Pharm. Innov.*, vol. 8, no. 3, pp. 147–162, 2013.

[4] S. Y. Park, S. C. Galbraith, H. Liu, H. W. Lee, B. Cha, Z. Huang, T. O'Connor, S. Lee, and S. Yoon, "Prediction of critical quality attributes and optimization of continuous dry granulation process via flowsheet modeling and experimental validation," *Powder Technol.*, vol. 330, pp. 461–470, 2018.

[5] J. Liu, Q. Su, M. Moreno, C. Laird, Z. Nagy, and G. Reklaitis, "Robust state estimation of feeding–blending systems in continuous pharmaceutical manufacturing," *Chem. Eng. Res. Des.*, vol. 134, pp. 140–153, Jun. 2018.

[6]   S.-H. Hsu, G. V. Reklaitis, and V. Venkatasubramanian, "Modeling and Control of Roller Compaction for Pharmaceutical Manufacturing. Part I: Process Dynamics and Control Framework," *J. Pharm. Innov.*, vol. 5, no. 1–2, pp. 14–23, 2010.

APPENDIX A. CHAPTER 2 EXPERIMENTAL PROCEDURE & INFORMATION FLOW

The experimental procedures for executing runs for cases E1 and E2 are quite straightforward. A run is initiated by turning on Feeders 1 and 2, which take 1 -2 minutes to arrive at steady-state as defined by the conditions shown in Table 1. With the blender operating at 200 rpm, an additional period 2-3 min was allowed from the moment the feeders reached steady-state [39] before reconciliation was initiated, corresponding to reaching a Q at or below the values in Table 1. In case E1 the process set-points are as shown in in Table 1. The measurements are collected and sent to MATLAB-2016a-64bit for execution, as shown in the system configuration, Fig. A1. During the runs, all variables are saved in the Emerson DeltaV v.13.3 process historian, operating using two Dell Precision T1700 64-bit computers with Windows 2008 RL2, configured as five virtual computers. The first server, houses App Station (AS), ProPlus (PP) and Terminal Server 1 (TS1). The second server houses Engineering station (ENG) and Terminal Server 2 (TS2). As configured, each virtual computer has a specific function: the ENG station is used for adding equipment and troubleshooting issues. Computer AS is used when other applications, such as MATLAB, need to communicate with the DeltaV system. The PP computer is used to add any figures in the main template for the operation of the pilot plant, process historian view diagrams and to add or modify the control modules in DeltaV. Finally, TS1 and TS2 are used to operate the process.

The codes for the SSDR-M and the SSDR-D frameworks are implemented in MATLAB 2015b 32-bit and installed in the AS computer. Since MATLAB serves as a client, by using the OPC toolbox, the code reads the measured flows and composition from the DeltaV historian, solve the optimization problem and writes the information directly into the DeltaV control module

"SSDR" All control modules and process historian view diagrams were created on the PP computer. For these experiments, the system was operated from the TS1 computer.

Another, physical computer: Dell Inspiron Windows 10 64-bit laptop using Intel Core^{TM} i7 and 8GB RAM called "PAT," is used to connect to the necessary PAT tools, such as the CDI-NIR or the Mettler Toledo balance, and to link the required information to the other computers. All composition and flow measurements are in the DeltaV historian. However, it has to be noted, that of these four subsystem measurements, only the feeder flow values are directly sent to the DeltaV system. For the composition measurements, the CDI NIR is connected via USB port to the "PAT" computer. By using the CDI-NIR software, called Spec32, the spectra measurement is collected as a '.spc file'. This file is generated every second and read by the multivariate statistical methods code. The code, which is in MATLAB 2016a 64-bit, executes the PLS model, created in "phi" software package, and loaded as a ".mat" file. This software was acquired through personal communication with Dr. Garcia-Munoz. The results are sent to KEPServerEX v6. This server, provided by Kepware, sends the information to LinkMaster v3, which is connected to the DeltaV system. In parallel, the mass flow MATLAB code is executed using the "parfool" function. This code calls the information from the APW-Link^{TM} software, which is connected to the Mettler Toledo balance through a COM port. It is important to mention that the acetaminophen composition and mass flow measurements are computed using an averaging time window of 10s.
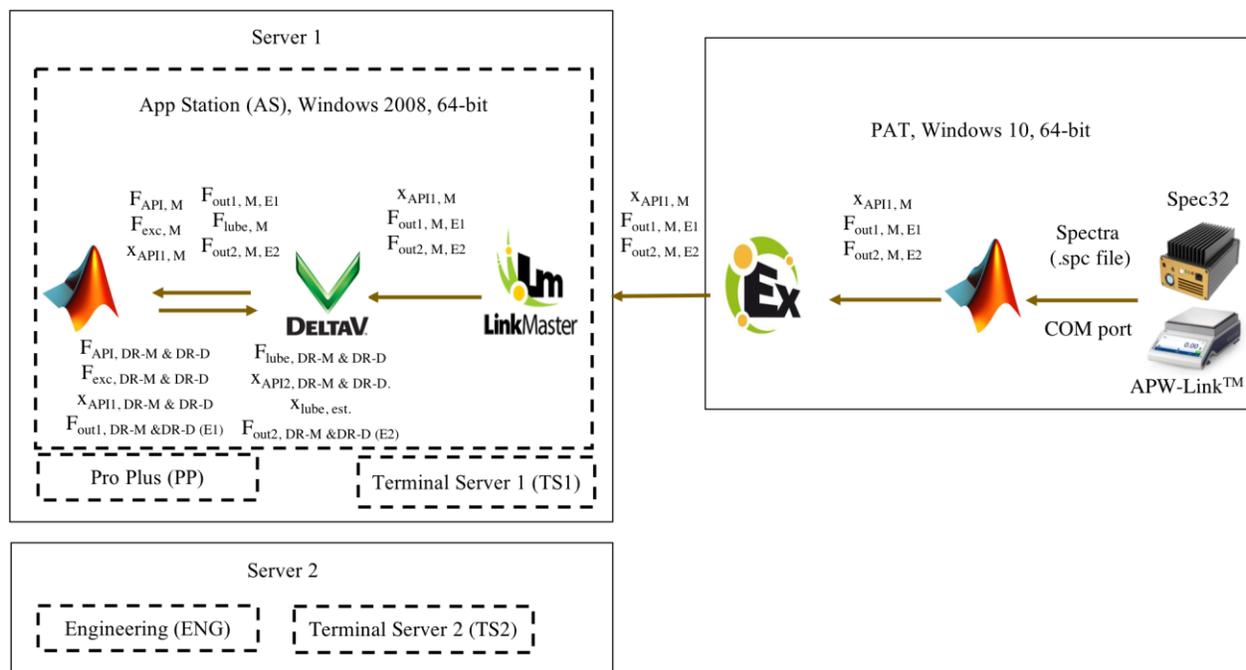
Figure A. 1. System configuration for the Continuous Tableting Line: Virtual Computers (--) and Physical Computer (-).

APPENDIX B. CHAPTER 3 EXAMPLE FOR SSDR APPLICATION

For each run, the SSDR framework was used. This section shows an example (Experiment 7) of how the framework performs for the entire run in an end-to-end SSDR. For the tablet weight, tensile strength and thickness, the red asterisk represents the point in time where the bias was corrected through the use of the at-line measurements. In the case of API composition, there is no bias correction because the measurement was done offline. The results of section 3.6.2 are based one second before the bias correction (red asterisk).

Experiment 7 feeder-blender subsystem reconciled values and measurement test (binary value) are shown in Figure B. 1.The left-hand side has the reconciled and measurement values, while the right-hand shows if the measurement test passed (1-pass, 0-fail). At the bottom of the measurement and reconciled values, the SPE and global test are used for gross error detection with a 99% confidence interval. In the right-hand side, at the bottom, the number of measurements used in the reconciliation is shown.

The total number of measurements has to be higher than six, or there is no redundancy for SSDR-M; while for SSDR-D at least 80% of the measurements active is needed. In these images, SSDR-D 6PC is used to have a comparison with SSDR-M. Through SSDR-M, a gross error was detected around 129s-133s in the Raman composition. This detection might be due to a small bias in the Raman measurement since it is slightly slower to correct the measurement due to its higher sampling time.
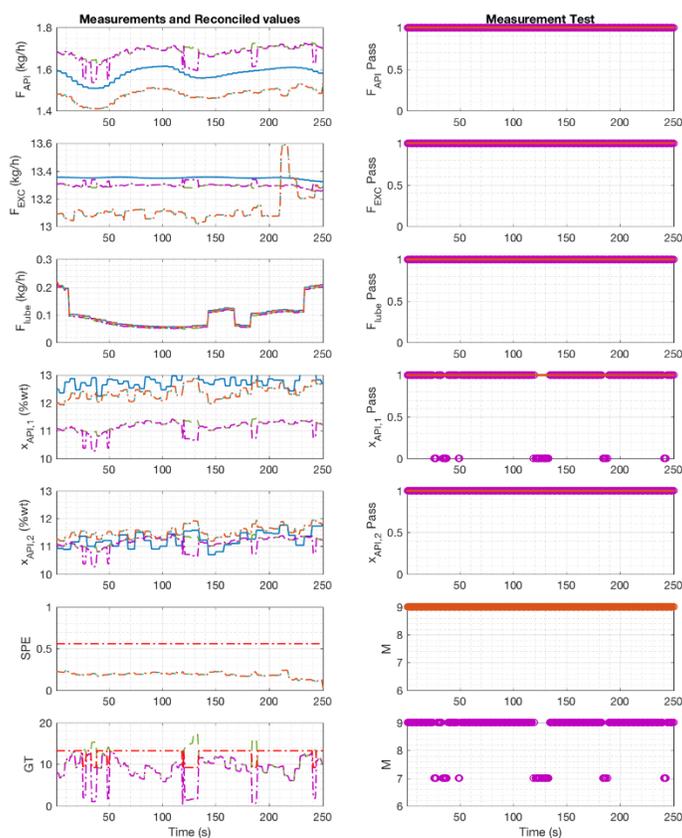
Figure B. 1. Reconciled values and gross error detection for the feeder-blender subsystem.

Figure B. 1. represents the reconciled values and measurement test (binary value) of the tablet press subsystem. The organization of this image is analogous to the organization of Figure B. 1.; the left-hand-side has the reconciled values, while the right-hand side shows if the measurement pass or not. The weight was considered a gross error from 129s to 133s by the SSDR-M algorithm. This detection might be a false positive since the measurement test is based on the linearized version of the model.
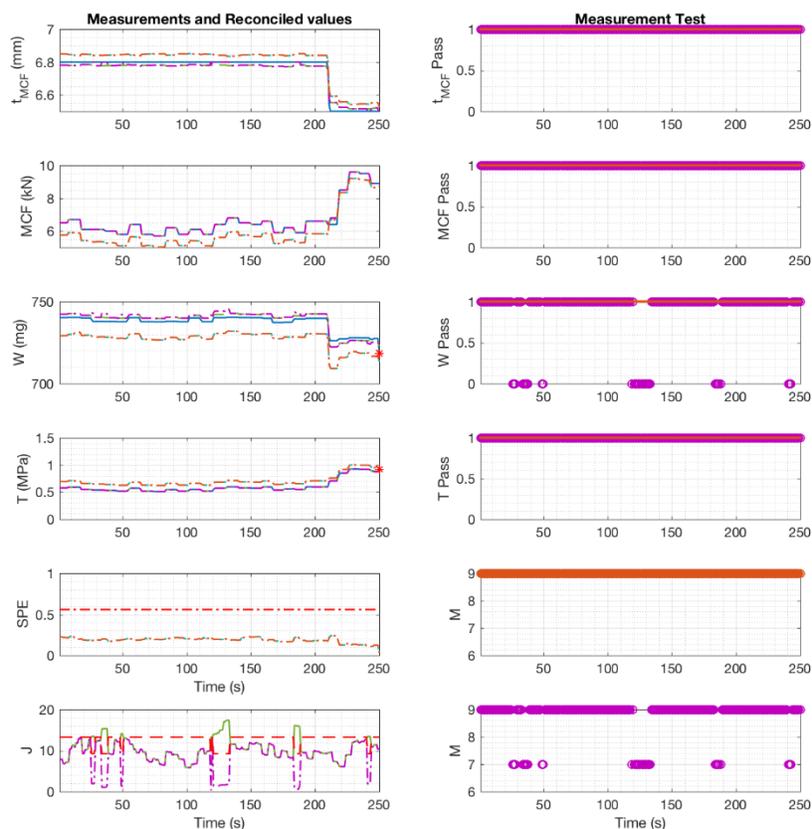
Figure B. 2. Reconciled values and gross error detection for the tablet press subsystem.

Furthermore, there are instances where steady-state is lost due to the change in the thickness at MCF (approx. 210s, see Figure B. 2) and the turret speed (approx.160s and 240s, see. Figure B. 3). The global test did not violate the limit in the SSDR-M algorithm because the parameters (*a* and *1/b*) were re-estimated at the new conditions. However, the results during those fractions of time cannot be trusted because the entire framework is based on a steady-state scenario. Figure B. 3. shows the parameters and the estimation of the unmeasured variables.
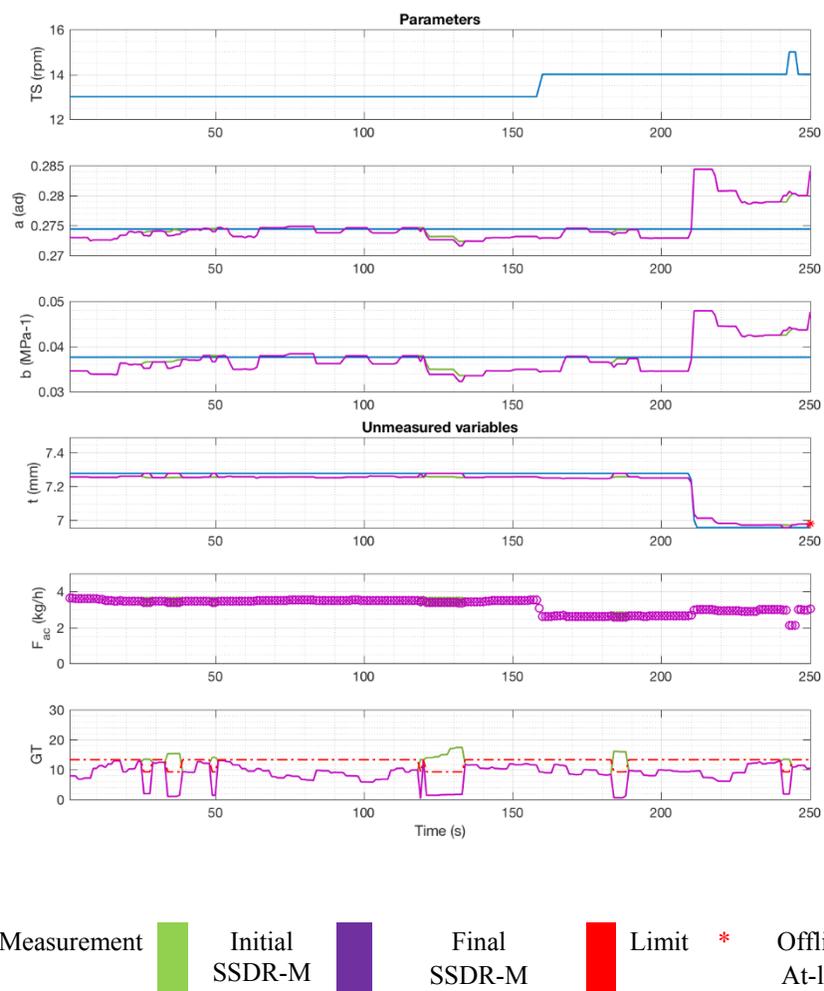
Figure B. 3. Parameters and unmeasured variables for SSDR-M

APPENDIX C. CHAPTER 4 COVARIANCE ESTIMATION

The covariance (Q) is estimated from the relative standard deviation of the measurements (RSD). The RSD is dependent on the material and NOC. For the direct compression line, the covariance was estimated at 10% APAP, 0.2% of silicon dioxide was used, and the rest was MCC200. For the dry granulation line, the estimation was done at 10% APAP and 90% MCC102. We used previous runs to have an idea of the Q that had to be used for the SSDR-M implementation. Table C1 shows the conditions used for the direct compression, while Table C2 shows the conditions used for the dry granulation line. Each estimation was done above 200s. For most of the variables, we selected the highest RSD.

Table C1. Direct compression covariance

Variable	NOC	STD	RSD (%)
F_{API} (kg/h)	1.00	0.009	0.9
F_{exc} (kg/h)	9.00	0.150	1.7
$x_{API,B1}$ (% wt)	10.0	0.445	4.4
F_{gli} (g/h)	22.7	3.548	15.7
$x_{API,B2}$ (% wt)	7.37	0.485	6.6
$F_{out,B2}$ (kg/h)	9.90	0.442	4.5
TS (rpm)	40.2	0.414	1.0
MCF (kN)	3.9	0.083	2.1

Table C2. Dry granulation tableting line covariance

Variable	NOC	STD	RSD (%)	NOC	STD	RSD (%)	NOC	STD	RSD (%)
F_{API} (kg/h)	1.00	0.01	1.45	1.00	0.02	1.59	-	-	-
F_{exc} (kg/h)	9.00	0.01	0.14	9.00	0.01	0.17	9.0	0.0271	0.30
$x_{API,B1}$ (% wt)	8.69	0.61	6.97	8.93	9.00	7.06	-	-	-
S (μm)	1822.84	16.94	0.93	1886.49	39.12	2.07	1824.4	56.0878	3.07
ρ_{RC} (kg/m^3)	1042.07	1.34	0.13	1029.75	1.80	0.17	-	-	-
$F_{out,B2}$ (kg/h)	8.74	0.54	6.22	9.83	0.74	7.49	-	-	-
TS (rpm)	-	-	-	29.81	0.41	1.38	-	-	-
MCF (kN)	-	-	-	11.39	0.36	3.17	-	-	-
$F_{out,TP}$ (kg/h)	9.27	0.41	4.44	9.06	0.45	5.02	-	-	-

For the roll gap measurement, we used the ribbon thickness RSD, shown in C3, instead of the Q shown in Table C2. The variance in the ribbon density is high because one side is thicker than the other. The RSD is between 6% and 10%, depending on the NOC.

Table C3. RSD for ribbon thickness

Sample	NOC (mm)	STD (mm)	RSD (%)
1	1.8	0.14	7.86
2	1.62	0.13	7.86
3	1.565	0.15	9.49
4	1.54	0.17	11.02
5	1.735	0.15	8.56
6	1.56	0.16	9.97
7	1.715	0.13	7.83
8	1.525	0.13	8.81
9	1.755	0.09	5.24
10	1.54	0.11	7.35
11	1.725	0.12	6.97
12	1.545	0.12	7.78
13	1.7	0.16	9.15
14	1.525	0.15	9.74
15	1.745	0.06	3.65
16	1.52	0.16	10.23
17	1.53	0.16	10.17
18	1.53	0.13	8.32
19	1.765	0.13	7.61
20	1.725	0.13	7.79
21	1.76	0.13	7.23
22	1.595	0.13	8.42
23	1.595	0.13	8.42
24	1.75	0.10	5.66
25	1.615	0.11	6.57
26	1.555	0.09	5.91

APPENDIX D. CHAPTER 5 ADDITIONAL INFORMATION

APPENDIX D1. Particle size distribution (PSD) correction

Assuming ellipsoid shape it is easy to imagine that the observed minor axis size ($D_{2,obs}$) is always equal to the real minor axis size (D_2). Therefore, the major axis length needs to be calculated as a function of the position of the ellipsoid. The impact of random orientation on observed major axis length is illustrated in Figure D.1.1. Equations for particle size distribution (PSD) correction were published by: Szilagyi, B.; Nagy, Zoltan K. (2018) *"Aspect ratio distribution and chord length distribution driven modeling of crystallization of two-dimensional crystals for real-time model-based applications"* Crystal Growth and Design.
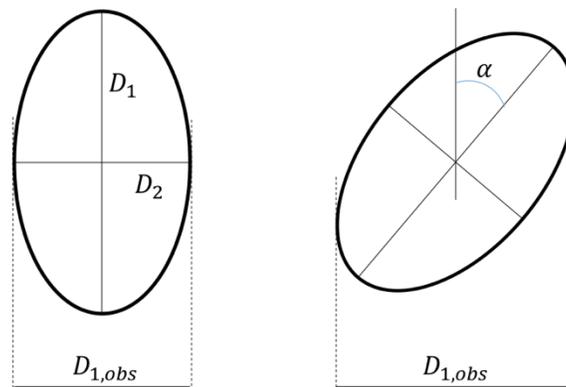

Figure D.1.1. Impact of crystal orientation on the observable major axis length, where ellipse position (α)

Using the trigonometric relationships, $D_{1,obs}$ can be expressed as a function of ellipse position (α) as:

$$D_{1,obs}(\alpha) = \sqrt{D_1^2 sin^2(\alpha) + D_2^2 cos^2(\alpha)} \qquad (D.1.1)$$

$$D_{2,obs}(\alpha) = D_2 \qquad (D.1.2)$$

Assuming random particle orientation i.e. that α is uniformly distributed random variable within the [0, $\pi/2$] interval, the expected value of $D_{2,obs}$ remains unchanged and $D_{1,obs}$ is given by the integral:

$$D_{1,obs}(\alpha) = \frac{1}{\pi/2}\int_0^{\pi/2}\sqrt{D_1^2 sin^2(\alpha) + D_2^2 cos^2(\alpha)}d\alpha \tag{D.1.3}$$

$$D_{2,obs}(\alpha) = D_2 \tag{D.1.4}$$

The particle shape is often characterized by the aspect ratio, expressing the ratio of major to minor axis length, defined as:

$$AR_{obs} = \frac{D_{2,obs}}{D_{1,obs}} = \frac{D_2}{\frac{1}{\pi/2}\int_0^{\pi/2}\sqrt{D_1^2 sin^2(\alpha) + D_2^2 cos^2(\alpha)}d\alpha} \tag{D.1.3}$$

The observed diameter, which is detected with the Eyecon, can be written as:

$$d_{obs} = \frac{D_{1,obs} + D_{2,obs}}{2} \tag{D.1.6}$$

Similarly, the real diameter is:

$$d_{real} = \frac{D_1 + D_2}{2} \tag{D.1.7}$$

The ratio between the observed and real diameter is:

$$\frac{d_{obs}}{d_{real}} = \frac{D_{1,obs} + D_{2,obs}}{D_1 + D_2} = \frac{\frac{1}{\pi/2}\int_0^{\pi/2}\sqrt{D_1^2 sin^2(\alpha) + D_2^2 cos^2(\alpha)}d\alpha + D_2}{D_1 + D_2}$$

$$= \frac{AR_{real}}{AR_{obs}}\frac{(AR_{obs} + 1)}{(AR_{real} + 1)} \tag{D.1.8}$$

Where the real aspect ratio of crystal can be defined similarly to the Eq. (D.1.5) as:

$$AR_{real} = \frac{D_2}{D_1} \qquad (D.1.9)$$

The direct outcome of Eq.(D.1.9) is that the mean crystal diameter underestimation due to the random orientations can be calculated if a relationship can be found between AR_{real} and AR_{obs} (which is the measured quantity). For this, Eq. (D.1.5) should be reformulated:

APPENDIX D2. Eyecon® images according to pressure

Images of the PSD, taken by the Eyecon, at different pressures.

25 bar
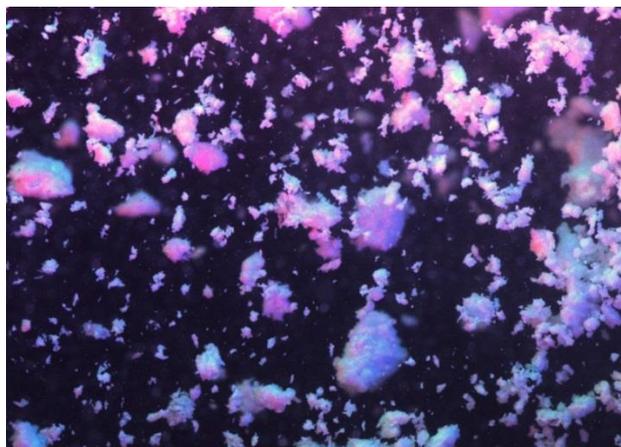
35 bar
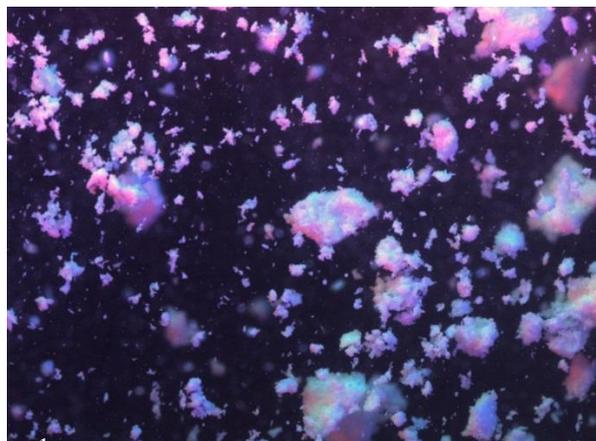
45 bar
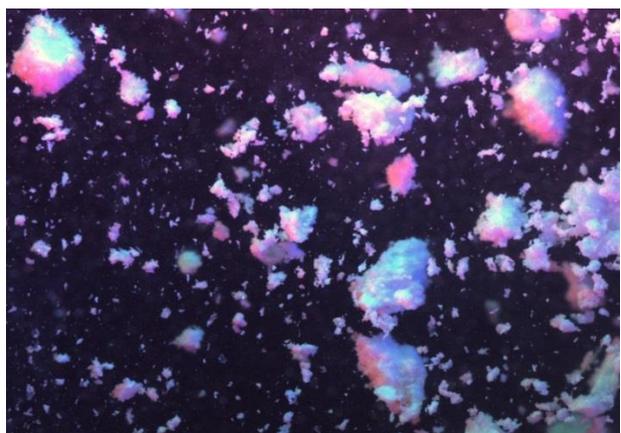
60bar
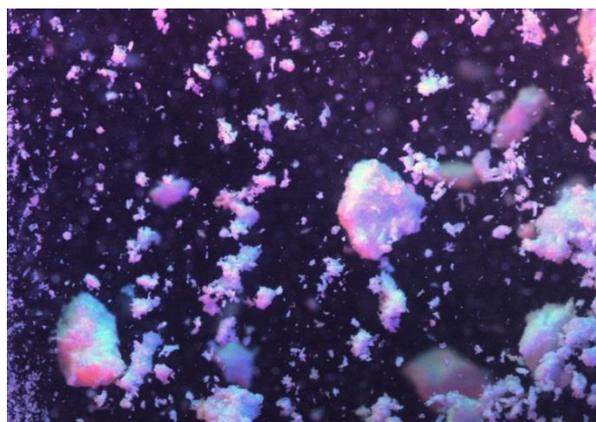
Figure D.2.1. Granules images according to the pressure.

APPENDIX D3. PSD based on particle count for the Eyecon® Camera

Cumulative and volume distribution for particle count given by Eyecon® offline and real-time. The particle count gives a similar trend. This image also shows the importance of selecting the right illuminations parameters.
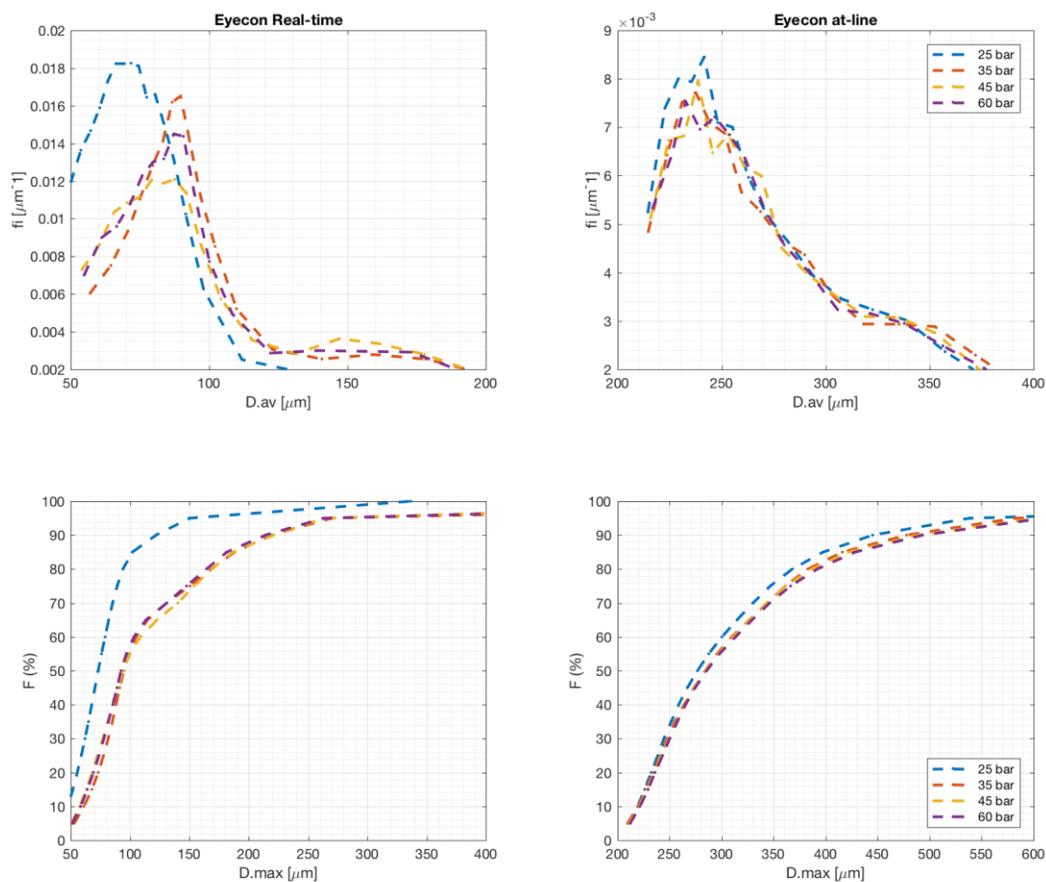

Figure D.3.1. Count-based cumulative distribution and volume-based cumulative distribution

VITA

Mariana Moreno was born in Brownsville, Texas and raised in Mexico City, Mexico. She attended Instituto Tecnológico y de Estudios Superiores de Monterrey, Campus Monterrey for her bachelor's degree in Chemical Engineering with a minor in Environmental Engineering. During her junior year, Mariana was a visiting scholar at Purdue University in the summer of 2010. Then, in the summer of 2011 and the summer of 2012, she worked for R&D, Pfizer at Groton, CT.

In Fall 2013, Mariana joined Purdue University to pursue the doctoral degree in Chemical Engineering. She focused on the process monitoring and control of continuous pharmaceutical processes. Her research was supervised by Prof. Zoltan K. Nagy and Prof. Gintaras V. Reklaitis. Mariana was part of the Engineering Research Center for Structured Organic Particulate Systems (ERC-SOPS). Furthermore, in the summer of 2016 and while in pursuit of her doctoral degree, Mariana took the opportunity to do an internship in ExxonMobil for the Applications Engineering department. During her Ph.D., Mariana was the recipient of the 2013-2014 teaching assistant award and the 2014 School of Chemical Engineering Eastman travel award.

In addition to her research work, Mariana took leadership positions. She served as Secretary of the Student Leadership Committee of ERC-SOPS and as Secretary of the Latino Graduate Student Organization. Mariana also served as co-curricular chair in the Purdue Chemical Engineering Graduate Student Organization. She was a graduate woman ambassador of the Davidson School of Chemical Engineering and in the organization committee of the 2017 Women in Chemical Engineering Seminar. Mariana will graduate with her Ph.D. in August 2018 and will join ExxonMobil in Houston, Texas.

PUBLICATIONS

Q. Su, **M. Moreno**, A. Giridhar, G. V Reklaitis, and Z. K. Nagy, "A Systematic Framework for Process Control Design and Risk Analysis in Continuous Pharmaceutical Solid-Dosage Manufacturing," *J. Pharm. Innov.*, Oct. 2017.

J. Liu, Q. Su, **M. Moreno**, C. Laird, Z. Nagy, and G. Reklaitis, "Robust state estimation of feeding–blending systems in continuous pharmaceutical manufacturing," *Chem. Eng. Res. Des.*, vol. 134, pp. 140–153, Jun. 2018.

S. Ganesh, **M. Moreno**, M. Gonzalez, J. Liu, M. Gonzalez, Z. K. Nagy, and G. V. Reklaitis, "Sensor Network for Continuous Tablet Manufacturing," *Proc. 13th Int. Symp. Process Syst. Eng. PSE 2018*, 2018.

S. Qinglin, **M. Moreno**, S. Ganesh, G. V. Reklaitis, and Z. K. Nagy, "Resilience and risk analysis of fault-tolerant process control design in continuous pharmaceutical manufacturing," *MKO Process Saf. Int. Symp.*, 2018.

M. Moreno, J. Liu, Q. Su, C. Leach, A. Giridhar, N. Yazdanpanah, T. O'Connor, Z. K. Nagy, and G. V. Reklaitis, "Steady-state Data Reconciliation Framework for a Direct Continuous Tableting Line," *J. Pharm. Innov.*, *accepted*, 2018.

M. Moreno, J. Li, B. Rentz, S. Panikar, Q. Su, G. Callegari, N. Yazdanpanah, S. Lee, F. Muzzio, Z. K. Nagy, and G. V. Reklaitis, "Case Study: Steady-State-Data Reconciliation for a Continuous Tableting Line via Direct Compression," *in prep.*, 2018.

M. Moreno, S. Ganesh, Y. Shah, Q. Su, M. Gonzalez, Z. K. Nagy, and G. V. Reklaitis, "Nonlinear Steady-State Data Reconciliation for Continuous Tableting Processes,", *submitted*, 2018.

M. Moreno, B. Rentz, B. Szilagyi., Z. K. Nagy, and G. V. Reklaitis, "Real-Time Monitoring of Particle-Size in a Continuous Dry Granulation-based Process," *in prep.*, 2018.