Data-Driven Dynamic Models for Nonlinear Process Optimization and Control

A dissertation submitted by

Zhenyu Wang

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Chemical Engineering

TUFTS UNIVERSITY

February 2018

Advisor: Prof. Christos Georgakis

Abstract

Mathematical models play an essential role for the purposes of process optimization and control. There are two major information sources for the development of these models: the knowledge of the process inner workings and the input-output data set. The model estimated using the detailed knowledge are called the knowledge-driven model. However, the inner workings of many industrial processes are not always fully understood to enable the development of accurate knowledge-driven models.

In such a situation, the data-driven model, relying on the input-output data, is an attractive alternative. Among varieties of data-driven modeling approaches, the Design of Dynamic Experiments (DoDE), a generalization of the traditional Design of Experiments (DoE) approach, has been demonstrated as an effective modeling methodology for optimizing nonlinear processes. When time-resolved data are obtainable during the experiments, developing a Dynamic Response Surface Methodology (DRSM) model is more favorable. As the estimated DRSM model with time-varying parameters captures the process dynamics, it has the potential to be applied for not only the process optimization but also the process control purposes.

The main goal of this research work is to further advance and improve the two data-driven methodologies, the DoDE and the DRSM, to model, optimize and control nonlinear processes. We first proposed ways to incorporate prior process knowledge to improve the design of the input domain, in which the time-varying input of the DoDE experiments are selected. Improved process performance has been achieved in the refined input domain. In addition, as process optimization is usually under budgetary and time constraint, we developed an evolutionary DoDE approach to

optimize the processes in a timely manner. The size of the initial set of experiments has been dramatically reduced while the achieved optimal process performance is similar to the one obtained using the original DoDE approach.

To extend the applicability of the original DRSM approach (DRSM-1) to deal with processes with various and infinite time duration, we proposed a new DRSM approach (DRSM-2). The novelty of the DRSM-2 rests on a nonlinear transformation of time, the independent variable. Comparing to the DRSM-1, the new method has the following advantages. It is capable of

1) Modeling both continuous as well as batch processes, handling semi-infinite as easily as finite time domains
2) Using data that are not equidistant in time
3) Using data segments that are of varied durations due to possible strong nonlinearities in dynamics

We also developed a single model approach, using the DRSM model, for both process optimization and control purposes. The proposed method reduces the experimental effort comparing to the current practices which use separate models for process optimization and control purposes, respectively. When the number of measurements is small, the proposed approach provides better control performance compared to the performance achieved using a model estimated with Pseudo Random Binary Signal (PRBS) data.

Acknowledgement

I always consider myself fortunate indeed to pursue my Ph.D in the Systems Research Institute (SRI) at Tufts University and under the guidance of Prof. Christos Georgakis. The SRI is such a great platform for me to conduct the research with industrial impact. It has been a wonderful learning experience with dedicated and talented professionals from both academia and industry who are passionate about leveraging data-driven modeling methodologies for better process performance. I owe deep gratitude to my advisor, professors, sponsors, colleagues, and my family. Without their support, this thesis would not have been possible.

First of all, I am particularly grateful to my advisor, Prof. Georgakis, for his continuous support and mentorship over the past five years. He has helped me lay firm foundation in process systems engineering, by tutoring me himself, helping me choose courses at and outside Tufts. He is always patient and provide me the freedom to work independently. He always encourages me to keep an eye on the industrial side, as the influential innovation usually inspired by the industrial demand and challenge. Prof. Georgakis's expertise, vision and passion towards the process systems engineering, and his willingness to cultivate the younger generation, is a constant inspiration to me. I do appreciate that he provides such a great role model for me. Under the guidance of Prof. Georgakis, I have improved myself significantly as an independent innovator and engineer.

Special thanks should go to the members of my thesis committee, Prof. Eric Miller, Prof. Manolis Tzanakakis and Dr. David Schmidt. They have followed and supported my thesis work and given their feedback along the way. Their suggestions and comments make my research more rigorous.

Moreover, I would like to acknowledge our industrial sponsors (in alphabetic order), Biogen, ExxonMobil, Merck, Pfizer, Sunovion and Teva, for their financial support and the feedback for my research. Some of their interests and demands, such as optimizing the process performance with reduced experimental effort, have motivated my research projects. I would like to also express my gratitude to several colleagues from our sponsors, Dr. Ke Wang at Pfizer and Dr. David Schmidt, Dr. Kiran Sheth and Dr. Cara Touretzky at ExxonMobil. Dr. Wang has provided insightful comments on the statistical methods examining the quality of the obtained data-driven models. Dr. Schmidt, Dr. Sheth and Dr. Touretzky have collaborated with me on the research project of modeling the polymerization process using the Dynamic Response Surface Methodology.

In addition, I would like to thank my internship supervisor, Dr. Lilong Huang at Biogen. It has been a precious experience to examine the data-driven modeling methodologies using real industrial data in bio-pharmaceutical industry. I would also like to thank my internship supervisors, Dr. Ali Esmaili and Dr. Ye Chen at Air Products and Chemicals for the opportunity to work closely with the engineers on two challenging projects. These intern experiences facilitate my preparation for the career as an engineer in process systems engineering.

Last but not least, I would like to thank my family. I am deeply indebted to my wife, Yan Huo, for her unreserved support and strong confidence in me. She is also the first and corresponding “author” of my two most important “papers”, my son Mason and my daughter Claire. I also owe my sincere gratitude to my beloved parents and parents-in-law, who give me their selfless love.

Table of Contents

List of Tables

List of Figures

Nomenclature

Roman Symbols

C	RSM or DRSM model class
d	Input time delay
K_m	Number of measurements in each experiment
M	Number of experiments
n	Number of factors
p	Number of parametric functions
R	Order of Shifted Legendre polynomial
t_c	Time constant
u	Input variable
x	Factor or dynamic sub-factor
y	Output variable
z	State variable

Greek Symbols

α	Input nonlinearities coefficients of H-W model
β	Parameter in RSM model
$\beta(\tau)$	Parametric function in DRSM model
γ	Parameter in DRSM model
κ	Output nonlinearities coefficients of H-W model
σ	Standard deviation
τ	Dimensionless time

ϑ	Exponentially transformed time

Abbreviations

ARX	Auto-Regressive with eXogenous input
ARMAX	Auto-Regressive and Moving Average with eXogenous input
BIC	Bayesian Information Criterion
DE	Differential Evolutions
DoDE	Design of Dynamic Experiments
DoE	Design of Experiments
DoF	Degree of Freedom
DRSM	Dynamic Response Surface Methodology
DRSM-1	The first version of DRSM
DRSM-2	The second version of DRSM
EA	Evolutionary Algorithms
GA	Genetic Algorithm
GBN	Generalized Binary Noise
H-W	Hammerstien-Wiener Model
LoF	Lack-of-Fit
LPV	Linear Parameter Varying
MBO	Model-Based Optimization
MFI	Melt-Flow Index
MOESP	Multivariable Output Error State Space Model Identification
MPC	Model Predictive Control
N4SID	Numerical Algorithms for Subspace State Space System Identification

NMPC	Nonlinear Model Predictive Control
ODE	Ordinary Differential Equation
PCA	Principal Component Analysis
PEM	Prediction Error Method
PLS	Partial Least Squares
PRBS	Pseudo Random Binary Signal
PSO	Particle Swarm Optimizations
RSM	Response Surface Model
SMF	Secondary Metabolites Fermentation
SPC	Statistical Process Control
TF	Transfer Function

Data-Driven Dynamic Models for Nonlinear Process Optimization and Control

1. Introduction

Mathematical models are essential to the success of process optimizations and control. Thanks to the progress in the computational methods and data collecting techniques, the application of mathematical models has become increasingly popular and useful. As shown in Figure 1-1, the number of papers with "modeling" in the title has increased exponentially over the past forty years, which reflects the increasing academic interest [1].

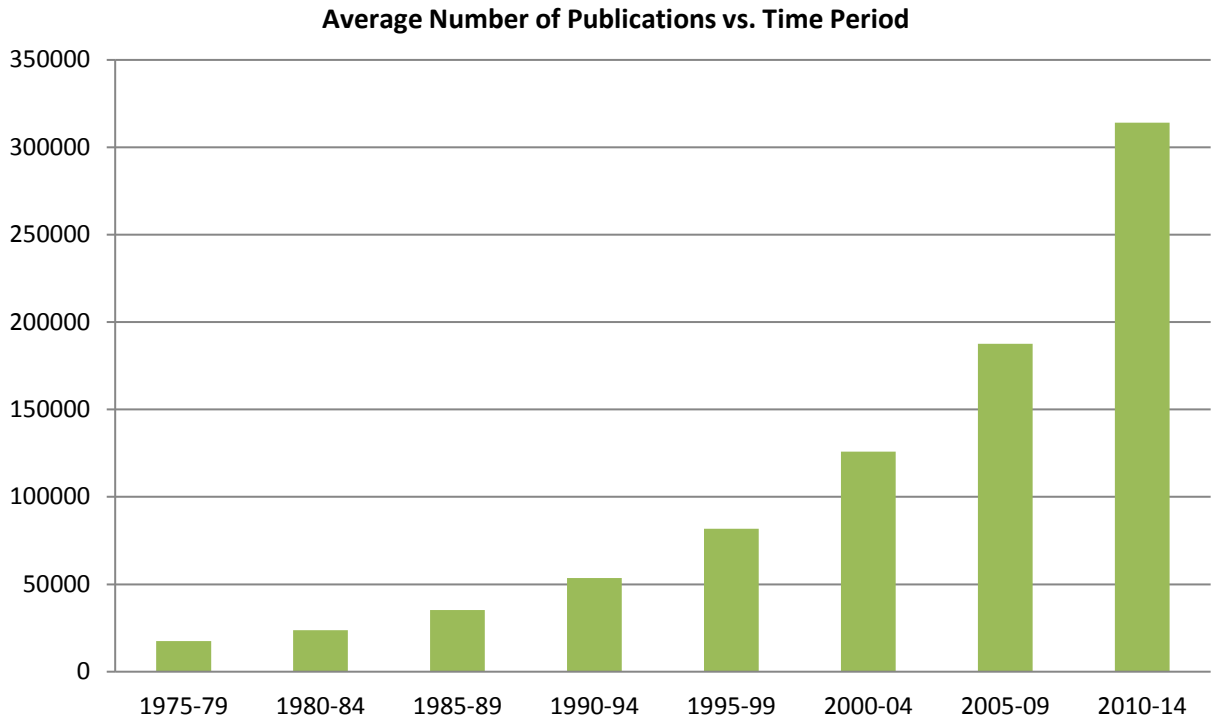

Figure 1-1: Number of papers with "modelling" in the title, 1975-2014

Based on the information used for their development, mathematical models fall into two categories: knowledge-driven and the data-driven models. The knowledge-driven model is developed based on detailed knowledge of the process inner workings, such as kinetics and mass and hear transfer

rates as well as the use of mass and energy balances [2-5]. Once the model structured is determined based on first principles, experimental or manufacturing data are used to estimate the values of the parameters in the model. However, for some complicated chemical and bio-pharmaceutical processes, such as mammalian cell culture in batch reactor [6-8], or polymerization processes [9, 10], some or all of the metabolic paths or reaction mechanisms are not well understood. Thus, the knowledge-driven models with imperfect understanding may not provide the expected accuracy. Sometimes, the lack of knowledge might be substantial to even make it impossible to build a knowledge-driven model. Even with a knowledge-driven model at hand, optimizing a very complex process using such a detailed model is computational costly, which makes real-time optimal control a challenge.

The data-driven models are developed using input-output data collected from a finite number of experiments or manufacturing activities. Though some empirical process knowledge may help to improve the quality of designed experiments [11], the detailed knowledge about the process inner workings is not required to develop an accurate data-driven model. Therefore, it is an attractive alternative to the knowledge-driven model. In addition, the simple structure of the data-driven model makes the optimization for the process much faster.

Several types of data-driven models have been proposed for modeling processes with different input and output relationships, including linear or nonlinear, static or dynamic ones. Multivariate analysis methodologies, such as Principle Component Analysis (PCA)[12, 13] and Partial Least Squares (PLS) [14, 15], are used to approximate linear and static relationship between the input and output variables. Time-series analysis models [16, 17] and transfer function models [18] are employed to represent the linear dynamics of the processes. Dynamic PLS modeling approaches

combining the PLS and time-series analysis approaches, have also been reported [19, 20]. The Response Surface Model (RSM) [21, 22] resulted from the Design of Experiments (DoE) [23, 24] is applied for estimating static and nonlinear input-output relationships.

Recently, Prof. Georgakis' group generalized the DoE methodology to incorporate time-variant inputs in the framework of Design of Dynamic Experiments (DoDE) [25, 26]. The key generalization allows for the time-varying input variables, such as the feeding flow of sugar source for batch fermentation processes. The DoDE methodology has also been successfully used in optimizing some additional batch processes [27-29] and it has been experimentally verified in an industrial hydrogenation reaction [30].

For the purposes of process optimization, static model, either linear or nonlinear, is sufficient, though optimization based on the models account for dynamic behavior might provide better process optimum. However, for the purpose of process control, a dynamic model is required. Besides the aforementioned dynamic models, Kelebanov and Gerogakis proposed the Dynamic Response Surface Methodology (DRSM) [31], a dynamic generalization of the RSM model for modeling batch processes with fixed duration.

The main goal of this research work reported in this thesis is to further advance and improve the two data-driven methodologies, the DoDE and the DRSM, to model, optimize and control nonlinear processes. We first propose ways to incorporate prior process knowledge to improve the design of the input domain, in which the time-varying input of the DoDE experiments are selected. Improved process performance has been achieved in the refined input domain [11]. In addition, as process optimization is usually under budgetary and time constraint, we propose an evolutionary DoDE approach to optimize the processes in a timely manner. The size of the initial set of

experiments has been dramatically decreased while the achieved optimal process performance is similar to the one obtained using the original DoDE approach.

To extend the applicability of the original DRSM approach (DRSM-1) to dealing with processes with infinite time duration, we proposed a new DRSM approach (DRSM-2). The novelty of the DRSM-2 rests on a nonlinear transformation of time, the independent variable. Comparing to the DRSM-1, the new method has the following advantages. It is capable of

1. Modeling both continuous as well as batch processes, handling semi-infinite as easily as finite time domains
2. Using data that are not equidistant in time
3. Using data segments that are of varied durations due to possible strong nonlinearities in dynamics

We also describe a single model approach, using the DRSM model, for both the process optimization and control purposes [32]. The optimal operating conditions are first be determined using the DRSM model. Then the time-series model, either a linear or a nonlinear one is identified in the vicinity of the optimal trajectory by sampling the DRSM model. The proposed method reduces the experimental effort required in comparison to the current practices which develop separate models for process optimization and control purposes, respectively. When the number of measurements is small, it provides better control performance of a model predictive controller compared to the performance achieved through a model estimated using Pseudo Random Binary Signal (PRBS) data [33].

The thesis is organized as follows. We first introduce the prior work upon which this research work builds on in the chapter of back ground information. Following that, we discuss the two

methodological improvements on the DoDE approach. In the chapter of a new DRSM methodology, we introduce the novel DRSM-2 approach. The proposed framework of using a single DRSM model for process optimization and control purposes is discussed in Chapter 5. The new methodologies discussed in each chapter are examined using two (or several) *in silico* nonlinear processes, including batch and continuous processes. The further work of each new methodologies are discussed in Chapter 6.

2. Background Information

2.1. Nonlinear Process of Interest

The aim of this research work is to develop data-driven modeling methodologies, based on the DoDE and DRSM approaches, for modeling, optimizing and controlling nonlinear processes. These nonlinear processes of interest include semi-batch processes and continuous process in transitions. The batch or semi-batch processes are widely adopted in pharmaceutical industries for producing antibiotics and antibodies. Fermentation and cell culture are two major type of semi-batch processes used in pharmaceutical industry. Due to the small production rates and complicated metabolic mechanisms, it is quite challenging to develop knowledge-driven model for such batch processes.

Another nonlinear process of interest is continuous processes in transition. In olefin polymerization industry, polymers of different specifications are made in the same single plant. To meet demands for different products, the steady-state operation of a plant should be changed frequently. Model-based Optimization (MBO) has been applied to determine the optimal input profile to complete the transition in shortest time and with minimum off-spec product [10]. However, the process inner workings are highly complicated and are not always been fully understood to enable the development of the knowledge-driven model to be used in MBO.

In the aforementioned cases, the data-driven methodologies are attractive alternatives. To examine the efficacy of the proposed approaches, we selected several representative simulations based on open literatures, including a semi-batch penicillin fermentation [34], a semi-batch cell culture

processes [35, 36] and a continuous propylene polymerization process [9]. The mathematical descriptions of those processes and the values of the key parameters are given in Appendix A**.**

2.2. Knowledge-Driven Model and Model-Based Optimization

Knowledge-driven models are developed based on the understanding of inner workings of processes. Changes caused by reaction kinetics as well as mass and energy transfer are accounted for by energy, mass, or momentum, conservation laws and are described by differential equations. Many knowledge-driven models have been successfully applied for modeling of continuous and batch processes. If the process is thoroughly understood and the model is properly developed, a knowledge-driven model will provide accurate predictions in the entire input and output spaces. However, this is not always the case as the process might be too complicated to be fully understood. Even if the development of knowledge-driven model is feasible, the optimization requires substantial computational effort [37], the larger the process under consideration is. In general, the optimization problem, consisting of a set of algebraic and differential equations with constraints on the states and the decision variables, is formulated as follows,

$$\begin{aligned} &\min_{\boldsymbol{u}} J(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{u}, t) \\ s.t. &\frac{d\boldsymbol{z}}{dt} = f(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{u}, t); \\ &g(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{u}, t) \leq \boldsymbol{0} \\ &\boldsymbol{u^L} \leq \boldsymbol{u} \leq \boldsymbol{u}^U;\ \boldsymbol{z}^L \leq \boldsymbol{z} \leq \boldsymbol{z}^U \end{aligned} \tag{2-1}$$

Here, $\boldsymbol{z}$ represent the state variables of the system. $\boldsymbol{p}$ are the parameters of the model; and $\boldsymbol{u}$ are the decision variable with respect to which we wish to maximize (or minimize) the system's

performance, quantified by the index *J*. For example, if we aim to maximize the total weight of product at the end of the batch, t_b, the performance index is defined as

$$J[z(t_b)] = -z_1\,(t_b) \times z_2(t_b) \tag{2-2}$$

where $z_1(t_b)$ and $z_2(t_b)$ are the volume of the liquid phase and the concentration of the product at the end of the batch, respectively. The function $f(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{u}, t)$ represents the knowledge-driven model, which is usually a set of ordinary differential equations (ODEs). $g(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{u}, t) \leq 0$ are the set of inequality constraints as function of $\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{u}$, and t.

Among the different ways such a model-based optimization can be performed, we choose to follow the simultaneous approach, advocated by Professor Biegler's group [38-40]. In such an approach, the interval $(0, t_b]$ is divided into a number of finite elements and inside each element, the method of orthogonal collocations [41] is used to convert the set of ODEs into a large set of algebraic equations. Then, an optimization algorithm, such as sequential quadratic programming (SQP) [42] or interior point method [43], is used to calculate the optimum. The original optimization problem, given in eq. (2-1), is rewritten as:

$$\begin{aligned}
&\min_{\boldsymbol{u}_{j,k}, \Delta\zeta_j} J\left(\boldsymbol{z}_{j,k}, \boldsymbol{p}, \boldsymbol{u}_{j,k}, \Delta\zeta_j\right) \\
s.t. &\sum_{k=0,l}^{K} \boldsymbol{z}_{j,k}\phi_{j,l}(\tau_k) - \Delta\zeta_j f\left(\boldsymbol{z}_{j,k}, \boldsymbol{u}_{j,k}\right) = 0; \\
&\boldsymbol{z}_{j+1,0} - \boldsymbol{z}_{j,k} = 0; \\
&\boldsymbol{z}_{1,0} - \boldsymbol{z}_{j,K} = 0; \\
&g(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{u}, t) \leq 0;
\end{aligned} \tag{2-3}$$

$$\boldsymbol{u}_{j,k}^{L} \leq \boldsymbol{u}_{j,k} \leq \boldsymbol{u}_{j,k}^{U};\ \boldsymbol{z}_{j,k}^{L} \leq \boldsymbol{z}_{j,k} \leq \boldsymbol{z}_{j,k}^{U}$$

where $\boldsymbol{z}_{j,k}$ and $\boldsymbol{u}_{j,k}$ represent the state and input variables at k^{th} collocation point in j^{th} finite element and $\Delta\zeta_j$ the length (in time) of the j^{th} finite element. The first constraint forces the approximated state profiles equals to the simulated profiles. The second constraint imposed the continuity to state variables, but not to input variables. The third constraint is on the initial conditions and the inequality constraints confine the variables within the space defined by function $g(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{u}, t)$ and the lower and upper bounds. It is clear that the number of variables for the optimization problem is roughly $N \times K$ times larger than the original problem, if N finite elements and K collocation points are used. This cost lots of computational effort to solve the large-scale computational problem.

Other optimization algorithms for optimizing fed-batch biopharmaceutical processes, such as Evolutionary Algorithms (EA) [44], Differential Evolutions (DE) [45] [46] and Particle Swarm Optimizations (PSO) [47], are reviewed by [48]. A complete comparison between these bio-inspired algorithms and the simultaneous approach has not been presented yet. However, Riascos and Pinto [49] reported that the simultaneous approach gave a slightly better result than EA in optimizing product weight of the penicillin fermentation process, examined in detail in this thesis. In our study of hybridoma cell culture [35], reported later in Chapter 3, we obtained a little higher product weight using the simultaneous approach than that calculated by a Genetic Algorithm(GA) [50], one of the EA approaches.

2.3. Design of Dynamic Experiments and Process Optimization

Design of Dynamic Experiments (DoDE), a generalization of the classic Design of Experiments (DoE) [23, 51, 52] approach, allows for time-varying input. In the DoDE framework, we first define traditional time-invariant DoE factors, such as the experiment duration, amount of the reactants loaded in the reactor at the beginning of each experiment, etc. For example, we here define a traditional DoE factor, w, that relates to the batch duration t_b, and assume that we are unsure about its optimal value.

$$t_b = t_{b,r} + \Delta t_{b,r} w \text{ with } -1 \leq w \leq +1 \tag{2-4}$$

In the above equation, the batch duration varies between $t_{b,r} - \Delta t_{b,r}$ and $t_{b,r} + \Delta t_{b,r}$ with $t_{b,r}$ being the reference value.

The time-varying input variable, $u(\tau)$ is given by

$$u(\tau) = u_0(\tau) + \Delta u(\tau) w(\tau) \text{ with } -1 \leq w(\tau) \leq +1 \tag{2-5}$$

Where $u_0(\tau)$ is the reference input profile while the $\Delta u(\tau)$ is the deviation from the reference. $u(\tau)$ varies in the input domain from $u_0(\tau) - \Delta u(\tau)$ to $u_0(\tau) + \Delta u(\tau)$. The time dependency within each experiment is defined in terms of the dimensionless time, $\tau = t/t_b$ where t_b is the batch time for the corresponding experiment. Now let the coded time-varying factor, $w(\tau)$ be expanded in a series of shifted Legendre polynomials [53].

$$w(\tau) = x_1 P_0(\tau) + x_2 P_1(\tau) + x_3 P_2(\tau) \cdots \tag{2-6}$$

with

$$P_0(\tau) = 1;\ P_1(\tau) = 2\tau - 1;\ P_2(\tau) = 6\tau^2 - 6\tau + 1; \cdots \quad (2\text{-}7)$$

Here the coefficients x_1, x_2, and x_3 are the dynamic sub-factors, which will be varied to define the coded time-varying dynamic factor $w(\tau)$. $P_{i-1}(\tau)$ is the i^{th} shifted Legendre polynomial, an $(i-1)^{th}$ order polynomial in the dimensionless time τ. Here we use 3 dynamic sub-factors and the first 3 shifted Legendre polynomials as an example. A different number of shifted Legendre polynomials, or dynamic sub-factors, parameterizing $w(\tau)$ might be selected by the process practitioners based on their needs. In general, the more dynamic sub-factors, the more complex the input profile will be allowed to be and the obtained process optimum might be better. However, the number of experiments increases with the number of dynamic sub-factors.

A set of DoDE experiments will be selected by varying the traditional factors and the dynamic sub-factors, w, x_1, x_2, and x_3 in the input domain. To minimize the experimental effort and maximize the information obtained through those experiments, optimal design criteria [23], such as D-optimal design, is usually applied.

After collecting the data from the designed experiments, a RSM model in the following form is estimated via linear regression. Here we use a quadratic RSM model with n factors as an example.

$$y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \sum_{i<j}^{n} \beta_{ij} x_i x_j + \sum_{i=1}^{n} \beta_{ii} x_i^2 \quad (2\text{-}8)$$

Where y is the output variable of interest and the x's are the traditional factors and the dynamic sub-factors. The model parameters, β's are estimated via linear regression. For most processes, the quadratic RSM is sufficiently complex to provide an accurate model.

Optimizing the process performance using the estimated RSM model is similar to the optimization of the process using the knowledge-driven model. The only difference is to substitute the knowledge-driven model in the eq.(2-1) with the RSM model given in eq.(2-8).

2.4. Dynamic Response Surface Methodology for Process Modeling

The original DRSM [31] methodology (DRSM-1) was proposed to analyze the time-varying output variables of batch processes with fixed durations. The obtained DRSM model has a similar form to the RSM model, with the DRSM model parameters being functions of time. For example, a quadratic DRSM model with n factors is given as follows:

$$y(\tau) = \beta_0(\tau) + \sum_{i=1}^{n} \beta_i(\tau) x_i + \sum_{i=1}^{n} \sum_{i<j}^{n} \beta_{ij}(\tau) x_i x_j + \sum_{i=1}^{n} \beta_{ii}(\tau) x_i^2 \qquad (2\text{-}9)$$

Where $y(\tau)$ is the output variable of interest and x_i's are the traditional DoE factors and (or) dynamic sub-factors parameterizing the input profile of each DoDE experiment. The time-varying input profile, $u(\tau)$, is given by eq. (2-5).

Here $\beta_q(\tau)$ will be used to represents any of the time-varying parametric functions, $\beta_0(\tau), \beta_i(\tau), \beta_{ij}(\tau), \beta_{ii}(\tau)$, in eq. (2-9). We select the first $R+1$ shifted Legendre polynomials as the orthogonal basis for the parameterization of each $\beta_q(\tau)$ function given by

$$\beta_q(\tau) = \gamma_{q,1} P_0(\tau) + \gamma_{q,2} P_1(\tau) + \cdots + \gamma_{q,R+1} P_R(\tau) \qquad (2\text{-}10)$$

With $q = 0, i, ij, ii$ for $i = 1, 2, \cdots, n$ and $j > i$. The model parameters, γ's, are estimated via linear regression. The fixed value of t_b limits the application of the DRSM-1 approach to the batch

process with fixed operating duration. This also prevents the DRSM-1 approach from using historical data, which inevitably, will be over different time durations. Such limitations motivate the DRSM-2 methodology presented in Chapter 4.

2.5. Recursive Models and Model Predictive Control

With the DRSM model capturing the process dynamics, one can potentially use it for both optimization and control. However, the DRSM model cannot be directly applied for control purpose. The Model Predictive Control (MPC) relies on the recursive models and this will be discussed in this sub-section. We will present our approach of using a transformed form of the DRSM model for control purposes in Chapter 5. This is a recursive model for control purposes which is estimated from the DRSM model at hand.

2.5.1. Identification of Recursive Models

In this sub-section, we introduce the recursive models and ways to identify these models. There are two classes of recursive models, linear and nonlinear models, approximating the linear and nonlinear dynamic input-output relationships. The state-space model and the time-series model (including ARX or ARMAX type of models) are two major forms of the recursive model defined in time domain. The mathematical expression of an Auto-Regressive with eXogenous input (ARX) model, is given by

$$y_{k+1} = a_1 y_k + a_2 y_{k-1} + \cdots + a_n y_{k-n+1} + b_1 u_k + b_2 u_{k-1} + \cdots + b_m u_{k-m} \qquad (2\text{-}11)$$

This input-output relationship indicates that the current output value is the linear combination of previous *n* outputs and previous *m* inputs. A generalized form of ARX model is the ARMAX model, which also including the effect of moving average as shown in the following equation.

$$y_{k+1} = \sum_{i=1}^{n} a_i y_{k-i+1} + \sum_{i=1}^{m} b_j u_{k-i+1} + \sum_{i=1}^{q} c_i e_{k-i+1} \tag{2-12}$$

e_k is a nonzero value at time instant k, representing the moving average of the process being modeled. Both ARX and ARMAX model can be rearranged into canonical state-space form [33], given as follows.

$$\boldsymbol{z}_{k+1} = \mathbf{A}\boldsymbol{z}_k + \mathbf{B}\boldsymbol{u}_k \tag{2-13}$$

$$\boldsymbol{y}_k = \mathbf{C}\boldsymbol{z}_k \tag{2-14}$$

Where $\boldsymbol{z}_k$ is the state variable. When the $\mathbf{C}$ matrix is an identity matrix, the state-space model is equivalent to the ARX model.

Several recursive model structures representing the nonlinear dynamics of the batch processes have been proposed as well, including Linear Parameter-Varying (LPV) model and Hammerstein-Wiener (H-W) model. The LPV model [54] introduces the scheduling parameters varying with evolution of state variables in order to approximate bilinear dynamics. The H-W models, consisting of two static nonlinear blocks in the inputs and outputs and a dynamic linear block in between [55-57], has been applied for the modelling batch processes with linear kinetics and static nonlinear functions on the output, such as pH neutralization [58, 59] and distillation column [60]. The schematic representation of a Hammerstein-Weiner model is given as below.

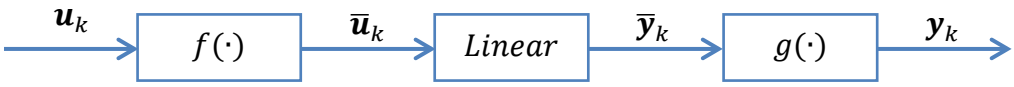

Figure 2-1: Hammerstein-Wiener model

$f(\cdot)$ and $g(\cdot)$ blocks represent the static input and output nonlinearities while the $Linear$ block as given in eq. (2-11) or eq. (2-13) & (2-14), represents the linear dynamics between the intermediate input $\bar{\boldsymbol{u}}_k$ and intermediate output $\bar{\boldsymbol{y}}_k$. $\boldsymbol{u}_k$ is the input exiting the process while $\boldsymbol{y}_k$ is the corresponding process output.

The aforementioned linear and nonlinear recursive models are identified via algorithms such as Prediction Error Method (PEM) [17], Multivariable Output Error State Space Model Identification (MOESP) [33], and Numerical Algorithms for Subspace System Identification (N4SID) [61]. These algorithms utilize the data collected from Pseudo Random Binary Signal (PRBS) or Generalized Binary Noise (GBN) [62] experiments in the vicinity of a pre-determined trajectory, possibly an optimal one. However, when the available measurements in a single batch are infrequent, the estimation of such a linear or nonlinear dynamic model of satisfactory accuracy is not feasible. In addition, to determine the optimal trajectory around where the recursive model is identified, a separate model may be estimated with the cost of additional experiments.

2.5.2. Model Predictive Control

Model Predictive Controller (MPC) has been applied widely in controlling multivariable and constrained processes effecting changes in set points and countering the effect disturbances [63-65]. In general, MPC calculates future control actions, so that it minimize the differences between the expected future values of the outputs and the corresponding set-points. The Kalman Filter [66,

67] is usually applied to update the estimation of the state variables after the present measurements have been collected. A schematic datagram of a MPC by Lima [68] is given in Figure 2-2.

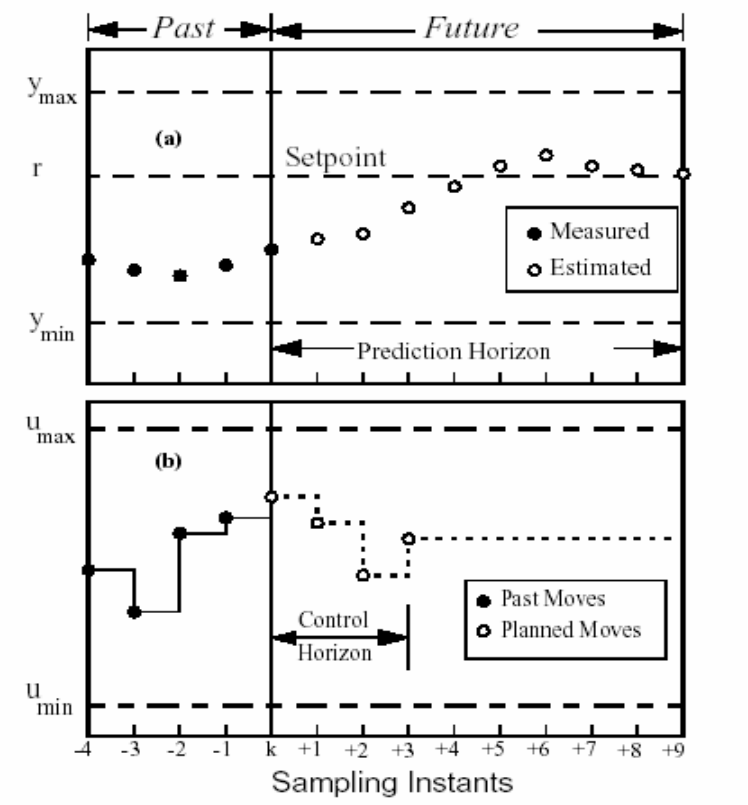

Figure 2-2: Constrained MPC strategy with prediction horizon = 9 and control horizon = 4

The mathematical formulation of a Single-Input-Single-Output (SISO) MPC is given by

$$\begin{aligned} &\min_{\Delta u_{i|k}} \left\{ q \sum_{i=k+1}^{k+N_p-1} \tilde{y}_{i|k}^2 + w \sum_{i=k+1}^{k+N_p-1} \delta u_{i|k}^2 \right\} \\ &s.t\ \delta u_{i|k} = \Delta u_{i|k} - u_{i-1|k};\ \tilde{y}_{i|k} = \Delta y_{sp,i} - \Delta \hat{y}_{i|k} \\ &\hat{\boldsymbol{z}}_{i+1|k} = \mathbf{A}\hat{\boldsymbol{z}}_{k|k} + \boldsymbol{b}^T \Delta u_{i|k};\ \ \Delta y_{i|k} = \boldsymbol{c}^T \hat{z}_{i|k}; \\ &u^L \le u_{0i} + \Delta u_{i|k} \le u^U \end{aligned} \tag{2-15}$$

Here $y_{sp,i}$ is the set point at time instant i. w is the control action weight, u^L and u^U are the lower and upper control limit. The subscript $i|k$ indicates the estimation of the variable value of future instance i while the estimation is done at instant k. N_p and N_u are the prediction and control horizons, respectively. The tuning parameters for MPC controller, are N_p, N_u, q, and w. Guidelines for tuning the MPC has been discussed in the publications [69-73]. Nonlinear MPC (NMPC) using nonlinear recursive models has also been reported [74-76], including the control strategy based on block oriented models [77-80].

3. Methodological Improvements to DoDE for Process Optimization

In this chapter, the new methodical improvements to the DoDE approaches have been discussed. Batch processes often characterized by small production rates resulting in economic constraints that do not afford the necessary time and effort to develop an accurate knowledge-driven model describing their inner workings in detail. To accommodate such a lack of detailed understanding, we recently introduced a new data-driven approach, the Design of Dynamic Experiments (DoDE) [25, 26] The DoDE defines a finite number of time-varying input profiles within an input domain. From the collected data at the end of each batch, it develops a data-driven Response Surface Methodology (RSM) model that enables the optimization of the process. The DoDE methodology has been successfully used in optimizing several batch processes [27-29].

Even though the input domain has a significant impact on the performance of the DoDE approach [81], discussion in the literature about how to improve the design domain is nonexistent. We here incorporate *a priori* process knowledge to aid the design of the input domain. We will show that the change in the design domain guided by prior process knowledge results in improved optimal process performances in two classes of representative biopharmaceutical processes examined in this chapter. In addition, because the model development and process optimization for batch or semi-batch processes are always under tight budgetary and time constraints, we consider how the number of initial experiment can be reduced and still obtain a process optimum close to the one obtained using the more numerous set of experiments in the original DoDE approach.

The proposed two improvements of the DoDE approach will be examined *in silico* against two classes of representative biopharmaceutical processes. This includes the multi-feed Hybridoma

cell culture[82] and a set of eight secondary metabolite fermentations (SMF). These SMF processes are defined by varying the four most significant kinetic parameters (μ_{max}, ρ_{max}, K_p , and K_{in}) of the B&R model over a wide range of values.

3.1. Ordinary DoDE Experiments as a Reference Case

In this section, we design a set of DoDE experiments, without incorporating the process knowledge. This resulting design domain is denoted as Domain A. The obtained process optimum will serve as a reference case. In section 3.2 where we incorporate the prior additional process knowledge, we will define a different design domain, Domain B, by arguing that the co-reactants flow rates should maintain a non-zero value at the end of the batch. We will compare the obtained optimal process performances within Domains A and B in the results section and demonstrate the benefits of incorporating the process knowledge that motivates the use of Domain B,

We first define the reference input profile, $u_0(t)$. As the total volume of a batch reactor is usually fixed, the selected input profile has to satisfy the total volume constraint, which is given by:

$$\int_0^{t_{b,r}} u_0(t)dt = V\left(t_{b,r}\right) - V(0) = V_f - V_0 \equiv \Delta V \tag{3-1}$$

where V_0 and V_f are the initial and final volumes of the liquid phase in the reactor, respectively. The simplest choice for $u_0(t)$. is with a linear dependence on time, which will be used here. For many batch processes and, in particular, the biopharmaceutical processes examined in this chapter, when the initial substrate concentration is low, the feeding flow rate of the co-reactant, or substrate, should be high at the beginning of the batch and low at the end of the batch. Without the aid of

any further process knowledge, we simply impose $u_0(t_b) = 0$, as the substrate is a co-reactant. The resulting $u_0(t)$, with a linear dependence on t, is

$$u_0(t) = 2(1 - t/t_{b,r})\Delta V/t_{b,r} \tag{3-2}$$

We then let all substrate inflows vary within the domain $u_0(t) \pm \Delta u_0(t)$, with $u_0(t) = \Delta u_0(t)$. This allows $u(t)$ to vary between zero and $2u_0(t)$ but at the same time it forces the flow rates to be zero at the end of the batch time. We denote this input domain as Domain A.

We now define the coded time-varying factor, $w(\tau)$, related to the substrate inflow as given in eq.(2-5)

$$u(\tau) = u_0(\tau) + \Delta u(\tau)w(\tau) \quad \text{with} \quad -1 \le w(\tau) \le +1 \tag{2-5}$$

The dimensionless time, τ, is defined by $\tau = t/t_b$ and the batch duration t_b, is given by eq. (2-4).

$$t_b = t_{b,r} + \Delta t_{b,r} w \text{ with } -1 \le w \le +1 \tag{2-4}$$

Where w is the traditional DoE factor. Now let the coded time-varying factor, $w(\tau)$, be expanded using three dynamic sub-factors and first three shifted Legendre polynomials.

$$w(\tau) = x_1 P_0(\tau) + x_2 P_1(\tau) + x_3 P_2(\tau) \tag{3-3}$$

Here the coefficients x_1, x_2, and x_3 are the dynamic sub-factors, which will be varied to define the coded time-varying dynamic factor $w(\tau)$ and therefore define the co-reactant's inflow in eq.(2-5). We limit the series expansion of $w(\tau)$ to only the first three polynomials so that the set of dynamic sub-factors, x_i's and the corresponding number of experiments is limited. Substituting $u(\tau)$ into the total volume constraint in eq. (3-1) we have:

$$t_b \int_0^1 [u_0(\tau) + \Delta u_0(\tau)] d\tau = \Delta V \tag{3-4}$$

Solving eq. (3-4) for the traditional DoE factor, w, we obtain the following relationship in which x_3 is not present.

$$w = \frac{(3x_1 - x_2)t_{b,r}}{[3 + (3x_1 - x_2)]\Delta t_{b,r}} \tag{3-5}$$

This relationship reduces the degree of freedom by one, as w is dependent on x_1 and x_2. The constraints that the designed experiments need to satisfy are the inequalities in eq. (2-4) and the following set of inequalities, which ensure that the constraint in eq. (2-5) is satisfied.

$$-1 \leq x_1 \pm x_2 \pm x_3 \leq +1 \tag{3-6}$$

The inequality of eq. (2-4) translates, though eq. (3-5), into the following constraint on x_1 and x_2:

$$-1 \leq \frac{(3x_1 - x_2)t_{b,r}}{[3 + (3x_1 - x_2)]\Delta t_{b,r}} \leq +1 \tag{3-7}$$

3.2. Incorporating Prior Knowledge

Even though many batch processes might not be understood completely to enable the development of an accurate knowledge-driven model, some *a priori* knowledge about the process characteristics is usually available. In this section, we discuss how to incorporate such knowledge to improve the design of the DoDE experiments and the resulting process performance. In general, the prior knowledge will be used to aid the design of the reference time-varying input profile and the selection of the corresponding input domain. We here consider two kinds of prior knowledge. In

the first subsection, we utilize the knowledge that sufficient amount of the nutrient is needed to maintain the product production at the end of the batch in order to compensate the product destruction [83, 84]. In the second subsection, we incorporate one additional piece of prior knowledge related to the sequential scheduling of multiple co-reactants, such as substrate feeds. In both of these two cases, the prior knowledge is used to modify the DoDE input domain.

3.2.1. Maintenance of Product

It is understood that in secondary metabolite fermentations (SMF) the product might degrade during the entire batch duration. If the glucose supply is too low, the production rate will be low and the amount of product will decrease. This motivates the need for the glucose feed rate to be larger than zero at the end of the batch. Based on this information, we define the reference input with the property that $u_0(t_{b,r}) = 0.2u_0(0)$. The resulting $u_0(t)$, with a linear dependence on t, is

$$u_0(t) = (5 - 4t/t_{b,r})\Delta V/(3t_{b,r}) \tag{3-8}$$

Again, the substrate flow profile $u(t) = u_0(t) + \Delta u_0(t)w(\tau)$ can vary in the domain $u_0(t) \pm \Delta u_0(t)$. This allows $u(t_b)$ to vary between zero and $0.4u_0(0)$, instead of being forced to become zero as in Domain A. We will call this domain B.

The selection of the ratio of the ending flow rate, $u_0(t_b)$, to the beginning flow rate, $u_0(0)$, in the reference input, here taken as 20%, is mostly intuitive. A different value of the ratio can be selected by the process expert based on a qualitative understanding on how strong is the maintenance demand of the product. These assumptions on the reference input can be ameliorated later with additional experiments by further modifying the input domain.

Substituting eq. (3-8) in eq. (3-4) and solving for w, we obtain the following counterpart to eq. (3-5) in domain B.

$$w = -\frac{(9x_1 - 2x_2)t_{b,r}}{[9 + (9x_1 - 2x_2)]\Delta t_{b,r}} \tag{3-9}$$

The independent factors and the number of experiments for developing a quadratic RSM model are the same in both domain A and B. In Figure 3-1, some of the feeding profiles of the substrate inflow are given for one of the biopharmaceutical process of interest, SMF process #5, which will be defined later. In this figure, we also depict a thick line, the optimal feeding profile, to be also discussed later in this chapter. The dashed and dotted lines are the reference input profile and the upper limit of the input domain, respectively. In the upper sub-plot we depict some of the input profiles constrained in Domain A. The corresponding ones in Domain B are given in the lower sub-figure. In each domain, the batch time and the incoming flow profiles are considerably different between two experiments in each of the DoDE sets of experiments. All the inflow rates become zero at the end of the batch in Domain A. The superiority of the design of Domain B can be argued by examining whether the optimal values of the factors (and dynamic sub-factors) for Domain A lie on the boundary of A. A similar examination, of whether the optimal values of the factors in domain B lie on the corresponding boundary, would indicate whether a third domain C, larger than B, would have improved the process optimum further. We will discuss these further in the conclusions section.

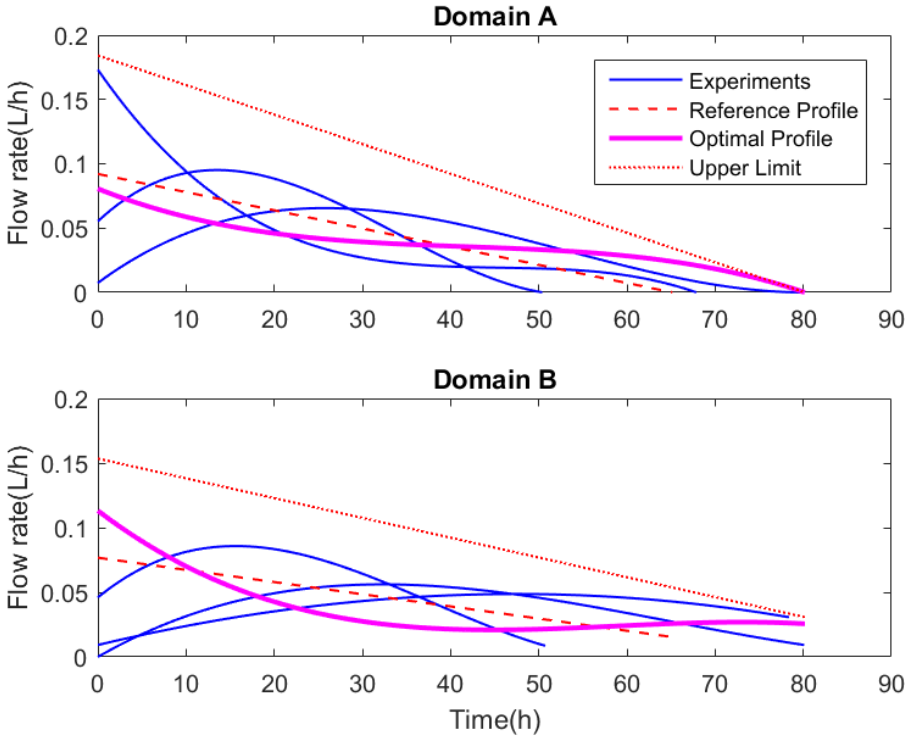

Figure 3-1: Some of the feeding profiles (—) for SMF #5 in Domain A (upper) and B (lower), along with the center point reference profile (----), the calculated optimal one (▬) and the input domain upper limit (.....)

3.2.2. Scheduling of Feeding Substrates

For processes with multiple feeding streams, the process performance may not only depend on the individual feeding profiles but also on whether these profiles are sequential or simultaneous. If we have *a priori* knowledge that one of the two co-reactants, or substrates in the case of biological processes, should be fed ahead of the other, we can improve the effectiveness of the DoDE methodology by incorporating this knowledge in the DoDE design. For the Hybridoma cell culture examined here, the feed of glucose should be scheduled after the feed of glutamine [82], in a sequential manner. Without such prior knowledge, one would most certainly choose to feed simultaneously both streams.

We now define the reference feeding profiles and the corresponding input domains for the feeding of the two co-reactants, both for the simultaneous and the sequential modes. However, we will consider only domain B here. A similar study can also be presented in domain A. As in the previous definition of domain B, we let $u_{i,0}(t = t_{b,r}) = 0.2u_{i,0}(0)$ for $i = 1, 2$. The total feed volume, ΔV, is split into the feed volume of glutamine (ΔV_1) and the feed volume of glucose (ΔV_2). We assume $\Delta V_1 = \alpha \Delta V$, implying $\Delta V_2 = (1 - \alpha)\Delta V$ and we select $\alpha = 0.8$ for both the simultaneous and sequential feeding strategies. If one were to explore the impact of different α values, an additional factor could be introduced by $\alpha = 0.8 + 0.3w_3$ with $-1 \leq w_3 \leq 0.5$ to explore α values between 0.5 and .0.95. This additional factor will increase the number of experiments but will also help select the optimal α value.

For the simultaneous feeding, the batch time is defined as in eq. (2-4), and the feed streams of glucose and glutamine both last for the entire batch duration. Either of the two feeding profiles is parameterized using three dynamic sub-factors, $\{x_i | i = 1, 2, 3\}$ and the counterparts of eq. (3-8) and eq. (3-9) can be easily derived. The total number of independent factors is six. The desired quadratic RSM model has 28 $(1 + 6 + 5 \times 6/2 + 6)$ parameters requiring at least that many experiments. We again add three replicated experiments for the estimation of the process variability and three additional experiments to estimate the LoF statistics. Therefore, a set of 34 experiments in total is needed for the case of the simultaneous feeding.

For the sequential feeding of the two co-reactants, we define an additional factor for the switching time, t_s, the time when we stop feeding the first co-reactant (glutamine for the application here) and start to feed the second co-reactant (glucose). We define the switch time, t_s, by $t_s = t_{s,r} +$

$\Delta\, t_s w_2$ with $-1 \le w_2 \le +1$ in a similar manner as we defined the batch time, t_b, by $t_b = \ t_{b,r} + \Delta\, t_b w_1$ with $-1 \le w_1 \le +1$.

Because of the two distinct feeding intervals we define two dimensionless times, as follows.

$$\tau_1 = t/t_s \text{ and } \tau_2 = (t - t_s)/(t_b - t_s) \tag{3-10}$$

The feeds of glutamine and glucose, in the reference case, last for $t_{s,r}$ and for $t_{b,r} - t_{s,r}$ units of time, respectively. Here $\Delta V_1 = \alpha \Delta V$ with $\alpha = 0.8$ and $u_{i,0}\big(t = t_{b,r}\big) = 0.2u_{i,0}(0)$ for $i = 1, 2$ as mentioned before. The linear reference input profiles for these two feeds are then given as

$$\begin{aligned} u_{1,0}(\tau_1) &= \alpha \Delta V(5 - 4\tau_1)/(3t_{s,r}) \\ u_{2,0}(\tau_2) &= (1 - \alpha)\Delta V(5 - 4\tau_2)/[3(t_{b,r} - t_{s,r})] \end{aligned} \tag{3-11}$$

We let $\Delta u_{i,0}(\tau_i) = u_{i,0}(\tau_i)$ and $u_i(\tau_i) = u_{i,0}(\tau_i) + \Delta u_{i,0}(\tau_i) w_i(\tau_i)$ with $w_i(\tau_i) = x_{i1}P_0(\tau_i) + x_{i2}P_1(\tau_i) + x_{i3}P_2(\tau_i)$ for $i = 1, 2$. Then the total volume constraint is given by

$$t_s \int_0^1 u_1(\tau_1) d\tau_1 + (t_b - t_s) \int_0^1 u_2(\tau_2) d\tau_2 = \Delta V \tag{3-12}$$

Manipulating eq. (3-12), we can express w_1, defining the total batch time, in term of other factors and reduce the degrees of freedom by one.

$$w_1 = -\frac{(t_{b,r} - t_{s,r})\left[1 - h_1 - h_2 - \left(\frac{\Delta t_s}{t_{s,r}} h_1 - \frac{\Delta t_s}{t_{b,r} - t_{s,r}} h_2\right) w_2\right]}{h_2 \Delta t_b} \tag{3-13}$$

where $h_1 = \alpha(1 + x_{11} - x_{12}/9)$ and $h_2 = (1 - \alpha)(1 + x_{21} - x_{22}/9)$. Then we design the 42 DoDE experiments, perform them and fit a quadratic RSM with seven independent

factors, $w_2, x_{11}, x_{12}, x_{13}, x_{21}, x_{22},$ and x_{23}. Obviously the number of experiments required to estimate the above quadratic RSM is not small. This is because we allowed three dynamic sub-factors and aimed for a quadratic RSM. We will discuss in the next section ways to reduce the number of experiments.

3.3. Reduction in the Initial DoDE Experiments

In industrial practice, the practitioner is always interested in optimizing his process while satisfying tight budgetary and development time constraints. In such a case, the data-driven optimization approach using a significantly smaller number of experiments is of great interest. We follow up here the discussion of the previous section in which the initial set of experiment required for estimating a quadratic RSM are 16 and 42 for SMF processes and the Hybdridoma cell culture respectively. These might appear quite numerous to the practitioner, especially in cases where the batch duration might extend to several hours, if not days. There are two ways to reduce the number of experiments. The first way is by minimizing the number of sub-factors considered, from three to two, and the second is by aiming for the simplest RSM, the linear one, neglecting both quadratic and two-way interaction model terms. We will focus here on the benefits of initially estimating a linear RSM but still consider three dynamic sub-factors.

For the sequential feeding of the Hybridoma cell culture, only 14 experiments are required if we simply aim for a linear RSM. Here, the number of model parameters is eight and six extra runs are added for reasons explained above. For the SMF processes, a linear RSM with three factors can be estimated using ten experiments in total. The number of experiments can be further reduced by only adding two, instead of three, extra runs for the LoF statistics and only two, instead of three,

replicate experiments for the process variability. The last item, reduction on the number of replicated runs might reduce significantly the accuracy of estimation of the normal variability of the process. This way, the number of runs decreases to 12 and eight (8) for the Hybridoma cell culture and the SMF processes, respectively. However this might compromise the soundness of the statistical analysis of the collected data, especially the calculation of the uncertainty characteristics of the estimated model. It will not be used here.

We now discuss the modelling and optimization for the processes of interest here, using this reduced number of experiments. Because the linear RSM may not accurately model the nonlinear process, the calculated optimal conditions will be tentative, with associated prediction intervals that might be large. We thus need to worry whether the predicted uncertainty of the planned next experiment is too large. If the prediction interval is larger than a desired upper bound, a suboptimal operation might be determined and run so that the uncertainty constraint is satisfied, following in principle the Tendency Modelling approach [85-88]. This approach accepts the direction of process change implied by the preliminary but uncertain tendency model and then selects a suboptimal operating point for the next experiment at a location along the line starting from the best of the initial experiments and parallel to the direction of the steepest accent predicted by the model. The exact location is decided by the maximum allowed predication uncertainty. An approach to estimate this bound follows.

After the model has been estimated from the available data, we can easily calculate the model prediction's uncertainty at any point $\boldsymbol{x}$ in the domain. The uncertainty is quantified by the half width of the prediction interval [23] and is defined by:

$$\Delta y_i = t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2[1 + \boldsymbol{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{x}_i]} \tag{3-14}$$

Here $\boldsymbol{x}$ is the information matrix of the initial set of experiments [23], $\hat{\sigma}$ is the estimated standard deviation of the process normal variability, while the vector x_i represents the operating conditions of interest. Also $t_{\alpha/2,n-p}$ is the *t*-statistics at a confidence level of $100(1-\alpha)\%$. At first, we calculate the above half-width of the prediction interval for each of the N initial experiments, $\Delta y_{exp,j}$ for $j = 1, 2, \cdots, N$, using eq. (3-14). Then the average of these values is given by $\Delta\bar{y}_{exp} = \sum_{j=1}^{N} \Delta y_{exp,j}/N$ and the estimate of the corresponding standard deviation is $\hat{\sigma}_{\Delta y_{exp}} = 1/(N-1)\sqrt{\sum_{j=1}^{N}(\Delta y_{exp,j} - \Delta\bar{y}_{exp})^2}$. We are going to use these two values ($\Delta\bar{y}_{exp}$, $\hat{\sigma}_{\Delta y_{exp}}$) to establish the allowed maximum prediction half-interval for any new optimal or sub-optimal operation.

$$\Delta y_{opt} \leq \Delta\bar{y}_{exp} + \lambda\hat{\sigma}_{\Delta y_{exp}} \tag{3-15}$$

Here Δy_{opt} is half of the 95% certainty prediction interval at the tentative optimal or sub-optimal operating point. If we wish to be very conservative we might require $\lambda = 2$, but a more appropriate value will be higher than 2 and less than 3. As the number of experiments has been significantly reduced comparing to the number for a quadratic RSM, the estimated linear RSM may have a large uncertainty in some regions far away from the experimental points. In particular, the predicted optimum by a linear RSM is on the boundary of the input domain where the prediction uncertainty is the largest. Therefore, the selection of the optimal operating conditions should be done with great caution. Because the initial model is linear, the optimal point will be along a line that starts from the best experiment and is parallel to the steepest accent direction. We will call this the

"tendency line". If we were not concerned about the uncertainty constraint in eq. (3-15), the calculated optimum with a very large prediction uncertainty would be of minimal practical value. Consequently, we should estimate a suboptimal operation, which satisfies the uncertainty constraint of eq.(3-15). A simple way to estimate such a suboptimal point is to consider the operating points along the "tendency line". Along this line we select the suboptimal point for the next experiment so that it does not violate the uncertainty constraint in eq. (3-15).

3.4. Biopharmaceutical Processes of Interest

To examine the two proposed methodological improvements to the DoDE approach we select two quite different sets of biopharmaceutical processes: i) eight secondary metabolite fermentations (SMF) and ii) a Hybridoma cell culture [82]. The set of eight SMF processes is constructed by modifying the four most significant parameters of the initial model reported by B&R [34]. For the SMF processes, K_p and K_{in} determine at what substrate concentration the specific production rate reaches its maximum while μ_{max} and ρ_{max} affect the maximum specific growth and production rates, respectively. Therefore, the optimal feeding strategies substantially depend on those four parameters. The diverse parameter sets used to define the SMF processes are given in columns 2-5 of Table 3-1. The last row reports the parameter values for the B&R process, the base case. We do not examine all the possible combinations of the high and low values of the four significant parameters varied. This would have resulted in 16 ($=2^4$) processes, a full factorial design. Instead, we have designed a resolution IV 2^{4-1} fractional factorial design with eight process examples. The diversity of these fermentations is quite apparent if one looks at processes #1, 3 and 5, for example. Process #1 is characterized by a relatively slow growth rate for biomass and product. In contrast,

process #5 is characterized by fast growth rates for both biomass and product. Process #3 has a slow biomass growth rate and a relatively fast one for the product.

Table 3-1: Eight diverse sets of model parameters representing a multitude of possible fermentation processes and the process optima via the DoDE and the MBO approach

# Process	μ_{max} × 10^2	ρ_{max} × 10^3	K_P × 10^5	K_{in} × 10^2	Domain A			Domain B			% Diff btw Domain A&B	MBO in Domain B	% Diff btw MBO and DoDE
					P_pred	PI Width	P_sim	P_pred	PI Width	P_sim			
Units	h^{-1}	$\frac{g_P}{g_X \cdot h}$	$\frac{g_S}{L}$	$\frac{g_S}{L}$	g_P	g_P	g_P	g_P	g_P	g_P		g_P	
1	5	3	5	32	35.8	1.7	36.2	43.4	1.7	43.6	20.4	45.0	3.1
2	5	3	20	3	24.2	2.8	24.1	33.5	1.4	32.3	34.3	35.2	8.0
3	5	11	5	3	120.2	20.2	115.6	138.8	19.9	152.7	32.1	155.3	1.7
4	20	3	5	3	37.2	1.9	37.6	45.6	3.2	44.6	18.7	46.8	4.6
5	20	11	20	3	115.6	13.4	115.6	133.9	11.0	138.2	19.6	147.8	6.5
6	20	11	5	32	153.2	9.5	144.7	167.3	10.0	162.9	12.6	166.1	2.0
7	20	3	20	32	33.1	1.9	31.7	39.7	2.4	39.4	24.3	41.5	5.0
8	5	11	20	32	104.4	15.4	111.8	162.5	13.6	149.3	33.6	159.3	6.2
B&R	10	5.5	10	10	75.0	9.0	71.5	88.1	5.4	85.9	20.1	87.9	2.3

The Hybridoma cell culture has a significantly different structure from that for the SMF processes. Two main substrates, glucose (Glc) and glutamine (Gln), are required to be fed for cell growth and antibody production. Meanwhile, two toxic byproducts, lactate (Lac) and ammonia (Amm), are released during the production of the desired metabolites. This leads to a more involved multi-feed optimization problem.

In all cases, the only available process measurements are the product concentration and bioreactor volume at the end of the batch. Since these experiments are *in silico*, we simulate the normal variability of the process by adding a 4% measurement error to the simulated value of the product concentration, y_{sim}, i.e. $y \sim N(y_{sim}, 0.02y_{sim})$. The resulting values are the ones used to build the RSMs. The same noise has been added to all the processes discussed in this chapter.

3.5. Results and Discussions

In this section, we examine the effectiveness of the proposed improvements to the DoDE approach in the selected classes of biopharmaceutical processes. We first demonstrate the incorporation of prior process knowledge as discussed in the previous sections. In the second subsection, we confirm that similar optimal results can be achieved using the reduced number of experiments.

For the eight SMF processes studied here, we simply assume that the appropriate reference batch time depends on the biomass and product growth rates. While in practice, the reference batch time will be selected according to the knowledge about the process. In the initial B&R process for penicillin, the biomass needed about 30 h to grow and the penicillin production extended for another 100 h. For each of the eight processes, the reference batch time is calculated by

$$t_{b,r} = t_{g,r}\frac{\mu_{max,r}}{\mu_{max}} + t_{p,r}\frac{\rho_{max,r}}{\rho_{max}} \tag{3-16}$$

With $\mu_{max,r} = 0.1h^{-1}, \rho_{max,r} = 0.0055\ g_p(g_x \cdot h)^{-1}, t_{g,r} = 30h, t_{p,r} = 100h$, the reference values from the original B&R process. The value for $\Delta t_{b,r}$ in eq. (2-4) for each of the eight other processes will be set by $\Delta t_{b,r} = 30t_{b,r}/130$ as the values of $\Delta t_{b,r}$ and $\mathrm{t_{b,r}}$ for the original B&R process is 30 and 130 h. For all the SMF processes, we will assume that the initial volume, V_0, of

the bioreactor is 7 L and the maximum possible value, V_f, is 10 L. For the Hybridoma cell culture, the values of $\Delta t_{b,r}$ and $t_{b,r}$ are 1 and 9 days, respectively. The switch time $t_{s,r}$ and its deviation $\Delta t_{s,r}$ are 3 days and 1 day, respectively. We assume that the initial and final volume be 1.2 L and 2 L.

3.5.1. Improvement through *a priori* Knowledge

For each of the eight SMF processes, we design two sets of DoDE experiments for a quadratic RSM in Domain A and B, respectively. We simulate these experiments, add the 4% measurement error in the results, fit the RSM with the collected data of product weight at the end of the batch and, then, calculate the optimal operating conditions by solving a constrained optimization problem with the fitted RSMs. The predicted maximum amount of product (P_pred) and the half-width of the prediction interval (PI Width) are reported in columns 6 & 7 for domain A and in columns 9 & 10 for domain B in Table 3-1. For example, for SMF process #1 it is predicted that the optimum amount of product is 35.8 ± 1.7 g in domain A and 43.4 ± 1.7 g in domain B. The estimated optimal feeding policy is then verified by a follow-up *in silico* experiment. The resulting amount of product (P_sim) is reported in column 8 and column 11 for domain A and B, respectively. We will see that for all eight SMF processes examined here, the P_sim is inside the prediction interval predicted by the quadratic RSM, which indicates that all RSMs are sufficiently accurate. In all the eight processes, the optimal product weights obtained (P_sim) in Domain B are higher than those obtained in Domain A. The percentage differences between the optima in the two domains are calculated using the equation, $diff_{AB} = 100|P_{D-B} - P_{D-A}|/P_{D-A}$ %, and the results are listed in column 12 in Table 3-1. On average, the difference is 24.4%, while the process specific values range from 12.6% to 34.3%. The process optima have been improved significantly

by incorporating the process knowledge about the product degradation that aided the design of the input domain.

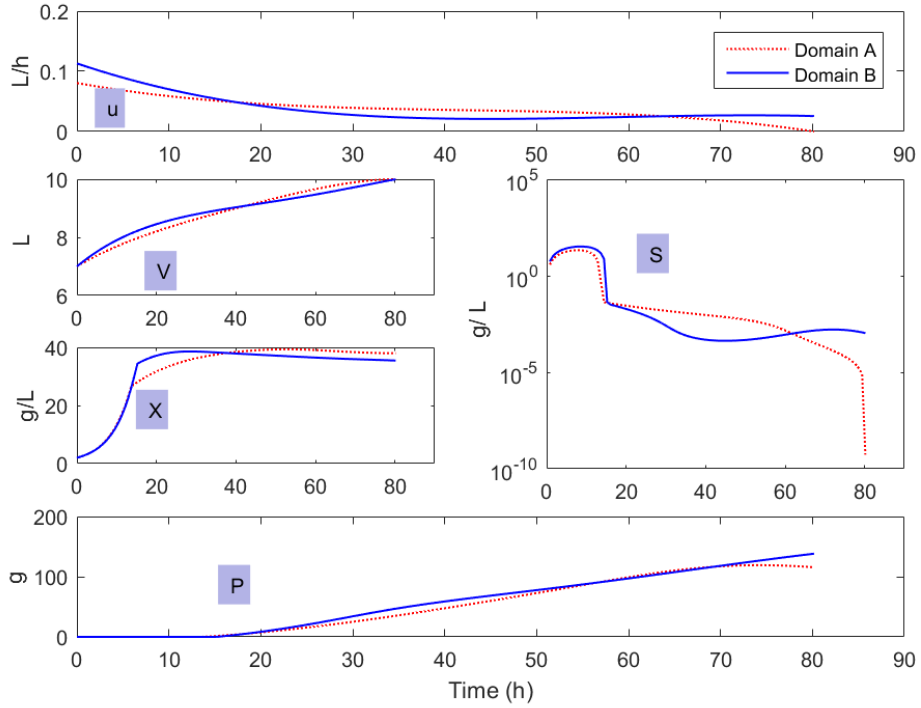

Figure 3-2: The time-evolution of the state and input variables of the optimal operation for SMF # 5 in Domain A (.....) and B (——)

We compare the feeding profiles (u) in the two domains as well as the resulting time evolutions of other species, including liquid phase volume (V), substrate concentration (S), biomass concentration (X) and product weight (P*V), in Figure 3-2. The optimal feeding profiles obtained in Domain A and B are similar in shape during the first 65 hours. However, the one in Domain A decreases to zero after that. This results in a much lower substrate concentration, shown in logarithmic scale in the figure, than that in Domain B during the same period. Due to this characteristic of product degradation, the grams of product in Domain A stops increasing after 65

hours and even starts to decrease. While the grams of product in Domain B keeps increasing till the end of batch.

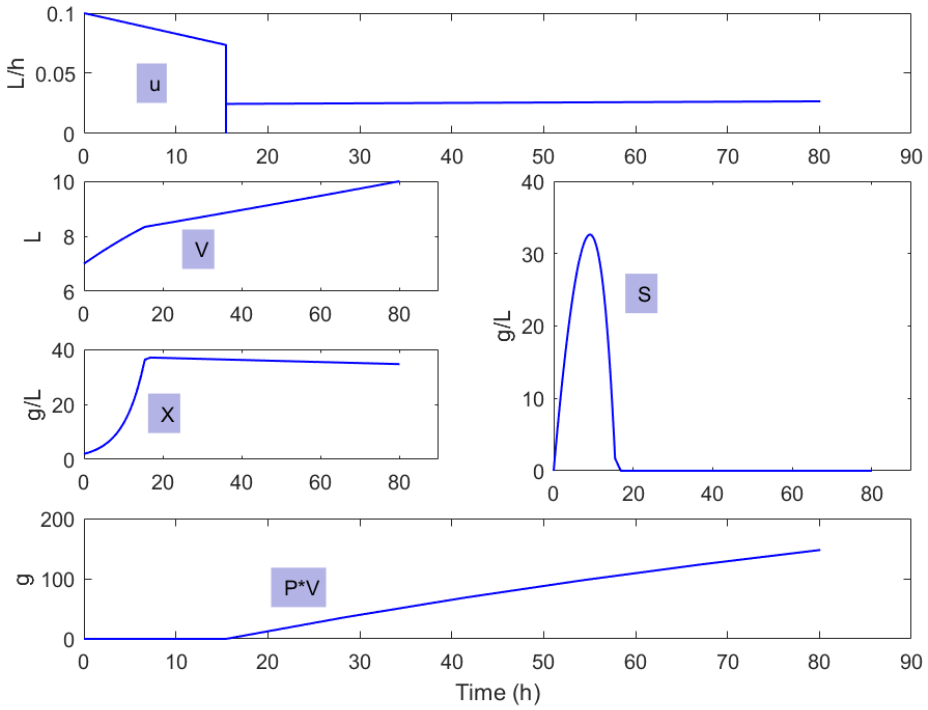

Figure 3-3: The model-based optimal time-evolution of the state and input variables for SMF # 5 in Domain B

We also calculate the process optimum for the eight SMF processes using the Model-Based Optimization (MBO) approach and assume that the perfectly accurate model is at hand. We take these MBO optima as the maximum obtainable product weight. If the optima obtained using the DoDE approach are close to the MBO optima, the power of the DoDE approach for process optimization is confirmed. The MBO optimum is calculated using the simultaneous approach [38-40]. The obtained optimal grams of the product are shown in column 13 in Table 3-1. The percent difference between the DoDE and the MBO optima in domain B, is calculated by $diff_{DoDE-MBO} =$

$100|P_{DoDE} - P_{MBO}|/P_{MBO}$ % and listed in column 14. The DoDE and MBO optima are very close to each other, the former, naturally, a bit smaller that the latter. In all eight processes the difference is less than 8%. The time-evolutions of the state and input variables obtained by the MBO approach for SMF process #5 are depicted in Figure 3-3. In this process the data-driven and MBO optima have a 6.5% difference. If we compare this figure with Figure 3-2, presenting the DoDE optimal operations, we will see that both the obtained substrate concentrations by the MBO and the DoDE approaches in Domain B remain constant and non-zero at the end of the batch while the substrate concentration of the DoDE in Domain A case drops very close to zero.

In the above optimization tasks, we have focused on maximizing the total grams of product at the end of the batch, independently of how long the batch might last. Maximizing the process productivity, e.g. the ratio of the product grams produced to the batch duration, is also possible in the DoDE framework. A simpler example has been discussed elsewhere[27].

For the Hybridoma cell culture, we compare the resulting optimal product weight using the two feeding strategies, the simultaneous and sequential feeding of the two co-reactants, in domain B. The first one is selected without the aid of the prior knowledge on the desired feeding sequence. While the second one is defined by incorporating the information that glutamine should be fed first and then followed by the feed of glucose. The predicted maximum amount of product for the simultaneous feeding is 355.5±17.0 *mg*. and the optimal batch time is 10 days. We perform such optimal semi-batch operation and the resulting product weight is 372.1 *mg*, inside the above prediction interval. This again confirms the prediction accuracy of the RSM.

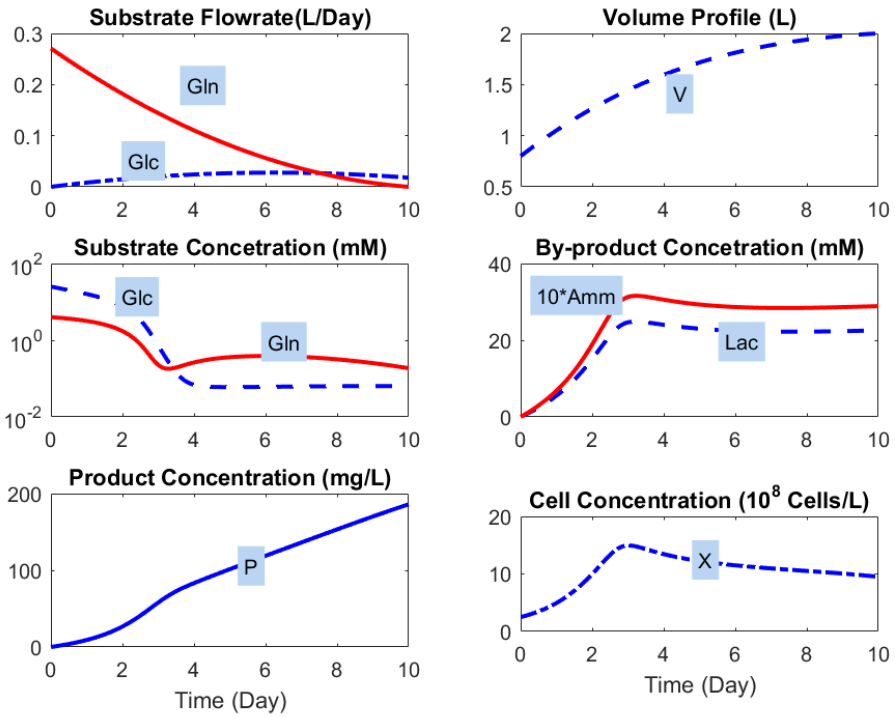

Figure 3-4: The optimal time-evolution of the state and input variables for Hybridoma cell culture using the simultaneous feeding

Figure 3-4 presents the time evolution of the main compositions and the inflows of the two substrates for the optimal operation of the cell culture. The profiles in the middle left sub-figure are presented in logarithmic scale, due to the large variation of the concentrations of glucose and glutamine. The concentration of ammonia (Amm) in the middle-right sub-figure is multiplied by 10, for better visibility.

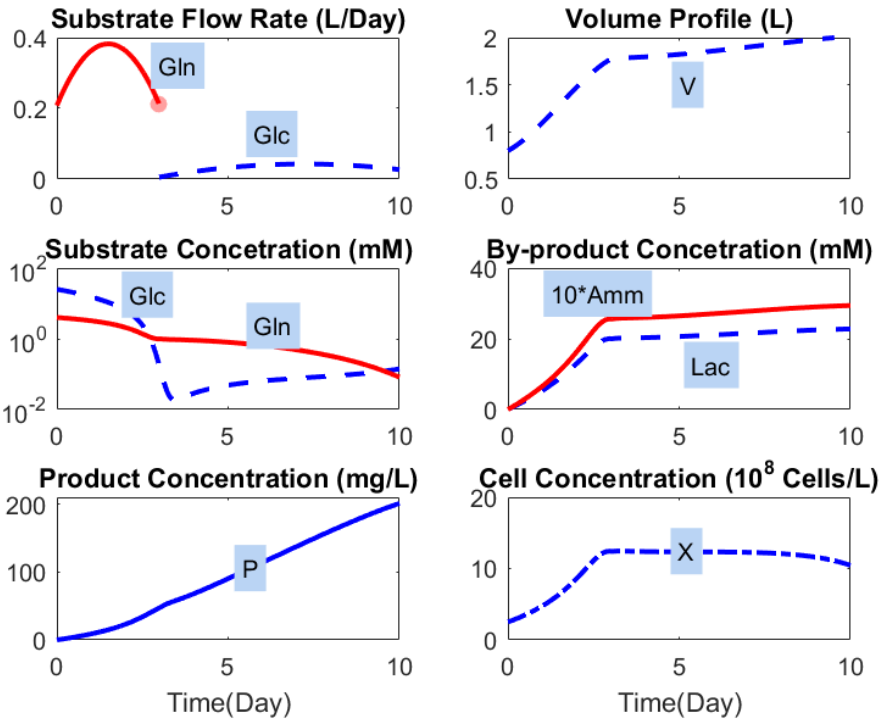

Figure 3-5: The optimal time-evolution of the state and input variables for Hybridoma cell culture using the sequential feeding

For the case of sequential feeding, the DoDE optimum estimated from the RSM model is 429.6±51.8 *mg* and the estimated optimal operation of the sequential feeding produces 400.1 *mg*, which is a 7.5% improvement compared to the one obtained using the simultaneous feeding. The optimal switch time is 3 days. The time-evolution of the inflows for the two substrates and the state variables (volume and concentrations) are depicted in Figure 3-5.

The MBO optimum calculated by us using the simultaneous approach is 407.6 *mg*. The DoDE optima by the simultaneous feeding and sequential feeding are 8.7% and 1.8% less than the MBO optimum, respectively. The time-evolutions of the inflows and the state variables obtained using the MBO approach are plotted in Figure 3-6.

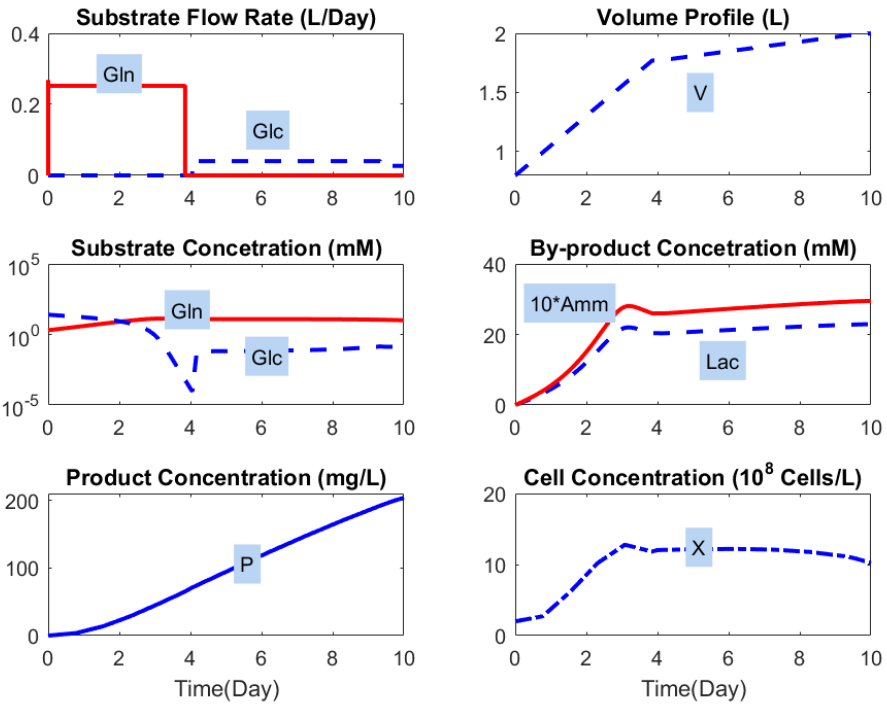

Figure 3-6: The model-based optimal time-evolution of the state and input variables for Hybridoma cell culture

The above process examples illustrate how to incorporate some *a priori* knowledge into the DoDE design and to improve the optimal process performance. We have here used a key motivator for the design of domain B, the need for maintaining the substrate concentrations at non-zero values at the end of the batch to compensate the substantial degradation of the product for the SMF processes. In addition, we schedule the feeds of the two feeding streams using knowledge on the desired sequence. In other similar uses of the DoDE methodology, one will end up using whatever is the most appropriate prior knowledge on the process to appropriately select the general characteristics of the reference input profile and the design of the input domain, which might provide the best optimal results. In processes that are not being developed from ground zero, prior experience will influence significantly the definition of $u_0(t)$ and the corresponding input domain

within which the experiment will be performed. Very often we might select a $u_0(t)$ to be the present operating conditions and define the domain to represent a certain limited deviation from such present conditions.

3.5.2. Optimization with Fewer Experiments

In this subsection, we discuss the results of the optimization of the aforementioned biopharmaceutical processes using fewer initial experiments. We examine SMF processes #1 and #5 and the sequential feeding case of the Hybridoma cell culture as examples. The number of experiments in the initial set for each process is minimal because the set of experiments is designed to only estimate a linear RSM model. For such a model, one needs a DoDE design with only ten (10) experiments for each of the two SMF processes and 14 experiments for the Hybridoma cell culture. These sets of experiments include as many as the number of model parameters plus three more experiments for estimating the LoF statistic and three replicate runs for the estimation of the normal variability of the process, if it is not know *a priori*. This is a much smaller set of experiments compared to the ones used to develop the quadratic RSM in the previous section, which consisted of 16 and 42 experiments for the SMF processes and the Hybridoma cell culture, respectively.

With the estimated linear RSM, we calculate the "tentative" data-driven optimum for each of the three processes. The prediction uncertainties of the estimated optimal conditions for SMF processes #1 and #5 do not violate inequality (3-15). Then the calculated optimum need not be called "tentative" and the process should be operated at such a point. As shown in row 2 and 3 of Table 3-2, the predicted optima for these two processes are 45.0±4.9 g and 126.0±21.4 g,

respectively. Operating each process in such manner, the resulting value for the process output, in grams of product, is 43.6 g and 127.7 g, respectively. Both values lie inside the corresponding prediction interval of the RSMs.

Table 3-2: Comparison between data-driven optima obtained using reduced and full sets of experiments

Process	P_sim by Full Experiments	Reduced Experiments			% Diff btw Reduced and Full Experiments	A
		P_pred	PI Width	P_sim		
	g_P	g_P	g_P	g_P		
SMF 1	43.6	45.0	4.9	43.6	0[I]	1
SMF 5	138.2	126.0	21.4	127.7	7.6	1
Hybridoma	400.1	393.0	49.5	387.8	3.1	0.2

[I] The two DoDE optima compared here are identical as the optimal values of the dynamic sub-factors they both refer to are the same, a vertex of the domain.

For the Hybridoma cell culture, the predicted optimum is 565.3 ± 80.7 *mg* and the corresponding uncertainty violates inequality (3-15). Indeed if one runs the simulated process under these uncertain optimal conditions the resulting product amount is 153.0 *mg*, a very different value from the predicted one. This is a direct consequence of the tentative character of this initial model and the substantial uncertainty in the prediction of this calculated optimum. Indeed the prediction interval at the expected optimum point is characterized by a value of $\lambda = 14.7$ for eq. (3-15). Instead, we will seek out the conditions under which the amount of the product is maximal while inequality (3-15) is satisfied, for $\lambda = 3$.

To verify the choice of $\lambda = 3$ for inequality (3-15) as a reasonable one and illustrate the evolution of the width of the prediction interval as the operating conditions approach the "tentative" optimal conditions from the best one of the previous experiments, we define a series of candidate operating conditions $\boldsymbol{x}_i$ along a straight line between the best previous experiment, x_0, and the "tentative" optimal conditions x_{opt}, as given below:

$$\boldsymbol{x}_i = \boldsymbol{x}_0 + \beta \times \left(\boldsymbol{x}_{opt} - \boldsymbol{x}_0\right) for\ \beta \in [0,1] \tag{3-17}$$

The expected product weight and the related predicted uncertainty is plotted in Figure 3-7 for 10 values β (0.1, 0.2, … 1.0). We plot the estimated values of the half-width of the prediction intervals (y-axis) versus the predicted values of the product grams. In this plot we denote the characteristics of the prior experiments and we draw the $\pm\sigma, \pm 2\sigma$ *a*nd $\pm 3\sigma$ lines around the mean value of the half-width of the prediction interval of the prior experiments. The plotted red diamonds, for the 10 values of $\beta = 0.2$, indicate that as these values increase so does the expected grams of the product as well as the size of the corresponding half prediction interval. We observe that if we are to impose the constraint in eq. (3-15) with $\lambda = 3$ we should limit this fractional distance about $\beta = 0.2$. At such point the expected product weight are 393.0 ± 49.5 mg and the corresponding simulated operation of the process yields 387.8 *mg.* We can now claim that by searching for the optimal operating conditions along the tendency line and also bounding the model uncertainty we have substantially optimized this process.

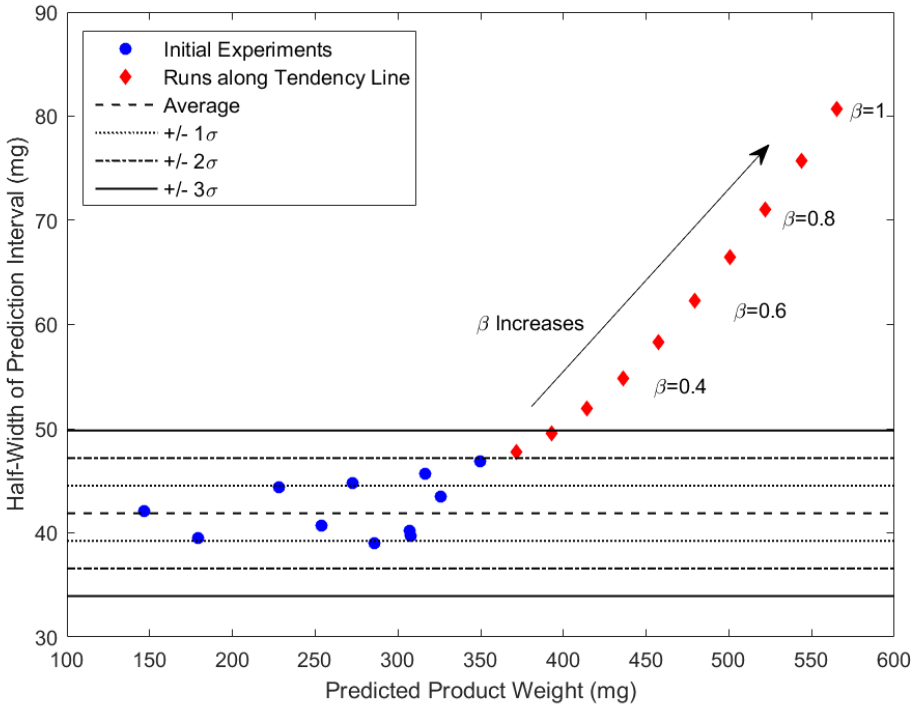

Figure 3-7: Prediction intervals of experiments and new runs calculated using linear RSM for Hybridoma cell culture

In Figure 3-8, we plot the predicted values (blue dashed line) and the corresponding prediction intervals (green dashed lines) by the tentative linear RSM model for β varying from 0 to 1. The red diamonds are the simulated value at $\beta = 0, 0.1, \cdots, 1$. It has been seen that when the β increases, the differences between the simulated and predicted values becomes larger. When $\beta > 0.4$, the obtained simulated value falls outside the prediction interval, which indicates the extrapolation of the RSM model becomes inaccurate. If we stick to $\beta = 0.2$, corresponding to $\lambda = 3$, we achieved the highest product weight.

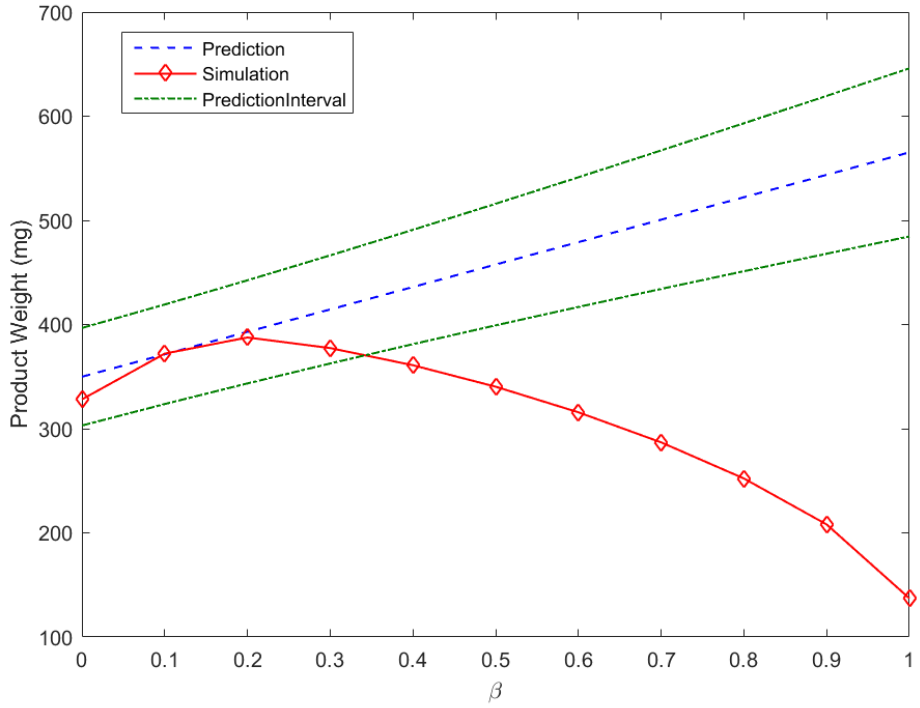

Figure 3-8: The predicted and simulated product weight vs β

In all three processes examined here, the optimal product weights based on the linear RSM estimated from a smaller number of experiments is lower than that obtained when enough experiments were performed to estimate a quadratic RSM. This is expected. However the difference, as reported in column 6 of Table 3-2, is small. The largest difference is 7.6% in SMF process #5. Considering the 37.5% reduction in the number of experiments (from 16 experiments to 10 experiments), the proposed approach is quite attractive in optimizing the process in a timely manner. Moreover, one may further improve the data-driven optimum by conducting additional experiments, which augment the initial design so that one can estimate a more complex RSM model. These experiments might also be selected to be in a region where the initial model predicts process improvements.

3.6. Conclusions

In this chapter, we have discussed two improvements to enhance the originally proposed DoDE methodology. The first one uses the prior knowledge about the process characteristics to aid the design of the reference input profile and the corresponding input domain. We discussed two different types of *a priori* knowledge. The proposed approaches are examined in two representative classes of the biopharmaceutical processes. In the set of eight SMF processes, our process insight motivates the design of Domain B, resulting in a 24.4% average improvement over Domain A, with the largest improvement equal to 34.3%. For the Hybridoma cell culture, we incorporate a different kind of process knowledge, introducing the multiple inputs sequentially instead of simultaneously. We thus obtain a 7.5% improvement in product weight. These case studies demonstrate that by incorporating the process knowledge into the DoDE framework, one can further improve the optimal performance of the process at hand.

The second methodological improvement of the original DoDE methodology addresses the minimization of the initial number of experiments. To minimize the number of the initial experiments, one should initially aim for a linear RSM, without any two-factor interaction terms. The predicted optimum is considered “tentative” until its prediction uncertainty is estimated and is shown to satisfy an uncertainty upper limit. If the uncertainty of the optimal point is larger than the maximum allowed, a suboptimal point of operation is sought that will satisfy the uncertainty constraint. This is activated only in the Hybridoma cell culture example.

We also confirm, with the two representative classes of biopharmaceutical processes, an earlier finding that the DoDE optimum is very close to the model-based optimum obtained by using a

detailed and perfectly accurate knowledge-driven model of the process, which is rarely available with great accuracy, if at all.

Concerning the examination whether the present domain used is the best and provides the final credible possibility for process improvement, the following idea might be of interest. We touch on this briefly here but we also recognize that a more thorough and separate examination is required. We can state it in the form of a question: *Could the calculation of the optimum in a given domain (say A here) provide a hint about changes of the initial domain into a new domain (say B, here) which will allow further improvement in the process?*

To answer this question, one can investigate whether the optimal operating conditions in Domain A, characterized by values of the related factors and dynamic sub-factors, activate some of the constraints defining the boundary of Domain A. By identifying the activated constraints, which limit the obtained optimal operating conditions, one can appropriately modify the current domain by relaxing the activated constraints to allow the further optimization of the process. For the SMF process # 5 the optimal values of the three dynamic sub-factors in Domain A are $x_1 = 0, x_2 = 0.56, x_3 = 0.44$. These values indicate that they activate the constraint $x_1 + x_2 + x_3 = 1$ which implies that the coded dynamic variable,$w(\tau)$, is attaining its maximum value at the end of the batch, i.e. $w(1) = 1$,. This hints that the pinching of domain A at the end of the batch serves as a limitation, definitely motivating a move from domain A to domain B. If the same constraint were activated by the optimal operating conditions obtained in domain B, we would have understood that the initially allowed width at the end of the batch, set to be 20% of the width of the domain at time zero, is restrictive and needs to be enlarged. However, the calculated optimal values in domain

B are $x_1 = -0.16, x_2 = -0.10, x_3 = 0.74$ which result to $x_1 + x_2 + x_3 = 0.78$ a value that is less than 1. Therefore we conclude that the allowed domain size is wide enough.

4. The New Dynamic Response Surface Methodologies (DRSM-2)

Accurate mathematical models are essential for the success of process optimization and control for industrial manufacturing processes. Such models are developed either based on our knowledge of the inner workings of the process or by analyzing the input-output data from a set of designed experiments or historical data of past manufacturing activities. The models estimated using the first approach are known as the *knowledge-driven models* while the ones developed via the second are called *data-driven models*. In many cases, the inner workings of a process are not fully understood to enable the development of an accurate knowledge-driven model. Especially for the processes with low production rates, such as batch and semi-batch processes employed widely by the pharmaceutical and specialty chemicals industries, the development of a knowledge-driven model is often prohibitively expensive and time consuming.

Data-driven modeling approaches are attractive alternatives as they utilize only input and output data from an existing process. To analyze and model these time-resolved outputs, we proposed the Dynamic Response Surface Methodology (DRSM) [31]. Unlike the time-invariant parameters in the RSM model, the parameters estimated in the DRSM model are functions of time. With the process dynamics captured in the DRSM model, one can achieve both optimization as well as control purposes for batch processes using the single data-driven DRSM model [32], saving experimental cost of developing separate models for optimization and control purposes.

The previously proposed DRSM methodology requires fixed operating durations of the DoDE experiments. Consequently, it cannot be applied for the modeling of continuous processes, when the operating duration is various and can be even semi-infinite. In this paper, we present a new

DRSM methodology, denoted by DRSM-2, which is capable of dealing with the data collected over semi-infinite time horizons. With the new approach, we are able to accurately model the time-varying output variables of continuous processes as well as those of batch processes. In addition, the new approach is capable of accounting for time-delayed input changes, and is able to use non-equidistant data.

This chapter is organized as follows. We first introduce the DRSM-2 methodology. Here multiple decision variables of the DRSM-2 model affecting the model accuracy need to be selected. We propose a systematic procedure to select their best values so that the final data-driven model is the most accurate possible. We then examine the efficacy of the proposed approach in two representative and challenging nonlinear processes. The first one is a continuous propylene polymerization process [9, 89] while the second one is a semi-batch penicillin fermentation [34]. In both cases, the obtained DRSM-2 models accurately represent the time-varying process outputs.

4.1. The DRSM-2 Methodology

As mentioned previously, the limitation of the fixed batch time required by the DRSM-1 approach is originated from the definition of the dimensionless time $\tau = t/t_b$. In the DRSM-2 methodology, we define a new independent variable θ, as an exponential function of time, t.

$$\theta = 1 - exp(-t/t_c) \tag{4-1}$$

Here t_c is an approximate value for the time constant characterizing the long-term dynamic changes in the output variable to the input changes. The definition of θ is motivated by our basic understanding of linear process dynamic, whose responses to simple inputs are of exponential

rather than polynomial character. For cases where an input change takes place at a time instant different from zero, say at d_i, the corresponding time transformation will be given by

$$\theta_i = 1 - exp[-(t - d_i)/t_c] \tag{4-2}$$

As the time t increases from zero to infinity, the transformed time variable θ increases from zero to one, i.e. $\theta \in [0, 1]$, the interval in which the shifted Legendre polynomials are defined. Therefore, we can use the shifted Legendre polynomials in θ as the functional basis for the expression of the $\beta_q(\theta)$ function. While the input time delay d_i may be different for each factor, we will assume that the t_c value is of the same for the effects of all factors. The t_c value aims to characterize the slowest dynamics of interest. Faster dynamics will be modeled by higher powers of θ corresponding to exponential terms with smaller time constants. We illustrate this with the second order Shifted Legendre polynomial. By substituting eq. (4-1) or (4-2) in the polynomial $P_2(\cdot)$, we obtain

$$P_2(\theta) = 1 - 6\theta + 6\theta^2 = 1 - 6\,exp(-t/t_c) + 6exp(-t/0.5t_c) \tag{4-3}$$

Or

$$\begin{aligned} P_2(\theta_i) &= 1 - 6\theta_i + 6\theta_i^2 \\ &= 1 - 6\,exp[-(t - d_i)/t_c] + 6\,exp[-(t - d_i)/0.5t_c] \end{aligned} \tag{4-4}$$

The last term, $6exp(-t/0.5t_c)$ or $6exp[-(t - d_i)/0.5t_c]$, is from θ^2 and can be seen to correspond to a time constant of $0.5t_c$. In general, the R^{th} order shifted Legendre polynomial, $P_R(\theta)$ contains terms with a time constant equals to t_c/R. This property allows the DRSM-2 model to accurately represent the time-varying outputs affected by both fast as well as slow

dynamics using a single sufficiently large t_c. One the other hand, if the t_c is selected to be too large, we will need higher order polynomials to make t_c/R small enough to represent the fast dynamics of the process. This will lead to a DRSM-2 model with too many γ parameters. To avoid the unnecessary large amount of parameters, the value of t_c should be properly selected. The appropriate selection of t_c value will be made using the Bayesian Information Criterion (BIC)[90] and will be discussed in detail later.

4.1.1. Definition of the DRSM-2 Variables

With the new independent variable θ or θ_i, the quadratic DRSM model given in eq (2-9) is re-written as follows

$$y(\theta) = \beta_0(\theta_0) + \sum_{i=1}^{n} \beta_i(\theta_i)x_i + \sum_{i=1}^{n}\sum_{i<j}^{n} \beta_{ij}(\theta_{ij})x_i x_j + \sum_{i=1}^{n} \beta_{ii}(\theta_{ii})x_i^2 \qquad (4\text{-}5)$$

The parametric function, $\beta_q(\theta_q)$, in the DRSM model is given by

$$\beta_q(\theta_q) = \gamma_{q,1}P_0(\theta_q) + \gamma_{q,2}P_1(\theta_q) + \cdots + \gamma_{q,R+1}P_R(\theta_q) \qquad (4\text{-}6)$$

with $q = 0, i, ij, or\ ii$. Here we use the quadratic model class as an example. However, other model classes, such as linear, two-factor interaction (2FI), or cubic, are of interest. For many cases the effect of each factor commences at time zero and in such cases we can assume that all $\theta_q{}'$s are identical. The need for non-identical $\theta_q{}'$s will be discussed later in the case study of the propylene polymerization process. Before solving the linear regression problem to estimate the values of the γ's,, we list here the definition of the model variables:

C : DRSM model class, *e.g.* linear, 2 factor interaction (2FI), quadratic, cubic, *etc.*

n : Total number of factors, including classic factors as well as dynamic sub-factors

p : Number of parametric functions $\beta_q(\theta)$ which depends on the selection of C

R_q : Order of Shifted Legendre polynomial parameterizing a corresponding $\beta_q(\theta)$

M : Number of experiments

K_m : Number of data records in the m^{th} experiment

K_T: Total number of data records of all experiments

In the DRSM-2 approach, the operating duration and the number of data record in each experiment can vary. Also the data need not be equidistant and the *K* value need not be the same for each experiment. These important programming generalizations are also applicable in the DRSM-1 methodology.

4.1.2. The Estimation of DRSM-2 Model via Linear Regression

To estimate the values of the model parameter γ's, we rewrite the DRSM-2 model in eq (4-5) in matrix form as follows

$$\boldsymbol{y} = \mathbf{G}\boldsymbol{\gamma} \tag{4-7}$$

Where the $K_T \times 1$ column vector $\boldsymbol{y} = [\boldsymbol{y}_1^T \quad \boldsymbol{y}_2^T \quad \cdots \quad \boldsymbol{y}_M^T]^T$ is formed by stacking the measurements of the M experiments. Its m^{th} element $\boldsymbol{y_m} = [y_{m,1} \quad y_{m,2} \quad \cdots \quad y_{m,K_m}]^T$ is a $K_m \times 1$ column vector with the K_m measurements of the output variable in m^{th} data set.

The $K_T \times R_T$ matrix **G** is the Hadamard product [91] of the two matrices (**Z** and **P**) with the same dimensions,

$$\mathbf{G} = \mathbf{Z} \circ \mathbf{P} \tag{4-8}$$

R_T is the total number of γ's in the DRSM-2 model. As the number of γ's parameterized each $\beta_q(\theta_q)$ is $R_q + 1$, R_T equals to the value $\sum R_q + 1$.The matrix $\mathbf{Z}$ is constructed using the values of the factors and is defined as

$$\mathbf{Z} = [\mathbf{X}_1^{\mathbf{T}} \quad \mathbf{X}_2^{\mathbf{T}} \quad \cdots \quad \mathbf{X}_{\mathrm{M}}^{\mathrm{T}}] \tag{4-9}$$

The m^{th} block ($m = 1, 2, \cdots, M$) in $\mathbf{Z}$, *e.g.* for estimating a quadratic DRSM model, is defined as

$$\mathbf{X_m} = \begin{bmatrix} \mathbf{1} & \mathbf{x_{m,1}} & \cdots & \mathbf{x_{m,n}} & \mathbf{x_{m,12}} & \cdots & \mathbf{x_{m,(n-1)n}} & \mathbf{x_{m,11}} & \cdots & \mathbf{x_{m,nn}} \end{bmatrix} \tag{4-10}$$

It is a $K_m \times R_T$ matrix, in which each element is a $K_m \times (R_q + 1)$ matrix with identical elements in all its entries as given below.

$$\mathbf{1} = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}, \ \mathbf{x_{m,i}} = x_{m,i}\mathbf{1}, \ \mathbf{x_{m,ij}} = x_{m,i}x_{m,j}\mathbf{1}, \ \mathbf{x_{m,ii}} = x_{m,i}^2\mathbf{1} \tag{4-11}$$

The scalar $x_{m,i}$ in the above matrices is the values of the i^{th} factor for the m^{th} experiment. The $\mathbf{X}_{\mathrm{m}}$ matrices for estimating the DRSM-2 models of other model classes can be obtained in a similar manner.

Matrix $\mathbf{P}$ is defined as $\mathbf{P} = [\mathbf{P}_1^{\mathrm{T}} \quad \mathbf{P}_2^{\mathbf{T}} \quad \cdots \quad \mathbf{P}_{\mathbf{M}}^{\mathbf{T}}]^{\mathbf{T}}$. It is a $K_T \times R_T$ matrix with the entries as the values of the Shifted Legendre polynomials at the time instants when the measurements are taken. As mentioned previously, these instants can take any values and they do not need to be equidistant in time. The m^{th} block of $\mathbf{P}$ is a $K_m \times R_T$ matrix defined as $\mathbf{P_m} =$

$\left[\mathbf{P_{m,0}} \quad \mathbf{P_{m,1}} \quad \cdots \quad \mathbf{P_{m,n}} \quad \mathbf{P_{m,12}} \quad \cdots \quad \mathbf{P_{m,(n-1)n}} \quad \mathbf{P_{m,11}} \quad \cdots \quad \mathbf{P_{m,nn}}\right]$, in which the element is a $K_m \times (R_q + 1)$ matrix given as follows.

$$\mathbf{P_{m,q}} = \begin{bmatrix} P_0(\theta_{q,1}) & P_1(\theta_{q,1}) & \cdots & P_{R_q}(\theta_{q,1}) \\ P_0(\theta_{q,2}) & P_1(\theta_{q,2}) & & P_{R_q}(\theta_{q,K_m}) \\ \vdots & \vdots & \ddots & \vdots \\ P_0(\theta_{q,K_m}) & P_1(\theta_{q,K_m}) & \cdots & P_{R_q}(\theta_{q,K_m}) \end{bmatrix} \text{ with } q = 0, i, ij, ii \tag{4-12}$$

The value of $\theta_{q,k}$ is determined by eq. (4-1) and (4-2) with the selected $\mathrm{t_c}$ and the corresponding input time delay, d_q. As mentioned previously, we take $d_q = 0$ for all factors and therefore $\theta_{q,k} = \theta_k$ for $\forall q$.

All the model parameters to be estimated is listed in the $\mathrm{R_T} \times 1$ column vector defined as

$$\boldsymbol{\gamma} = \left[\boldsymbol{\gamma}_0^T \quad \boldsymbol{\gamma}_1^T \quad \cdots \quad \boldsymbol{\gamma}_n^T \quad \boldsymbol{\gamma}_{12}^T \quad \cdots \quad \boldsymbol{\gamma}_{(n-1)n}^T \quad \boldsymbol{\gamma}_{11}^T \quad \cdots \quad \boldsymbol{\gamma}_{nn}^T\right]^T \tag{4-13}$$

Where each block is a $(R_q + 1) \times 1$ column vector $\boldsymbol{\gamma}_q = (\gamma_{q,1} \quad \gamma_{q,2} \quad \cdots \quad \gamma_{q,R_q})^T$, with entries the parameters of the polynomial representation of $\beta_q(\theta)$. We left-multiply both sides of eq. (4-8) by $\mathbf{G}^* = (\mathbf{G^T G})^{-1}\mathbf{G^T}$ and obtain the least squares solution for γ given below.

$$\boldsymbol{\gamma} = \mathbf{G}^* \boldsymbol{y} \tag{4-14}$$

In practice, a set of DRSM-2 models with different choices of the three decision variables are estimated via stepwise regression [92]. The decision variables includes the time constant t_c, the order of Shifted Legendre polynomials R_q, for each parametric function, $\beta_q(\theta)$, and the model class *C*. The procedure of selecting the most appropriate values for the decision variables resulting in the best DRSM-2 model is discussed in the following sub-section.

4.1.3. Iterative Procedure for Best DRSM-2 Model

We here propose an iterative procedure, shown in Figure 4-1, to determine the optimal values of the decision variables and estimate the best DRSM-2 model.

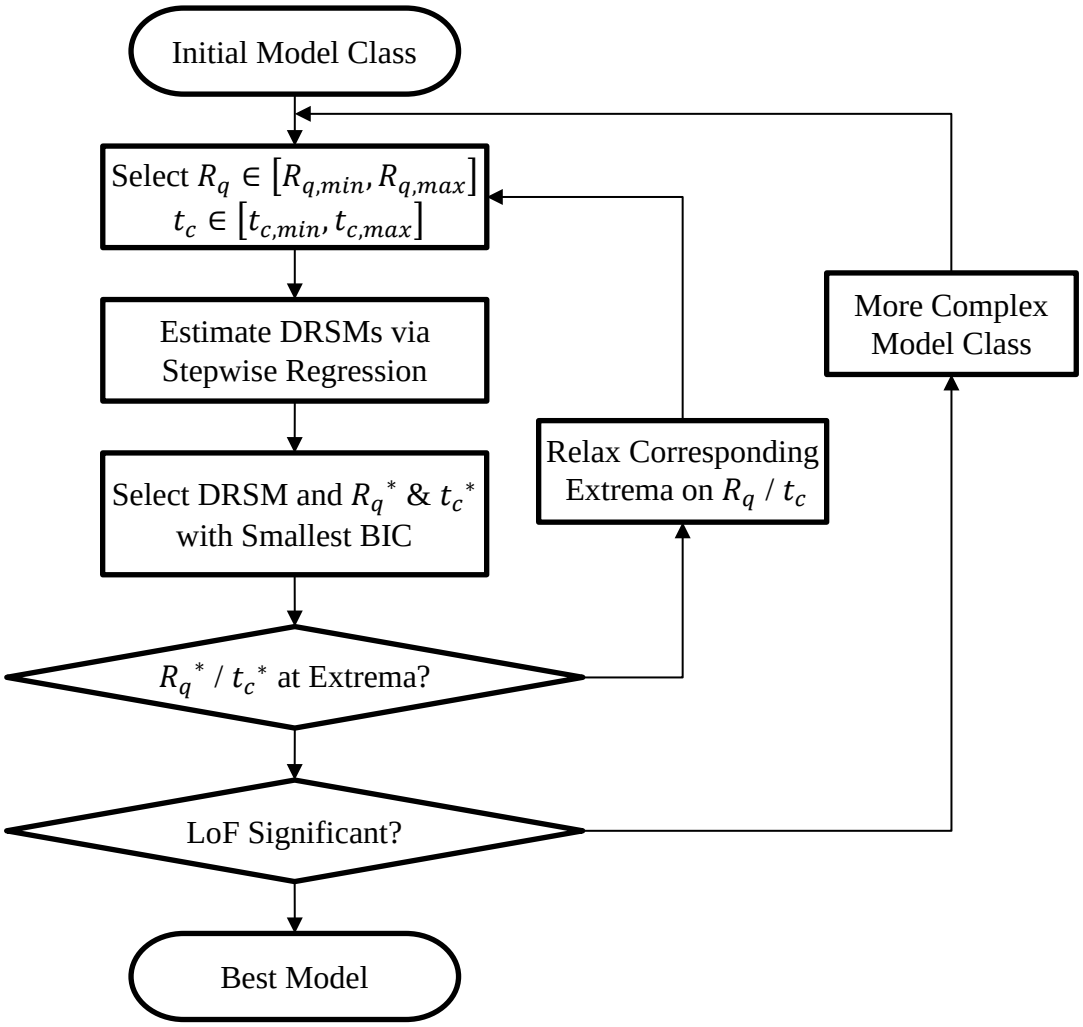

Figure 4-1: Iterative procedure for the selection of decision variables and the best DRSM-2 model

For simplicity, we choose the same value of R_q for all $\beta_q(\theta)$ functions, i.e. $R_q = R$, to formulate the linear regression problem in eq. (4-7). The stepwise regression algorithm with a threshold of

p-value equals to 0.05 will retain only the significant γ's. This could result in a different number of the shifted Legendre polynomials parameterizing each $\beta_q(\theta)$ function. In some specific applications, one may have prior knowledge on the potential optimal values of each $R_q{}'$s. In such a case, the proposed new algorithm allows the utilization of $R_q{}'$s of different values by following the same procedure described in Figure 4-1.

We first choose the initial model class, the one for which the experiment have been designed for. This might be the linear one if there is a strong preference in minimizing the number of experiments. It may be also determined by the prior knowledge on the process nonlinearity. If the process is known to be nonlinear, one might start with the 2FI model class or even a more complex one. In the selected model class, we estimate a set of DRSM-2 models using different combinations of the values for t_c and R_q within their corresponding initial ranges of interest, i.e. $t_c \in [t_{c,min}, t_{c,max}]$ and $R_q \in [R_{q,min}, R_{q,max}]$. For a continuous process, the initial choice of $t_{c,min}$ and $t_{c,max}$ are 10% and 40% of the longest settling time [18]. This can be estimated, for example, from some step responses. For nonlinear processes, the settling time varies with the input magnitude. As the DRSM-2 model can easily account for dynamics that are faster than the one characterized by the selected time constant, the longest settling time is used to determine the upper value in the t_c interval. For batch processes, where the settling time might not be estimable, we will choose $t_{c,min}$.and $t_{c,max}$.to be 10% and 40% of the longest batch time. The initial choices of $R_{q,min}$ and $R_{q,max}$ for both batch and continuous processes we have used here are 6 and 10, respectively. We feel that this range of polynomial order provides sufficient choices for the accurate representation of the time-varying output. Since the model parameters are estimated via stepwise regression, an initial polynomial order of 6 or higher could be reduced to a smaller

polynomial order if it is sufficient in representing the data at hand. If the process is expected to exhibit a more complicated dynamic behaviors, one could select larger values for $R_{q,min}$ and $R_{q,max}$. Meanwhile, the selection of the R_q and t_c. has to satisfy the following sampling constraint.

$$t_c/R_q > \Delta t \tag{4-15}$$

Where t_c/R_q is an approximate time constant for the fastest dynamics represented by the DRSM-2 model while Δt is the smallest sampling interval. We should not try to model dynamic phenomena that have a time constant (t_c/R_q) smaller than the sampling interval (Δt).

For each obtained model with the various values of R_q and t_c, we calculate the Bayesian Information Criterion (BIC) [90].

$$\text{BIC} = -2lnPr(\boldsymbol{y}|\widehat{\boldsymbol{\gamma}}) + R_s lnK_T \tag{4-16}$$

This is a model selection criterion that maximizes the likelihood function and penalizes for overfitting. The model with the smaller BIC value is preferred. Here $Pr(\boldsymbol{y}|\widehat{\boldsymbol{\gamma}})$ is the likelihood function, the probability of observing the measured data set y given the estimated DRSM-2 model parameters $\widehat{\boldsymbol{\gamma}}$. R_s in the second term of eq. (4-16) is the total number of the significant $\hat{\gamma}$'s estimated via stepwise regression and K_T, as defined previously, is the total number of measurements. Assuming that the error in the measured data is normal distributed, the log likelihood value is estimated via the formula[93]

$$lnPr(\boldsymbol{y}|\widehat{\boldsymbol{\gamma}}) = -\frac{1}{2}(\boldsymbol{y}-\widehat{\boldsymbol{y}})^T\hat{\sigma}_y^{-2}(\boldsymbol{y}-\widehat{\boldsymbol{y}}) - \frac{K_T}{2}ln\hat{\sigma}_y^2 - \frac{K_T}{2}ln(2\pi) \tag{4-17}$$

Where the estimated variance is $\hat{\sigma}_y^2 = \frac{1}{K_T}(\boldsymbol{y} - \hat{\boldsymbol{y}})^T(\boldsymbol{y} - \hat{\boldsymbol{y}})$. As the number of significant $\hat{\gamma}$ parameters, R_s, increases resulted from an higher value of R_q, the value of the second term in eq. (4-16) increases and might cause an increase in the BIC value. We will select the values of t_c and R_q which lead to the DRSM-2 model with the smallest BIC value. This, together with the stepwise regression, will keep the DRSM model sparse and accurate. If the selected values, t_c^* and R_q^*, are at the boundaries of the examined intervals (e.g. $t_c^* = t_{c,max}$ and $R_q^* = R_{q,max}$) we will relax the corresponding boundary and will estimate additional DRSM-2 models using the new combinations of t_c and R_q values. We repeat this procedure until the obtained best values for t_c^* and R_q^* are inside the examined interval.

Once the task is completed, we examine if the selected model represents all the non-random information in the data. This is quantified through the Lack-of-Fit (LoF) test. In this F-test, we examine whether the Sum of Squares due to the LoF (SS_{LoF}) is significantly larger than the one related to the normal variability of the process (SS_{PE}), also known as the Sum of Squares due to pure error. If the LoF statistic is significant, quantified by a *p*-value smaller than 0.05, a more complex model class should be considered to further increase the model accuracy. To perform the LoF test, we first calculate the total residual Sum of Squares (SS_E) associated with the estimated model using the following equation.

$$SS_E = \sum_{m=1}^{M} \sum_{k=1}^{K_m} (\hat{y}_{DRSM,m,k} - y_{m,k})^2 \tag{4-18}$$

Where $\hat{y}_{DRSM,m,k}$ and $y_{m,k}$ are the model predicted and measured values at time instance *k* in the m^{th} experiment, respectively. The degree of freedom (DoF) of SS_E is $K_T - R_s$, where K_T is the

total number of measurements of all M data sets while R_s is the total number of significant γ parameters in the DRSM-2 model. SS_E can be partitioned as the sum of SS_{LoF} and SS_{PE}as follows

$$SS_E = SS_{LoF} + SS_{PE} \tag{4-19}$$

To estimate the value of SS_{PE}, some replicate data are needed. Assume that M_C out of the M experiments are replicates at some experimental conditions and the measurements are taken at K_C time instants in each experiment. The value of the SS_{PE} is calculated as follows

$$SS_{PE} = \sum_{m=1}^{M_c} \sum_{k=1}^{K_c} (y_{m,k} - \bar{y}_{m,k})^2 \tag{4-20}$$

Here $\bar{y}_{m,k}$ is the average of $y_{m,k}$, the M_C replicate values at each of the K_C time instants. Therefore, the DoF of SS_{PE} is $(M_C - 1)K_C$, and the SS_{LoF} is estimated as follows

$$SS_{LoF} = SS_E - SS_{PE} \tag{4-21}$$

The DoF of the SS_{LoF} is calculated as the difference between those for SS_E and SS_{PE}, *i.e.* $(K_T - R_s) - (M_C - 1)K_C$. The ratio of the mean SS_{LoF} and SS_{PE} values follows an F distribution

$$F_0 = \frac{SS_{LoF}/n_1}{SS_{PE}/n_2} = \frac{SS_{LoF}/[(K_T - R_s) - (M_C - 1)K_C]}{SS_{PE}/[(M_C - 1)K_C]} \tag{4-22}$$

The LoF statistics is significant if $F_0 > F_{\alpha,n_1,n_2}$ or, equivalently, if the *p*-value corresponding to F_0 is smaller than α. Here we use $\alpha = 0.05$, corresponding to a 95% confidence level.

The desired result of this test is that the LoF is not significant, and the corresponding *p*-value is larger than α (0.05). In such a case we will accept the model and the choices of the decision

variables. Otherwise, we will turn to a more complex model class, e.g. from the linear to the two-factor interaction (2FI) model class or from the 2FI to the quadratic class and so on. Additional experimental data might be needed to estimate a more complex DRSM-2 model. These new data are obtained by augmenting the previous design of experiments with new runs. The augmented experiments as well as the initial ones can be designed using commercial software, such as MATLAB [94] and JMP [95] For the related algorithm please refer to the publications [96, 97]. To minimize the number of the new experiments while obtaining the maximum information, optimal experiment design criteria [23], such as D-Optimality, might be used for the design of the new experiments. Then a new set of DRSM-2 models of a more complex model class are estimated via stepwise regression and are assessed using BIC and LoF statistics. The iterative procedure will be repeated until the best DRSM-2 model is arrived at. We will examine the proposed DRSM-2 approach using two representative and quite different nonlinear processes in the following section.

4.2. Separate Models for Continuous Polypropylene Grade Transitions

In this section, we examine the performance of the DRSM-2 methodology in modeling the time-varying outputs of a continuous polymerization process. We will verify the capability of the DRSM-2 approach to model a highly complicated continuous process during a variety of transitions between an initial and a finial steady state.

In the polymerization industry, the operating conditions of a plant are changed frequently to produce polymers of different specifications measured, for example, by the Melt Flow Index (also known as Melt Flow Rate) of the product. Strong economic incentives motivate the minimization of the transition time and thus the amount of the off-spec product. To achieve this, an accurate

dynamic model predicting the time-varying process characteristics is needed. As the detailed process behavior is not always fully understood, a data-driven modeling approach is of great interest. Here we apply the DRSM-2 approach to model the continuous polymerization process where the data are obtained from a dynamic simulation based on the published models [9, 89]. As shown in Table 8-3 in the Appendix A, the simulation for the propylene polymerization process includes five types of reactions; chain initiation, chain propagation, chain transfer, site activation and site deactivation. The reactions are described in column 2 while the kinetic constants and the corresponding units are given in column 3 and 4, respectively.

4.2.1. DRSM-2 Model for Increasing MFI Transitions

As the process dynamics for transitions to higher Melt Flow Index (MFI) are different from those for the transitions to lower MFI values, separate DRSM-2 models should be developed for each case, respectively. We here explain in detail the development of the DRSM-2 model for the increasing MFI case. The decreasing MFI case is modeled in a similar manner.

The feed concentration of hydrogen, affecting the chain length and thus the MFI value of the produced polymer, is the input variable. Here we parameterize the time-varying input profile in a series of segments following the industrial practice. Each profile consists of three ramps and a steady-state value of the input variable for the remaining time. This finial input value is related to the finial steady state for the output. The mathematical expression of the input profiles for the increasing MFI operations are given below.

$$u(t) = \begin{cases} u_1 + at, & 0 \le t \le t_1 \\ u_2 - at, & t_1 < t \le t_2 \\ u_3 + at, & t_2 < t \le t_3 \\ u_4, & t_3 < t \le \infty \end{cases} \quad (4\text{-}23)$$

Here u_1 and u_4 are the necessary steady-state input values to achieve the corresponding output values at the initial and final steady-state operating points. The *a* value is fixed from process consideration and represents the maximum allowed rate of the change in the input. The u_2 and u_3 values need to be properly selected to achieve the most desirable transition between the initial and final steady states. The four values, u_1, u_2, u_3 and u_4 will be the factors for the experiments that we soon design. Using the selected u_i values and the fixed *a* value, the time instants t_1, t_2, and t_3 can be calculated by

$$t_1 = \left|\frac{u_2 - u_1}{a}\right|, t_2 = t_1 + \left|\frac{u_3 - u_2}{a}\right|, and \;\; t_3 = t_2 + \left|\frac{u_4 - u_3}{a}\right| \quad (4\text{-}24)$$

In the first row of Figure 4-2, we show two example input profiles and their corresponding outputs profiles in the second row for the increasing MFI case. The range of interest for $u_i's$ is between 0.001 and 0.030 *gmol/L*. We estimate that it takes 4 hours to increase the hydrogen concentration from its minimum to its maximal value. Thus the ramp rate is $a = 0.0073$ *gmol/L/h*. For each operation with different input profile, the transient period is of different length. As shown in Figure 4-2, the two runs last 14.9 and 12.4 hours, respectively. We will demonstrate later that the DRSM-2 methodology is capable to dealing with data sampled from runs with different durations.

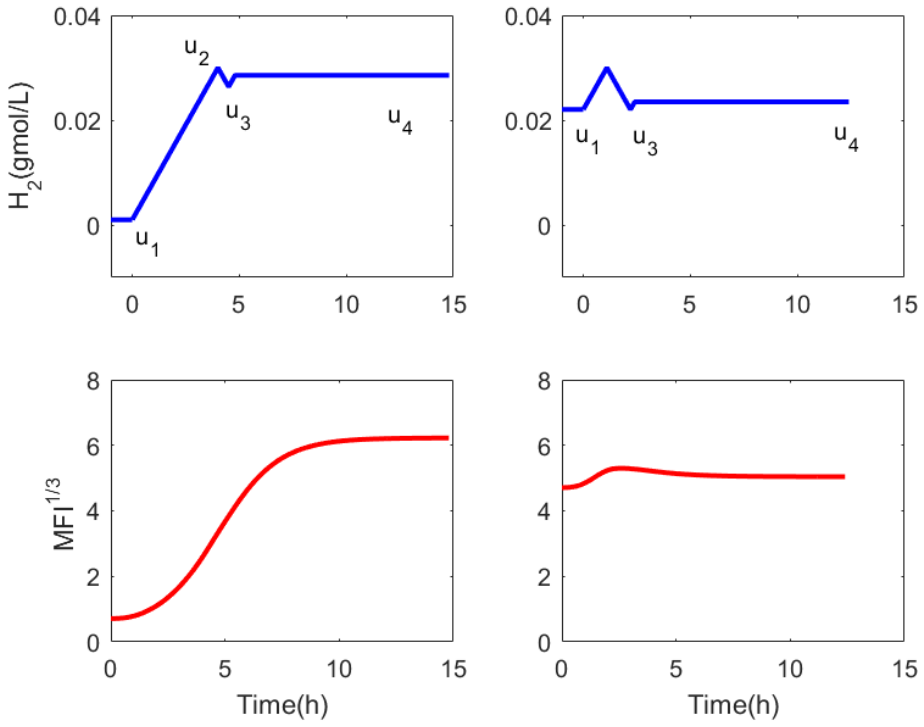

Figure 4-2: Example of the process responses (lower) to two different input profiles (upper) for the increasing MFI case. The responses in the right indicate an overshoot.

To design the set of experiments, we express each u_i using a coded factor, x_i, by

$$u_i = 0.0155 + 0.0145x_i, \qquad i = 1,2,3,4 \tag{4-25}$$

The values of x_i's are in the [-1, +1] interval. In addition we wish to impose the following constraints on the four coded factors to constrain the input profiles so that they lead to a higher MFI value than the initial one.

$$\begin{cases} x_4 > x_1 \\ x_2 > x_4 \\ x_4 > x_3 \end{cases} \tag{4-26}$$

Because design of experiments software cannot enforce strict inequalities and thus avoid having some of the x_i values equal to each other, *e.g.* $x_1 = x_2 = x_3$. We modify the above constraints to the following ones.

$$\begin{cases} x_4 - x_1 \geq \varepsilon \\ x_2 - x_4 \geq \varepsilon \\ x_4 - x_3 \geq \varepsilon \end{cases} \tag{4-27}$$

We here let $\varepsilon = 0.1$ but other values could be selected as well. Some of the possible input profiles lead to an overshoot in the output MFI, as shown in the lower right of Figure 4-2. This is because we have no prior knowledge on the dynamic characteristics of the process we are trying to model. If additional process information exists, we could design the experiments over a narrower domain using a tighter set of inequalities. Because we do not have prior process information, the task of modeling in this larger domain is more challenging as the responses with overshoot require additional γ parameters to be accurately represented. As we will see later in this section, the DRSM-2 models developed over this larger than necessary input domain are very accurate.

We design four sets of experiments to estimate four different classes of DRSM-2 models: linear, two-factor interactions (2FI), quadratic and cubic. The total number of experiments in each set is 11, 17, 21, and 41, respectively. Among them, the minimum number of experiments corresponding to the number of $\beta_q(\theta_q)$ parametric functions in the model is 5, 11, 15 and 35 for the four classes of models. To this minimum amount of experiments, one adds three distinct additional experiments for the estimation of the LoF statistic and three replicate runs for the estimation of the SS_{PE}, quantifying the normal process variability. During each experiment, the MFI is sampled every 0.1 *h*. We use equidistant sampling in *t*, which translates into a non-equidistant sampling in θ

When the u(t) changes from 0.001 to 0.030 *gmol/L*, the MFI value varies from 0.4 to 282.2 *gram/10 min*, a change of almost three orders of magnitude; a substantial modeling challenge. To narrow the range of values in the output variable, we choose to model the cubic root of MFI, $cMFI = (cMFI)^{\frac{1}{3}}$, in the DRSM-2 model. With the cubic root transformation, the output variable varies in the range [0.7, 6.2]. To the results of the simulated experiments, $cMFI_s$, we add a normally distributed fractional noise to arrive the measured $cMFI$ value for estimating the DRSM-2 model.

$$cMFI_m = cMFI_s[1 + \sigma N(0,1)] \; with \; \sigma = 0.02 \tag{4-28}$$

As each of the factors starts to affect the output at different times, there are different time delays associated with each factor. Because x_1 and x_2 define the first ramp of the input profile which affects the output variable at the beginning of the polymer grade transition, the corresponding time delays for these two factors are zero. The x_3 factor defines the minimum value of the second ramp, which affects the output after the end of the first ramp. So the time delay associated with x_3 is the time instant the second ramp starts. For a similar reason, the time delay for the x_4 factor is the time instant the third ramp starts. Consequently, the four time delays for the four factors are summarized as follows:

$$d_1 = d_2 = 0; \; d_3 = t_1; \; d_4 = t_2 \tag{4-29}$$

The time delays used to calculate the θ_q for each parametric function $\beta_q(\theta_q)$ in the DRSM-2 model is determined based on the above four factors' time delays with the following additional rules.

1) The time delay, d_q, for θ_q associated with terms involving a single factor (*e.g.* $q = i, ii, iii \cdots$ and $i = 1,2,3,4$) has the same time delay of the corresponding factor, *i.e.* $d_q = d_i$.
2) The time delay, d_q, for θ_q associated with multiple factors equals to largest time delay of the related factors. For a two-factor interaction term ($q = ij$) we have $d_q = max(d_i,\ d_j)$. Similarly for a three-factor interaction term ($q = ijr$) we have $d_q = max(d_i,\ d_j,\ d_r)$ and so on. This is because the interaction effect of the related factors starts to affect the output behavior when all of the related factors are acting on the process.
3) The time delay for θ_0, associated with $\beta_0(\theta_0)$, is zero, as $\beta_0(\theta_0)$represents the output profiles of the reference case, in which all factors are equal to zero ($x = 0$).

Then the calculated $\theta_q{}'$s are substituted into the Shifted Legendre polynomials, care should be exercised to evaluate a non-zero value only when θ_q is positive. In particular, the value of the Shifted Legendre polynomial, $P_k(\theta_q)$ ($k = 0, 1, \cdots, R_q$), used to parameterize the parametric function $\beta_q(\theta_q)$ is given by

$$P_k(\theta_q) = \begin{cases} 0, & if\ \theta_q \leq 0 \\ P_k(\theta_q), & if\ 0 \leq \theta_q \leq 1 \end{cases} \tag{4-30}$$

We know that the transient behavior of the polymer grade transition is nonlinear. However, for the purpose of illustrating the overall procedure for the selection of the best DRSM-2 model, we assume that such knowledge is not at hand. To minimize the number of initial experiments, we start by estimating DRSM-2 models of the linear class. For a continuous process, we choose the initial range of the time constant, t_c, based on the estimate of the settling time possibly from a set of step response experiments.

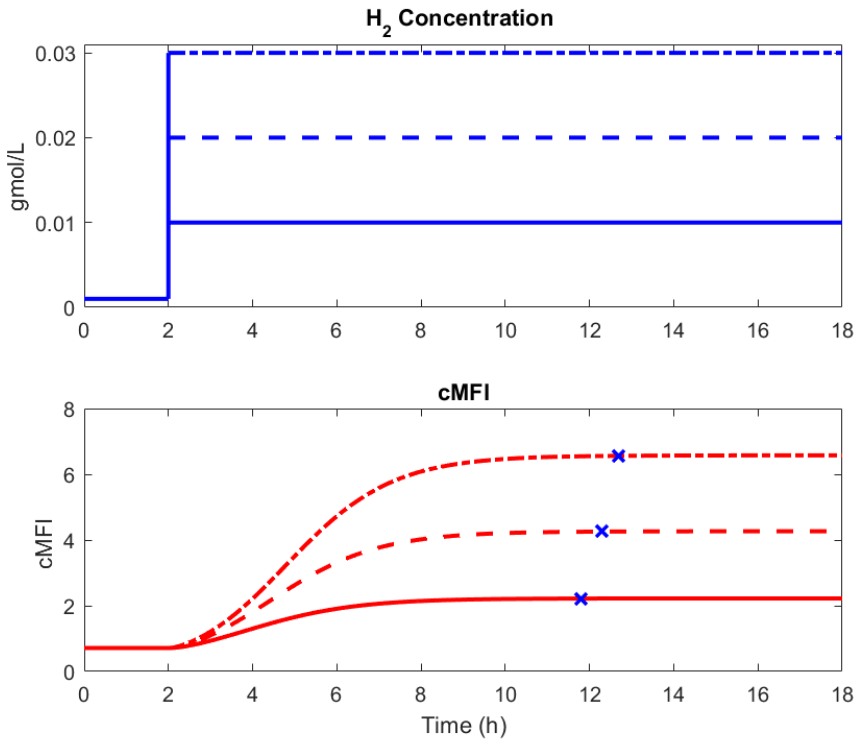

Figure 4-3: Three step responses of the cubic root transformed MFI (lower) corresponding to the different magnitudes of inputs change (upper). Each input and output pair is in the same type of line.

In Figure 4-3, we show three such step responses of the polymerization process, corresponding to three different input magnitudes. The settling time, marked with "×" in the lower figure on each output profile, is calculated as the time when 99.33% of the final change of the output variable is achieved. Because the process is nonlinear, the settling time increases as the input change increases. The longest settling time is 10.7 hours when the input variable changes from its minimal value of 0.001 *gmol/L* to its maximal value of 0.03 *gmol/L*.

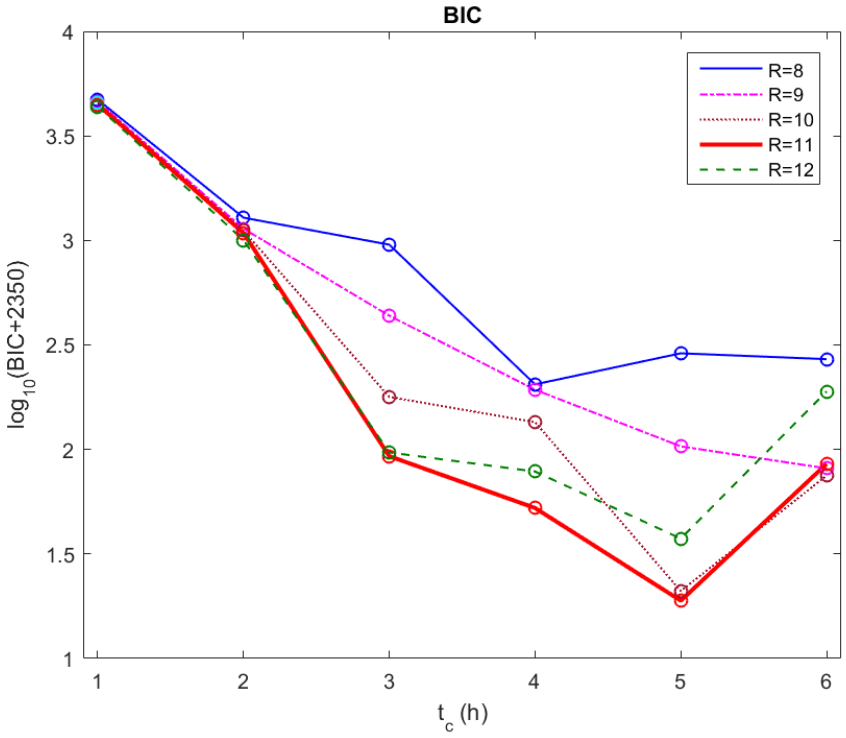

Figure 4-4: Transformed BIC values (y-axis) of linear DRSM-2 models vs t_c for various R values for the increasing MFI case of propylene polymerization process. Markers (○) indicate the model has significant LoF (p-value<0.05). All the DRSM-2 models satisfy the sampling constraint.

Therefore, we choose to assign t_c the following values $t_c = 1, 2, 3, 4, 5$ in hours. The initial range for R is $[6, 7, \cdots, 10]$. The calculated BIC values of the corresponding estimated DRSM-2 models are plotted in Figure 4-4. The BIC values shown in the y-axis are the $log_{10}(BIC + 2350)$ values for better visibility of where the minimal BIC value is achieved. The x-axis is the value of the time constant, t_c. Each line plotted in the figure represents the models estimated using the same value of R but different value of t_c. To keep the figure concise, the BIC values with R =6 and 7 are not shown, as those values are larger than the ones with other R's. In our initial round of estimating the DRSM-2 models, we realize that the model with the smallest BIC value results from $R = 10$ and $t_c = 5$, which lie on the corresponding upper limits. Therefore we enlarge the range of t_c and

R to also include $t_c = 6$ and $R = 12$.The BIC values of the corresponding DRSM-2 models, estimated with the new values of the decision variables, are also shown in the same figure. We can see now that $R = 11$ and $t_c = 5$ have resulted in the model with the smallest BIC value. Following this we examine the LoF statistic of this model. The corresponding *p*-value is almost zero, indicating that a more complex model class should be considered. In fact, all the obtained models of the linear class have significant LoF and correspond to *p*-values quite smaller than 0.05.

Before we proceed to design and perform the experiments to estimate DRSM-2 models in the class of two-factor interaction (2FI), we can eliminate some choices of the decision parameters which are unlikely to provide an accurate model based on the results for the linear models. We observe that for $t_c = 1\ h$, one obtain very large BIC values. This indicates that such a choice of time constant is too small. In addition, the BIC values of the DRSM models with $R = 6, 7$ are always larger than those with $R \geq 8$. Because the process dynamics are quite complex, one is not surprised that several Shifted Legendre polynomials are needed for an accurate approximation of the process behavior. So in the next round of DRSM-2 model estimation, we will explore value of R that are higher than 7. The related calculations, not presented here, for the best DRSM-2 model of the 2FI model class reveal that the LoF statistic is still significant. This motivates us to examine model classes with higher complexity, such as Quadratic and Cubic models. The best models in each class examined (Linear, 2FI, Quadratic and Cubic) and their related statistics are listed in row 3-6 of Table 4-1. It has been seen in the last column of the table that only the Cubic model class provides the desired result for the insignificant LoF statistic, with a *p*-value of 0.25. The best values of the decision variables are $R = 11$ and $t_c = 3\ h$ as reported in row 6 of the table. In addition

t_c/R ($= 0.21\ h$) is larger than the sampling interval ($0.1\ h$). Thus the sampling constraint given in eq.(4-15) is satisfied. Therefore, we choose this model as the best one.

Table 4-1: Statistics of the DRSM-2 models with smallest BIC values in four model classes for the propylene polymerization process

The increasing MFI case				
Model Class	R	t_c	BIC(x10^4)	LoF p-value
Linear	11	5	-0.2331	0
2FI	11	5	-0.2382	0
Quadratic	10	4	-0.5415	0.02
Cubic	11	3	-1.3349	0.25
The decreasing MFI case				
Model Class	R	t_c	BIC(x10^4)	LoF p-value
Linear	13	4	-0.1679	0
2FI	11	4	-0.2451	0
Quadratic	11	3	-0.5574	0.03
Cubic	11	3	-1.3198	1.00

The BIC value of the DRSM models in the Cubic model for different combinations of t_c and R values are shown in Figure 4-5, from which we can tell that the smallest BIC is achieved at $R = 11$ and $t_c = 3\ h$. For higher t_c values, the BIC value increases, sometimes dramatically. This is because more Shifted Legendre polynomials have to be included in the obtained model to represent the fast dynamics when the time constant is chosen too large.

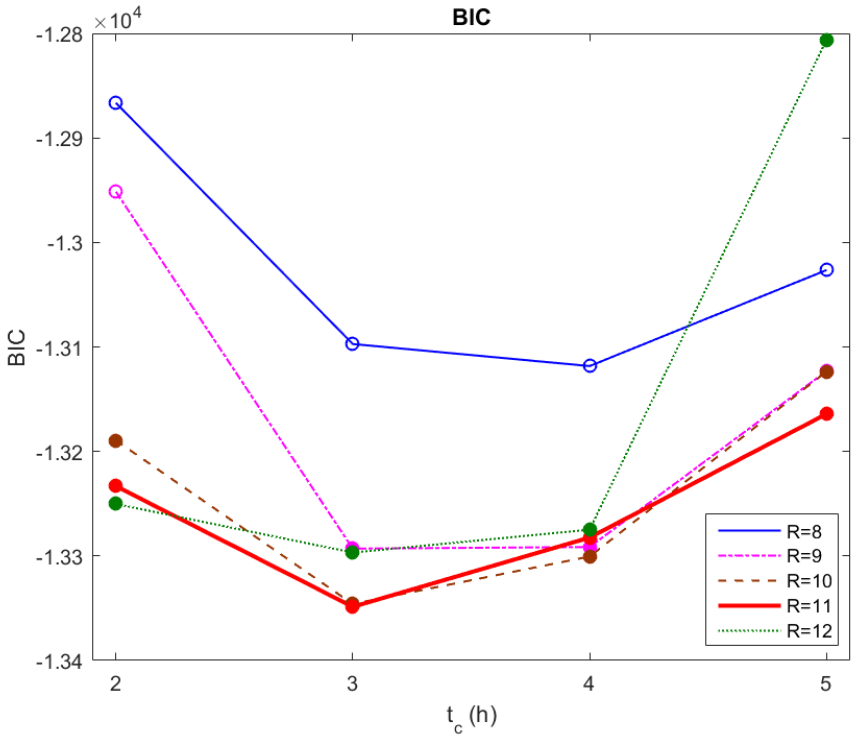

Figure 4-5: BIC values of cubic DRSM-2 models vs t_c for various R values for the increasing MFI case of propylene polymerization. Hollow marker (○) indicates the model has significant LoF (p-value<0.05). Solid markers (●) represent the model with insignificant LoF (p-value>0.05). All the models satisfy the sampling constraint.

We apply analysis of residuals to examine the quality of the selected best model. The residuals between the data, y, and the predicted values by the DRSM-2 model,$\hat{y}$, is calculated by $error = y - \hat{y}$. The model with desired quality should have residuals that are normally distributed and independent from the factors and dynamic sub-factors.

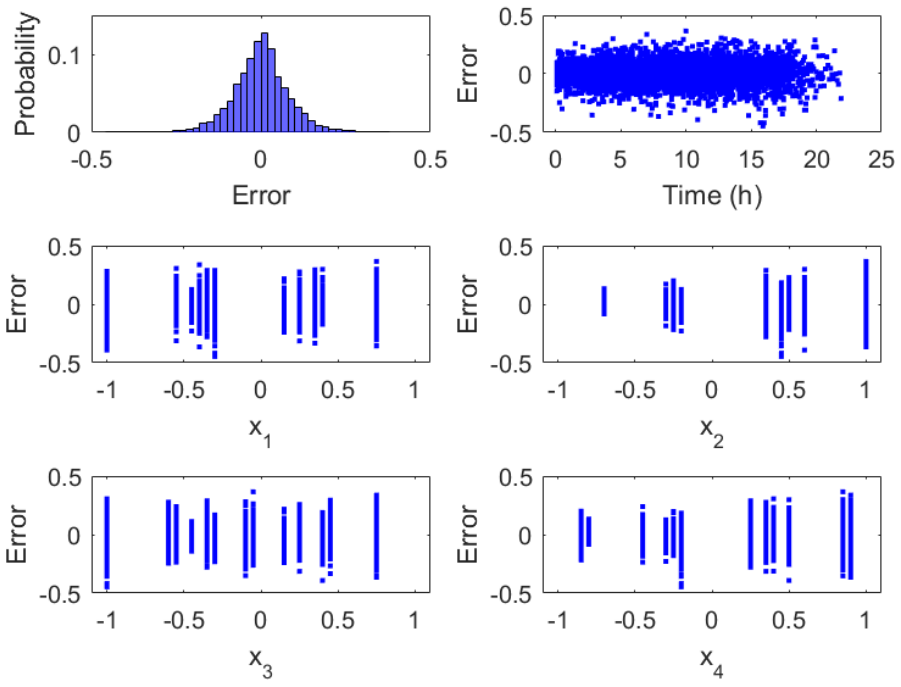

Figure 4-6: Analysis of residuals of the selected DRSM-2 model for the case of increasing MFI.

In the top-left sub-figure of Figure 4-6, we plot the histogram. The x-axis is the value of residuals, error, while the y-axis is the probability of the corresponding residual value. It has been shown that the residuals are normally distributed. We plot the residual against the time in the top-right sub-figure. We can tell that the residuals are independent of time. In the rest for sub-figures, we plot the residuals against each of the four sub-factors, x_1, x_2, x_3, and x_4. The values of the residuals are independent of the values of the sub-factors as well except for x_2 which usually determines the highest MFI value during each transition in the case of increasing MFI. As the noise added to the simulated data is fractional noise, it is reasonable that the variance of the residuals is smaller when the MFI value and the x_2 value is smaller. Based on the results of residual analysis, the obtained DRSM-2 model exhibits consistent prediction accuracy over the input domain.

To further examine the accuracy of the obtained data-driven model, we define a cross-validation set of four new experiments which are not used for estimating the DRSM-2 model. The comparison of the DRSM-2 model predictions and the measured values are shown in Figure 4-7.

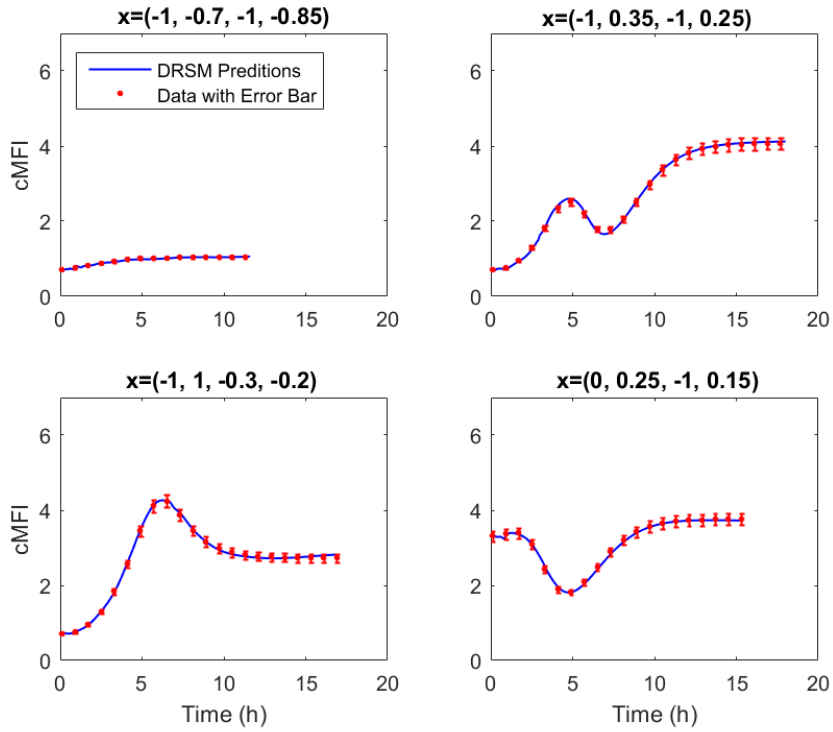

Figure 4-7: Comparison between the measurements (•) and the DRSM-2 model predictions (—) in the cross-validation set for the increasing MFI case of propylene polymerization. Only one out of eight measured values are marked in red dots along with their corresponding error bars.

The values of the factors parameterizing each input profile is shown in the title of each sub-figure. The durations of the four runs are quite different ranging from 11.6 *h* (upper left) to 17.9 *h* (upper right). So that we do not crowd the figure, we only plot every 8^{th} measured data point, along with the corresponding error bars. The corresponding DRSM-2 model predictions are plotted in a blue line. In all four runs, the predicted values overlap with the data and always lie within the error bars, which indicates that the DRSM model is very accurate. Whether such a very accurate model is

necessary largely depends on the task that it will be used for. Quite possibly the models that will be used for control purposes would not require such high accuracy.

4.2.2. DRSM-2 Model for Decreasing MFI Transitions

We also developed the DRSM-2 model for the case of decreasing MFI transitions. The time-varying input profile consisting of three ramps and one final steady-state value is in the same input range as the one for the increasing MFI operations. The mathematic expression is similar to eq. (4-23) but with the new ramp rate, a', defined by $a' = -a$. The time instants, t_{1-3}, and the time delays, d_{1-4} are calculated using eq.(4-24) and eq. (4-29) with the new ramp rate, a', respectively. The factors, $x_i's$, are defined in eq. (4-25) as well. Four set of experiments for estimating the DRSM-2 models in four model classes, linear, 2FI, quadratic, cubic, are designed in the domain defined by the following constraints.

$$\begin{cases} x_1 - x_4 \geq \varepsilon \\ x_4 - x_2 \geq \varepsilon \\ x_3 - x_4 \geq \varepsilon \end{cases} \text{ with } \varepsilon = 0.1 \tag{4-31}$$

Two of the input and corresponding output profiles among the designed experiments are plotted in Figure 4-8. As discussed previously, this can be mitigated by a more considered design domain than the one in eq.(4-31). We will show later again that we have obtain accurate DRSM-2 model for the decreasing MFI case using the data with overshoot. The same fractional noise has been added to the simulated value as given in eq. (4-28).

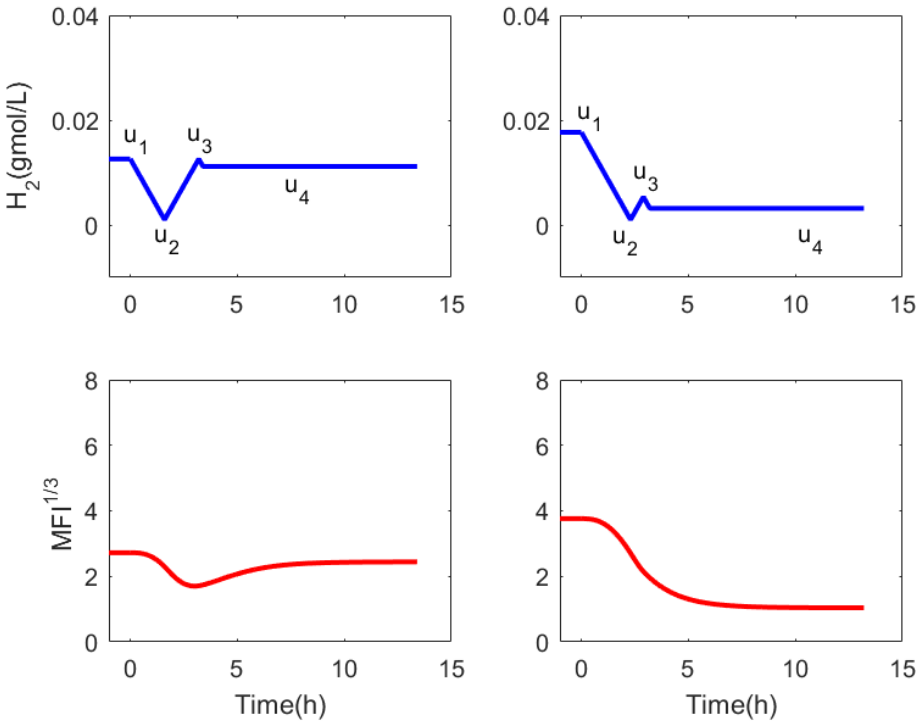

Figure 4-8: Example of the process responses (lower) to two different input profiles (upper) for the decreasing MFI case of propylene polymerization. The responses in the left indicate an overshoot.

The statistics of the best obtained model of each model class are given in row 9-12 of Table 4-1. Based on the *p*-value, we arrive at the best DRSM-2 model of the cubic model class. The corresponding values of R and t_c are 11 and 3 h, respectively. We note that these values are the same ones arrived at for the increase MFI model. We verify the accuracy of the obtained DRSM-2 model using four cross-validation runs. As shown in Figure 4-9, the durations of the four runs are quite different ranging from 11.6 h (upper left) to 18.3 h (upper right). We here again plot every 8^{th} measured data point, along with the corresponding error bars. The corresponding DRSM-2 model predictions are plotted in a blue line. In all four runs. The predicted values almost overlap

with the data and almost lie within the error bars, which indicates that the DRSM-2 model for the decreasing MFI case is very accurate as well.

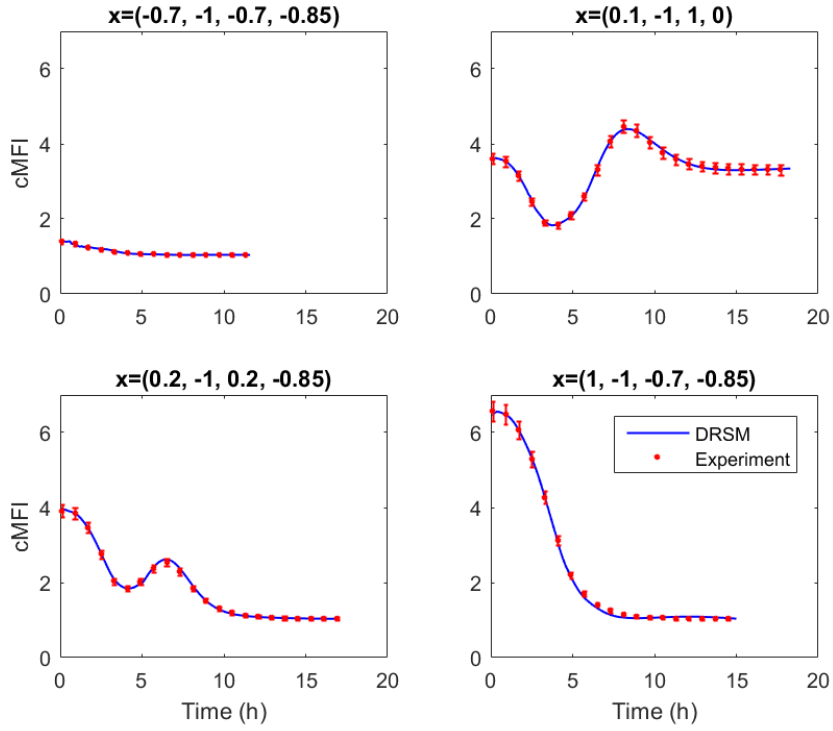

Figure 4-9: Comparison between the measurements (•) and the DRSM-2 model predictions (—) in the cross-validation set for the decreasing MFI case of propylene polymerization. Only one out of eight measured values are marked in red dots along with their corresponding error bars

4.3. Modeling Semi-Batch Penicillin Fermentation

In the previous case study, we applied the DRSM-2 methodology to the modeling of continuous processes. We here demonstrate that the new approach is also capable in modeling batch processes using the example of the penicillin fermentation. For the pharmaceutical industry, the process production rate is usually small, comparing to the continuous chemical processes, which makes the development of the knowledge-driven model frequently economically unaffordable. A data-

driven modeling approach, such as the DRSM-2 methodology is an efficient and economic alternative. With such a data-driven dynamic model at hand, one may use it to optimize and control the process to achieve the desired product quality. In the penicillin fermentation, the amount of penicillin produced is mainly determined by the feeding profile of the substrate and the initial concentration of biomass. We here take them as the input variables and the penicillin concentration measured at different time instants as the output variable.

To estimate the DRSM-2 model, we first design and conduct a set of DoDE experiments. We define the factor related to initial biomass concentration varying between 1 and 2 *g/L* as follows

$$x(0) = 1.5 + 0.5w_1 \; with - 1 \leq w_1 \leq +1 \tag{4-32}$$

The time-varying input variable, the flow rate of the substrate feeding stream, is parameterized using three dynamic sub-factors and is defined as [31]

$$u(\tau) = u_0(\tau) + \Delta u(\tau)[x_1 P_0(\tau) + x_2 P_1(\tau) + x_3 P_2(\tau)] \tag{4-33}$$

where the dimensionless time is defined as $\tau = t/t_b$, and $u_0(\tau) = \Delta u(\tau) = 2\Delta V(1-\tau)/t_b$. The $u_0(\tau)$ is selected to ensure that the constraint on the reactor volume, given below, is met.

$$\int_0^{t_b} u_0(t)dt = \Delta V = V(t_b) - V(0) \tag{4-34}$$

Furthermore, $\Delta u(\tau)$ is so selected to allow $u(\tau)$ to vary between zero and $2u_0(\tau)$. In each experiment, we will fill the maximum allowed capacity of the bio-reactor at 10 *L*, from the initial value of 7 *L* (*i.e.*$\Delta V = 3L$). As our purpose here is to show that the DRSM-2 approach is applicable for modeling batch processes, we will fix the batch time $t_b = 130\ h$ to reduce the total number of

factors and the resulting number of experiments. We impose two additional equality constraints, given as below, which further reduce the number of required experiments.

$$t_b \int_0^1 [u_0(\tau) + \Delta u(\tau) z(\tau)\, u(\tau)] d\tau = \Delta V$$
$$[x_1 P_0(1) + x_2 P_1(1) + x_3 P_2(1)] = 0 \tag{4-35}$$

The first constraint in eq. (4-35) is motivated by the total volume constraint of the bioreactor. The second one is introduced because $u_0(\tau) = \Delta u(\tau) = 0$ at $\tau = 1$. Consequently, there is no reason for $[x_1 P_0(1) + x_2 P_1(1) + x_3 P_2(1)]$ to have any other value than zero. By solving the constraints in eq. (4-35) for x_2 and x_3, we obtain $x_2 = 3x_1$ and $x_3 = -(x_1 + x_2)$. Therefore the number of independent factors is reduced to two, w_1 and x_1. As the process is nonlinear, a D-optimal design of experiments is selected for estimating a quadratic DRSM-2 model. This consists of a minimum of six experiments to estimate the six parametric functions, three additional experiments to assess the LoF statistic, and three replicates to estimate the inherent process variability.

The product concentration is sampled in every 6.5 hours. So the total number of measurements in each experiment is 20. We add to each simulated output value, y_s, a fractional error that is normally distributed.

$$y_m = y_s + N(0, \sigma y_s)\ with\ \sigma = 0.04 \tag{4-36}$$

For the semi-batch process, the initial values of $\mathrm{t_c}$ examined are in the range of 10% to 40% of the batch time of 130 h. We thus estimate the DRSM-2 models with $t_c = 13, 23, 33, 42,$ and $52\ h$. The values of R examined are $R = 6, 7, 8, 9\ 10$. The BIC value of each DRSM-2 model is plotted in Figure 9. According to the figure, the values of t_c and R of the model with the smallest BIC are

42 h and 9, respectively. The corresponding LoF statistic with a p-value = 0.95 strongly indicates that all the non-random information in the data indeed been modeled. Thus, the obtained DRSM-2 model is sufficiently accurate. However, the sampling constraint described in eq. (4-15) is not satisfied. The value of $t_c/R = 4.7\ h$ while the sampling interval $\Delta t = 6.5\ h$. The resulted DRSM-2 model may exhibit oscillatory behavior in the interval between the two samples.

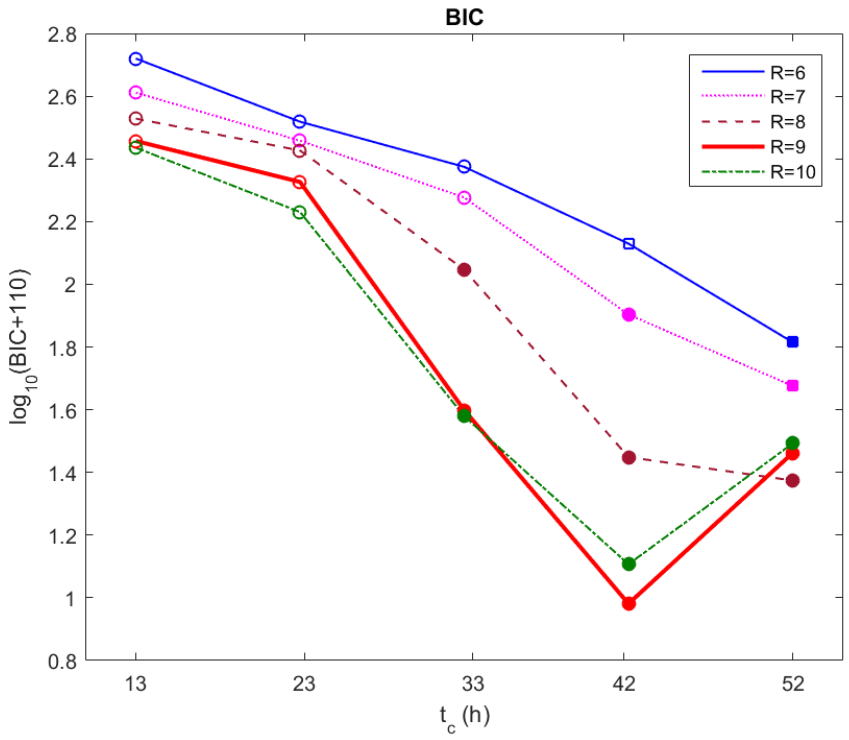

Figure 4-10: Transformed BIC values (y-axis) of quadratic DRSM-2 models vs t_c and R for penicillin fermentation. Hollow marker (○, or □) indicates the model has significant LoF (p-value<0.05). Solid marker (● or ■) represents the model with insignificant LoF (p-value>0.05). Square marker (□ or ■) means the models satisfy the sampling constraint

To illustrate this, we use the DRSM-2 model with $t_c = 42$ hours and $R = 9$ to predict the time-varying profiles of the penicillin concentrations of two new cross-validation runs. The values of the factors for the two new runs are shown in the titles of the sub-figures in Figure 4-10. The

predicted values by the DRSM-2 model are shown in blue line while the simulated values and their corresponding error bars are represented using red diamonds. In general, the DRSM-2 model predictions are quite accurate, as the predicted values almost overlap with the simulated ones at times larger than 6.5 *h* after the start of the batch. However, we observe oscillatory predictions by the DRSM-2 model during the first 6.5 *h*.

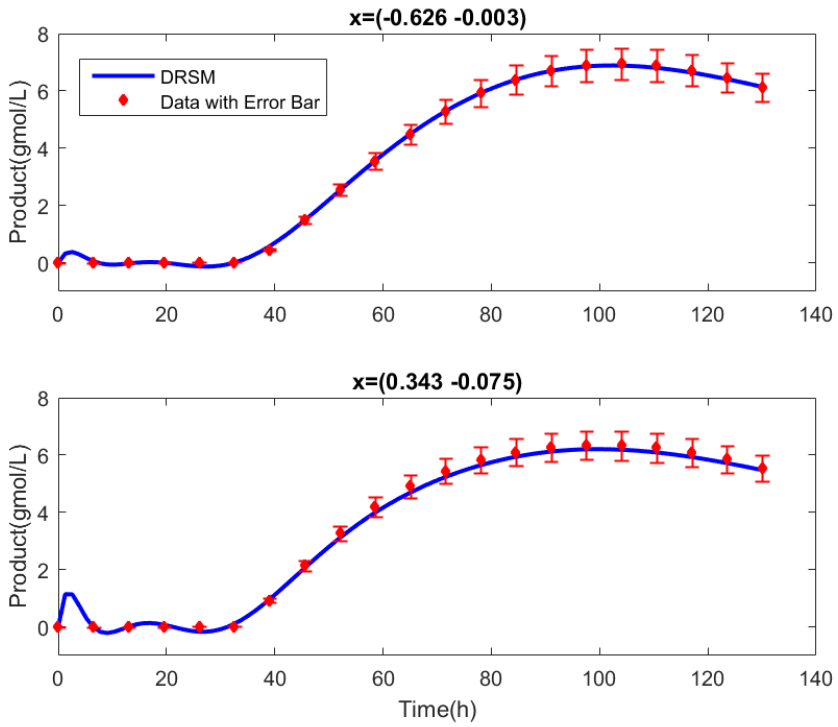

Figure 4-11: Comparison of the simulated data sampled every 6.5 h (♦) and the predicted values (—) by the DRSM-2 model(R=9, t_c=42 h) for penicillin fermentation. The error bars of two standard deviation of the simulated value are also plotted in the figure.

In Table 4-2, we list the DRSM-2 models examined that satisfy the sampling constraint given in eq. (4-15) and their corresponding BIC values. If we choose the model with smallest BIC value from those three models, the selected DRSM-2 model has $t_c = 52$ and $R = 7$. The ratio $t_c/R = 7.4 > 6.5$ *h*. The corresponding LoF statistic is not significant, with a *p*-value=0.44.

Table 4-2: Statistics of the DRSM-2 models satisfying the sampling constraint for the penicillin fermentation

Model #	R	t_c	t_c/R	BIC	LoF p-value
1	6	42	7.0	24.7	0.01
2	6	52	7.7	-44.7	0.21
3	7	52	7.4	-62.6	0.44

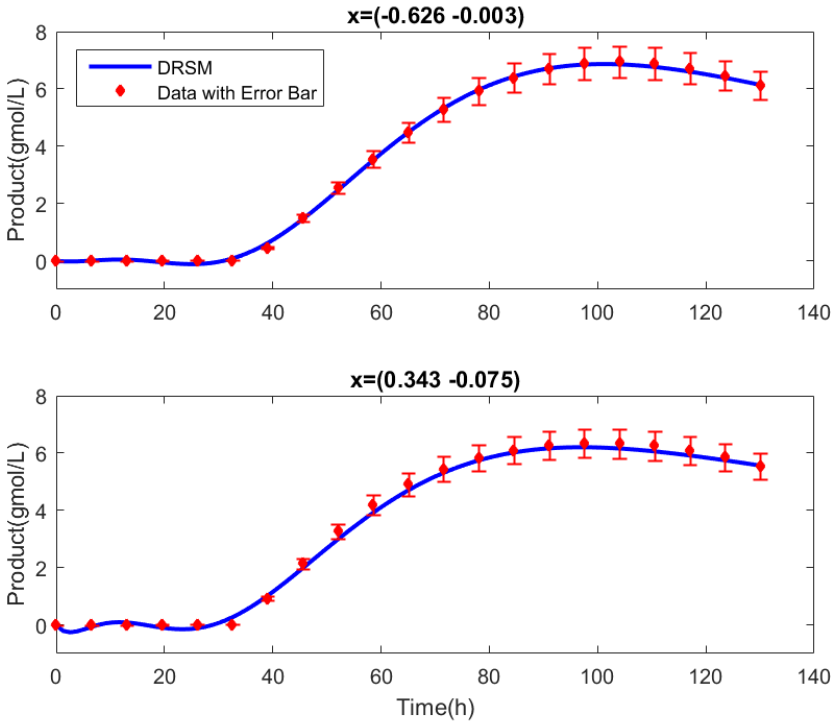

Figure 4-12: Comparison of the simulated data sampled every 6.5 h (♦) and the predicted values (—) by the DRSM-2 model(R=7, $t_c = 52$ h) for penicillin fermentation. The error bars of two standard deviation of the simulated value are also plotted in the figure.

The predictions for the output profiles of the same two cross-validation experiments are plotted in Figure 4-12. It is seen that the oscillations have been dramatically reduced. This validate the the

sampling constraint shown in eq. (4-15) should be satisfied when we select the values for t_c and R and the resulting best DRSM-2 model. Therefore, we should choose the DRSM-2 model with $t_c = 52\ h$ and $R = 7$ instead of the one with $t_c = 42\ h$ and $R = 9$.

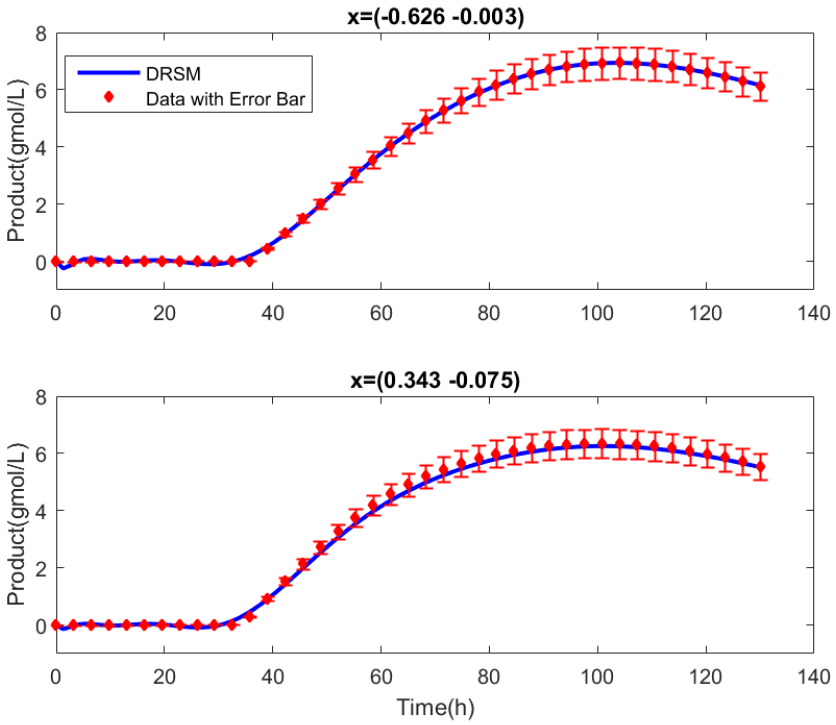

Figure 4-13: Comparison of the simulated data sampled every 3.25 h (♦) and the predicted values (—) by the DRSM-2 model(R=9, t_c=42 h) for penicillin fermentation. The error bars of two standard deviation of the simulated value are also plotted in the figure.

The violation of this sampling constraint can be taken as an indicator that the sampling interval is not sufficiently small. When the data-driven model with the smallest BIC value is achieved with a combination of (t_c, R) for which $t_c/R \leq \Delta t$, we should consider whether it is possible to reduce the sampling interval. To demonstrate this, we reduce the sampling interval from 6.5 h to 3.25 h, taking 40 measurements during each experiment. The obtained best DRSM-2 model has $t_c = 42\ h$,

$R = 9$ and a LoF p-value = 0.22. We compare the predicted and simulated values of the same two cross-validation experiments in Figure 4-11 and plotted them in Figure 4-13.

It is seen that the oscillatory predictions in the first several hours have been significantly reduced to the level even lower than the one shown in Figure 4-12, confirming the benefit of the decrease in the sampling interval. If more frequently measured data is not available, the DRSM-2 model with the smallest BIC value among the ones who satisfy the sampling constraint should be selected.

4.4. Conclusions

We have presented here a new Dynamic Response Surface Methodology, DRSM-2, for the modeling of nonlinear continuous processes and we have also demonstrated that the DRSM-2 approach is also applicable to batch processes. The main idea is to define a new independent variable θ as an exponential transformation of time. With the new variable θ, we can model the time-varying output over a semi-infinite time horizon as easily as over a finite one. The new method also allows the utilization of the data that are not measured equidistantly.

In addition, we proposed a systematic procedure to determine the most appropriate values of the decision variables that influence the structure of the DRSM model. In the original DRSM-1 approach [31], we only considered the polynomial order, R, on the model accuracy. In the new DRSM-2 methodology, three decision variables, the model class C, the time constant t_c as well as R are examined. The choice of the most appropriate values of these decision variables within each model class is determined using two statistics, the BIC and the LoF. We select the DRSM-2 model with the smallest BIC value and an insignificant LoF. Moreover, we pointed out in this paper that

the choices of t_c and R_q, determining the fastest dynamics represented by the DRSM-2 model are also constrained by the sampling interval.

To illustrate the power and versatility of this new methodology, we examined the accuracy of the proposed approach in two nonlinear processes, a continuous propylene polymerization and a semi-batch penicillin fermentation. The presented results clearly demonstrate that the DRSM-2 model provides an accurate representation of the dynamic behaviors of the processes examined. With such a detailed model at hand, one can readily proceed in optimizing the operations of the process with respect to a performance characteristic in a similar manner as discussed previously [32]. The purpose of process control can be achieved using the DRSM model as well. The approach of converting the DRSM model to linear state-space model and nonlinear block-oriented model has been discussed in the same publication [32]. Alternative approaches of utilizing the obtained DRSM-2 model for control purposes will be discussed in forthcoming paper.

5. DRSM-1 for Process Optimization and Control

Data-Driven modelling of batch processes for process optimization and control purposes has attracted substantial interest from both academic and industrial researchers in the past decades. Several recursive model structures representing the nonlinear dynamics of the batch processes have been proposed, including Hammerstein-Wiener (H-W) model and Linear Parameter-Varying (LPV) model. The H-W models, consisting of two static nonlinear blocks in the inputs and outputs and a dynamic linear block in between [55-57], has been applied for the modelling batch processes with linear kinetics and static nonlinear functions on the output, such as pH neutralization [58]. The LPV model[54] introduces the scheduling parameters varying with evolution of state variables in order to approximate bilinear dynamics. The aforementioned models are identified locally through Pseudo Random Binary Signal (PRBS) or Generalized Binary Noise (GBN) [62] experiments in the vicinity of a pre-determined trajectory, possibly an optimal one. However, when the available measurements in a single batch are infrequent, the estimation of such a linear or nonlinear dynamic model of satisfactory accuracy is not feasible. In addition, to determine the optimal trajectory around where the recursive model is identified, a separate data-driven model may be estimated with the cost of additional experiments.

We here proposed process optimization and control method relying on a single DRSM model. The estimation of the DRSM model does not require excessive number of measurements during each batch. The DRSM model will be used to calculate the optimal input trajectory and will also be used here to estimate a recursive model for control purposes in a Model Predictive Controller (MPC) [63]. Because the DRSM model captures both the linear and nonlinear dynamics of the process quite accurately, it can be used to develop either a linear or a nonlinear recursive model.

Therefore, we can achieve both the process optimization and process control using a single DRSM model, which saves the experimental cost of developing separate models for each purpose, respectively. Moreover, it has been shown that when the number of measurements is limited, the MPC using a DRSM model provides improved control performance compared to the one using a state-space model estimated via PRBS experiments.

This chapter is organized as follows. We first discuss the optimization of a nonlinear process using a DRSM model. After the optimal input profile as well as the optimal operating duration is determined, we estimate linear state space model and Hammerstein-Wiener model for control purposes via the DRSM model. The nonlinear MPC controller using Kalman Filter [66] is described in the following section. We examine the proposed approach using an *in silico* isothermal batch reactor with three nonlinear reactions in the Results section and obtained satisfying control performance when measurements are infrequent.

5.1. Process Optimization Using DRSM Model

As the DRSM model captures the process dynamics, the process optimization using a DRSM model can determine the optimal batch durations as well as the optimal input profile without including the operating duration as a factor. In contrast, the optimization for batch duration using a static RSM model requires such a factor which costs more experiments[81]. The optimal input profile and the optimal batch duration, are determined by solving the following constrained optimization problem.

$$(\boldsymbol{x}^*, \tau^*) = \underset{x,\tau}{argmin}\, J(y, \boldsymbol{x}, \tau)$$

$$s.t\ y(\tau) = \beta_0(\tau) + \sum_{i=1}^{3} \beta_i(\tau) x_i + \sum_{i=1}^{3} \sum_{i<j}^{3} \beta_{ij}(\tau) x_i x_j + \sum_{i=1}^{3} \beta_{ii}(\tau) x_i^2 \qquad (5\text{-}1)$$

$$g(\boldsymbol{x}, \tau) \leq 0, lb \leq \boldsymbol{x}, \tau \leq ub$$

Where $J(\boldsymbol{x}, \tau)$ is the cost function. $\boldsymbol{x}$ and τ are the dynamic sub-factors parameterizing the input profile and the dimensionless time, respectively. $g(\cdot)$ are inequality constraints. lb and ub are lower and upper limit of the variables $(\boldsymbol{x}, \tau)$. Once the optimal solution, $\boldsymbol{x}^*$ and τ^*, are obtained, the optimal batch time is given by $t_b^* = \tau^* \times t_b$ while the input profile is determined by substituting them into eq.(2-4).

5.2. Transforming DRSM Models to Hammerstein-Wiener Models

Here we detail the transformation of the nonlinear DRSM model to a nonlinear H-W recursive model appropriate for the design of feedback controllers. This is necessary because the DRSM model is an input-output model and not a recursive one for real-time online use. The advantage of the DRSM model is that it is a non-linear one between the process inputs and its output but linear in its parameters and thus can be estimated using linear regression methods. Its role is also especially important when the number of output measurements during a batch process is very limited, for example when we measure the concentration of reaction species such as products or intermediates. Then the direct estimation of a recursive model is almost infeasible. However, the limited number of the output measurements is not a limitation for the estimation of a DRSM model. With such a model at hand, a much larger set of in silico data can be generated for the estimation of the recursive model. For example, 100 to 500 or more in silico measurements per batch can

easily generated by the DRSM model which in turn might have used as few as 10 to 20 real process measurements for its estimation. The recursive dynamic models that we will identify here will represent deviations from a reference batch profile. We denote the input and output variables of the reference profile by $u^*(\tau)$ and $y^*(\tau)$, respectively. This reference profile might be the optimal one, just calculated through the DRSM model, or any other profile of interest. Initially we will ignore the nonlinear character of the DRSM model and a linear recursive state space model will be estimated for comparison purposes with the Hammerstein-Weaner models to be examined later. In such an estimation, we will use the MOESP [23] subspace identification algorithm to relate $\Delta u(\tau)$ and $\Delta y(\tau)$ in the following discrete model:

$$\boldsymbol{z}_{k+1} = \mathbf{A}\boldsymbol{z}_k + \boldsymbol{b}\Delta u_k \tag{5-2}$$

$$\Delta y_k = \boldsymbol{c}^T \boldsymbol{z}_k \tag{5-3}$$

Where $\boldsymbol{z}_k$ is the state variable at the time instant, $\tau = \tau_k$. The type of linear model will only be accurate enough when the nonlinear component of the DRSM model are not dominant.

As the general character of the DRSM model is a nonlinear one, we will address the estimation of a nonlinear recursive model next. The interest here is the Hammerstein-Wiener (H-W) type of model. The H-W models consisting of static nonlinear input and output blocks and a dynamic linear block in between are shown in Figure 2-1. The functional blocks $f(\cdot)$ and $g(\cdot)$ represent the static input and output nonlinearities while the linear block represents the dynamics between the intermediate input $\Delta\bar{u}_k$ and intermediate output $\Delta\bar{y}_k$. The input exciting the process is Δu_k and Δy_k is the corresponding process output. Here we will assume that the nonlinear bocks $f(\cdot)$ and $g(\cdot)$ are of polynomial form.

The mathematic representations of each block for SISO case are shown in eq. (5-4)-(5-7).

$$\Delta\bar{u}_k = f(\Delta u_k) = \sum_{i=1}^{I} \alpha_i \Delta u_k^i \tag{5-4}$$

$$\boldsymbol{z}_{k+1} = \mathbf{A}\boldsymbol{z}_k + \boldsymbol{b}\Delta\bar{u}_k \tag{5-5}$$

$$\Delta\bar{y}_k = \boldsymbol{c}^T \boldsymbol{z}_k \tag{5-6}$$

$$\Delta y_k = g(\Delta\bar{y}_k) = \sum_{j=1}^{J} \kappa_j \Delta\bar{y}_k^j \tag{5-7}$$

As given in eq (5-4) and (5-7), the static input and output nonlinearities are in the polynomials of order I^{th} and J^{th} order respectively. The coefficients α_i and κ_j representing the input and output nonlinearities are going to be estimated using the DRSM model. For simplicity and without losing generality, we let $I = J$ the rest of the chapter. For the case of $I > J$ (or $I < J$), the derived results still apply if we let $\kappa_{j>J} = 0$ (or $\alpha_{i>I} = 0$).

With a DRSM model at hand, we identify a Hammerstein-Wiener model in two steps: 1) estimating nonlinearity coefficients, $\alpha_i{}'$s and $\kappa_j{}'$s, from the DRSM model; 2) identifying linear block using the subspace identification method [33]. As discussed in literature [98], we let $\alpha_1 = \kappa_1 = 1$, to obtain the unique set of the nonlinearity coefficients of the identified H-W model. The nonlinearity of the identified H-W model is up to the nonlinearity of the DRSM model. For example, with a quadratic or a cubic DRSM model at hand, we can identify an H-W model with quadratic or cubic nonlinearities, respectively.

To avoid burdening the reader with excessive algebra, we present the algorithm using a quadratic DRSM model with three dynamic sub-factors parametrizing the input, as given in (2-9) with n=3.

The deviations of the input and output profiles from the corresponding optimal trajectories in discrete forms are given by

$$\Delta u_k = \Delta u_{0,k}(\Delta x_1 P_{0,k} + \Delta x_2 P_{1,k} + \Delta x_3 P_{2,k}) \tag{5-8}$$

$$\Delta y_{DRSM,k} = \sum_{i=1}^{3} \beta_{i,k} \Delta x_i + \sum_{i=1}^{3} \sum_{i<j}^{3} \beta_{ij,k} \Delta x_i \Delta x_j + \sum_{i=1}^{3} \beta_{ii,k} \Delta x_i^2 \tag{5-9}$$

where $\Delta x_i = x_i - x_i^*$ for $i = 1, 2, 3$ is the deviations of the dynamic sub-factors from their optimal values. The quadratic input and output nonlinearities coefficients of the H-W model, $\boldsymbol{h} = (\alpha_2 \quad \kappa_2)^T$, are going to be estimated from the DRSM model via least square regression as follows,

$$\boldsymbol{h} = argmin \|\mathbf{\Delta Y_{DRSM}}(\Delta \boldsymbol{x}) - \mathbf{\Delta Y_{H-W}}(\Delta \boldsymbol{x}, \boldsymbol{h})\|_F^2 \tag{5-10}$$

$\|\cdot\|_{\mathrm{F}}^2$ is the Frobenius norm[99]. $\Delta\mathbf{Y}_{\mathrm{DRSM}}$ and $\Delta\mathbf{Y}_{\mathrm{H-W}}$ are the $K \times M$ matrices of the outputs predicted by DRSM and H-W model to be identified, respectively, assuming that there are M simulated runs and K samples during each run. In the case study examined in section 5.3, we choose $K = 100$ and $M = 16$. To estimate the input and output nonlinearity coefficients of the H-W model, we will express $\mathbf{\Delta Y}_{\mathrm{DRSM}}$ and $\mathbf{\Delta Y}_{\mathrm{H-W}}$ as a function of $\Delta \boldsymbol{x}$ and $\boldsymbol{h}$ in the following sub-sections. As it will become apparent in the following subsections, the identification of the H-W model will be done in an iterative fashion. Initially, a linear dynamic state space model will be identified ignoring the H-W nonlinear blocks. Then the nonlinear parameters of the vector $\boldsymbol{h}$ of the H-W blocks will be identified through a modified version of eq. (5-10). Once an estimate of the two nonlinear static blocks is at hand, then the intermediate inputs $\Delta\bar{u}_k$ can be calculated from the Δu_k values through eq. (5-4) and the values of $\Delta\bar{y}_k$ can be *back*-calculated from the Δy_k values through eq.(5-5). With the $\Delta\bar{u}_k$ and $\Delta\bar{y}_k$ values at hand an updated linear state space model can

be estimated. These two interacting steps are repeated until the algorithm converges and no update in the model is obtained. In the following we present the details of this iterative algorithm.

5.2.1. Formulation of the ΔY_{DRSM}

We first detail the dependence of $\mathbf{\Delta Y}_{\mathrm{DRSM}}$ on the process inputs by rewriting the eq (5-9) in matrix form

$$\mathbf{\Delta Y_{DRSM}} = (\mathbf{\Psi \Gamma})\mathbf{\Delta X} \tag{5-11}$$

The $(p-1) \times M$ matrix $\mathbf{\Delta X}$ consisting of the deviation values of the dynamic sub-factors in the simulated runs is given by

$$\mathbf{\Delta X} = [\mathbf{\Delta x}_1 \quad \mathbf{\Delta x}_2 \quad \cdots \quad \mathbf{\Delta x}_{\mathrm{M}}] \tag{5-12}$$

Each column in $\mathbf{\Delta X}$ consists of $p-1$ elements and is given as follows

$$\mathbf{\Delta x}_{\mathrm{m}} = [\Delta x_{m,1} \cdots \Delta x_{m,3} \ \Delta x_{m,12} \cdots \Delta x_{m,23} \ \Delta x_{m,11} \cdots, \Delta x_{m,33}]^{\mathrm{T}} \tag{5-13}$$

With $\Delta x_{ij} = \Delta x_{m,i}\Delta x_{m,j}$ and $\Delta x_{ii} = \Delta x_{m,i}\Delta x_{m,i}$. The scalar $x_{m,i}$ in the above matrices is the values of the i^{th} factor for the m^{th} experiment. $\mathbf{\Psi}$ is the $K \times (R+1)$ matrix consists of the values of shifted Legendre polynomials defined as follows

$$\mathbf{\Psi} = \begin{bmatrix} P_{0,0} & P_{1,0} & \cdots & P_{R,0} \\ P_{0,1} & P_{1,1} & \cdots & P_{R,1} \\ \vdots & \vdots & \ddots & \vdots \\ P_{0,K-1} & P_{1,K-1} & \cdots & P_{R,K-1} \end{bmatrix} \tag{5-14}$$

$\boldsymbol{\Gamma}$ is a $(R+1)\times(p-1)$ matrix consisting the DRSM model parameters and is defined in as follows

$$\boldsymbol{\Gamma} = [\boldsymbol{\gamma_1} \quad \boldsymbol{\gamma_2} \quad \boldsymbol{\gamma_3} \quad \boldsymbol{\gamma_{12}} \quad \boldsymbol{\gamma_{13}} \quad \boldsymbol{\gamma_{23}} \quad \boldsymbol{\gamma_{11}} \quad \boldsymbol{\gamma_{22}} \quad \boldsymbol{\gamma_{33}}] \tag{5-15}$$

Where each column, $\boldsymbol{\gamma_q}$, is a $(R+1)\times 1$ column vector $\boldsymbol{\gamma_q} = (\gamma_{q,1} \quad \gamma_{q,2} \quad \cdots \quad \gamma_{q,R})^T$ for $q = i, ij, ii$.

5.2.2. Formulation of the ΔY_{H-W}

To obtain the expression of $\boldsymbol{\Delta Y_{H-W}}$, we first rewrite the H-W model in the input-output relation form as given below by substituting eq. (5-4)-(5-6) and eq. (5-8) into eq. (5-7)and letting the initial state variable $\mathbf{z_0} = \mathbf{0}$,

$$\Delta y_{H-W,k}(x,h) = \sum_{j=1}^{2} \kappa_j \left(\boldsymbol{c}^T \sum_{\eta=1}^{k} \mathbf{A}^k \boldsymbol{b} \sum_{i=1}^{2} \alpha_i \left(\Delta u_{0,k} \sum_{s=1}^{3} \Delta x_s P_{s-1,k-\eta} \right)^i \right)^j \tag{5-16}$$

Where $\Delta y_{H-W,k}$ is the output predicted by the H-W model at time instant k. By expanding eq. (5-16) and letting $\alpha_1 = \kappa_1 = 1$, we rearrange the input-output relation of the H-W model into the similar form of the DRSM model.

$$\begin{aligned} \Delta y_{H-W,k} = \Delta u_{0,k} \sum_{j=1}^{3} \varphi_{j,k} \Delta x_j + 2\Delta u_{0,k}^2 \sum_{j=1}^{3} \sum_{s>j}^{3} \left(\alpha_2 \varphi_{js,k} + \kappa_2 \varphi_{j,k} \varphi_{s,k}\right) \Delta x_j \, \Delta x_s \\ + \Delta u_{0,k}^2 \sum_{j=1}^{3} \left(\alpha_2 \varphi_{jj,k} + \kappa_2 \varphi_{j,k}^2\right) \Delta x_j^2 + O(\Delta \boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\kappa}) \end{aligned} \tag{5-17}$$

Where $O(\Delta \boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\kappa})$ is the higher order (>2) term of $\Delta x_i{}'$s which cannot be estimated through the intermediary DRSM model. The scalars $\varphi_{j,k} = \boldsymbol{\psi}_{j,k}\boldsymbol{w}$, $\varphi_{js,k} = \boldsymbol{\psi}_{js,k}\boldsymbol{w}$, $\varphi_{jj,k} = \boldsymbol{\psi}_{jj,k}\boldsymbol{w}$, with $\boldsymbol{w} = (\boldsymbol{c}^T\mathbf{A}^0\boldsymbol{b} \quad \boldsymbol{c}^T\mathbf{A}^1\boldsymbol{b} \quad \dots \quad \boldsymbol{c}^T\boldsymbol{A}^{K-1}\boldsymbol{b})^T$ is a $K \times 1$ column vector. $\boldsymbol{\psi}_{j,k}$ and $\boldsymbol{\psi}_{ij,k}$ are the k^{th} row of the upper triangular matrices $\boldsymbol{\Psi}_{\mathbf{j}}$ and $\boldsymbol{\Psi}_{\mathbf{ij}}$ which are given by

$$\boldsymbol{\Psi}_{\mathbf{j}} = \begin{bmatrix} P_{i,0}\Delta u_{0,0} & 0 & \cdots & 0 \\ P_{i,1}\Delta u_{0,1} & P_{i,0}\Delta u_{0,0} & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ P_{i,K-1}\Delta u_{0,K-1} & P_{i,K-2}\Delta u_{0,K-2} & \cdots & P_{i,0}\Delta u_{0,0} \end{bmatrix} \text{ and } \boldsymbol{\Psi}_{\mathrm{ij}} = \boldsymbol{\Psi}_{\mathrm{i}} \circ \boldsymbol{\Psi}_{\mathrm{j}} \tag{5-18}$$

Here $\circ$ is Hadamard product[91] and K is the number of the simulated data in each experiment.

By rewriting the eq. (5-17) in matrix form, the $\Delta\mathbf{Y}_{\mathrm{H-W}}$ is expressed as follows,

$$\Delta\mathbf{Y}_{\mathrm{H-W}} = \boldsymbol{\Phi}\Delta\mathbf{X} \tag{5-19}$$

Where the matrix $\boldsymbol{\Phi}$ is defined as follows

$$\boldsymbol{\Phi} = [\boldsymbol{\varphi}_1 \ \boldsymbol{\varphi}_2 \ \boldsymbol{\varphi}_3 \ (\boldsymbol{\varphi}_{12}\boldsymbol{h}) \ (\boldsymbol{\varphi}_{13}\boldsymbol{h}) \ (\boldsymbol{\varphi}_{23}\boldsymbol{h}) \ (\boldsymbol{\varphi}_{11}\boldsymbol{h}) \ (\boldsymbol{\varphi}_{22}\boldsymbol{h}) \ (\boldsymbol{\varphi}_{33}\boldsymbol{h})] \tag{5-20}$$

With $\boldsymbol{\varphi}_i = \boldsymbol{\Psi}_{\mathrm{i}}\boldsymbol{w}$ is $K \times 1$ column vector while $\boldsymbol{\varphi}_{ij} = 2[\boldsymbol{\Psi}_{ij}\boldsymbol{w} \ \boldsymbol{\Psi}_i\boldsymbol{w} \circ \boldsymbol{\Psi}_j\boldsymbol{w}]$ and $\boldsymbol{\varphi}_{ii} = [\boldsymbol{\Psi}_{\mathrm{ii}}\boldsymbol{w} \ \boldsymbol{\Psi}_{\mathrm{i}}\boldsymbol{w} \circ \boldsymbol{\Psi}_{\mathrm{i}}\boldsymbol{w}]$ are $K \times 2$ matrices.

5.2.3. Estimation of the Nonlinearity Parameters in h

As we are interested to select the nonlinearity coefficients, $\boldsymbol{h}$, that minimizes the $\|\Delta\mathbf{Y}_{\mathrm{DRSM}} - \Delta\mathbf{Y}_{\mathrm{H-W}}\|_{\mathrm{F}}^2 = \|(\boldsymbol{\Psi\Gamma} - \boldsymbol{\Phi})\Delta\mathbf{X}\|_{\mathrm{F}}^2$ regardless the choice of $\Delta\mathbf{X}$, we estimate the value of h by solving the following optimization problem.

$$(\boldsymbol{h},\boldsymbol{w}) = argmin\, J(\boldsymbol{h},\boldsymbol{w}) = \|(\boldsymbol{\Psi\Gamma}) - \boldsymbol{\Phi}\|_F^2 \tag{5-21}$$

According to the submultiplicative property of the Frobenius norm[100], $\|(\boldsymbol{\Psi\Gamma})\Delta\mathbf{X} - \boldsymbol{\Phi}\Delta\mathbf{X}\|_{\mathrm{F}}^2 \leq \|(\boldsymbol{\Psi\Gamma}) - \boldsymbol{\Phi}\|_{\mathrm{F}}^2 \times \|\Delta\mathbf{X}\|_{\mathrm{F}}^2$. The obtained $\boldsymbol{h}$ by eq. (5-21) minimizes the upper bond of the mismatch between the DRSM model and the H-W model for any given $\Delta\mathbf{X}$.

As the above optimization problem is nonlinear on $\boldsymbol{h}$ and $\boldsymbol{w}$, we will solve $\boldsymbol{w} = argmin[J(\boldsymbol{w}|\boldsymbol{h})]$ and $\boldsymbol{h} = argmin[J(\boldsymbol{h}|\boldsymbol{w})]$ for $\boldsymbol{w}$ and $\boldsymbol{h}$ iteratively. The procedure for estimating the nonlinearities are summarized as follows:

Step 1, let $\boldsymbol{h} = 0$, estimate the initial $\boldsymbol{w}^{(0)}$ by solving $\boldsymbol{w} = argmin[J(\boldsymbol{w}|\boldsymbol{h})]$. The obtained solution is $\boldsymbol{w}^{(0)} = \left[(\boldsymbol{\Psi}_1^T\ \boldsymbol{\Psi}_2^T\ \boldsymbol{\Psi}_3^T)^T\right]^* [(\boldsymbol{\Psi\gamma}_1)^T\ (\boldsymbol{\Psi\gamma}_2)^T\ (\boldsymbol{\Psi\gamma}_3)^T]^T$.

Step 2, solve $\boldsymbol{h} = argmin[J(\boldsymbol{h}|\boldsymbol{w})]$ for $\boldsymbol{h}$ with $\boldsymbol{w} = \boldsymbol{w}^{(0)}$. The obtained solution is, $\boldsymbol{h}^{(0)} = \left[\left(\boldsymbol{\varphi}_{12}^T\ \boldsymbol{\varphi}_{13}^T\ \boldsymbol{\varphi}_{23}^T\ \boldsymbol{\varphi}_{11}^T\ \boldsymbol{\varphi}_{22}^T\ \boldsymbol{\varphi}_{33}^T\right)^T\right]^* \times$

$$[(\boldsymbol{\Psi\gamma}_{12})^T\ (\boldsymbol{\Psi\gamma}_{13})^T\ (\boldsymbol{\Psi\gamma}_{23})^T\ (\boldsymbol{\Psi\gamma}_{11})^T\ (\boldsymbol{\Psi\gamma}_{22})^T\ (\boldsymbol{\Psi\gamma}_{33})^T]^T$$

Step 3, update $\boldsymbol{w}$ by solving $\boldsymbol{w}^{(k)} = argmin[J(\boldsymbol{w}^{(k)}|\boldsymbol{h}^{(k-1)})]$

Step 4, update $\boldsymbol{h}$ by solving $\boldsymbol{h}^{(k)} = argmin[J(\boldsymbol{h}^{(k)}|\boldsymbol{w}^{(k)})]$

Step 5, Repeat Step 3-4 until $\|\boldsymbol{h}^{(k)} - \boldsymbol{h}^{(k-1)}\| < \varepsilon$ where ε is a small threshold value for determining convergence.

Though we presents identifying a quadratic SISO Hammerstein-Wiener model via a quadratic DRSM example here, it can be generalized to dealing with higher order nonlinearities if a DRSM model of corresponding nonlinearity is at hand.

5.2.4. Model Predictive Control

For batch processes, we are interested in achieving the desired product concentration or qualities at the end of the each run. Therefore, we design the following receding horizon MPC controller, in which the prediction horizon decreases as the process proceeds in time.

$$\begin{aligned}
&\min_{\Delta u_{i|k}} \left\{ q_1 \tilde{y}_{N|k}^2 + q_2 \sum_{i=k+1}^{k+N_p-1} \tilde{y}_{i|k}^2 + w \sum_{i=k+1}^{k+N_p-1} \delta u_{i|k}^2 \right\} \\
&s.t\ \delta u_{i|k} = \Delta u_{i|k} - u_{i-1|k};\ \tilde{y}_{i|k} = \Delta y_{sp,i} - \Delta \hat{y}_{i|k} \\
&\quad \Delta \hat{y}_{i|k} = g\left(\Delta \bar{y}_{i|k}\right);\ \ \Delta u_{i|k} = f^{-1}(\Delta \bar{u}_{i|k}) \\
&\quad \hat{\boldsymbol{z}}_{i+1|k} = \mathbf{A}\hat{\boldsymbol{z}}_{k|k} + \boldsymbol{b}\Delta \bar{u}_{i|k};\ \ \Delta \bar{y}_{i|k} = \boldsymbol{c}^T \hat{\boldsymbol{z}}_{i|k};\ \ u^L \le u_{0i} + \Delta u_{i|k} \le u^U
\end{aligned} \tag{5-22}$$

Here $y_{sp,i}$ is the set points, the optimal trajectory at time instance i. w is the control action weight, u^L and u^U are the lower and upper control limit. The subscript $i|k$ indicates the estimation of the variable value of future instance i while the estimation is done at instance k. A Kalman Filter[66] is used to update $\hat{\boldsymbol{z}}_{k|k}$ using the new measurement made at k. The function $g\left(\Delta \bar{y}_{i|k}\right)$ in the output nonlinearity function while $f^{-1}(\Delta \bar{u}_{i|k})$ is the inverse function of the input nonlinearity function. The nonlinear MPC here requires that the input nonlinear function to be invertible. In the MPC using linear state-space model, the intermediate input and output are identical with the input and output variables, i.e. $\Delta \hat{y}_{i|k} = \Delta \bar{y}_{i|k}$ and $\Delta u_{i|k} = \Delta \bar{u}_{i|k}$. So the controller discussed in eq. (5-22) is still applicable.

As the recursive models, include the linear and H-W model are identified by "sampling" the DRSM model with a much higher frequency than the actual measurement frequency, the state variable will be updated when there is an available measurement. For example, the recursive models are identified by sampling the DRSM every 9 *min*, while the real measurements are available every 45 *min*. Consequently, the state variables will be updated every 45 minutes. The tuning parameters, for MPC controller, are N_p, N_u, q_1, q_2 and w. We use $N_p = N - k$, $N_u = min(10, N_p)$ and w=0.05. Since MPC controller the aims to control the end product concentration, the weight for mismatch at the end of the batch is ten times larger than those during the batch. Therefore $q_1 = 100$ and $q_2 = 10$. For the MPC with linear model identified using PRBS experiments (sampling every 45 *mins*), the equivalent controller settings are $N_p' = 0.2N_p$, $N_u' = min(2, N_p')$ and $q_2 = 50$.

5.3. Results and Discussions

We demonstrate the efficacy of the proposed method *in silico* by considering a reaction example in a semi batch reactor where one of the reactants is fed over a time period during the batch. We consider two cases of optimal operation; the first one maximizes the product weight at the end of the batch and the second one maximizes the product yield, defined as the moles of product per moles of reactants fed. We will control the process so it reaches the desired product concentration in the above two optimal operating conditions. The batch process is described as follows:

Rxn1: $A + B \rightarrow C$, $r_1 = k_1[A][B]$ with $k_1 = 0.13$ l gmol h^{-1}

Rxn2: $2B \rightarrow D$, $r_2 = k_2[B]^2$ with $k_2 = 0.08$ l gmol h^{-1}

Rxn3: $C \rightarrow E$, $r_3 = k_3[C]$ with $k_3 = 0.07$ h^{-1}

We assume a reactor volume of 10 L. Reactant A is loaded at the beginning with an initial concentration of 5.0 *gmol/L* while reactant B is fed in semi-batch mode to reduce the production of by-product D. The batch time, t_f, is 15 h. In practice, one will select the batch time base on the available knowledge of the process. If the optimal batch time is one the boundary of the constraint, one can increase the batch time and improve the process further using augmented experiments.

The decision variables for designing the experiments are the amount of reactant B to be fed and its feeding profile. We set 75 *gmol* as the reference value for the total amount of B fed, and we will vary it in the range between 50 and 100 *gmol*. Then the total amount of B fed, B_T is given by $B_T = 75 + 25a$ with $-1 \leq a \leq +1$. We define a reference feeding profile, $u_0(t)$, which uses the reference amount of B fed. In order to satisfy these conditions, $u_0(t)$, will be constrained by $\int_0^{t_f} u_0(t)dt = 75$. There are many possible choices for $u_0(t)$, which satisfy the above constraint. Without any prior information, it is reasonable to choose the simplest meaningful profile, a linear one. As B is a reactant, we choose a decreasing linearly profile in which the initial feed of B is 3 times that at the end of the batch, i.e. $u_0(0) = 3u_0(t_b)$. This results into the following reference feeding profile: $u_0(\tau) = 7.5 - 5\tau$, in *gmol/h*. We select $\Delta u_0(\tau) = u_0(\tau)$.The flow rate of B for the experiments is defined by substitute $u_0(\tau)$ and $\Delta u_0(\tau)$ in eq.(2-5). Because $-1 \leq w(\tau) \leq +1$, the input profile will vary in the domain of $[0,\ 2u_0(\tau)]$. The dynamic factor, $w(\tau)$, will be parameterized using the first three shifted Legendre polynomials as given in eq. (2-6). Besides the constraints, $-1 \leq w(\tau) = x_1 \pm x_2 \pm x_3 \leq +1$, one additional constraint must be imposed on $u(\tau)$, to ensure that the amount of reactant B, fed is the desired one:

$$t_f \int_0^1 \left(u_0(\tau) + \Delta u(\tau) w(\tau)\right) dt = 75 + 25a \tag{5-23}$$

This results into $a = 3x_1 - 0.5x_2$ and reduces the number of independent factors from 4 to 3, namely x_1, x_2. and x_3. Since $-1 \leq a \leq +1.$, we have $-1 \leq 3x_1 - 0.5x_2 \leq +1$. We design a set of 16 experiments for fitting a quadratic DRSM with 3 factors by systematically varying the values of $x_1, x_2$and x_3 satisfying constraints: aforementioned constraints.

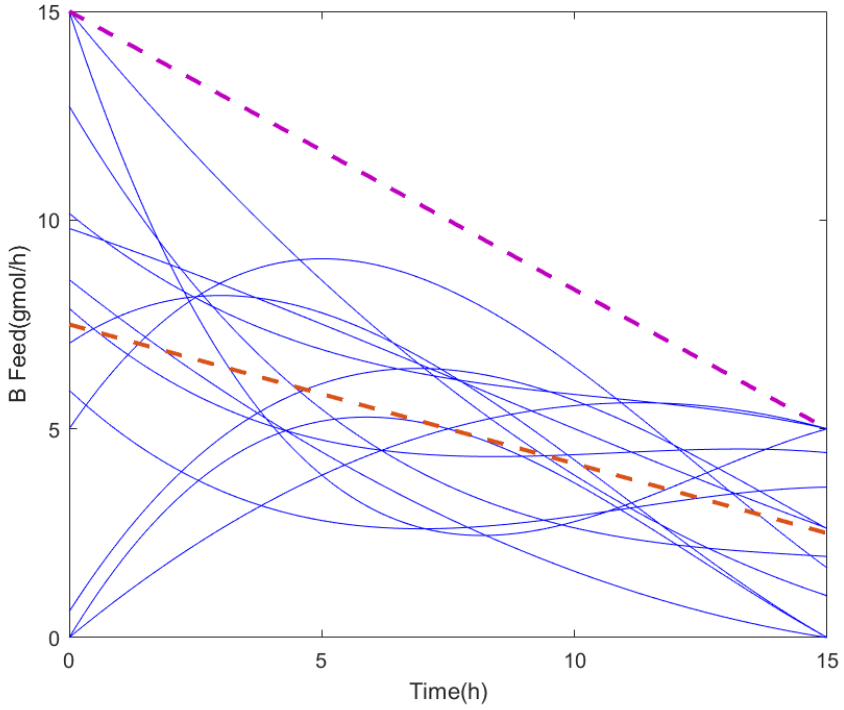

Figure 5-1: Input profiles (—) of reactant B parameterized by 3 dynamic sub-factors. The upper limit (--) and reference case (--) are also plotted as well.

The experiments are determined using a D-Optimal design. Among the 16 experiments, 10 are needed to estimate the 10 parametric functions, $\beta_q(\tau)'$s, 3 experiments to estimate the Lack-of-Fit (LoF) statistic and another 3 are replicates for the estimation of the normal variability of the process.

The designed input profiles of the selected experiments are shown in Figure 5-1. The dashed lines are reference (in the middle) input profile and the upper limit (at the top) of the design domain. While the solid blue lines are the selected input profiles to estimate the DRSM model. In these experiments, functions $u_0(\tau)$ and $\Delta u(\tau)$ have a linear dependency on time.

The DRSM model is estimated using the data collected from the designed experiments. In section 5.3.1 and 5.3.2, we illustrate the proposed single DRSM model approach for process optimization and control for two objectives, maximizing the product concentration and maximizing the product yield, respectively. In each case, we measure the concentration of C every 45 min, resulting in 20 measurements for each batch experiment. We further study the effect of lower sampling frequencies (10 and 5 measurements in each experiment) on the optimization and control performances in section 5.3.3. The control performances for the MPC controllers under different disturbances magnitudes are discussed in section 5.3.4.

5.3.1. Maximizing and Controlling the Production

With the collected measurements, 20 measurements/batch, we develop a DRSM model as discussed previously using 9 Shifted Legendre polynomials ($R + 1 = 9$) parameterizing each $\beta_q(\tau)$. The predicted output values by the DRSM model are compared with the *in silico* experimental data in Figure 5-2. The DRSM predictions are plotted in blue solid line while the experimental data are given in red diamonds. One can observe that the DRSM model approximates the nonlinear batch process very accurately.

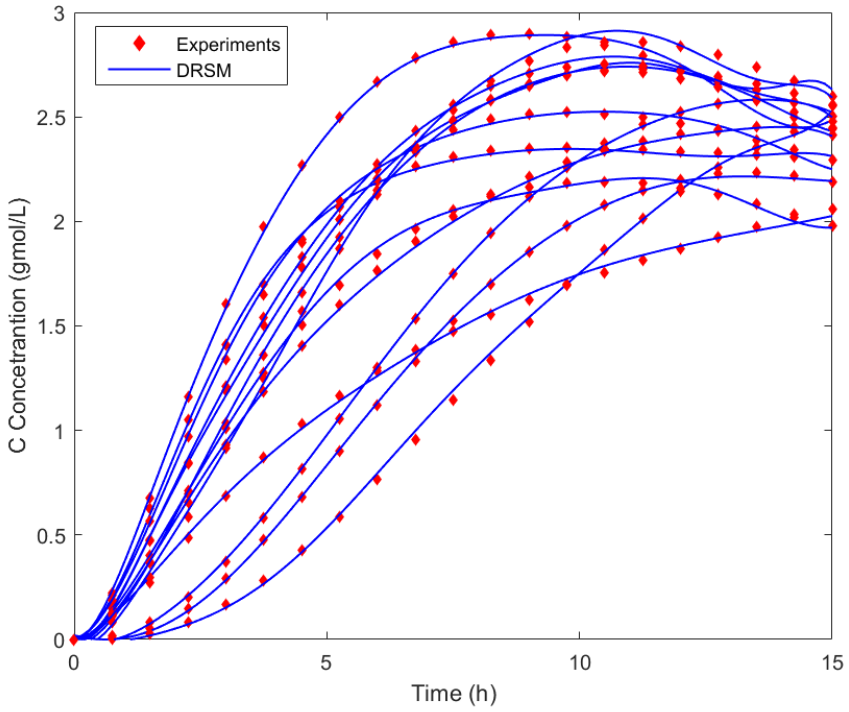

Figure 5-2: Comparison of DRSM model predictions (—) and the experiment measurements (♦)

We determine the optimal input profile of B maximizing the production of C as described in eq. (5-1). The cost function $J(\boldsymbol{x}, \tau)$ is defined as $J(\boldsymbol{x}, \tau) = -y(\tau) \times V$, where V is the reactor volume. The obtained optimal values for the dynamic sub-factors are $\boldsymbol{x}^* = (0.2, -0.8,\ 0)$ and the optimal batch time is $t_f^* = 8.7$ h. The maximum product concentration is 2.90 gmol/L and the maximum product is 29.0 gmol. The determined optimal input profile (upper part) is plotted in Figure 5-3. The maximum value for the input profiles and the reference profiles are plotted in dashed lines as well. To further confirm the obtained optimization result, we run the batch process with the determined optimal input profile and compare the resulting optimal output with the one predicted by the DRSM model. Both output profiles, the simulated (in blue) and predicted (in red), are

plotted in Figure 5-3 and they are indistinguishable, which verifies again that the DRSM model is quite accurate.

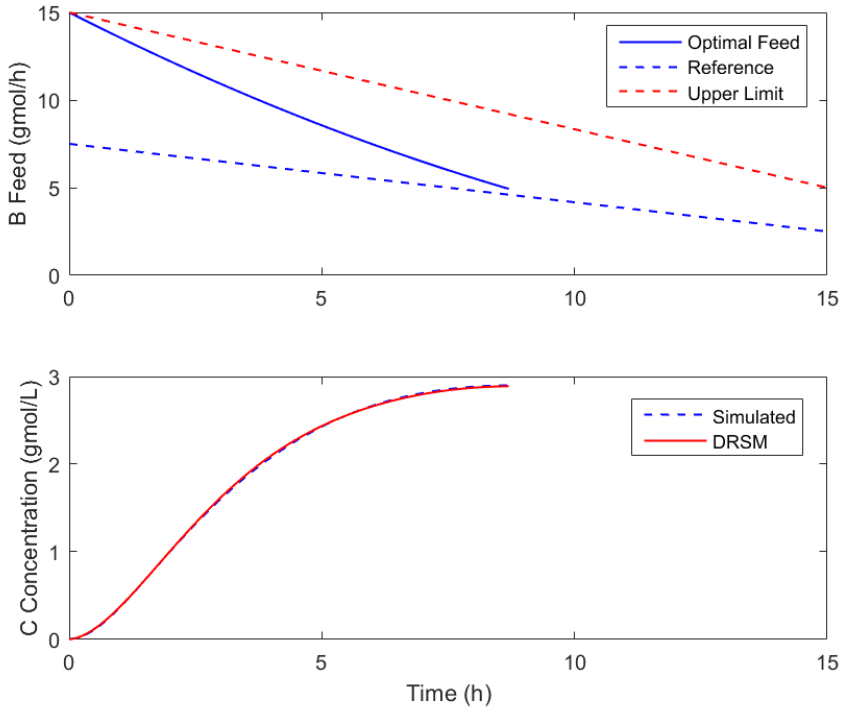

Figure 5-3: Optimal input (upper) and output (lower) profiles for the case of maximizing product concentration.

To test the MPC based on the identified models, we introduce a severe step disturbance in the rate constant k_1. The rate constant for the main reaction, k_1, is decreased by 30% at $t = 1.5\ h$. This can be caused by some toxic substrates that entered the reactor along with the feeding of reactant *B*. The controller is designed to make the concentration of product *C* meet the desired concentration at the end of the batch by manipulating the feed of reactant *B* using the receding horizon control given in eq. (5-22). The optimal batch duration is 8.7 *h*. The sampling interval for the state-space model estimated using the DRSM model is 9 *min*. Therefore the total number of time instances

during the batch is $N = 59$. However, the measurement is available every 45 min, so the number of measurements during the batch is 11. The upper limit for the manipulated input is 15 *gmol/h*, the highest flow of *B* allowed by the actuator. The lower limit is zero. Here we identify three local models, a linear and a Hammerstein-Wiener model from the DRSM model, notated as Linear and H-W respectively. And a linear model by conducting PRBS experiment round the optimal input trajectory with 11 measurements, notated as PRBS.

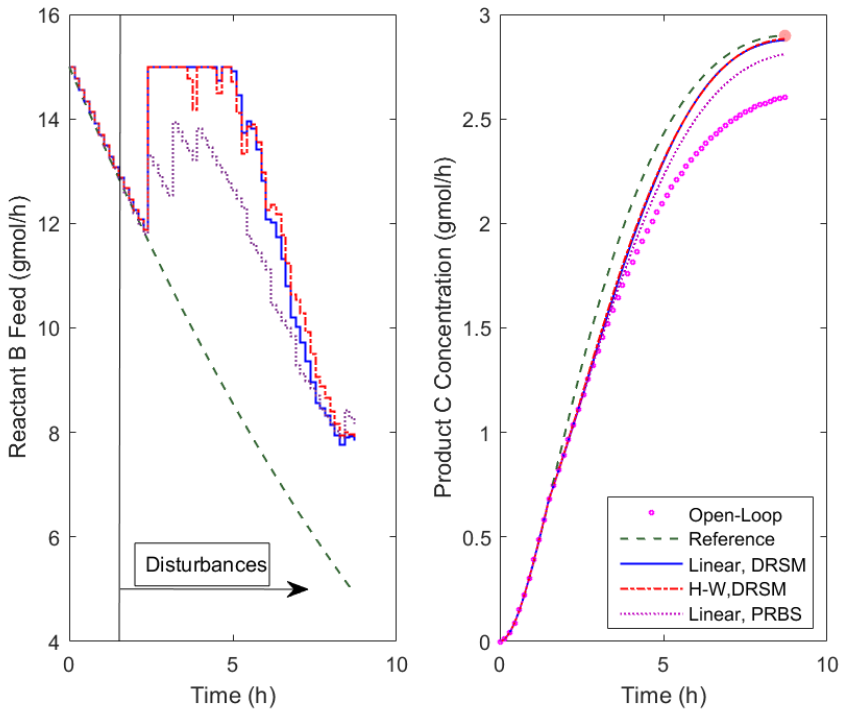

Figure 5-4: Input (left) and output (right) profiles under MPC for the case of maximizing product concentration with k_1 decreases by 30% at t=1.5 hours

The inputs calculated by MPCs based those models are plotted in the left of Figure 5-4 while the corresponding output profiles are plotted in the right. The final product concentration, defined as the controller's set point, is 2.90 *gmol/L* and is denoted by a red dot. The time evolution of the

output under the disturbance without the MPC is plotted in magenta circles with an end product concentration at 2.60 *gmol/L*, substantially less than the desired value. The evolution of the output values, when the MPC controller is active, is plotted in a blue continuous line (Linear) and red dash-dotted line (H-W) in the right of Figure 5-4. Both the linear and nonlinear controller have increased the inflow of *B* (left) and lead to a final concentration at 2.88 *gmol/L*, quite close to the set point. In contrast, the MPC based on the model identified via PRBS experiment leads to a final concentration of 2.81 *gmol/L* (in purple dotted line), less than those based on the models estimated from the DRSM model. Due to limited number of measurements, PRBS model is not estimated in desired accuracy. Therefore, the identified model based on DRSM is more favorable

5.3.2. Maximizing and Controlling Product Yield

For the case of the maximizing the product yield, we still use the same DRSM model estimated in section 5.3.1. As the DRSM model accurate represents the process behaviors in the entire design domain of interest, we can apply the same estimated DRSM model for different optimization and control aims. We only need to update the cost function for the optimization problem. Here we define the cost function as $J(\boldsymbol{x}, \tau) = -2C(\tau^*)/[A_0 + B_T(\tau^*)]$, the negative yield of product. Since we are going to maximize the yield, we use the negative sign in the cost function. The product yield is the ratio of moles of the product to the total moles of reactant fed into the reactor. Here A_0 is fixed at 50 gmol, $B_T(\tau^*) = \int_0^{\tau^*} u(\tau)d\tau$ and $C(\tau^*) = y(\tau^*) \times V$, where $y(\tau^*)$ is the concentration of C at time instance τ^*. By maximizing the yield, we obtain the optimal solution as $\boldsymbol{x}^* = (0, -0.4,\ 0.6)$ and the optimal batch time $t_f^* = 7.05$ h. The maximum yield is 46.8%. Since the reaction is $A + B \rightarrow C$, two moles of reactant are needed for one mole of product. This justifies

the multiplication by 2 in the yield defined in the cost function. The corresponding product concentration is 2.50 gmol/L, while 56.7 gmol B is used for producing 25 gmol C. In contrast, for the maximum production case, 29 gmol C is obtained with 82.8 gmol B consumed. The determined optimal input profile and both the simulated and predicted output profiles are plotted in Figure 5-5. It has been seen that the predicted optimal output profile (in red solid line) and the simulated one (in blue dashed line) are almost identical. which indicates that the DRSM prediction for the optimal output is very accurate.

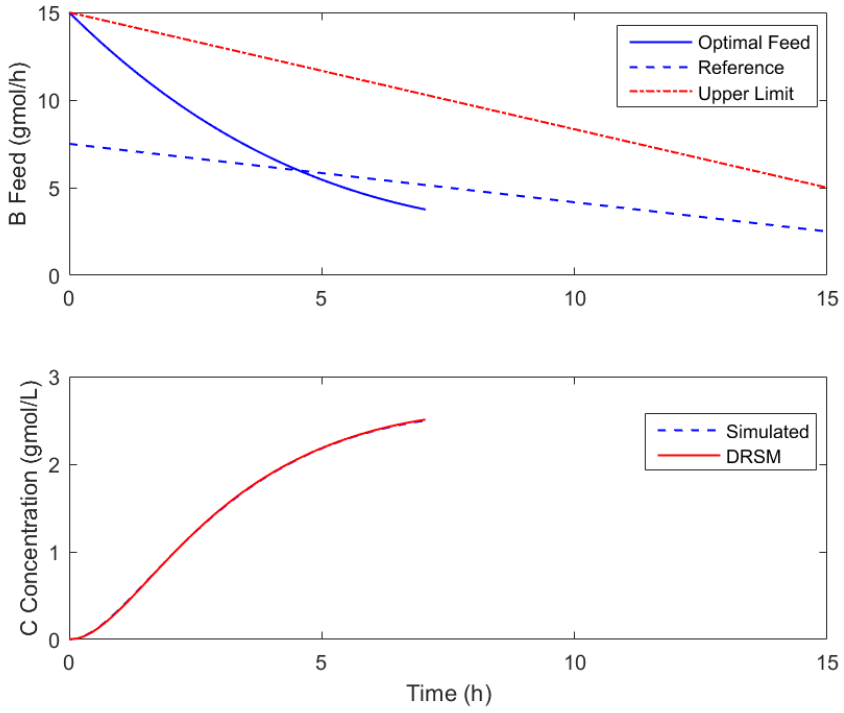

Figure 5-5: Optimal input (upper) and output (lower) profiles for the case of maximizing product yield.

The disturbances and the controller settings are the same with the ones in Section 5.3.1. The batch duration now is 7.1 h and only 9 measurements are available during the batch. Three local models

(Linear, H-W and PRBS) are identified and the control performances based on these models are compared. As shown in Figure 5-6, the final product concentration controlled using H-W model meets the target at 2.46 *gmol/L*. However, the one controlled based on linear model is 2.40 *gmol/L*, slightly lower than the target value. The one controlled using PRBS model is even lower at 2.34 *gmol/L* but higher than the open-loop one (2.21 *gmol/L*). The control performances based on the model identified from DRSM again are better when the measurements are limited.

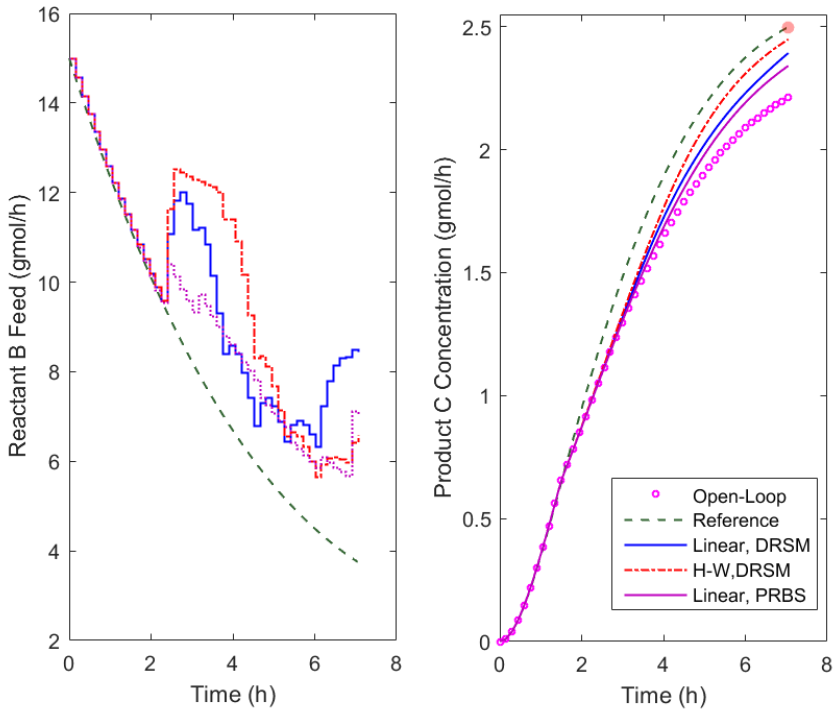

Figure 5-6: Input (left) and output (right) profiles under MPC for the case of maximizing product yield with k_1 decreases by 30% at t=1.5 hours

5.3.3. Performances of the Proposed Method with Fewer Measurements

For the measurements of certain species, the sampling frequency might be very low. It is of great interest to examine the performances of the proposed approach when the number of experiments

are fewer than the previously discussed cases. Here we reduces the number of measurements taken during each batch from 20 to 10 and 5. The corresponding sampling interval increased form 45 *mins* to 1.5 and 3.0 *hours*. In each case, we estimate a quadratic DRSM model. The number of shifted Legendre polynomials of each DRSM model is given in column 2 of Table 1. As the number of measurements obtained in each batch decrease, the allowed numbers of shifted Legendre polynomials parameterizing the parametric function in the DRSM models decreases to 5 and 4, respectively.

Table 5-1: Process optima obtained using the DRSM estimated with different number of measurements in each experiment for the cases of maximizing product weight and maximizing yield

Samples/Batch	R+1	Maximizing Product		Maximizing Yield		
		C_C (gmol/L)	Time (h)	Yield (%)	C_C (gmol/L)	Time (h)
20	9	2.90	8.70	46.8	2.50	7.05
10	5	2.90	8.40	46.7	2.50	7.20
5	4	2.90	8.70	46.5	2.50	6.75

With each of the two DRSM model, we maximize the product concentration and the product yield as discussed previously. The product concentration at the end of the batch, C_C, for the case of maximizing product concentration is reported in column 3 of Table 5-1 and the corresponding optimal batch times are listed in column 4. It has been seen that the obtained process optima of the DRSM models estimated using different sampling frequencies (20, 10 and 5 per batch) are identical, i.e. 2.90 *gmol/L*. This demonstrates that the process optimization using a DRSM model provides satisfying process optimum even when the sampling frequency becomes very low. As we

aimed to maximize the product concentration itself, the optimal batch time does not need to be the shortest. Therefore, we can see that the obtained optimal batch times are a little different from each other, varying from 8.40 to 8.70 *hours*.

The obtained yield for the case of maximizing product yield has been listed in column 5 of Table 5-1. As the number of measurements in each batch for estimating the DRSM model decreases from 20 to 5, the resulted optimal yield becomes slightly lower, from 46.8% to 46.5%. Therefore, we confirm that the achieved optimization performances using the DRSM model are almost identical, regardless the sampling frequencies. The corresponding product concentration at the end of each batch and the optimal batch time are reported in column 6 and 7 respectively.

We here again identify in the vicinity of the obtained optimal trajectories three recursive models, the linear state-space model using PRBS experiment, the linear model and H-W model from the DRSM model. The close-loop performances of the MPC controller based on these three models are shown in column 2-4 of Table 5-2 for the case of maximizing the product concentration and in column 6-8 for the case of maximizing product yield. The optimal product concentrations reported in column 3 and 6 of Table 5-2 serve as the set points for the corresponding cases. The open-loop product concentrations obtained without any control action are given in column 5 and 9, respectively. The disturbance considered here is that k_1 drops by 30% at t=1.5 hours.

Table 5-2: Comparison of control performances using different reclusive models with different sampling frequencies and under different disturbances.

$\Delta k_1 = -30\%, t_{dis} = 1.5\ h$								
Samples/Batch	Maximizing Product				Maximizing Yield			
	H-W	Linear	PRBS	OL	H-W	Linear	PRBS	OL
20	2.88	2.88	2.81	2.60	2.46	2.40	2.34	2.21
10	2.84	2.82	2.77	2.59	2.41	2.34	NA	2.21
5	2.78	2.77	NA	2.60	2.36	2.35	NA	2.21
$\Delta k_1 = -20\%, t_{dis} = 1.5\ h$								
20	2.89	2.88	2.85	2.72	2.47	2.44	2.41	2.33
10	2.88	2.87	2.83	2.72	2.46	2.41	NA	2.33
5	2.84	2.83	NA	2.72	2.42	2.41	NA	2.33
$\Delta k_1 = -40\%, t_{dis} = 1.5\ h$								
20	2.82	2.82	2.75	2.46	2.40	2.33	2.25	2.08
10	2.75	2.74	2.69	2.45	2.34	2.26	NA	2.07
5	2.70	2.68	NA	2.46	2.26	2.26	NA	2.08
$\Delta k_1 = -30\%, t_{dis} = 0.5\ h$								
20	2.88	2.88	2.82	2.59	2.44	2.39	2.34	2.19
10	2.86	2.84	2.79	2.58	2.41	2.34	NA	2.18
5	2.79	2.78	NA	2.59	2.36	2.36	NA	2.19
$\Delta k_1 = -30\%, t_{dis} = 3\ h$								
20	2.88	2.88	2.81	2.60	2.45	2.39	2.34	2.21
10	2.84	2.82	2.77	2.59	2.41	2.34	NA	2.21
5	2.78	2.77	NA	2.60	2.36	2.35	NA	2.21

According to the results reported in the row 3-6 of Table 5-2, if there is no control action, the product concentrations at the end of each run are 2.60 gmol/L and 2.21 gmol/L for the maximizing product concentration and yield, respectively. As the number of measurements decreases from 20 to 10, the close-loop product concentration obtained by a MPC using H-W model reduces from 2.88 gmol/L to 2.84 gmol/L for the case of maximizing product concentration. However, it still provides best control performances among the three recursive models examined here. In addition, the MPC using the model identified with PRBS experiment, provides the worst control performances. When the number of measurements during each run further reduces to 5, the control performances of the MPCs using the H-W model and linear model identified from the DRSM model becomes worse, from 2.84 and 2.82 *gmol/L* to 2.78 and 2.77 *gmol/L*, respectively. However, the linear model with PRBS experiment is not identifiable when the measurement is so few. As the sampling interval is 3.0 hours, there are only two measurements taken during the 8.4 or 8.7 hours batch time. Therefore a "NA" is filled in the cell of the table. In fact, if there is no recursive model and the corresponding MPC, we can assume that the obtained product concentration is 2.60 g*mol/L*, a much lower value than the ones achieved using the recursive models by the DRSM model.

The differences in control performance between the controllers based on H-W model and linear model identified from DRSM model becomes more significant for the case of maximizing yield than the case of maximizing product concentration. This may be because that the former case is more nonlinear than the latter. When there are 10 measurements, the concentration obtained using the H-W model is 2.41 gmol/L while the one obtained using the linear model by the DRSM model is 2.34 gmol/L. As the sampling frequencies further decrease, the obtained product concentration

at the end of the batch using the two model are about the same. This may be because that the number of measurements during each run is too few. When there are only 5 measurements in each experiment, the sampling interval is 3 hours. The optimal batch time for maximizing yield case is about 7 hours. Consequently, the state variables are only updated twice during the operation. As the Kalman filter is used only to update the state variable instead of adapting the model parameters to the disturbances, even the MPC using the H-W model does not provide a final product concentration close to the set point value. However, it is still much better than the MPC based on the model identified using PRBS experiments, which is not identifiable due to the too few measurements. The optimal batch time for the maximizing yield case is about 7.0 *hours*, while the sampling intervals corresponding to the sampling frequencies of 10 and 5 measurements/ batch are 1.5 and 3.0 *hours*. As a result, the number of measurements during each optimal run is 4 and 2, too few for identifying a linear state-space model. In such a case, the control using a DRSM model is the only feasible way to achieve a product concentration higher than the open-loop concentration.

5.3.4. Control Performances under Different Disturbances Magnitudes

We also examine the effect of different disturbances introduced at different time instants on the control performances. In row 8-10 and 12-14 of Table 5-2, we examined the disturbances of k_1 drops by 20% and 40% at t=1.5 hours. Comparing these results with the ones listed in row 4-6 for the case that k_1 drops by 30% at t=1.5 hours, we can see that as the disturbances increase, i.e. k_1 decreases by 20% to 40%, the obtained product concentrations using all three kinds of models become lower. However, the controller using the H-W model from the DRSM model still provides the best control performances while the one using the linear model by the PRBS experiments provides the worst performances.

In the last 8 rows of Table 5-2, we examine the effect of the disturbance introduced at different time instants, i.e. at t=0.5 and 3 *hours*. As the open-loop results are not affected much by the starting time instants of the disturbance, the control performance are almost identical. However, the results have demonstrated again the controller using the H-W model provides the product concentrations closest to the desired values while the controller using the linear model by PRBS experiments, provides the least close values.

5.4. Conclusions

In this chapter, we present a new data-driven dynamic model, the DRSM, for optimizing and controlling batch processes. Based on the DRSM accurately approximating the batch processes, one can determine the optimal operation conditions. Then a local recursive models, including linear state-space model and Hammerstein-Wiener model, can be identified by directly "sampling" the DRSM in the vicinity of the optimal input profiles without conducting another set of experiments. In this way, process optimization and control are achieved based on a single data-driven dynamic model, which saves experimental effort to identify local dynamic models, especially when the desired optimal operating conditions are changed, for example, from maximum production to maximum yield. In addition, when the measurement during the batch operation is too limited to estimate state-space model, sampling the DRSM may be the most attractive way to identify dynamic models with satisfactory accuracy for control purposes.

The effectiveness of this proposed method has been verified by the *in silico* nonlinear reactions in semi-batch reactor. The obtained process optimum compromises little as the number of measurements during each experiment decreases. The control performances based on the recursive

models, including the linear model via PRBS experiments, the linear and nonlinear models estimated from the DRSM, become worse. However, in all cases examined, the MPC based on H-W model estimated using the DRSM model always provides the best control performance. In addition, when the sampling frequency becomes quite low, say 5 measurements per batch, the recursive model is still identifiable through the DRSM model, but not via the PRBS experiment. Therefore, the identification via DRSM is a more favorable approach in the situation when the sampling frequency is low.

6. Conclusions and Future Work

In this thesis, we advance the two data-driven modeling concepts, the DoDE and the DRSM, for the purposes of process optimization and control. There are three major contributions, improvements to DoDE of this research work: 1) we have made two methodological improvements to the DoDE approach; 2) we have developed a new DRSM methodology for modeling both continuous and batch processes; 3) we have proposed the single model approach, using a DRSM model, for both process optimization and control purposes. All the proposed approaches have been examined using representative nonlinear processes, which demonstrates the efficacy the methodologies as data-driven tools for process optimization and control purposes. In this chapter, we summarized the three aforementioned contributions. The future work related to each methodological improvement has been discussed in the corresponding section as well. .

6.1. Methodological Improvements to DoDE

6.1.1. Conclusions for the Improved DoDE

We have discussed two improvements to the DoDE methodology in Chapter 3. The first one focuses on improving the design of the input domain using the prior knowledge about the process characteristics. Though the input domain has a significant impact to the obtained process optimum, the ways to select better input domain has not been discussed yet. We use two different types of *a priori* knowledge to illustrate the proposed approaches and examine it in two representative classes of the biopharmaceutical processes, the set of eight SMF processes and the Hybridoma cell culture. The refined input domains lead to higher process optima in both two classes of biopharmaceutical

processes. These case studies demonstrate that by incorporating the process knowledge into the DoDE framework, one can further improve the optimal performance of the process at hand.

The second methodological improvement of the original DoDE methodology addresses the minimization of the initial number of experiments. It is motivated by the fact that the process optimization task in industry has to be completed within tight time and budgetary constraints. To minimize the number of the initial experiments, one should initially aim for a linear RSM. The predicted optimum is considered "tentative" until its prediction uncertainty is estimated and is shown to satisfy an uncertainty upper limit. If the uncertainty of the optimal point is larger than the maximum allowed, a suboptimal point of operation is sought that will satisfy the uncertainty constraint.

6.1.2. Future Work related to DoDE

The ways to reduce the initial number of DoDE experiments could be further developed into an evolutionary DoDE approach. After one runs the process at the initial optimal point obtained using the initial RSM as discussed in chapter 3, a new data point is obtained, which can be used to update the RSM model and a new optimal point will be obtained using the updated model. If we repeated these steps, we expect to continuously improve the process performance with minimal number of experiments, which meets the industrial demand for completing the process optimization tasks under tight time and budgetary constraints. We have presented such an evolutionary DoDE approach, described in Figure 6-1, at AICHE 2016[101].

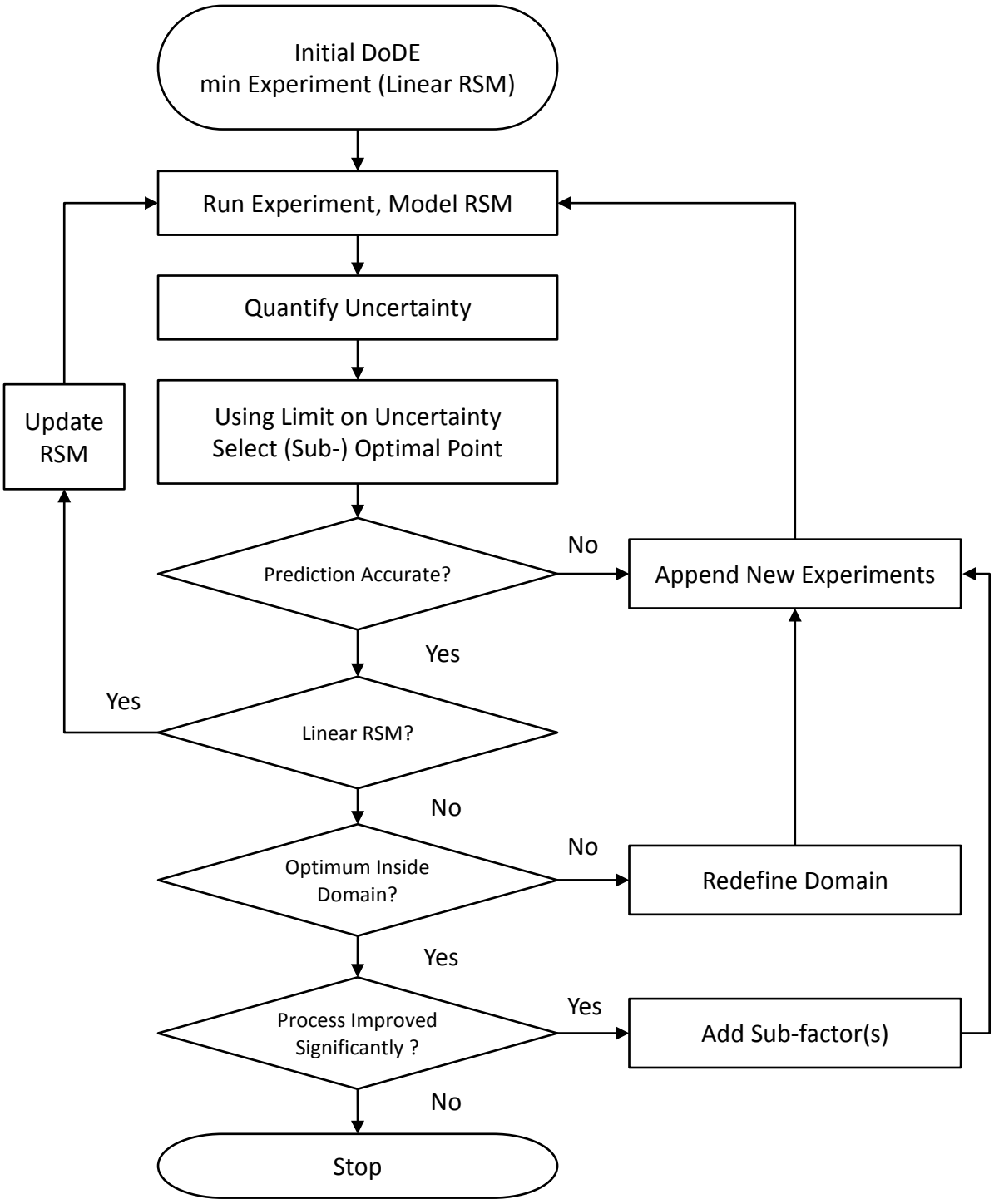

Figure 6-1: Schematic diagram for evolutionary DoDE operations

It has been shown in the two case studies, the penicillin fermentation and hybridoma cell culture, examined in the presentation that the evolutionary DoDE approach leads to close process optimum to the ones obtained using the model-based optimization.

To complete the evolutionary DoDE approach, one may consider to develop a systematic procedure to refine the input domain when one notices that the present domain is not the best one. As the input domain affects significantly the obtained process optimum, it is of great interest to update the input domain appropriately. In the previous AICHE presentation, we consider the relaxation of the activated constraint (boundary of the input domain) by 20%. However, further investigation should be made concerning whether such a choice of relaxation is an optimal one.

6.2. New DRSM Methodology

6.2.1. Conclusions for DRSM-2

In Chapter 4, we have presented here a new Dynamic Response Surface Methodology, DRSM-2, for the modeling of both continuous and batch processes with various duration. In contrast, the original DRSM methodology is only capable of modeling batch processes with fixed duration. The key idea of the DRSM-2 approach is to define a new independent variable θ as an exponential transformation of time. The new method also allows the utilization of non-equidistant data or missing data which makes the DRSM-2 approach have the potential to use historical data to save the experimental cost for the model development.

In addition, we proposed a systematic procedure to determine the most appropriate values of the decision variables that influence the structure of the DRSM model. Three decision variables, the

time constant, t_c, the polynomial order, R, as well as the model class, C, are considered. The choice of the most appropriate values of these decision variables within each model class is determined using two statistics, the BIC and the LoF. We select the DRSM-2 model with the smallest BIC value and an insignificant LoF. Moreover, we pointed out that the choices of t_c and R_q, determining the fastest dynamics represented by the DRSM-2 model are also constrained by the sampling interval.

The DRSM-2 approach has been applied to model two nonlinear processes, a continuous propylene polymerization and a semi-batch penicillin fermentation. The results clearly demonstrate that the DRSM-2 model provides an accurate representation of the dynamic behaviors of the processes examined.

6.2.2. Future Work related to DRSM-2

The estimation of the DRSM-2 model utilizes the unconstrained linear regression in each iteration of the stepwise regression. It may be of interest to use regularized regression, such as ridge regression[102, 103] or LASSO regression[104, 105], to estimate the model parameters. As we pointed out in Chapter 4, when the sampling constraint given in eq.(4-15) is violated, the DRSM-2 model predictions will exhibit oscillations between the two measurements. We consider such oscillation is caused by the DRSM-2 model representing the dynamics faster than the sampling frequency. As the regularization methods are known for the capability of reducing the effect of overfitting, it is of interest to refine the DRSM-2 model estimation using the regularized regression.

In addition, we are interested to see if a single DRSM model can represent the increasing and decreasing MFI cases of the polymerization process discussed in Chapter 4. The first step for

developing a single model is to combine the input domain. The preliminary thoughts on the design of a single input domain for both increasing and decreasing MFI cases are given as follows

$$\begin{gathered}(x_4 - x_1)(x_2 - x_4) > \varepsilon_1 \\ (x_2 - x_4)(x_4 - x_3) > \varepsilon_1 \\ (x_4 - x_3)(x_3 - x_1) > \varepsilon_1 \\ (x_3 - x_1)(x_4 - x_1) > \varepsilon_1 \\ (x_4 - x_1)[x_2 - (1+\delta)x_3 + \delta x_1] < \varepsilon_2\end{gathered} \tag{6-1}$$

The above input domain is obtained by combining the input domains given by eq.(4-27) and eq.(4-31). We here propose to choose $\varepsilon_1 = 0.01$ and $\varepsilon_2 = -0.01$. The ramp rate, a, has the same magnitude as discussed previously, but the its sign is determined by the values of u_1 and u_4 by

$$\begin{cases} a > 0, & if\ u_4 > u_1 \\ a < 0, & if\ u_4 < u_1 \end{cases} \tag{6-2}$$

The factors x_i's for the DoDE experiments are defined as in eq. (4-25) and in the interval of $[-1, +1]$. The ratio, δ, reduces the overshoot/undershoot resulting from the selected input profile. The value of δ can be selected by prior process knowledge. We here assume $\delta = 1$. By combining the input domain, we can reduce the number of experiments required for estimating the DRSM-2 model by half. Further examination of the proposed method worth the thorough discussion in a journal paper.

6.3. A Single DRSM Model for Optimization and Control

6.3.1. Conclusions for Optimization and Control using DRSM

We have proposed a single model approach, using a DRSM model, for optimizing and controlling batch processes in Chapter 5. Based on the DRSM model accurately approximating the batch processes, one can determine the optimal operation conditions. Then a local recursive models, including linear state-space model and Hammerstein-Wiener model, can be identified by directly by "sampling" the DRSM in the vicinity of the optimal input profiles without conducting another set of experiments. In addition, when the measurement during the batch operation is too limited to estimate state-space model, sampling the DRSM model may be the most attractive way to identify dynamic models with satisfactory accuracy for control purposes.

The effectiveness of this proposed method has been verified by the *in silico* nonlinear reactions in semi-batch reactor. The obtained process optimum compromises little as the number of measurements during each experiment decreases. The control performances based on the recursive models, including the linear model via PRBS experiments, the linear and nonlinear models estimated from the DRSM model, become worse. However, in all cases examined, the MPC based on H-W estimated using the DRSM model always provides the best control performance. In addition, when the sampling frequency becomes quite low, say 5 measurements per batch, the recursive model is still identifiable through the DRSM model, but not via the PRBS experiment. Therefore, the identification via DRSM is a more favorable approach in the situation when the sampling frequency is low.

6.3.2. Future Work Related to Control using DRSM Model

The identification of nonlinear recursive model using a DRSM model discussed in this research work is limited to Hammerstein-Wiener model, which accounts for input and output nonlinearities, but not the nonlinearities of the interactions between input and output variables. In the future, we are interested to identify the nonlinear recursive model with more flexible form of nonlinearities. An example quadratic SISO nonlinear dynamic model is given as follows,

$$\boldsymbol{y}_k = \boldsymbol{a}^T \boldsymbol{y}_{k-1} + \boldsymbol{a}_2^T \boldsymbol{y}_{k-1}^2 + \boldsymbol{b}^T \boldsymbol{u}_{k-1} + \boldsymbol{b}_2^T \boldsymbol{u}_{k-1}^2 + \boldsymbol{c}^T \boldsymbol{y}_{k-1} \otimes \boldsymbol{u}_{k-1} \tag{6-3}$$

Where the scalar y_k is the output at time instant k. $\boldsymbol{y}_{k-1}$ is a $n \times 1$ column vector defined as $\boldsymbol{y}_{k-1} = [y_{k-1}\ y_{k-2}\ \cdots\ y_{k-n}]^T$. $\boldsymbol{y}_{k-1}^2 = \boldsymbol{y}_{k-1} \otimes \boldsymbol{y}_{k-1}$, and $\otimes$ is the Kronecker product[106]. $\boldsymbol{u}_{k-1}$ is a $m \times 1$ column vector defined as $\boldsymbol{u}_{k-1} = [u_{k-1}\ u_{k-2}\ \cdots\ u_{k-m}]^T$ and $\boldsymbol{u}_{k-1}^2 = \boldsymbol{u}_{k-1} \otimes \boldsymbol{u}_{k-1}$. The model parameters to be estimated are $\boldsymbol{a}$, $\boldsymbol{a}_2$, $\boldsymbol{b}$, $\boldsymbol{b}_2$, and $\boldsymbol{c}$. They are $n \times 1$, $n^2 \times 1$, $m \times 1$, $m^2 \times 1$ and $mn \times 1$ column vectors, respectively. The estimation of the model parameters can be achieved using use the regularized identification methods[107, 108]. With the new identification method, we expect to 1) identify nonlinear recursive model with nonlinear interactions between the input and the output variables; 2) estimate higher order nonlinearities using lower order DRSM model. For example, we can estimate cubic nonlinearities using a quadratic DRSM model.

7. Bibliography

1. Bonvin, D., et al., *Linking models and experiments.* Industrial & Engineering Chemistry Research, 2016. **55**(25): p. 6891-6903.

2. Rasmuson, A., et al., *Mathematical Modeling in Chemical Engineering*. 2014: Cambridge University Press.

3. Bird, R.B., E.N. Lightfoot, and W.E. Stewart, *Transport Phenomena*. 2002: Wiley.

4. Sandler, S.I., *Chemical, Biochemical, and Engineering Thermodynamics*. 2006: John Wiley & Sons.

5. Welty, J., *Fundamentals of Momentum, Heat and Mass Transfer, 5th Edition*. 2007: Wiley Global Education.

6. Mulukutla, B.C., M. Gramer, and W.-S. Hu, *On metabolic shift to lactate consumption in fed-batch culture of mammalian cells.* Metabolic Engineering, 2012. **14**(2): p. 138-149.

7. Nolan, R.P. and K. Lee, *Dynamic model of CHO cell metabolism.* Metabolic Engineering, 2011. **13**(1): p. 108-124.

8. Nolan, R.P. and K. Lee, *Dynamic model for CHO cell engineering.* Journal Of Biotechnology, 2012. **158**(1-2): p. 24-33.

9. Zacca, J.J. and W.H. Ray, *Modeling of the liquid-phase polymerization of olefins in loop reactors.* Chemical Engineering Science, 1993. **48**(22): p. 3743-3765.

10. Nie, Y.S., et al., *Reactor modeling and recipe optimization of polyether polyol processes: Polypropylene glycol.* Aiche Journal, 2013. **59**(7): p. 2515-2529.

11. Wang, Z. and C. Georgakis, *An in silico evaluation of data-driven optimization of biopharmaceutical processes.* AIChE Journal, 2017: p. doi:10.1002/aic.15659.

12. Wold, S., K. Esbensen, and P. Geladi, *Principal component analysis.* Chemometrics and intelligent laboratory systems, 1987. **2**(1-3): p. 37-52.

13. Abdi, H. and L.J. Williams, *Principal component analysis.* Wiley interdisciplinary reviews: computational statistics, 2010. **2**(4): p. 433-459.

14. Wold, S., M. Sjöström, and L. Eriksson, *PLS-regression: a basic tool of chemometrics.* Chemometrics and intelligent laboratory systems, 2001. **58**(2): p. 109-130.

15. Chiang, L.H., et al., *Diagnosis of multiple and unknown faults using the causal map and multivariate statistics.* Journal of Process Control, 2015. **28**: p. 27-39.

16. Box, G.E., et al., *Time series analysis: forecasting and control*. 2015: John Wiley & Sons.

17. Ljung, L., *System identification*. 1999: Wiley Online Library.

18. Seborg, D.E., et al., *Process dynamics and control*. 2010: John Wiley & Sons.

19. Ku, W., R.H. Storer, and C. Georgakis, *Disturbance detection and isolation by dynamic principal component analysis.* Chemometrics and intelligent laboratory systems, 1995. **30**(1): p. 179-196.

20. Shi, R. and J.F. MacGregor, *Modeling of dynamic systems using latent variable and subspace methods.* Journal of Chemometrics, 2000. **14**(5-6): p. 423-439.

21. Box, G.E.P. and N.R. Draper, *Response Surfaces, Mixtures, and Ridge Analysis*. 2007, Hoboken, NJ: Wiley.

22. Bezerra, M.A., et al., *Response surface methodology (RSM) as a tool for optimization in analytical chemistry.* Talanta, 2008. **76**(5): p. 965-977.

23. Montgomery, D.C., *Design and Analysis of Experiments*. 8th ed. 2013, New York: Wiley.

24. Box, G.E.P. and K.B. Wilson, *On the experimental attainment of optimum conditions.* Journal of the Royal Statistical Society Series B-Statistical Methodology, 1951. **13**(1): p. 1-45.

25. Georgakis, C., *A Model-Free Methodology for the Optimization of Batch Processes:Design of Dynamic Experiments*, in *IFAC Symposium on Advanced Control of Chemical Processes (ADCHEM)*. 2009: Istanbul, Turkey, IFAC.

26. Georgakis, C., *Design of Dynamic Experiments: A Data-Driven Methodology for the Optimization of Time-Varying Processes.* Industrial & Engineering Chemistry Research, 2013. **52**(35): p. 12369-12382.

27. Troup, G.M. and C. Georgakis, *Process systems engineering tools in the pharmaceutical industry.* Computers & Chemical Engineering, 2013. **51**: p. 157-171.

28. Fiordalis, A. and C. Georgakis, *Data-driven, Using Design of Dynamic Experiments, versus Model-driven Optimization of Batch Crystallization Processes.* Journal Of Process Control, 2013. **23**(2): p. 179-188.

29. Georgakis, C., et al. *Optimizing an Industrial Batch Process Using the Design of Dynamic Experiments Methodology*. in *AICHE Spring Meeting*. 2016. Houston,TX.

30. Makrydaki, F., C. Georgakis, and K. Saranteas. *Dynamic Optimization of a Batch Pharmaceutical Reaction Using the Design of Dynamic Experiments (DoDE): The Case of an Asymmetric Catalytic Hydrogenation Reaction*. in *Proceedings of the 9th International*

Symposium on Dynamics and Control of Process Systems. 2010. Oude Valk College, Belgium.

31. Klebanov, N. and C. Georgakis, *Dynamic Response Surface Models: A Data-Driven Approach for the Analysis of Time-Varying Process Outputs.* Industrial & Engineering Chemistry Research, 2016. **55**(14): p. 4022-4034.

32. Wang, Z., N. Klebanov, and C. Georgakis. *DRSM Model for the Optimization and Control of Batch Processes*. in *Dynamics and Control of Process Systems, including Biosystems/IFAC*. 2016. Trondhelm,Norway.

33. Verhaegen, M. and V. Verdult, *Filtering and system identification A least squares approach*. 2007: Cambridge University Press.

34. Bajpai, R.K. and M. Reuss, *A Mechanistic Model for Penicillin Production.* Journal Of Chemical Technology And Biotechnology, 1980. **30**(6): p. 332-344.

35. Jang, D.J. and J.P. Barford, *An unstructured kinetic model of macromolecular metabolism in batch and fed-batch cultures of hybridoma cells producing monoclonal antibody.* Biochemical Engineering Journal, 2000. **4**(2): p. 153-168.

36. Kiparissides, A., et al., *Design of In Silico Experiments as a Tool for Nonlinear Sensitivity Analysis of Knowledge-Driven Models.* Industrial & Engineering Chemistry Research, 2014. **53**(18): p. 7517-7525.

37. Wang, Z. and C. Georgakis. *On the Identification of Meta-Models for the Optimization of Grade Transition in Polymerization Processes*. in *AICHE Annual Meeting*. 2016. San Francisco, CA.

38. Biegler, L.T., *An overview of simultaneous strategies for dynamic optimization.* Chemical Engineering and Processing, 2007. **46**(11): p. 1043-1053.

39. Kameswaran, S. and L.T. Biegler, *Convergence rates for direct transcription of optimal control problems using collocation at Radau points.* Computational Optimization And Applications, 2008. **41**(1): p. 81-126.

40. Kameswaran, S. and L.T. Biegler. *Simultaneous dynamic optimization strategies: Recent advances and challenges*. in *7th International Conference on Chemical Process Control (CPC 7)*. 2006. Lake Louise, CANADA.

41. Biegler, L.T., *Solution of Dynamic Optimization Problems by Successive Quadratic Programming and Orthogonal Collocation.* Computers & Chemical Engineering, 1984. **8**(3-4): p. 243-247.

42. Edgar, T.F., D.M. Himmelblau, and L.S. Lasdon, *Optimization of chemical processes*. 2001: McGraw-Hill.

43. Bertsimas, D. and J.N. Tsitsiklis, *Introduction to Linear Optimization*. 1997: Athena Scientific.

44. Mandli, A.R. and J.M. Modak, *Evolutionary Algorithm for the Determination of Optimal Mode of Bioreactor Operation.* Industrial & Engineering Chemistry Research, 2012. **51**(4): p. 1796-1808.

45. Chiou, J.P. and F.S. Wang, *Hybrid method of evolutionary algorithms for static and dynamic optimization problems with application to a fed-batch fermentation process.* Computers & Chemical Engineering, 1999. **23**(9): p. 1277-1291.

46. Angira, R. and A. Santosh, *Optimization of dynamic systems: A trigonometric differential evolution approach.* Computers & Chemical Engineering, 2007. **31**(9): p. 1055-1063.

47. Mariano, A.P., et al., *Optimisation of a fermentation process for butanol production by particle swarm optimisation (PSO).* Journal Of Chemical Technology And Biotechnology, 2010. **85**(7): p. 934-949.

48. Rocha, M., et al., *Optimization of fed-batch fermentation processes with bio-inspired algorithms.* Expert Systems with Applications, 2014. **41**(5): p. 2186-2195.

49. Riascos, C.A.M. and J.M. Pinto, *Optimal control of bioreactors: a simultaneous approach for complex systems.* Chemical Engineering Journal, 2004. **99**(1): p. 23-34.

50. Nguang, S.K., L.Z. Chen, and X.D. Chen, *Optimisation of fed-batch culture of hybridoma cells using genetic algorithms.* Isa Transactions, 2001. **40**(4): p. 381-389.

51. Fisher, R.A., *Statistical Methods for Research Workers,* . 13th ed. 1958: Oliver and Boyd: Edinburgh, U.K.

52. Fisher, R.A., *The Design of Experiments*. 8th ed. 1966, New York: Hafner Publishing.

53. Arfken, G.B. and H.J. Weber, *Mathematical methods for physicists*. 1999, AAPT.

54. Verdult, V. and M. Verhaegen, *Subspace identification of multivariable linear parameter-varying systems.* Automatica, 2002. **38**(5): p. 805-814.

55. Bai, E.W., *A blind approach to the Hammerstein-Wiener model identification.* Automatica, 2002. **38**(6): p. 967-979.

56. van Wingerden, J.W., M. Verhaegen, and Ieee, *Closed-loop subspace identification of Hammerstein-Wiener models.* Proceedings Of the 48th Ieee Conference on Decision And Control, 2009 Held Jointly with the 2009 28th Chinese Control Conference (Cdc/Ccc 2009), 2009: p. 3637-3642.

57. Gomez, J.C. and E. Baeyens, *Identification of block-oriented nonlinear systems using orthonormal bases.* Journal of Process Control, 2004. **14**(6): p. 685-697.

58. Norquay, S.J., A. Palazoglu, and J.A. Romagnoli, *Application of Wiener model predictive control (WMPC) to a pH neutralization experiment.* Ieee Transactions on Control Systems Technology, 1999. **7**(4): p. 437-445.

59. Gomez, J.C., A. Jutan, and E. Baeyens, *Wiener model identification and predictive control of a pH neutralisation process.* Iee Proceedings-Control Theory And Applications, 2004. **151**(3): p. 329-338.

60. Bloemen, H.H.J., et al., *Wiener model identification and predictive control for dual composition control of a distillation column.* Journal Of Process Control, 2001. **11**(6): p. 601-620.

61. Van Overschee, P. and B. De Moor, *N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems.* Automatica, 1994. **30**(1): p. 75-93.

62. Tulleken, H., *Generalized binary noise test-signal concept for improved identification-experiment design.* Automatica, 1990. **26**(1): p. 37-49.

63. Rawlings, J.B. and D.Q. Mayne, *Model Predictive Control: Theory and Design*. 2009, Madison: Nob Hill Publishing.

64. Morari, M. and J.H. Lee, *Model predictive control: past, present and future.* Computers & Chemical Engineering, 1999. **23**(4): p. 667-682.

65. Garcia, C.E., D.M. Prett, and M. Morari, *Model predictive control: theory and practice—a survey.* Automatica, 1989. **25**(3): p. 335-348.

66. Kalman, R.E., *A New Approach to Linear Filtering and Prediction Problems.* Transactions of ASME, Journal of Basic Engineering, 1960. **87**: p. 35-45.

67. Lewis, F.L., *Optimal estimation: with an introduction to stochastic control theory*. 1986: Wiley.

68. Lima, F.V., *Interval Operability: A Tool to Design the Feasible Output Constraints for Non-Square Model Predictive Controllers*. 2007, Tufts University: Ann Arbor: ProQuest.

69. Shah, G. and S. Engell. *Tuning MPC for desired closed-loop performance for SISO systems*. in *Control & Automation (MED), 2010 18th Mediterranean Conference on*. 2010. IEEE.

70. Garriga, J.L. and M. Soroush, *Model predictive control tuning methods: A review*. Industrial & Engineering Chemistry Research, 2010. **49**(8): p. 3505-3515.

71. Lee, J. and Z. Yu, *Tuning of model predictive controllers for robust performance.* Computers & chemical engineering, 1994. **18**(1): p. 15-37.

72. Jørgensen, J.B., J.K. Huusom, and J.B. Rawlings. *Finite horizon MPC for systems in innovation form*. in *50th IEEE Conference on Decision and Control and European Control Conference*. 2011.

73. Shah, G. and S. Engell. *Tuning MPC for desired closed-loop performance for MIMO systems*. in *Proceedings of the 2011 American Control Conference*. 2011.

74. Grüne, L. and J. Pannek, *Nonlinear model predictive control*, in *Nonlinear Model Predictive Control*. 2011, Springer. p. 43-66.

75. Allgower, F., R. Findeisen, and Z.K. Nagy, *Nonlinear model predictive control: From theory to application.* Journal of the Chinese Institute of Chemical Engineers, 2004. **35**(3): p. 299-315.

76. Allgöwer, F. and A. Zheng, *Nonlinear model predictive control*. Vol. 26. 2012: Birkhäuser.

77. M'Sahli, F., R.B. Abdennour, and M. Ksouri, *Nonlinear model-based predictive control using a generalised Hammerstein model and its application to a semi-batch reactor.* International Journal Of Advanced Manufacturing Technology, 2002. **20**(11): p. 844-852.

78. Hong, M. and S. Cheng, *Model Predictive Control Based on Kalman Filter for Constrained Hammerstein-Wiener Systems.* Mathematical Problems in Engineering, 2013: p. 6.

79. Khani, F. and M. Haeri, *Robust model predictive control of nonlinear processes represented by Wiener or Hammerstein models.* Chemical Engineering Science, 2015. **129**: p. 223-231.

80. Bloemen, H.H.J., T.J.J. van den Boom, and H.B. Verbruggen, *Model-based predictive control for Hammerstein-Wiener systems.* International Journal Of Control, 2001. **74**(5): p. 482-495.

81. Wang, Z. and C. Georgakis. *On the Performance of DoDE in a class of in silico Fermentation Processes and the Impact of the Input Domain*. in *12th IFAC Symposium on Computer Applications in Biotechnology*. 2013. Mumbai, India.

82. De Tremblay, M., et al., *Optimization of Fed-Batch Culture of Hybridoma Cells Using Dynamic Programming Single and Multifeed Cases.* Bioprocess Engineering, 1992. **7**(5): p. 229-234.

83. Constantinides, A., J.L. Spencer, and E.L. Gaden, *Optimization of Batch Fermentation Processes. 1. Development of Mathematical Models for Batch Penicillin Fermentations* Biotechnology and Bioengineering, 1970. **12**(5): p. 803-830.

84. Pirt, S.J. and R.C. Righelato, *Effect of Growth Rate on Synthesis of Penicillin by Penicillium Chrysogenum in Batch and Chemostat Cultures* Applied Microbiology, 1967. **15**(6): p. 1284-1290.

85. Fotopoulos, J., C. Georgakis, and H.G. Stenger, *Effect of process-model mismatch on the optimization of the catalytic epoxidation of oleic acid using tendency models.* Chemical Engineering Science, 1996. **51**(10): p. 1899-1908.

86. Fotopoulos, J., C. Georgakis, and H.G. Stenger, *Uncertainty issues in the modeling and optimization of batch reactors with tendency models.* Chemical Engineering Science, 1994. **49**(24B): p. 5533-5547.

87. Filippi, C., et al., *Tendency modeling of semibatch reactors for optimziation and control.* Chemical Engineering Science, 1986. **41**(4): p. 913-920.

88. Filippibossy, C., et al., *Batch reactor optimziation by use of tendency models* Computers & Chemical Engineering, 1989. **13**(1-2): p. 35-47.

89. de Lucca, E.A., et al., *Modeling and Simulation of Liquid Phase Propylene Polymerizations in Industrial Loop Reactors.* Macromolecular Symposia, 2008. **271**(1): p. 8-14.

90. Schwarz, G., *Estimating the dimension of a model.* The annals of statistics, 1978. **6**(2): p. 461-464.

91. Styan, G.P.H., *Hadamard products and multivariate statistical analysis.* Linear Algebra and its Applications, 1973. **6**: p. 217-240.

92. Draper, N.R. and H. Smith, *Applied regression analysis*. 1998: Wiley.

93. Bishop, C.M., *Pattern Recognition and Machine Learning*. 2006: Springer.

94. MathWorks, *Statistics and Machine Learning Toolbox™ User's Guide (2017a). Natick,MA, MathWorks Inc.* 2017.

95. *SAS Institute Inc. JMP® 10 Design of Experiments Guide.* 2012, Cary, NC: SAS Institute Inc.

96. Meyer, R.D., D.M. Steinberg, and G. Box, *Follow-up designs to resolve confounding in multifactor experiments.* Technometrics, 1996. **38**(4): p. 303-313.

97. Atkinson, A.C. and A.N. Donev, *Optimum Experimental Designs*. 1992: Clarendon Press.

98. Bai, E.W., *An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems.* Automatica, 1998. **34**(3): p. 333-338.

99. Golub, G.H. and C.F.V. Loan, *Matrix computations* 3rd ed. 1996: Johns Hopkins University Press. 694.

100. Meyer, C., *Matrix Analysis and Applied Linear Algebra*. 2000: Society for Industrial and Applied Mathematics.

101. Wang, Z. and C. Georgakis. *Data-Driven Optimization Using an Evolutionary Design of Dynamic Experiments for Biopharmaceutical Processes*. in *AICHE Annual Meeting*. 2016. San Francisco, CA.

102. Marquardt, D.W. and R.D. Snee, *Ridge Regression in Practice.* The American Statistician, 1975. **29**(1): p. 3-20.

103. Ng, A.Y. *Feature selection, L 1 vs. L 2 regularization, and rotational invariance*. in *Proceedings of the twenty-first international conference on Machine learning*. 2004. ACM.

104. Tibshirani, R., *Regression shrinkage and selection via the lasso.* Journal of the Royal Statistical Society. Series B (Methodological), 1996: p. 267-288.

105. Friedman, J., T. Hastie, and R. Tibshirani, *A note on the group lasso and a sparse group lasso.* arXiv preprint arXiv:1001.0736, 2010.

106. Henderson, H.V. and S.R. Searle, *The vec-permutation matrix, the vec operator and Kronecker products: A review.* Linear and multilinear algebra, 1981. **9**(4): p. 271-288.

107. Kukreja, S.L., J. Löfberg, and M.J. Brenner, *A least absolute shrinkage and selection operator (LASSO) for nonlinear system identification.* IFAC proceedings volumes, 2006. **39**(1): p. 814-819.

108. Johansen, T.A., *On Tikhonov regularization, bias and variance in nonlinear system identification.* Automatica, 1997. **33**(3): p. 441-446.

109. Reginato, A.S., J.J. Zacca, and A.R. Secchi, *Modeling and simulation of propylene polymerization in nonideal loop reactors.* AIChE Journal, 2003. **49**(10): p. 2642-2654.

110. Bremner, T., A. Rudin, and D.G. Cook, *Melt flow index values and molecular weight distributions of commercial thermoplastics.* Journal of Applied Polymer Science, 1990. **41**(7-8): p. 1617-1627.

8. Appendix A: Process Simulations

8.1. Penicillin Fermentation

The penicillin fermentation examined in the thesis are from the work by Bajpai and Reuss[34]. Substrate (s) is fed in semi-batch mode to produce penicillin (p) and support the growth of biomass (x). The mathematical descriptions with 4 ordinary differential equations are given as follows

$$\frac{dV}{dt} = u \diagup s_f$$

$$\frac{dx}{dt} = \mu x - \frac{x}{s_f V}$$

$$\frac{ds}{dt} = -\frac{\mu x}{Y_{x/s}} - \rho \frac{x}{Y_{p/s}} - \frac{m_s s}{k_m + s} x + \left(1 - \frac{s}{s_f}\right)\frac{u}{V}$$

$$\frac{dp}{dt} = \rho x - kd - \frac{p}{s_f V}\mathrm{u}$$

with $\mu = \mu_{\max} \frac{s}{k_x x + s}$ and $\rho = \rho \max\left(\frac{s}{k_p + s + s^2 / k_{in}}\right)$.

The rate constants reported by Riascos and Pinto[49] are listed in the following table.

Table 8-1: Parameters of Penicillin fermentation

Parameter	Definition	Value
μ_{max}	Max biomass specific growth rate (h^{-1})	0.11
ρ_{max}	Maximum specific production rate ($g_P/g_X h$)	0.0055
K_X	Saturation parameter for biomass growth (g_S/g_X)	0.006

K_P	Saturation parameter for production (g_s/L)	0.0001
K_{in}	Inhibition parameter for production (g_s/L)	0.1
K_d	Product degradation rate (h^{-1})	0.01
K_m	Saturation parameter for maintenance consumption (g_s/L)	0.0001
m_s	Maintenance consumption rate ($g_s/g_X h$)	0.029
$Y_{X/S}$	Yield factor for substrate to biomass (g_X/g_S)	0.47
$Y_{P/S}$	Yield factor for substrate to product (g_P/g_S)	1.2

8.2. Hybridoma Cell Culture

The Hybridoma cell culture examined in the thesis are from the work by Nguang. et. al[50]. Two substrates, glucose (Glc) and glutamine (Gln) are fed in semi-batch mode to support the growth of cells (x_v) and produce antibody (MAb). Ammonia (Amm) and Lactate (Lac) are toxic byproduct. The mathematical descriptions consisting of 7 ordinary differential equations are given as follows

$$\frac{dV}{dt} = u_{Glc} + u_{Gln}$$
$$\frac{dx_v}{dt} = (\mu - k_d)x_v - \frac{(u_{Glc} + u_{Gln})}{V}x_v$$
$$\frac{dGlc}{dt} = \frac{u_{Glc}}{V}Glc_f - \frac{u_{Glc} + u_{Gln}}{V}Glc - q_{Glc}x_v$$
$$\frac{dGln}{dt} = \frac{u_{Gln}}{V}Gln_f - \frac{u_{Glc} + u_{Gln}}{V}Gln - q_{Gln}x_v$$
$$\frac{dLac}{dt} = q_{Lac}x_v - \frac{(u_{Glc} + u_{Gln})}{V}Lac$$
$$\frac{dAmm}{dt} = q_{Amm}x_v - \frac{u_{Glc} + u_{Gln}}{V}Amm$$
$$\frac{dGln}{dt} = q_{MAb}x_v - \frac{u_{Glc} + u_{Gln}}{V}MAb$$

$$\mu = \mu_{\max}\left[\frac{Glc}{K_{Glc} + Glc}\right]\left[\frac{Gln}{K_{Gln} + Gln}\right] \qquad q_{Glc} = \frac{\mu}{Y_{x_v/Glc}} + m_{Glc}\left[\frac{Glc}{k_{m_{glc}} + Glc}\right]$$

With $k_d = k_{d_{\max}}\left(\mu_{\max} - k_{d_{Lac}}Lac\right)^{-1} \times \left(\mu_{\max} - k_{d_{Amm}}Amm\right)^{-1}\left[\frac{k_{d_{Gln}}}{k_{d_{Gln}} + Gln}\right]$

$$q_{Gln} = \frac{\mu}{Y_{x_v/Gln}} \qquad q_{Lac} = Y_{Lac/Glc}q_{Glc} \qquad q_{Amm} = Y_{Amm/Gln}q_{Glc} \qquad q_{MAb} = \left[\frac{\alpha_0}{k_\mu + \mu}\right]\mu + \beta$$

The rate constants reported in the paper [50] are listed in the following table.

Table 8-2: Parameters of Hybridoma cell culture

Parameter	Value	Parameter	Value
μ_{max}	1.09 day^{-1}	$k_{d_{max}}$	0.69 day^{-1}
$Y_{Xv/glc}$	1.09×10^8 cells $mmol^{-1}$	$Y_{Xv/gln}$	3.8×10^8 cells $mmol^{-1}$
m_{glc}	0.17 mmol 10^{-8} cells day^{-1}	$K_{m_{glc}}$	19.0 mM
K_{glc}	1.0 mM	K_{gln}	0.3 mM
α_0	2.57 mg 10^{-8} cells day^{-1}	k_μ	0.02 day^{-1}
β	0.35 mg 10^8 cells day^{-1}	$k_{d_{lac}}$	0.01 day^{-1} mM^{-1}
$k_{d_{amm}}$	0.06 day^{-1} mM^{-1}	$k_{d_{gln}}$	0.02 mM
$Y_{lac/glc}$	1.8 mmol $mmol^{-1}$	$Y_{amm/gln}$	0.85 mmol $mmol^{-1}$

8.3. Propylene Polymerization

The propylene polymerization process used to examine the data-driven modeling approach is developed based on the literatures [9, 89]. The polymerization process is operated with high recycle ratio, which makes it behave like a CSTR. Therefore, we assume the polymerization process proceed in a non-isothermal CSTR. We here consider five reaction steps including the site activation, chain initiation, chain propagation, chain transfer and site deactivation. The reactions are listed in the second column of the following table. The reacting species are potential catalyst sites (S_P), co-catalyst (A), activated site (P_0), by-product (B), monomer (M) , live polymer with length i (P_i), hydrogen (H_2),dead polymer with length i (Q_i) and dead catalyst site (S_D). The rate constant and the corresponding units are given in column 3 and 4, respectively.

Table 8-3: Kinetic mechanism and rate constants for the propylene polymerization

Reaction Step	Reaction	Rate Constant	Unit
Site Activation	$S_P + A \overset{k_A}{\rightarrow} P_0 + B$	$\mathrm{k_A} = 7.04 \cdot 10^2 \mathrm{e}^{-12/\mathrm{R_g T}}$	l/gmol/s
Chain Initiation	$P_0 + M \overset{k_P}{\rightarrow} P_1$	$\mathrm{k_P} = 6.30 \cdot 10^8 \mathrm{e}^{-10/\mathrm{R_g T}}$	l/gmol/s
Chain Propagation	$P_i + M \overset{k_P}{\rightarrow} P_{i+1}$	$\mathrm{k_P} = 6.30 \cdot 10^8 \mathrm{e}^{-10/\mathrm{R_g T}}$	l/gmol/s
Chain Transfer	$P_i + H_2 \overset{k_{TH}}{\longrightarrow} P_0 + Q_i$	$\mathrm{k_{TH}} = 2.22 \cdot 10^{10} \mathrm{e}^{-14/\mathrm{R_g T}}$	l/gmol/s
	$P_i + M \overset{k_{TM}}{\longrightarrow} P_0 + Q_i$	$\mathrm{k_{TM}} = 2.76 \cdot 10^7 \mathrm{e}^{-14/\mathrm{R_g T}}$	l/gmol/s
	$P_i \overset{k_{TS}}{\longrightarrow} P_0 + Q_i$	$\mathrm{k_{TS}} = 1.72 \cdot 10^3 \mathrm{e}^{-14/\mathrm{R_g T}}$	1/s
Site Deactivation	$P_i \overset{k_D}{\rightarrow} S_D + Q_i$ $P_0 \overset{k_D}{\rightarrow} S_D$	$\mathrm{k_D} = 7.92 \cdot 10^3 \mathrm{e}^{-12/\mathrm{R_g T}}$	1/s

The ordinary differential equations used to describe the non-isothermal polymerization process in a CSTR are given below.

$$\frac{dC_M}{dt} = \frac{F}{V}\left(C_{M,in} - C_M\right) - \left(k_A C_{S_P} + k_P C_{P_0} + k_P Y_0 + k_{TM} Y_0\right) C_M$$

$$\frac{dY_0}{dt} = k_P C_{P_0} C_M - \left(k_{TH} C_{H_2} + k_{TM} C_M + k_{TS} + k_D\right) Y_0 - \frac{F}{V} Y_0$$

$$\frac{dY_1}{dt} = k_P C_{P_0} C_M - \left(k_{TH} C_{H_2} + k_{TM} C_M + k_{TS} + k_D\right) Y_1 + k_P C_M Y_0 - \frac{F}{V} Y_1$$

$$\frac{dY_2}{dt} = k_P C_{P_0} C_M - \left(k_{TH} C_{H_2} + k_{TM} C_M + k_{TS} + k_D\right) Y_2 + k_P C_M \left(2Y_1 + Y_0\right) - \frac{F}{V} Y_2$$

$$\frac{dX_0}{dt} = \left(k_{TH} C_{H_2} + k_{TM} C_M + k_{TS} + k_D\right) Y_0 - \frac{F}{V} X_0$$

$$\frac{dX_1}{dt} = \left(k_{TH} C_{H_2} + k_{TM} C_M + k_{TS} + k_D\right) Y_1 - \frac{F}{V} X_1$$

$$\frac{dX_2}{dt} = \left(k_{TH} C_{H_2} + k_{TM} C_M + k_{TS} + k_D\right) Y_2 - \frac{F}{V} X_2$$

$$\frac{dC_{S_P}}{dt} = \frac{F}{V}\left(C_{S_P,in} E_{cat} - C_{S_P}\right) - k_A C_A C_{S_P}$$

$$\frac{dC_{P_0}}{dt} = k_A C_A C_{S_P} + \left(k_{TH} C_{H_2} + k_{TM} C_M + k_{TS}\right) Y_0 - \left(k_P C_M + k_D Y_0\right) C_{P_0} - \frac{F}{V} C_{P_0}$$

$$\frac{dC_A}{dt} = \frac{F}{V}\left(C_{A,in} - C_A\right) - k_A C_{S_P} C_A$$

$$\frac{dC_{H_2}}{dt} = \frac{F}{V}\left(C_{H_2,in} - C_{H_2}\right) - k_{TH} Y_0 C_{H_2}$$

$$\frac{dT}{dt} = \frac{F}{V}\left(T_{in} - T\right) + \frac{-\Delta H_r \left(k_P C_M Y_0\right)}{\rho c_P} + \frac{4U}{D\rho c_P}\left(T_c - T\right)$$

$$\frac{dC_B}{dt} = k_A C_A C_{S_P} - \frac{F}{V} C_B$$

$$\frac{dC_{Sol}}{dt} = \frac{F}{V}\left(C_{Sol,in} - C_{Sol}\right)$$

$$\frac{dT_c}{dt} = \frac{F_c}{V_c}\left(T_{c,in} - T_c\right) - \frac{4U}{D\rho_c c_{P,c}}\left(T_c - T\right)$$

The parameters for the polypropylene simulations are given in Table 8-4.

Table 8-4: Values of the parameters for the propylene polymerization in a CSTR

V (L)	D (dm)	F(L/s)	F_c(L/s)
50303[9]	6[9]	11.55[9]	220[9]
T_{in}(K)	$T_{c,in}$(K)	ρ_c (gmol/L)	ρ (gmol/L)[a]
318.15[9]	327.15[9]	55.56	14.64
E_{cat}(gmol/g)	ΔH_r (kcal/gmol)	U (kcal/dm^2 K s)	cp (kcal/gmol K)[a]
0.0188[9]	-20[109]	0.1[109]	1.6548
cp_c (kcal/gmol K)	$C_{cat,in}$(g/L s)	$C_{M,in}$(gmol/L s)	$C_{A,in}$(gmol/L s)
18	0.26[9]	10.1[9]	0.105[9]

[a]Here take the heat capacity (density) of propylene as the average heat capacity (density) for all reactants and products.

The melt flow index are calculated by $MFI = a(\bar{M}_w)^b$ with $a = 3.39 \times 10^{22}$ and $b = -3.92$ [110]. The number average molecular weight is calculated by $\bar{M}_n = (X_1 + Y_1)/(X_0 + Y_0)$ while the weight average molecular weight is calculated by$\bar{M}_w = (X_2 + Y_2)/(X_1 + Y_1)$.

The nomenclature for the species and parameters involved in the simulation is listed as follows:

C Concentration

cp Specific heat capacity

D Diameter of the reactor

E_a Activation energy

F Inlet/outlet flow

ΔH	Reaction heat
H_2	Hydrogen
T	Temperature
U	Heat transfer coefficient
V	Volume
X_i	i-th moment of dead chain
Y_i	i-th moment of live chain
ρ	Density

9. Appendix B: Publications and Presentations

9.1. Publications

Wang, Z. and C. Georgakis, *New Dynamic Response Surface Methodology for Modeling Nonlinear Processes over Various Time Horizons.* Industrial & Engineering Chemistry Research. DOI: 10.1021/acs.iecr.7b02381

Wang, Z. and C. Georgakis, *An in silico evaluation of data‐driven optimization of biopharmaceutical processes.* AIChE Journal, 2017. **63**(7): p. 2796-2805.

Wang, Z., N. Klebanov, and C. Georgakis. *DRSM Model for the Optimization and Control of Batch Processes*. in *Dynamics and Control of Process Systems, including Biosystems/IFAC*. 2016. Trondhelm, Norway

Wang, Z. and C. Georgakis. *On the Performance of DoDE in a class of in silico Fermentation Processes and the Impact of the input Domain*. in *12th IFAC Symposium on Computer Applications in Biotechnology*. 2013. Mumbai, India.

9.2. Presentations

Wang, Z. and C. Georgakis. *Data-Driven Optimization Using an Evolutionary Design of Dynamic Experiments for Biopharmaceutical Processes*. in *AICHE Annual Meeting*. 2016. San Francisco, CA.

Wang, Z. and C. Georgakis. *Data-Driven Optimization Using an Evolutionary Design of Dynamic Experiments for Biopharmaceutical Processes*. in AICHE Annual Meeting. 2016. San Francisco, CA.

Wang, Z. and C. Georgakis. *A New Dynamic Response Surface Methodology for Modeling the Dynamics of Nonlinear Processes* in AICHE Annual Meeting. 2017. Minneapolis,MN.