UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI INGEGNERIA INDUSTRIALE

CORSO DI LAUREA MAGISTRALE IN INGEGNERIA CHIMICA E DEI PROCESSI INDUSTRIALI

Tesi di Laurea Magistrale in
Ingegneria Chimica e dei Processi Industriali

DESIGN OF DYNAMIC EXPERIMENTS FOR THE IDENTIFICATION OF DATA-DRIVEN DYNAMIC MODELS

Relatore: Prof. Massimiliano Barolo

Correlatore: Ing. Pierantonio Facco

Laureanda: SARA SBROGGIÒ

ANNO ACCADEMICO 2017 – 2018

Abstract

A data-driven approach commonly used to support the modelling of unknown systems is the design of experiments (DoE; Fisher 1971). The application of the DoE methodology is limited to the cases in which the input variables (or factors) are constant with time. In batch and semi-batch processes, the evolution of an operating condition with time can have a significant impact on the results (e.g. product specifications, reaction yield). The design of dynamic experiments (DoDE) is a new methodology developed by Georgakis (2013) to incorporate input variables that are time-dependent. The modeling approach, necessary to statistically estimate the relationship between inputs (both statics and dynamics) and the output variable is the response surface model (RSM; Box and Draper, 1987). However, the RSM describes the behavior of the output at the end-point of the process only. A generalization of RSM that captures the effects of the process inputs on time-resolved output has been proposed in Klebanov and Georgakis (2016). The latter is called dynamic response surface model (DRSM).

In this thesis, the above methodologies are reviewed and critically applied. Empirical models are built to describe the relationship between inputs variables (regardless of their nature), and the output, in order to calculate the operating optimum of a dynamic process. Its application to three case studies allows to compare the DoE methodology with the DoDE one and to define the utility of a DRSM in the optimization of the end-point of the process.

The results show how the DoDE methodology leads to an improvement in the definition of the optimal operating conditions, of compared to DoE. Instead, the use of a DRSM can only confirm the result obtained by DoDE.

Riassunto

Il design of experiments (DoE) è la tecnica di pianificazione degli esperimenti basata su dati più comunemente utilizzata per descrivere processi batch e semi-batch (Fisher 1971). Tra le strategie di pianificazione degli esperimenti considerate dal DoE è possibile ricordare quella fattoriale completa e quella cubica a facce centrate. L'applicazione di tale metodologia permette di ridurre al minimo il numero di esperimenti necessari alla definizione di un modello empirico in grado di descrivere l'influenza di una o più variabili in ingresso al processo su una variabile in uscita. L'utilizzo del DoE è però circoscritto ai casi in cui le variabili (dette anche fattori) in ingresso siano costanti nel tempo. Tuttavia, questa appare essere un'importante limitazione nei processi batch e semi-batch, sui quali la variazione temporale delle condizioni operating può avere un impatto significativo.

Una nuova metodologia, definita design of dynamic experiments (DoDE) è stata recentemente sviluppata da Georgakis (2013) con lo scopo di pianificare esperimenti che coinvolgano variabili in ingresso dinamiche. I profili in ingresso sono definiti come combinazione lineare di polinomi di Legendre e di coefficienti numerici chiamati sotto-fattori dinamici. Le stesse strategie di pianificazione che nel DoE determinano i livelli di variazione dei fattori statici, sono utilizzate nel DoDE per definire le diverse combinazioni di sotto-fattori dinamici. L'applicazione sia del DoE che del DoDE mira a costruire un modello empirico che stimi statisticamente la relazioni tra variabili in ingresso (sia statiche che dinamiche) e quella in uscita (chiamata anche variabile di risposta). Questi modelli sono definiti superfici di risposta (Box e Draper, 1987) e, in accordo con il numero di esperimenti eseguiti, possono essere di primo o secondo ordine. Sono modelli di primo ordine quelli che descrivono gli effetti lineari dei fattori in ingresso sulla variabile di risposta. Per definire invece la dipendenza quadratica è necessario costruire un modello del secondo ordine. Tuttavia, le superfici di risposta descrivono il processo in un preciso istante temporale, rappresentato, normalmente, dall'istante finale. Per questo motivo, si possono definire anche superfici di risposta statiche. Con lo scopo di descrivere l'intero profilo della variabile di risposta, un nuovo approccio di modellazione, chiamato superficie di risposta dinamica, è stato proposto (Klebanov e Georgakis 2016). Questi modelli dinamici definiscono la dipendenza della variabile di risposta sia dai fattori in ingresso, qualsiasi sia la loro natura, che dal tempo. Le Superfici di Risposta Dinamiche sono costruite come combinazioni dei polinomi di Legendre.

In questa tesi, viene definita una metodologia che, basandosi sulle metodologie fino a qui discusse, consente di costruire un modello empirico in grado di descrive la relazione tra le variabili in ingresso e una variabile in uscita, al fine di calcolare le condizioni operative ottimali di un sistema dinamico. Questo procedimento di ottimizzazione prevede cinque passaggi.

Il primo passaggio richiede la definizione della variabile in uscita al processo che si vuole ottimizzare e delle variabili in ingresso che maggiormente la influenzano. Per fare ciò, ci si può basare sia su dati di letteratura, sia su test di screening precedentemente eseguiti. È necessario inoltre definire l'intervallo di variazione delle variabili in ingresso e, nel caso di variabili dinamiche, il numero di sotto-fattori utilizzati.

Il passaggio successivo richiede di identificare la strategia di pianificazione degli esperimenti adatta a garantire la definizione del modello empirico desiderato. Un modello quadratico richiederà infatti un numero maggiore di esperimenti rispetto ad uno lineare.

Una volta definiti i valori assunti dai fattori statici e dai sotto-fattori dinamici, gli esperimenti sono eseguiti con l'obiettivo di raccogliere i dati relativi alla variabile in uscita necessari alla costruzione di una superficie di risposta statica o dinamica.

I coefficienti di una superficie di risposta statica sono calcolati considerando i valori assunti dalla variabile di risposta alla fine del processo. Per la costruzione di una superficie di risposta dinamica, la misurazione della variabile deve essere invece eseguita ad intervalli di tempo regolari durante il processo. Sia la superficie di risposta statica che quella dinamica sono utilizzate per determinare le condizioni operative ottimali del processo. Nel primo caso, l'ottimizzazione riguarderà il valore assunto dalla variabile di risposta alla fine del batch. Nel caso di una superficie di risposta dinamica il criterio di ottimalità prevede, invece, la massimizzazione di una funzione del profilo della variabile in uscita (per esempio, il suo integrale).

L'intera procedura è applicata a tre casi studio:

- un generico reattore batch con reazione reversibile;
- un reattore semi-batch;
- un processo di fermentazione della penicillina.

Nel primo caso studio si è cercato di ottimizzare il profilo della temperatura del reattore in modo tale da massimizzare la conversione finale del reagente. Nel caso studio del reattore semi-batch si vuole controllare la concentrazione in uscita di un prodotto manipolando la portata in ingresso di uno dei reagenti. Nel terzo caso invece sono considerati due fattori statici, rispettivamente la concentrazione di biomassa iniziale e la durata del batch e di un fattore dinamico, ovvero, la portata di substrato in ingresso. La variabile in uscita è la concentrazione finale di penicillina.

L’applicazione della procedura ai casi studio sopra citati ha dato modo di confrontare i risultati ottenibili attraverso DoE e DoDE e definire l’utilità delle superfici di risposta dinamiche nell’ottimizzazione del valore finale della variabile di risposta.
I risultati mostrano come la metodologia DoDE porti ad un miglioramento nella definizione delle condizioni operative ottimali, rispetto al DoE. Invece, l’uso di una superficie di risposta dinamica può solo confermare il risultato ottenuto dal DoDE.

Table of contents

Introduction

In batch and semi-batch processes, a data-driven modeling approach is often convenient since usually it is not possible to develop a detailed and accurate knowledge-driven model describing the process (Georgakis 2013). The classical data-driven approach to set up the experimental campaign is the design of experiments (DoE). This methodology, proposed by Fisher (1971), allows running a minimal number of experiments to characterize the behavior of a process output. DoE considers defining an empirical model, called response surface model (RSM; Box and Draper, 1987) that describes the relationship between the input variables, often called factors, and the output variable. However, the DoE methodology is limited by the fact that all the factors examined are invariant with time. In the case of batch and semi-batch processes, the time evolution of the input variables can have a substantial impact on the results. For this reason, a new methodology, for the description of dynamic experiments, has been developed. This methodology is called design of dynamic experiments (DoDE) and it has been proposed by Georgakis (2013). This approach is based on the design criteria used in DoE and allows defining a RSM that relates the time-variant input to a single output at the end-point process. Klebanov and Georgakis (2016) proposed the dynamic response surface model (DRSM) methodology. The latter is a generalization of the RSM that, starting from the results of DoDE, is able to describe the relationship between the input variables and the profile of the output along the process.

The objective of this thesis is to bring together the methodologies already explored (DoE and RSM) and the new ones (DoDE and DRSM) to determine the operating optimum of batch and semi-batch process. The aim is to understand if the new methodologies are able to define operating conditions that allows obtaining an improvement in the product specifications (e.g. product concentration, conversion of a reactant), with those achievable using the older ones. The thesis is divided in four chapters. In Chapter 1, the procedure useful to build an empirical model, which describes the relationship between manipulated inputs and process output, and to define the operating optimum of a dynamic process, is reported. In Chapter 2 three case studies, to which the procedure will be applied, are described. The case studies considered are:

- a batch reactor with reversible reaction,
- a semi-batch reactor with a network of three reactors,
- a penicillin fermentation process.

In Chapter 3, the optimization procedure is applied to the case study of a batch reactor with reversible reaction, while in the fourth chapter it is applied to the case study of the penicillin fermentation process. The development of the case study of a semi-batch reactor is reported in Appendix A.

Chapter 1

Mathematical background on design of dynamic experiments

In this chapter, a methodology for the optimization of the operating conditions of batch, semi-batch or fed-batch processes using data-driven models is described.

1.1 Introduction to a designed experimentation

The procedure presented in this chapter allows building an empirical model that describe the relationship between manipulated inputs and process output and defining the operating optimum of a dynamic process. It can be applied to batch and semi-batch processes in which first principle model is not available. First, it is necessary to define the process output that has to be optimized and the input variables, static or dynamic, that influence the output. Through some screening tests, it is possible to define the input variables that have a greater influence on the process and the appropriate ranges of variation within which the optimization analysis has to be concentrated. Once defined the variation interval of the input variables, a strategy for the design of the experiments must be identified. The experimental campaign is made by a series of tests, called runs, in which changes are made in the input variables in order to identify the reason for changes in the output response. The way in which the inputs are modified between experiments is defined using DoE criteria. The application of DoE for the definition of the design strategy allows obtaining the largest number of information, by minimizing the number of experiments to be performed. The data obtained from the experimental campaign, in terms of the value of the output variable, have to ensure the definition of an empirical model that describes the relationship between the input variables and the output. The design strategy has to consider both the complexity of the empirical model in terms of representativeness of the process, and the number of informative experiments that have to be done.

The empirical model defined will be used to calculate the operating optimum of the process.

Five general steps that compose the optimization procedure are schematized in Figure 1.1.

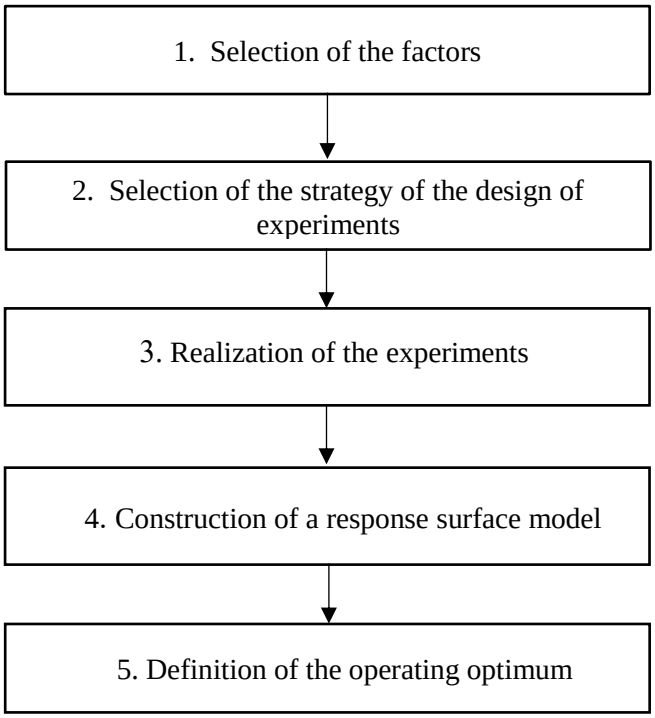


Figure 1.1 *Scheme of the optimization procedure*

As reported in Figure 1.1, the procedure for the process optimization can be schematized in the following steps:

- selection of factors;
- choice of the strategy of the design of experiments;
- realization of the experiments;
- construction of a response surface model;
- definition of the operating optimum.

The first part of the procedure consists in the design of dynamic experiments (DoDE). DoDE is a generalization of the statistical design of experiments (DoE), which is a widely used data-driven approach to explore the functional relation among output variables (e.g.: the product quality) and the input variables, usually called factors (e.g.: process operating conditions, initial settings, etc.). The factors considered in the DoE are static variables. Nevertheless, in a batch process, the time evolution of operating conditions has an important impact on the process output. DoDE can be used when a process is affected both by static and dynamic factors, so that it is possible to understand which is the profile of variation of a dynamic input that guarantees the optimal value of the process output. The application of this methodology allows the definition of static response surface model (RSM) and dynamic response surface model (DRSM).

RSM is a data-driven model structure that captures the relationship between process input (time-variant or time-invariant) and a single output, usually at the end-point of the process. DRSM, as an expansion of RSM, defines the correlation between input variables and time-resolved output variables.

1.1.1 Selection of the factors

The first step of the procedure consists in the selection of the factors that more influence the process. They can be: i) static (W_j), ii) or dynamic, namely time-varying. This is a standard step in the design of experiments, which needs to define also the variation range of the factors.

The selection of the factors needs screening tests in order to highlight those that have a greater influence on the output variable that must be optimized. These tests allow also to determine the variation range of the input variables on which the analysis may be focused on.

The variation of the static factors in the design of experiments (DoE) is described in Montgomery, D. (2013) and reported here in Subsection 1.1.2.

1.1.1.1 Static factors

DoE enables the design of a set of experiments, so that the maximum information is obtained or, conversely, the minimum number of experiments is performed to obtain the desired information. The procedure consists in the definition of the values that an input factor must assume during each experiment. The target is to collect the values of the process output in order to define the relationship between input and output variables. Numerical values among which the input factor changes are called levels. The variation of the factor W_j is usually codified to range its variation between -1 and +1. The codified version of factor W_j is indicated with a lowercase letter (w_j) and its levels $w_{i,j}$ are defined as:

$$w_{i,j} = \frac{W_{i,j} - DW_j}{dW_j} \quad , \tag{1.1}$$

$$-1 \leq w_{i,j} \leq +1 \tag{1.2}$$

with

$$DW_j = \frac{W_{max,j} + W_{min,j}}{2} \quad , \tag{1.3}$$

$$dW_j = \frac{W_{max,j} - W_{min,j}}{2} \quad . \tag{1.4}$$

The $W_{max,j}$ and $W_{min,j}$ represent the maximum and the minimum value of the variation range of factor W_j. The DoE methodology is usually applied to the case of static factors. In most cases, the input parameters must describe a time variable profile during the batch process; to design this type of processes the design of dynamic experiment is used.

1.1.1.2 Dynamic factors

For the sake of simplicity, consider one single factor X that varies inside the batch process with time. The factor can vary with different time profiles $X_m(t)$, whose set of *M* profiles is indicated with the bond letter $\mathbf{X}(\mathrm{t})$. As the static factors in DoE, the time profiles must be codified to range their variation between -1 and +1 in a time interval τ that is considered here a dimensionless time ranging from 0 (i.e., start of the process) to 1 (i.e., completion of the process). The set of codified profiles is defined by the lowercase letter $\mathbf{x}(\tau)$ while a single profile is defined using $x_m(\tau)$ and calculated as:

$$x_m(\tau) = \frac{X_m(t) - DX(t)}{dX(t)} \quad , \tag{1.5}$$

$$-1 \leq x_m(\tau) \leq +1 \quad , \tag{1.6}$$

and

$$DX(\mathrm{t}) = \frac{X_{max}(t) + X_{min}(t)}{2} \quad , \tag{1.7}$$

$$dX(t) = \frac{X_{max}(t) - X_{min}(t)}{2} \quad , \tag{1.8}$$

where the $X_{max}(t)$ and $X_{min}(t)$ are the maximum and minimum values that the factor X can assume. The novelty brought by DoDE consists in the way of defining the profiles $\mathbf{x}(\tau)$. It is assumed that all functions of time belong to the Hilbert space of squares integrable functions in $\tau \in [0, 1]$ interval. The profiles of the input variable must be defined as a linear combination of an ortho-normal set of functions that is a basis in the Hilbert space (Georgakis 2016) namely the shifted Legendre polynomials, and the dynamic subfactors $x_{n,m}$. The first five polynomials are reported in Table 1.1:

Table 1.1 *Shifted Legendre polynomials* $P_n(\tau)$

n	$\boldsymbol{Pn(\tau)}$:
n=0	1
n=1	$-1 + 2\tau$
n=2	$1 - 6\tau + 6\tau^2$
n=3	$-1 + 12\tau - 30\tau^2 + 20\tau^3$
n=4	$1 - 20\tau + 90\tau^2 - 140\tau^3 + 70\tau^4$

where *n* is the degree of the shifted Legendre polynomial $P_n(\tau)$. Each profile is defined as a combination of this set of functions:

$$x_m(\tau) = \sum_{n=1}^{N} x_{n,m} P_{n-1}(\tau) \quad . \tag{1.9}$$

The coefficients $\{x_{1,m}, x_{2,m}, \dots x_{N.m}\}$ are called dynamic subfactors and *N* is the total number of subfactors used. The profiles, whose total number *M* is related to the number of experiments scheduled, are differentiated according to the choice of the dynamic subfactors.

Since the value of each profile has to range between -1 and +1 according to Equation (1.6), the $x_{n,m}$ coefficients must observe the following inequalities:

$$-1 \leq x_{1,m} \pm x_{2,m} \pm x_{3,m} \pm \cdots \pm x_{N,m} \geq +1 \, . \tag{1.10}$$

The total number of dynamic subfactors *N* is also the total number of shifted Legendre polynomial used for the description of the input profiles while *N-1* is the maximum degree of polynomial used. If only the first Legendre polynomial P_0 is used, the profiles performed are constant with time while if the second or the third Legendre polynomials are added, the input factor can vary respectively in a linear or a quadratic way with time. The combinations of the dynamic subfactors that define the profiles of variation of the dynamic variable has to satisfy the constraint in Equation (1.10). However, other operating constraints must be considered in the design step.

1.1.2 Strategy for designing the experiments

Once the levels of variation are defined both of the static factors and the dynamic subfactors identified in the previous step, it is necessary to choose a proper strategy. It has to consider both the number of experiments that are able to perform, according to cost, time, or the availability of the facilities and the information that are aimed to collect from the experimental campaign.

In this paragraph, the design strategy that will be applied in the Chapter 3, to different case studies, are described. These are:

- Full Factorial (FFD)
- Central Composite Scheme (CCD)
- D-Optimal Design (DOD)

1.1.2.1 Full factorial design (FFD)

The full factorial design allows a complete collection of information because investigates all the possible combinations among all the levels for every factor. If each factor varies its value along I-levels, the number of runs necessary for a I-level full factorial design is *IJ*, where *J* is the number of factors. For example, in the case of 2 factors varied on 3-levels, a FFD requires 9 experiments. The 3 levels are usually defined as the lower, an intermediate and the higher value (usually equally spaced) that the factor can assume. (Montgomery, 2013)

If the variation of the factors is codified between -1 and +1, the -1 is considered the lower level, 0 the intermediate one, and +1 the higher level. The design matrix for two factors W_1 and W_2 codified as $w_1$and w_2 that vary among three levels (-1, 0, +1) is reported in Table 1.2:

Table 1.2 *Experimental plan in the case of three level full factorial design with two factors*

Run	Coded Variables	
	w_1	w_2
1	-1	+1
2	0	+1
3	+1	+1
4	-1	-1
5	0	-1
6	+1	-1
7	-1	0
8	0	0
9	+1	0

FFD is an orthogonal experimental design method because the scalar product of the columns of the design matrix in Table 1.2 is zero. The design points can be represented inscribed in a square as in Figure 1.2.

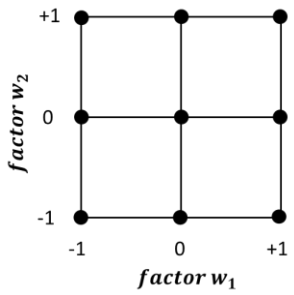


Figure 1.2. *Geometrical representation of a three-level full factorial design with two factors*

To the point defined by the structure of a FFD, it is necessary to add some replicated runs in order to estimate the accuracy of the measurements.

The performance of a FFD sometimes required high costs and long times to do all the tests. In fact, as the number of factors in a I-level factorial design increases, the number of runs necessary to do increases quickly. For example, in the case of three factors that vary among three levels 27 experiments are needed, 81 for 4 factors. In these cases, a fractional factorial design is preferable as it requires only half of those runs. Therefore, the result decreases the information that can be collected.

1.1.2.2 Central composite design

A more complicated experimental design, useful in the case of two or more than two factors, is a central composite design (CCD). This is an experimental design through which is possible to build a second order (quadratic) model for the response variable without needing to use a complete three-level factorial experiment.
Two types of CCD scheme are reported in the figure below:

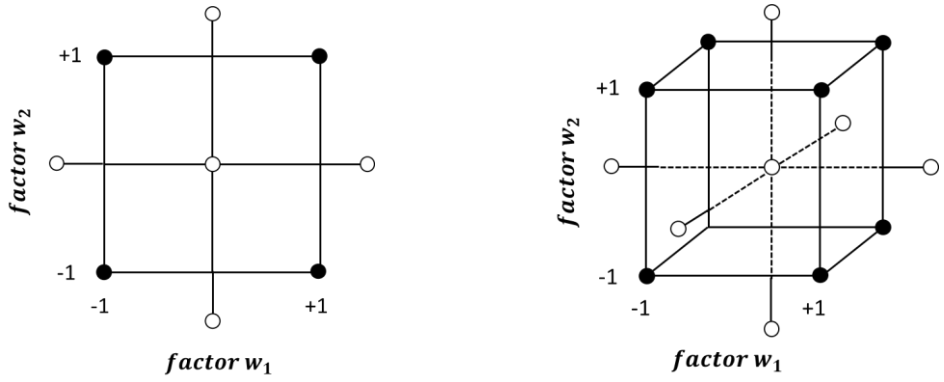


Figure 1.3. *Geometrical representation of a central composite design in the case of a) two factors and b) three factors.*

The CCD consists of:

- 2^j factorial design, that is described by the black points in Figure 1.2,
- a star design consisting of $2j + 1$ points.

In summary, CCD allows to obtain data to estimate first-order and interaction effects for each factor and additionally provides data to estimate second-order effects. One of these points is the central one, the experiment in this point is often repeated in order to define the precision of the experimental measurements (Trutna *et al*, 2012). The distance from the central point of the design space to a star point (in white in the Figure 1.3) is equal to α. This value is calculated according to the equation:

$$\alpha = [number\ of\ factorial\ runs]^{1/4} = \left[2^j\right]^{1/4}. \tag{1.11}$$

In the case of two factors a CCD requires the same number of trials (9) as 3^2 FFD, while in the case of three or four factors the CCD requires respectively only 15 and 25 trials.

1.1.2.3 D-optimal design

DOD is used when the latter requires too many runs for the amount of resources or time allowed for the experiment or when the design space is constrained. Unlike standard classical designs, such as factorials and fractional factorials, D-optimal design matrices are usually not orthogonal (Trutna *et al*, 2012). As a starting point in DOD, a set of candidate samples is needed and a restricted subset of them is selected according to the optimality criterion. This is defined using a full factorial experimental design with many levels for each factor. D-optimal designs depends on the adopted optimality criterion and on the response surface model which is intended to obtain (e.g. first order, first order plus interactions, full quadratic, cubic, etc.). **Z** is the matrix containing the set of candidate samples. The D-optimal design is built by looking for the submatrix **z** for which the value of the determinant of **z'z**, is maximized where **z'** is the transposed version of **z.** This optimality criterion aims at minimizing the variance of the parameters that will be calculated in the response surface models.

1.1.2.4 Example of design of dynamic experiments with linear profiles

DoDE allows the application of the DoE strategies to time-variant factors. In particular, it suggests building the experimental plan considering both the dynamic subfactors (x_n) and static factors (w_j). This means that the choice of the strategy, according to those presented in this paragraph, depends both on the number of dynamic subfactors and static factors.

In this subparagraph, some examples of the design of dynamic experiments, in the case of only one dynamic factor that can be described by a different number of dynamic subfactors, are reported. Consider for example the simplest set of DoDE experiments: this is obtained by selecting N=1 in the equation (1.11), meaning that the dynamic profile of $X(t)$, and accordingly its codified value $x(\tau)$ is a linear combination of the first two Legendre polynomials $P_0(\tau)$ and $P_1(\tau)$ reported in Table 1.1. The equation (1.9) can be rewritten in this form:

$$x_m(\tau) = x_{1,m}P_0(\tau) + x_{2,m}P_1(\tau) \ . \qquad (1.12)$$

The experiments are characterized by different profiles of $x(\tau)$ which variation is determined by the values of dynamic subfactors $x_{1,m}$ and $x_{2,m}$.

The values of the subfactors are varied on different levels, based on the design strategy that it is wanted to use. To perform a two-level full factorial design, the following four (2^2) experiments are described:

$$\begin{pmatrix} x_{1,m} \\ x_{2,m} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix} \quad . \tag{1.13}$$

However, these combinations do not meet the constrain in Equation (1.6) for all the experiments. For example, the case of $\left(x_{1,2}\,,\, x_{2,2} \right) = (1, -1)$ generates the following profile, that are indicated with $x_2(\tau)$:

$$x_2(\tau) = x_{1,2}\, P_0(\tau) + x_{2,2} P_1(\tau) = 1 - (1 - 2\tau) = 2\tau \tag{1.14}$$

which does not satisfy Equation (1.6) when τ is > 0.5. Therefore, the values of $x_{1,m}$ and $x_{2,m}$ are changed according to the inequality in equation (1.10), that in this case has the following formulation:

$$-1 \leq x_{1,m} \pm x_{2,m} \leq 1 \quad . \tag{1.15}$$

These lead to the following FF DoDE:

Table 1.3 *Combinations of two dynamic subfactors $x_{1,m}$, and $x_{2,m}$ in the case of two levels full factorial design with two factors*

Run	**Coded Variables**	
	$x_{1,m}$	$x_{2,m}$
1	-0.5	+0.5
2	-0.5	-0.5
3	+0.5	+0.5
4	+0.5	-0.5

The values in Table 1.3 define four linear profiles of variation of the dynamic factor that are reported in Figure 1.4.

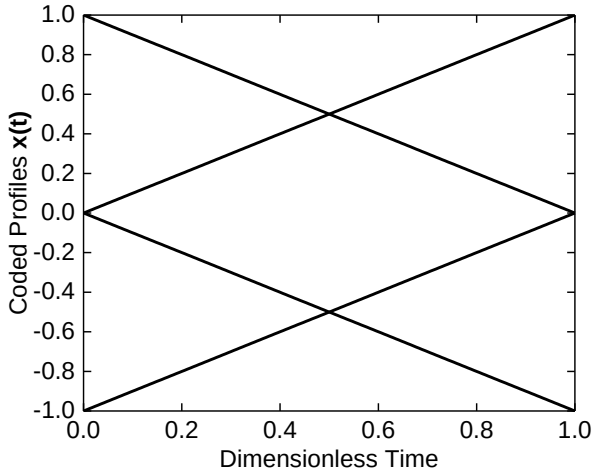


Figure 1.4 *Profiles of variation of the dynamic factor in the case of two level full factorial DODE with two dynamic subfactors*

1.1.2.5 Example of design of dynamic experiments with non-linear profiles

In the previous example linear profiles are used in DoDE. The profiles of a dynamic factors may not be linear. For example, consider the case of a quadratic time profile: in this case the first three Legendre polynomials: $P_0(\tau)$, $P_1(\tau)$ and $P_2(\tau)$ are considered. As a consequence, the values of three subfactors: $x_{1,m}$, $x_{2,m,}$ and $x_{3,m}$ are varied in a two levels full factorial DoDE. In order to define the correct combinations of the values that the subfactors assume, it is necessary to verify that each combination satisfies the inequality of Equation (1.6). In Table 1.4 the design matrix for a 2-level FF design in the case of three subfactors is reported.

Table 1.4 *Combinations of three dynamic subfactors x_1, x_2 and x_3 in the case of two level full factorial design with three factors.*

Run	**Coded Variables**		
	x_1	x_2	x_3
1	0,33	0,33	0,33
2	0,33	-0,33	0,33
3	0,33	0,33	-0,33
4	0,33	-0,33	-0,33
5	-0,33	0,33	0,33
6	-0,33	-0,33	0,33
7	-0,33	0,33	-0,33
8	-0,33	-0,33	-0,33

The quadratic profiles of the dynamic factor, related to the eight experiments, are shown in Figure 1.5

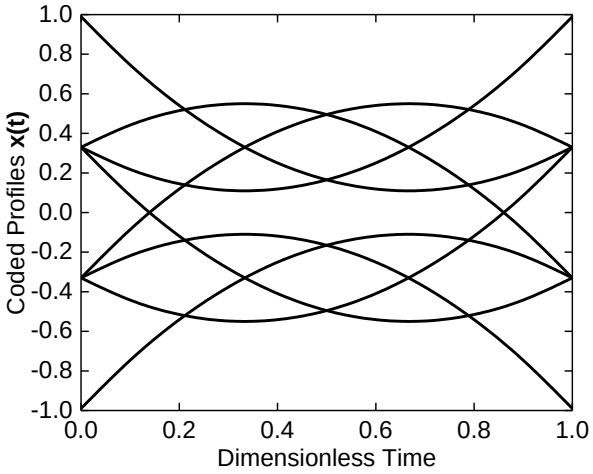


Figure 1.5 *Profiles of the dynamic factor in the case of two level full factorial DODE with three dynamic subfactors*

The design matrix in the Tables 1.3 and 1.4 illustrates the experimental plan for the description of the influence of a single dynamic factor on the process. In the case of DoE, it would necessary to consider the same number of experiments to describe two static factors, as shown in the Table 1.2. This means that the implementation of DoDE requires at least twice the number of experiments of DoE.

1.1.3 Execution of the experiments

Once defined the experimental plan, the practitioner has to perform the experiments in a random order. This part of the procedure has the purpose to collect the data related to the response variable that will be used in the next step to define a response surface model. The experiments are performed in silico using MATLAB® software.

1.1.4 Construction of a response surface model

Two strategies can be adopted for empirical model building: response surface model (RSM) and dynamic response surface model (DRSM)

1.1.4.1 Static response surface models (RSM)

The RSM is an empirical modeling approach to characterize the relationship between factors and outputs (Myers *et al*, 2009). The relationship can be modelled by a parametric equation that defines the influence of factors on the process output, that is the response variable such the conversion of a reactant or a concentration of a product at the end-point of the batch. The general equation that defines the relationship between a response variable y and two input factors x_1 and x_2 is the following:

$$y = \boldsymbol{f}(x_1, x_2)\boldsymbol{b} + \boldsymbol{\varepsilon} \tag{1.16}$$

where $\boldsymbol{f}(x_1, x_2)$ is the vector function of *d* elements that consists of powers and cross-product powers of x_1 and x_2 up to a certain degree, $\boldsymbol{b}$ is a vector of *d* unknown constant coefficients and $\boldsymbol{\varepsilon}$ is the random experimental error assumed to have zero mean. The vector function $\boldsymbol{f}(x_1, x_2)$ and the number of coefficients *d* depend on the degree of the model chosen to describe the influence of the input factors on the response variable. In the equation below, both the first and the second order terms are reported.

$$\hat{y} = \boldsymbol{f}(x_1, x_2)\boldsymbol{b} = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 + b_{11} {x_1}^2 + b_{22} {x_2}^2 \ . \tag{1.17}$$

The first terms are useful to define the linearity of the factors effects. If interaction terms are added to the first order, the model can represent some curvature in the response function. However, in some situations, the curvature of the response variable is not adequately modeled using only linear and interaction terms. In this case, the approximation of the response variable required second order terms to describe the quadratic effects. The model chosen depends on the design strategy according to which the experiments were performed. The most common first-order designs are 2-levels FF, while, a 3-level FFD or a CCD are usually required to define a second-order model. If *J* is the number of independent factors, the number of parameter *p* in the second-degree model is calculated as:

$$d = 1 + 2J + \frac{1}{2}J(J-1) \ . \tag{1.18}$$

Therefore, the number of distinct design points of a second order must be at least equal to *d*. (Montgomery, 2013)

The values of the response variables, necessary to define the empirical model, are those collected during the experimental campaign as the one in 1.1.3. If the total number of experiments is *M*, the same number of values of the response variable are collected. Using $\hat{y}$ the set of predicted values of y is indicated while with $\hat{y}_m$ is indicated the predicted value of the y_m experimental data. The estimation of the regression coefficients *d* in a multiple linear regression model is performed through the MATLAB function "*regress*" that uses the least square method. If y_m is characterized by the following equations:

$$y_m = \hat{y}_m + \varepsilon_m \ , \tag{1.19}$$

$$\hat{y}_m = b_0 + b_1 x_{1,m} + b_2 x_{2,m} + b_{12} x_{1,m} x_{2,m} + b_{11} {x_{1,m}}^2 + b_{22} {x_{2,m}}^2 \qquad (1.20)$$

the least squared method chooses $\boldsymbol{b}$ in Equation (1.20) so that the sum of squares error related to each measurement of the response variable y_m, is minimized. The adequacy of the approximation provided by the function (1.20) is defined by the value of the R^2-adjusted. The R^2-adjusted is a modified version of the R^2 that has to be adjusted according to the number of independent factors in the model. The number of independent factor determined the number of regressors *M*. The is R^2-adjusted calculated in this way:

$${R^2}_{adj} = 1 - \left(\frac{M-1}{M-d}\right)\frac{SSE}{(SSE + SSR)} \qquad (1.21)$$

SSE is the sum of squared error and SSR is the sum of squared regression calculated respectively as:

$$SSE = \sum_{m=0}^{M} (\hat{y}_m - \bar{y})^2 \qquad (1.22)$$

$$SSR = \sum_{m=0}^{M} (y_m - \hat{y}_m)^2 \qquad (1.23)$$

where y_m is the experimental value that has to be predicted using $\hat{y}_m$ and $\bar{y}$ is the mean value of the response variable. The value of ${R^2}_{adj}$ varies between 0 and 1, with larger numbers indicating better fit (Chatterjee and Hadi, 1989).

1.1.4.2 Dynamic response surface models (DRSM)

In processes with time varying nature, separate RSM models may be developed at different time instants during each batch to give a general view of the process.

To this end, a new procedure has been used: the dynamic response surface model (DRSM), described in Klebanov and Georgakis (2016).

DRSM is a data-driven model structure that captures the relationship between process input (time- invariant or time-variant) and time-resolved output. Its aim is to define the dependence of the process output profile to the input variables.

Consider for instance one dynamic factor x, whose variation is described by two dynamic subfactors: x_1 and x_2. Suppose that experimental plan was build, according to the DoDE rules and the experiments run, to collect the respective data. By measuring the process output *K* times along the process, *KM* values of the response variable can

be obtained, where *M* is the number of unique experiments performed. The choice in the number of *K* must be a compromise among carrying out few measurements and obtaining a larger data set to build a most informative response surface model. The *KM* is the data set containing all the values of the process output at different time instant t_k and in the case of x_m input profile.
A general relationship can be defined:

$$y_{DRSM}(t) = \{\beta_0(\tau)\} + \{\beta_1(\tau)\}x_1 + \{\beta_2(\tau)\}x_2 + \{\beta_{12}(\tau)\}x_1x_2 + \{\beta_{11}(\tau)\}{x_1}^2 + \{\beta_{22}(\tau)\}{x_2}^2 . \quad (1.24)$$

In order to describe the time profiles of the process output $y_{exp}(t)$, the time varying functions of the coefficients $\{\beta_0(\tau), \beta_1(\tau), \beta_2(\tau) \dots\}$ have to be defined. A finite approximation of $\{\beta_q(\tau)\}$ is achieved by a polynomial expansion using shifted Legendre polynomials, where the subscript [q] defines the index of variation of the parameters according to the used model (linear, linear plus interaction or quadratic). The $\{\beta_q(\tau)\}$ can be rewritten as a linear combination of the shifted Legendre polynomials $\{P_0(t), P_1(t), P_2(t), \dots\}$. For example, the following DRSM is obtained using the first three Legendre polynomials:

$$\begin{aligned} y_{DRSM}(\tau) = {} & \{g_{0,1}P_0(\tau) + g_{0,2}P_1(\tau) + g_{0,3}P_2(\tau)\} \\ & + \{g_{1,1}P_0(\tau) + g_{1,2}P_1(\tau) + g_{1,3}P_2(\tau)\}x_1 \\ & + \{g_{2,1}P_0(\tau) + g_{2,2}P_1(\tau) + g_{2,3}P_2(\tau)\}x_2 \\ & + \{g_{12,1}P_0(\tau) + g_{12,2}P_1(\tau) + g_{12,3}P_2(\tau)\}x_1x_2 \\ & + \{g_{11,1}P_0(\tau) + g_{11,2}P_1(\tau) + g_{11,3}P_2(\tau)\}{x_1}^2 \\ & + \{g_{22,1}P_0(\tau) + g_{22,2}P_1(\tau) + g_{22,3}P_2(\tau)\}{x_2}^2 . \end{aligned} \quad (1.25)$$

If *R* is the number of polynomials used, the parameters $\{\beta_q(\tau)\}$ will be approximated with polynomials up to degree *R-1*. The number of Legendre polynomials *R* determines also the number of parameters $g_{q,r}$, where the subscript [r] indicates that the parameter is related to the *(r-1)* degree polynomial.
The parameters $g_{q,r}$ are estimated using a stepwise regression method that retains only the statistically significant terms at the 95% significance level (Draper and Smith, 1998). This method is performed using MATLAB function "*stepwisefit*". The number of polynomials used is limited by the number of measurements *K* that can be collected from each experiment. For a given number of available measurements *K*, a model with several values of *R* up to *K-1* is developed. In order to identify the adequate model, characterized by a particular value of *R* and *K*, the sum of squares of the residuals is

defined as done in Klebanov and Georgakis (2016). The sum of squares is calculated using the following formulation:

$$\widehat{SS}_{un}(R,K) = \frac{\sum_{m=1}^{M}\left[\sum_{k=1}^{K}\left\{y_{DRSM,m}(\tau_k;R,K) - y_{exp,m}(\tau_k)\right\}^2\right]}{\sum_{m=1}^{M}\left[\sum_{k=1}^{K}\left\{y_{exp,m}(\tau_k)\right\}^2\right]} \tag{1.26}$$

where *M* is the number of the experiments performed and from which *MK* output values are collected. The $y_{DRSM,m}(\tau_k;R,K)$ is the value of the response variable in τ_k predicted by the DRSM. The $y_{exp,m}(\tau_k)$ is the value of the response variable in τ_k collected during the experiments and that has to be approximated. The best value of *R* is the one which provides the smallest value in the equation (1.26). However, if two values of *R* provide only slightly different values for $\widehat{SS}_{un}(R,K)$ the smaller *R* should be selected. Since the experiments are considered to be performed *in silico*, the value of $y_{exp,m}$ is available for the entire time interval τ. It is possible to verify the adequacy of the chosen model through the calculation of $\widehat{SS}_{un}(R,K)$, determining the following sum squares residuals:

$$SS_{un}(R,K) = \frac{\sum_{m=1}^{M}\left[\int_{\tau=0}^{\tau=1}\left\{y_{RSM,m}(\tau;R,k) - y_{exp,m}(\tau)\right\}^2\right]}{\sum_{m=1}^{M}\left[\int_{\tau=0}^{\tau=1}\left\{y_{exp,m}(\tau)\right\}^2\right]} \ . \tag{1.27}$$

The calculation of $SS_{un}(R,K)$ allows to verify if the DRSM, calculated using only K point for each profile, is able to approximate the complete profile. This calculation of the sum of squares is necessary but not sufficient to certify that the DRSM adequately represents the nonrandom variability of the data. To verify this, it is possible to examine whether the above-defined $\widehat{SS}_{un}$, is of the same order of magnitude of the SS_{err}, that describes the natural variability of the process. This variability can be calculated repeating the same experiment in such a way as to define the uncertainty on the measurements. The equation used is the following:

$$SS_{err} = \frac{\sum_{m=1}^{N_{CP}}\left[\sum_{k=1}^{K}\left\{y_{0,m}(\tau_k) - \overline{y_{0,m}}(\tau_k)\right\}^2\right]}{\sum_{i=1}^{N_{CP}}\left[\sum_{k=1}^{K}\left\{\overline{y_{0,i}}(\tau_k)\right\}^2\right]} \tag{1.28}$$

where $y_{0,m}(\tau_k)$ is the value of the response variable in τ_k calculated in the center point of the design space. The $\overline{y_{0,m}}(\tau_k)$ is the mean value of the response variable in τ_k calculated performing N_{CP}(= 10) repeated runs in the center point of the experimental region. The value of SS_{err} has to be compare to the $\widehat{SS}_{un}(R,K)$ calculating the following statistic ratio:

$$F_0 = \frac{\widehat{SS}_{un}(R,K)/n_1}{SS_{err}/n_2} = \frac{\widehat{SS}_{un}(R,K)/(MK-D)}{SS_{err}/(K(N_{CP}-1))} \quad . \tag{1.29}$$

The parameters n_1 and n_2 define the degrees of freedom and are calculated as shown in Equation (1.29). *MK* is the total amount of experimental data utilized and *D* is the number of significant parameters that are used in the definition of DRSM model via stepwise regression. According to the F-test, the *p*-value is calculated and the significance determined. If the *p*-value is lower than 0.95, the model can meaningfully represent the nonrandom variability of data.

1.1.5 Definition of the operating optimum

The empirical model defined in the RSM/DRSM allows defining the operating optimum of the process.

1.1.5.1 Optimization method using RSM

In the case of a RSM, the operating optimum corresponds to the values of the factors, that maximize / minimize the response variable. The calculation of the maximum point, for example, is performed minimizing the inverse of the equation that describes the RSM. Consider the following RSM:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + b_{11}{x_1}^2 + b_{22}{x_2}^2 \quad . \tag{1.30}$$

The optimization problem can be exemplified below:

$$\min_{(x_1,x_2)}(-\hat{y}) = \min_{x_1,x_2}(-(b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + b_{11}{x_1}^2 + b_{22}{x_2}^2)) \, . \tag{1.31}$$

The minimum value of the function $\hat{y}$ and the related value of $x_1$and x_2 are calculated using the MATLAB function "*fmincon*". Using the RSM, it is possible to optimize the response variable in a certain time instant.

1.1.5.2 Optimization method using DRSM

The optimization of a DRSM, instead, allows the definition of the optimum profile of the response variable along the process. It is possible to look for the values of the input factors $x_1$and x_2 that maximize the integral of the response variable profile. This calculation is performed maximizing the integral of the DRSM in the time domain as in Equation (1.32).

$$\max_{x_1,x_2} \int_{\tau=0}^{1} y_{DRSM}(\tau)d\tau \quad . \tag{1.32}$$

Equation (1.32) can be rewritten as a problem of minimization of the inverse of the integrated function as below:

$$\min_{x_1,x_2}\left(-\int_0^1 \beta_0(\tau)d\tau - x_1\int_0^1 \beta_1(\tau)d\tau - x_2\int_0^1 \beta_2(\tau)d\tau - x_1x_2\int_0^1 \beta_{12}(\tau)d\tau \right. \tag{1.33}$$
$$\left. - {x_1}^2\int_0^1 \beta_{11}(\tau)d\tau - {x_1}^2\int_0^1 \beta_{22}(\tau)d\tau \right) .$$

The values of $x_1$and x_2 that minimize the function in Equation (1.33) are calculated using the MATLAB function "*fmincon*".

Another criterion of optimality requests to define the profile of the response variable that is higher than all the others during the whole process.

First, it is necessary to verify if there is a combination of the factors that corresponds to this requirement. To do this, the values of the factors that maximize the response variable in different time instants of the process are calculated. This is done by choosing a few time instants(τ_h), in which the values of the factors that maximize the response variable in those precise moments are defined. The equation used is the following:

$$\max_{x_{1h},x_{2h}} y_{DRSM}(\tau_h) = \min_{x_{1h},x_{2h}} (-y_{DRSM}(\tau_h)) \tag{1.34}$$

where x_{1h} and x_{2h} are the values of the factors $x_1$and x_2 that maximize the value of the DRSM calculated in τ_h. If the values x_{1h} and x_{2h} change according to the time instant, then there is no unique combination of the factors that meets the optimization criterion. Instead, it is possible to define the optimum time-varying profile of the factors. In fact, in order to obtain a response variable profile that is always higher than all the others, it is necessary to change the value of the input factors during the batch.

In the case of a dynamic factor, the change of the values of its dynamic subfactors corresponds to modify its type of trend along the process.

Once define the optimal profile of the process output, either using the method in (1.33) or using the one in (1.34), it will be necessary to compare the profile predicted through the DRSM to the real one. The difference is defined calculating the root mean square error ($RMSE$) using Equation (1.35).

$$RMSE = \sqrt{\frac{\sum_{k=1}^{K'}(y_{DRSM}(t_k) - y_{\exp}(t_k))^2}{K'}} \tag{1.35}$$

where K' is the number of time instants in which the two profiles are compared.
This calculation will be a further confirmation of the accuracy of the DRSM calculated in the previous step and the results will be compared to the value of the measurement error, that the case of *in silico* experiments is known and added to the data collected as will be described in Chapter 2.

Chapter 2

Case Studies

In this chapter, the description of the case studies is reported. The same case studies are shown in Georgakis (2013) and Klebanov and Georgakis (2016).
The case studies considered are:

- a batch reactor with reversible reaction;
- a semi-batch reactor with three reactions;
- a penicillin fermentation process.

An exhaustive description of the model's equations and parameters is reported.

2.1 Batch reactor with reversible reaction

The first case study is the one of a batch reactor in which a reversible reaction between reactant A and product B takes place. The reaction has the following characteristics:

$$A \leftrightarrow B \tag{2.1}$$

$$r = k_A[A] - k_B[B] \tag{2.2}$$

where:

$$k_A = k_{A0} exp^{\left(-\frac{E_A}{R_g T}\right)} \tag{2.3}$$

$$k_B = k_{B0} exp^{\left(-\frac{E_B}{R_g T}\right)} . \tag{2.4}$$

The batch time is fixed at 2 hr. The values of the parameters are reported in Table 2.1.

Table 2.1 *Values of the kinetic parameters related to Equations (2.3), (2.4)*

Parameter	Value
k_{A0}	1.32×10^7 h^{-1}
k_{B0}	5.25×10^{13} h^{-1}
E_A	41840 J/mol
E_B	83680 J/mol
R	8.314 J/mol/h

Since the experiments are performed *in silico*, the simulated results (y) are perturbed adding an error, according to the following equation:

$$y_{exp} = y(1 + \sigma N(0,1)) \tag{2.5}$$

where $N(0,1)$ is a normally distributed number with zero mean and standard deviation equal to 1 and σ defines the percentage of measurement error added to the simulated experiments. In the case of DoE and DoDE, the value of σ in Equation (2.5) is equal to 0.005, corresponding to a measurement error of 1%. The error added is increased to 4% in the case of DRSM, in order to consider a higher variability of data. The results regarding this case study can be found in the Appendix.

2.2 Semi-batch reactor

In the second case study, a semi-batch reactor, in which a reaction network of three reactions takes place, is considered. The process in maintained at a constant temperature and volume. The reactions take place with the following characteristics:

$$A + B \rightarrow C \qquad r_1 = k_1[A][B] \tag{2.6}$$

$$2B \rightarrow D \qquad r_2 = k_2[B]^2 \tag{2.7}$$

$$C \rightarrow E \qquad r_3 = k_3[C] \tag{2.8}$$

The kinetic parameters related to the reactions are reported in the table below.

Table 2.2 *Values of the kinetic parameters related to Equations (2.6), (2.7), (2.8)*

Parameter	Value
k_1	2 L/gmol/h
k_2	1 L/gmol/h
k_3	1 h^{-1}

B is the coreactant and is fed in semi-batch mode in order to minimize the amount of by-product produced. The model is made by the following material balances:

$$\frac{d[A]}{dt} = -k_1[A][B] \tag{2.9}$$

$$\frac{d[B]}{dt} = \frac{u_B(t)}{V} - k_1[A][B] - 2k_2[B]^2 \tag{2.10}$$

$$\frac{d[C]}{dt} = k_1[A][B] - k_3[C] \tag{2.11}$$

$$\frac{d[D]}{dt} = k_2[B]^2 \tag{2.12}$$

$$\frac{d[E]}{dt} = k_3[C] \tag{2.13}$$

The reactor volume is fixed to 10 L, while the batch time is fixed to 1 hour. The initial concentration for A is fixed to 1 gmol/L, while the initial concentration of the other species is equal to 0 gmol/L. Since it is desirable to maximize the production of product C, the decision variable is the time dependence of the feeding flowrate of B. Since the experiments are performed *in silico*, a measurement error is added to the experimental data y according to Equation (2.5). The value of σ is equal to 0.005, corresponding to an error of 1%, in the case of DoE and DoDE, while it is equal to 0.02 in the case of DRSM.

2.3 Penicillin fermentation process

In the last case study, the penicillin fermentation process is simulated. This consists in two phases: the first one is a phase of growth and the second one is a phase of production.

During the first phase of the process, the biomass, which consists in an aggregate of bacteria, grows by consuming the substrate. The latter is a cultivation medium with a high glucose content, to allow a rapid growth of biomass. The concentration of the biomass has not to be excessive in order to avoid the consumption of oxygen. The second phase of the process is finalized to the production of penicillin, with a limitation in the growth of biomass. The fed-batch reactor is considered the optimal choice to avoid a rapid growth of biomass, because it allows a progressive substrate feeding. The model used for the description of the process is the one obtained by Bajpai and Reuss (1980). It consists of balances for the volume (V), biomass (b), substrate (s) penicillin (p) as follows:

$$\frac{dV}{dt} = u_s \tag{2.14}$$

$$\frac{db}{dt} = \mu b - \frac{bu_s}{V} \tag{2.15}$$

$$\frac{ds}{dt} = -\frac{\mu b}{Y_{XS}} - \frac{\rho b}{Y_{PS}} - \frac{m_s sb}{km + s} + \frac{(sf - s)u_s}{V} \tag{2.16}$$

$$\frac{dp}{dt} = \rho b - k_d p - \frac{pu_s}{V} \tag{2.17}$$

with:

$$\mu = \mu_{max}\left(\frac{s}{k_b + s}\right), \tag{2.18}$$

$$\rho = \rho_{max}\left(\frac{s}{k_p + s + \left(s^2/k_m\right)}\right). \tag{2.19}$$

The model parameters of Riascos and Pinto (2004) are reported in Table (2.3).

Table 2.3 *Values of parameters related to Equations (2.14)- (2.19)*

Parameter	Definition	Value
μ_{max}	maximum specific biomass growth rate	0.1 h^{-1}
ρ_{max}	maximum specific production rate	5.5 $\times 10^{-3}$ gr_p/gr_b h
k_B	saturation parameter for biomass growth	6 $\times 10^{-3}$ gr_b/gr_s
k_p	saturation parameter for penicillin production	0.1$\times 10^{-3}$ gr_s/L
k_{in}	inhibition parameter for penicillin production	0.1 gr_s/L
k_d	penicillin degradation	0.01 h^{-1}
k_m	saturation parameter for maintenance consumption	0.1$\times 10^{-3}$ gr_s/L
m_s	maintenance consumption rate	2.9 $\times 10^{-3}$ gr_s/gr_b h
Y_{XS}	yield factor, substrate (S) to biomass	0.47$\times 10^{-3}$ gr_b/gr_s
Y_{PS}	yield factor, substrate (S) to product	1.2 gr_p/gr_s
s_f	feed concentration of substrate (S)	500 gr_s/L

Other information useful for the implementation of the model are the value of the initial substrate concentration that is fixed equal to 500 gr_s/L. Since the experiments are performed *in silico*, the simulated results (y) are perturbed adding a measurement error, according to Equation (2.5). In the case of DoE and DoDE, the value of σ in that equation is equal to 0.005, corresponding to a measurement error of 1%. The error added is increased to 4% in the case of DRSM, in order to consider a higher variability of data

.

Chapter 3

Optimization of a batch reactor

In this chapter, the optimization procedure defined in Chapter 1 is applied to case study of a batch reactor that is described in Section 2.1. By applying DoE and DoDE, it is possible to define the profile of an input variable that optimizes the value of the output. The construction of a DRSM, instead, allows studying the dynamic behavior of the output variable along the process.

3.1 Optimization of a batch reactor using DoE methodology

The objective of the following procedure is to define an empirical model that describes the relationship between the input variables and an output factor, that, in this case, is the reactant A concentration. The model that has to be defined considers both linear and quadratic effects of the inputs on the output.

3.1.1 Selection of input factors

In this case, the input variable is the temperature of the reactor and it is considered a static factor. Temperature varies in the range between 15 °C and 50°C and is indicated using W_1 or, in the codified version, w_1.

3.1.2 DoE and experimentation

In the case of one independent factor w_1, three parameters must be estimated to define a quadratic model. DoE must be composed of at least three experiments; respectively at the highest, the intermediate and the lowest value of input variable range, the reactor temperature w_1. Two replicated experiments in the central point of the design space are added in order to represent the variability of the measurements. The results of the five experiments in terms of reactant A final conversion, are reported in the Table 3.1.

Table 3.1 *Reactant A conversion at the end of the batch regarding the experiments defined in DoE methodology*

	Run1	Run2	Run3	Run4	Run5
Temperature °C	15 °C	32.5 °C	32.5°C	32.5°C	50°C
Coded variable w_1	-1	0	0	0	1
Simulated conversion	48.18%	70.91%	70.91%	70.91%	59.27%
Measured conversion	48.17%	70.86%	71.14%	71.30%	59.60%

3.1.3 RSM and optimization

The relationship between temperature and conversion of reactant A can be described defining an empirical model, or a response surface model (RSM), with the following structure:

$$\hat{y} = b_0 + b_1 w_1 + b_2 w_1^2 \tag{3.1}$$

where $\hat{y}$ is the conversion of the reactant A at the end-point process and w_1 is the codified temperature. Using the experimental data, reported in the Table 3.1, the coefficients of Equation (3.1) can be calculated using a least square method as described in Sub-subsection 1.1.4.1. These are reported, along with their 95% confidence interval, in Table 3.2.

Table 3.2 *Coefficients of the response surface model in Equation (3.1) that describes the quadratic influence of the temperature on the reactant A final conversion.*

Coefficients		
b_0	b_1	b_2
71.1 % $\pm$ 0.56 %	5.71 % $\pm$ 0.69 %	-17.22 % $\pm$ 0.89 %

The optimization procedure is performed in order to define the reactor temperature that provide the higher conversion of the reactant at the end of the batch. Using the RSM in (3.1), it was possible to define the maximum value of the conversion of reactant A. According to the regression model the maximum conversion should be at $w_{1,OPT}$= 0.1658, corresponding to a constant temperature of 35.40 °C and at which the predicted conversion is 71.57%. The confidence intervals of the parameters define a range of variation of the predicted maximum, which varies in the range [70.88%, 72.28%]. It is necessary to verify that the reactant A process conversion related to the coded variable $w_{1,OPT}$, can be described by the RSM calculated. The real conversion, according to the suggested temperature $w_{1,OPT}$ is equal to 71.12% and this result allows confirming the adequacy of the empirical model defined.

The optimum conversion calculated can be compared with those in Georgakis (2013); the difference between the results is not significant since it is lower than 1% that is the value of the measurements error added to the simulated results according to (2.5).
Using DoE, the dynamic behavior of the temperature is not considered. To do this, DoDE methodology can be applied.

3.2 Optimization of a batch reactor using DoDE methodology with linear profiles

As in the previous case, the objective is to define a full quadratic RSM that describes the effects of the input variable, namely, the reactor temperature, on the final conversion of reactant A. The dynamic behavior of reactor temperature is considered.

3.2.1 Selection of input factors

In this case, only the constant and linear time dependence of reactor temperature is considered. According to the theory, the profiles of temperature are described using two dynamic subfactors x_1 and x_2. In the following equations, the general formulation of the temperature profile $X_m(t)$ and its codified version $x_m(\tau)$ are reported:

$$x_m(\tau) = x_{1,m}P_0(\tau) + x_{2,m}P_1(\tau) = x_{1,m} + x_{2,m}(-1 + 2\tau)\,, \tag{3.2}$$

$$X_m(t) = 32.5 + 17.5\left(x_{1,m} - x_{2,m} + 2x_{2,m}t\right), \tag{3.3}$$

where τ is the dimensionless time $\in [0, 1]$ and $x_{1,m}$ and $x_{2,m}$ are the values assumed by the dynamic subfactors in the *m*-th profile.

3.2.2 DoDE and experimentation

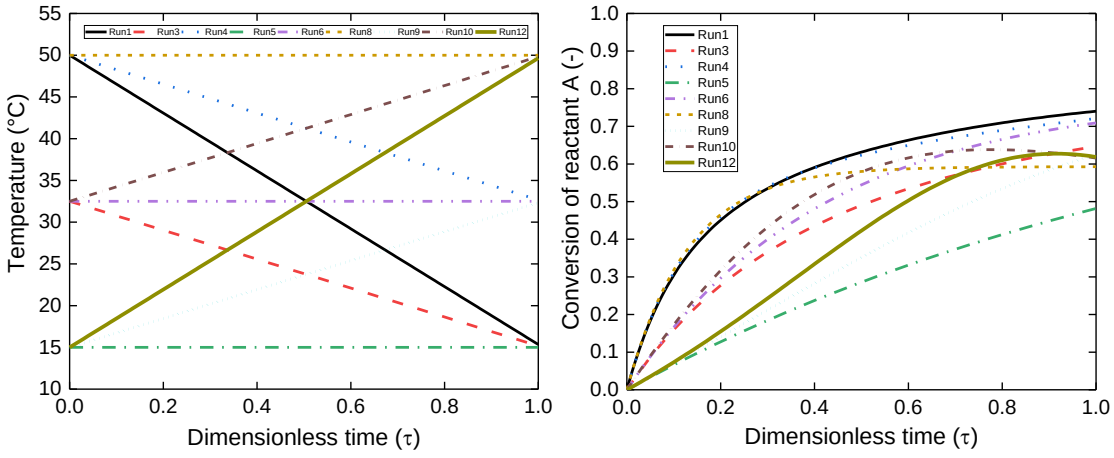
The number *M* of profiles and the design structure of the experiments are related to the type of RSM that must be obtained. The empirical model, in this case, has to describe the linear, interactions and quadratic effects of two independent factors x_1 and x_2 on the response variable, which is the end-point conversion of reactant A. The profiles of the reactor temperature, described by the combinations of dynamic subfactors x_1 and x_2, are defined according to a central composite design (CCD) with $\alpha = 2$. Since the combinations of subfactors have to satisfy the constraint (1.15), it is not possible to apply a full factorial design (FFD).

The values of the two dynamic subfactors, $x_{1,m}$ and $x_{2,m}$, that define the $\boldsymbol{x}(\tau)$ profiles, are reported in the second and third columns of Table 3.3. In the same table, the experimental data, in terms of the final conversion of the reactant A, are reported.

Table 3.3 *Temperature profiles defined according to a central composite design applied to the subfactors x_1 and x_2 and the related values of the reactant A final conversion.*

Run	Coded factors		T profile	Simulated conversion	Measured conversion
	x_1	x_2			
1	0	-1	50 °C > 15°C	74.01%	73.41%
2	0	-1	50 °C > 15°C	74.01%	74.07%
3	-0.5	-0.5	32.5 °C > 15 °C	64.66%	64.78%
4	0.5	-0.5	50 °C > 32.5°C	72.06%	71.98%
5	-1	0	15°C	48.18%	47.91%
6	0	0	32.5°C	70.90%	71.61%
7	1	0	50°C	59.26%	58.56%
8	1	0	50°C	59.26%	59.11%
9	-0.5	0.5	15 °C > 32.5°C	62.51%	62.10%
10	0.5	0.5	32.5°C > 50°C	61.56%	61.36%
11	0	1	15°C > 50 °C	61.92%	62.01%
12	0	1	15°C > 50 °C	61.92%	61.96%

The twelve temperature profiles and the related profiles of the conversion of reactant A are reported in the Figure (3.1):



a) *b)*

Figure 3.1 *Profiles of: a) temperature and b) the conversion of reactant A obtained by performing a CCD in the case of DoDE with linear profiles.*

3.2.3 RSM and optimization

Using the information in Table 3.3, it is possible to estimate the coefficients of the RSM in Equation (3.4) that describes the quadratic dependence of the reactant A final conversion ($\hat{y}$) from the two dynamic subfactors x_1 and x_2.

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 + b_{11} {x_1}^2 + b_{22} {x_2}^2 \quad (3.4)$$

The coefficients, calculated using a least squares algorithm for the estimation of the coefficients in the multiple linear regression empirical model, are reported in Table 3.4.

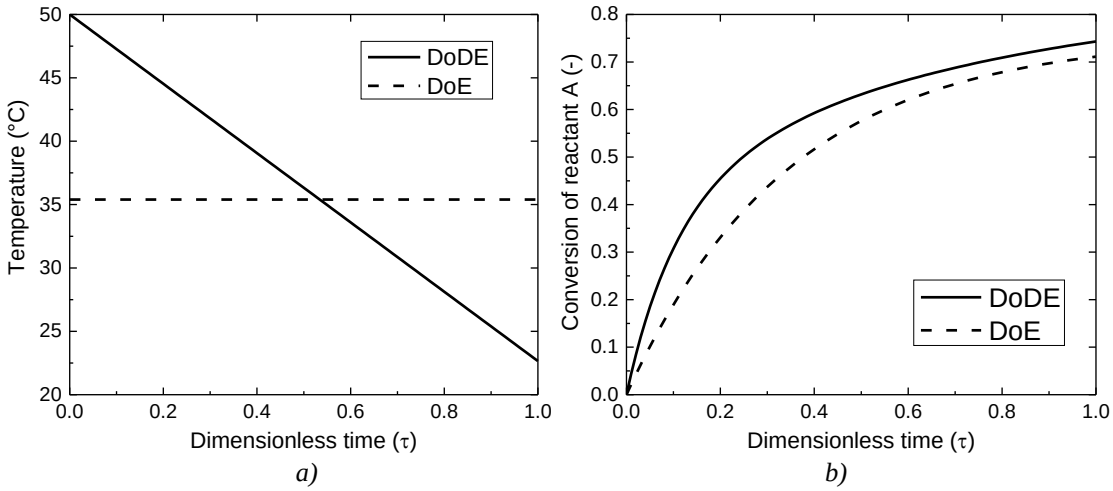
Table 3.4 *Coefficients of response surface model in Equation (3.4) that defines the full-quadratic relationship between the dynamic subfactors and the reactant A final conversion.*

Coefficients					
b_0	b_1	b_2	b_{12}	b_{11}	b_{22}
70.58 $\pm$ 1.9	4.9 $\pm$ 1.36	-6.3 $\pm$1.16	-7.94 $\pm$ 5.2	-17.19 $\pm$ 2.66	-2.85 $\pm$2.46

The R^2-adj for the above regression, calculated using Equation (1.21), is equal to 0.981. DoDE is applied in order to define the linear temperature profile that provides the maximum conversion of reactant A at the end of the process.

The profile, that provides the maximum conversion, is the one described by the values of subfactors $x_{1,OPT} = 0.2187$ and $x_{2,OPT}$ = -0.7813. These correspond to the linear temperature profile in Figure 3.2. The corresponding predicted conversion is 75.37 % while the conversion obtained by carrying out the process with the suggested optimal temperature profile is 74.32%. The value of the process conversion is inside the range [73.4%, 77.535%] that is calculated considering the coefficients of Table (3.4) at their extreme values of the confidence interval. The results obtained are comparable to those in Georgakis (2013). The predicted conversion calculated differs by the 1.2% while the process conversion is the same.

It is possible to compare also the optimal results obtained by applying the two methodologies: DoE and DoDE.



a) *b)*

Figure 3.2 *Comparison among the optimum calculated using DoE and the optimum obtained by DoDE: a) reactor temperature; b) conversion of reactant A.*

By comparing the result of DoDE to the one obtained performing DoE, it is clear that the first allows the definition of an optimum profile that provides a higher conversion.

3.3 Optimization of a batch reactor using DoDE methodology with non-linear profiles

The aim it to define a RSM that describes the relationship between reactor temperature and conversion of reactant A. In the previous section, DoDE considers linear temperature profiles. However, the optimal temperature profile might be more complex. For this reason, it was necessary to consider a design of experiments that involves nonlinear profiles.

3.3.1 Selection of input factors

The temperature profiles are characterized by at least three dynamic subfactors in order to add nonlinear time dependence. Consider for example a quadratic temperature profile.

The codified profiles $\boldsymbol{x}(\boldsymbol{\tau})$ are defined by using a linear combination of the first three shifted Legendre polynomials with three dynamic subfactors $\{x_1, x_2, x_3\}$ as in Equations (3.5) and (3.6).

$$\begin{aligned}\boldsymbol{x}(\tau) &= x_1 P_0(\tau) + x_2 P_1(\tau) + x_3 P_2(\tau) \\ &= x_1 + x_2(-1 + 2\tau) + x_3(+1 - 6\tau + 6\tau^2)\end{aligned} \tag{3.5}$$

$$\boldsymbol{X}(t) = 32.5 + 17.5(x_1 - x_2 + x_3 + 2t(x_2 - 3x_3) + 6x_3 t^2) \tag{3.6}$$

The dynamic subfactors may assume different numerical value $\{x_{1,m}, x_{2,m}, x_{3,m}\}$ according to the *m*-profile $x_m(\tau)$.

3.1.3.2 DoDE and experimentation

To obtain a full quadratic RSM, in the case of 3 independent factors, as x_1, x_2, and x_3 , it is necessary to perform at least 10 experiments, according to the number of parameters that are required and that represent both the linear, the interaction and the quadratic effects of the inputs on the output. Different combinations of the three dynamic subfactors characterize the experiments. A d-optimal design is used to define the combinations of subfactors, according to the constraint (1.10). This type of design aims to minimize the number of experiments compared to those required in the case of a FFD that, in this case, would have requested at least of 18 experiments. In the following table, the values of the conversion of reactant A for the 16 simulated profiles are reported.

Table 3.5 *Temperature profiles characterized by the values of the dynamic subfactors x_1, x_2 and x_3 defined according a d-optimal design and the related values of conversion of reactant A*

Run	Coded variables			Simulated conversion	Measured conversion
	x_1	x_2	x_3		
1	0	0	-1	70.63%	71.06%
2	0	0	-1	70.63%	70.92%
3	0	-0.5	-0.5	72.35%	72.73%
4	-0.67	0	-0.33	58.76%	58.54%
5	0	-1	0	74.02%	73.67%
6	0	-1	0	74.02%	73.55%
7	0.5	-0.5	0	72.07%	72.24%
8	0	0	0	70.91%	71.90%
9	1	0	0	59.27%	59.48%
10	0.5	0.5	0	61.56%	61.32%
11	0	1	0	61.87%	62.12%
12	0	1	0	61.87%	61.52%
13	-0.67	0	0.33	58.66%	58.24%
14	0.5	0	0.5	64.87%	65.10%
15	0	0.5	0.5	64.03%	63.78%
16	0	0	1	66.65%	71.06%

3.3.3 RSM and optimization

The desired RSM has the following structure:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{11}{x_1}^2 + b_{22}{x_2}^2 + b_{33}{x_3}^2. \quad (3.7)$$

The experiments in Table 3.5 allows defining the coefficients of Equation (3.7) along with their 95% confidence interval. These, calculated as described in Sub-subsection 1.1.4.1, are reported in Table 3.6.

Table 3.6 *Coefficients of response surface model in Equation (3.7) that defines the relationship between three dynamic subfactors and the reactant A final conversion.*

Coefficients				
b_0	b_1	b_2	b_3	b_{12}
70.14 %± 2.27 %	5.37 %± 2 %	-5.96 % ± 1.39 %	-2.49 % ± 1.53 %	-9.93 % ± 8.71 %
b_{13}	b_{23}	b_{11}	b_{22}	b_{33}
-4.73 % ± 7.62 %	-3.49 % ± 9.57 %	-17.06 %±4.25 %	-2.56 % ± 2.79 %	-1.46 % ± 2.95 %

The R^2-adj for the above regression is 0.9831. The profile that allows obtaining the maximum conversion is the one described by the values of subfactors $x_{1,OPT}$, $x_{2,OPT}$ and $x_{3,OPT}$, that are respectively 0.2451, -0.7549 and 0. The maximum predicted conversion is 75.32 %, in a range of [73.43%, 77.5%].

The conversion obtained by carrying out the process with the suggested optimal conditions is 74.25%. These results demonstrate that the addition of the quadratic time dependence of temperature profiles does not provide any improvement of the conversion. For this reason, it is possible to assert that the optimal temperature profile is the one calculated by DoDE with linear profiles.
As in the previous cases, the results of the optimization can be compared to those in Georgakis (2013). From the comparison, it is possible to assert that the process conversion does not differ in a significant way.

3.4 Definition of a dynamic response surface model

In the previous subsections, it was shown that is possible to define the input profile that provides the optimum value of the process output at end of the batch process. Nevertheless, in other application (for example in process control), it might be interesting to define the optimum profile of the output variable. Accordingly, a dynamic response surface model (DRSM) can be built following the procedure described in Sub-subsection 1.1.4.2. The definition of DRSM is based on the data, in terms of profiles of reactant A conversion, obtained performing the DoDE simulation. Since the previous analysis has demonstrated that it is possible to consider only the constant and linear time dependence of temperature, the profiles considered are those obtained by the 9 unique experiments in Table 3.3 and reported in Figure 3.1. From each profile, *K* value of the process output are collected in equidistant time intervals. Different DRSM can be defined with different combinations of *K* and *R*. The latter is the number of Legendre polynomials of which the DRSM. These DRSM can be characterized by a value of the sum of squares error $\widehat{SS}_{un}(R,K)$. The value of $\widehat{SS}_{un}(R,K)$ defines the accuracy of the DRSM calculated using *K* points and *R* polynomials and is calculated using the (1.26). In Table 3.7, the values of the $\widehat{SS}_{un}(R,K)$ are reported.Once defined the maximum number of measurements *KM* that can be done, the appropriate DRSM is the one that provides a small value of $\widehat{SS}_{un}(R,K)$.

Table 3.7 *Values of $\widehat{SS}_{un}(R,K)$ for dynamic response surface model for a variable K and R calculated in the case of batch reactor*

R	K=Number of experimental points for each profile							
	3	**4**	**5**	**6**	**7**	**8**	**9**	**10**
2	0.0625	0.0513	0.0455	0.0379	0.0338	0.0317	0.0304	0.0278
3		0.0052	0.0061	0.0062	0.0063	0.006	0.0063	0.006
4			7.36E-04	9.68E-04	0.0011	0.0013	0.0011	0.0012
5				4.92E-04	5.72E-04	4.26E-04	3.34E-04	3.09E-04
6					3.42E-04	6.79E-04	3.86E-04	4.63E-04
7						4.58E-04	2.71E-04	4.55E-04
8							4.62E-04	4.08E-04
9								4.36E-04

In Table 3.7, it can be noted that for *K*=10 there are several values of *R* (= 5, 6, 7, 8) that produce very low of $\widehat{SS}_{un}$. These are good candidates to build an appropriate DRSM models. However, it is necessary to compare the value of $\widehat{SS}_{un}$ with the value of SS_{err}, to asses if the DRSM defined is able to represent the nonrandom variability in the data. The calculation of SS_{err} allows carrying out an F-test that is traduced in the definition of a statistical ratio F_0 (1.29). In this case, the discriminant is the *p*-value related to the F_0, if its value is <0.95 the $\widehat{SS}_{un}$ calculated is a significant indicator of the accuracy of the DRSM and the latter is adequate for the description of the time variability of the process output. The *p*-value are reported for each combination of K and R in Table 3.8.

Table 3.8 *Values of p-values for dynamic response surface model for a variable K and R calculated in the case of batch reactor*

R	**K=Number of experimental points for each profile**							
	3	**4**	**5**	**6**	**7**	**8**	**9**	**10**
2	1	1	1	1	1	1	1	1
3		1	1	1	1	1	1	1
4			0.9637	1	1	1	1	1
5				0.7625	0.6771	0.2916	0.0221	0.0091
6					0.0769	0.949	0.1351	0.645
7						0.2509	0.0054	0.5255
8							0.171	0.0462
9								0.3969

An acceptable DRSM is obtained using a fourth order polynomial (*R*=5) with *K*=10 experimental points. The DRSM has the following formulation:

$$\begin{aligned} y_{DRSM}(\tau) = {} & g_{0,1}P_0 + g_{0,2}P_1 + g_{0,3}P_2 + g_{0,4}P_3 + g_{0,5}P_4 \\ & + \left(g_{1,1}P_0 + g_{1,2}P_1 + g_{1,3}P_2 + g_{1,4}P_3 + g_{1,5}P_4\right)x_1 \\ & + \left(g_{2,1}P_0 + g_{2,2}P_1 + g_{2,3}P_2 + g_{2,4}P_3 + g_{2,5}P_4\right)x_2 \\ & + \left(g_{12,1}P_0 + g_{12,2}P_1 + g_{12,3}P_2 + g_{12,4}P_3 + g_{12,5}P_4\right)x_1x_2 \\ & + \left(g_{11,1}P_0 + g_{11,2}P_1 + g_{11,3}P_2 + g_{11,4}P_3 + g_{11,5}P_4\right){x_1}^2 \\ & + \left(g_{22,1}P_0 + g_{22,2}P_1 + g_{22,3}P_2 + g_{22,4}P_3 + g_{22,5}P_4\right){x_2}^2 \end{aligned} \tag{3.8}$$

The coefficients of the DRSM in (3.8), calculated via stepwise regression, are reported in Table 3.9.

Table 3.9 *Coefficients of dynamic response surface model described by Equation (3.8) that define the relationship between the inputs and the time-variant output*

Coefficients				
g0,1	g0,2	g0,3	g0,4	g0,5
0.482	0.3204±0.0054	-0.1197 ±0.0064	0.0279 ±0.0066	-0.0181 ±0.0055
g1,1	g1,2	g1,3	g1,4	g1,5
0.1294 ±0.0043	-0.0319 ±0.0069	-0.0789 ±0.0087	0.0550 ±0.0085	-0.0241 ±0.0095
g2,1	g2,2	g2,3	g2,4	g2,5
-0.0914 ±0.0043	0.0387 ±0.0069	0.0338 ±0.0087	-0.0665 ±0.0085	0.0263 ±0.0095
g12,1	g12,2	g12,3	g12,4	g12,5
0.0385 ±0.0148	0	-0.1164 ±0.0303	-0.0291 ±0.0272	0.0458 ±0.0329
g11,1	g11,2	g11,3	g11,4	g11,5
-0.0941 ±0.0065	-0.1144 ±0.0108	0.0176 ±0.0118	0.0319 ±0.0132	0
g22,1	g22,2	g22,3	g22,4	g22,5
0	0	0	0	0

In Table 3.9, it is possible to notice that the stepwise regression has eliminated the coefficients that are not significant by assigning to them a value equal to 0. In particular, all the parameters that describe the quadratic dependence of the dynamic subfactors x_2 have been deleted, this means that the quadratic influence of x_2, on the profile of the reactant A conversion is not significant. The accuracy of the DRSM defined can be demonstrated calculating the error SS_{un} according to Equation (1.27). In this case, the experimental profiles of the conversion of the reactant A along the process are compared those approximated using a DRSM. The value of the SS_{un} is 6.2995e-04 and its magnitude is the same of $\widehat{SS}_{un}$. In Figure 3.3 the DRSM predictions are plotted against the experimental profiles.

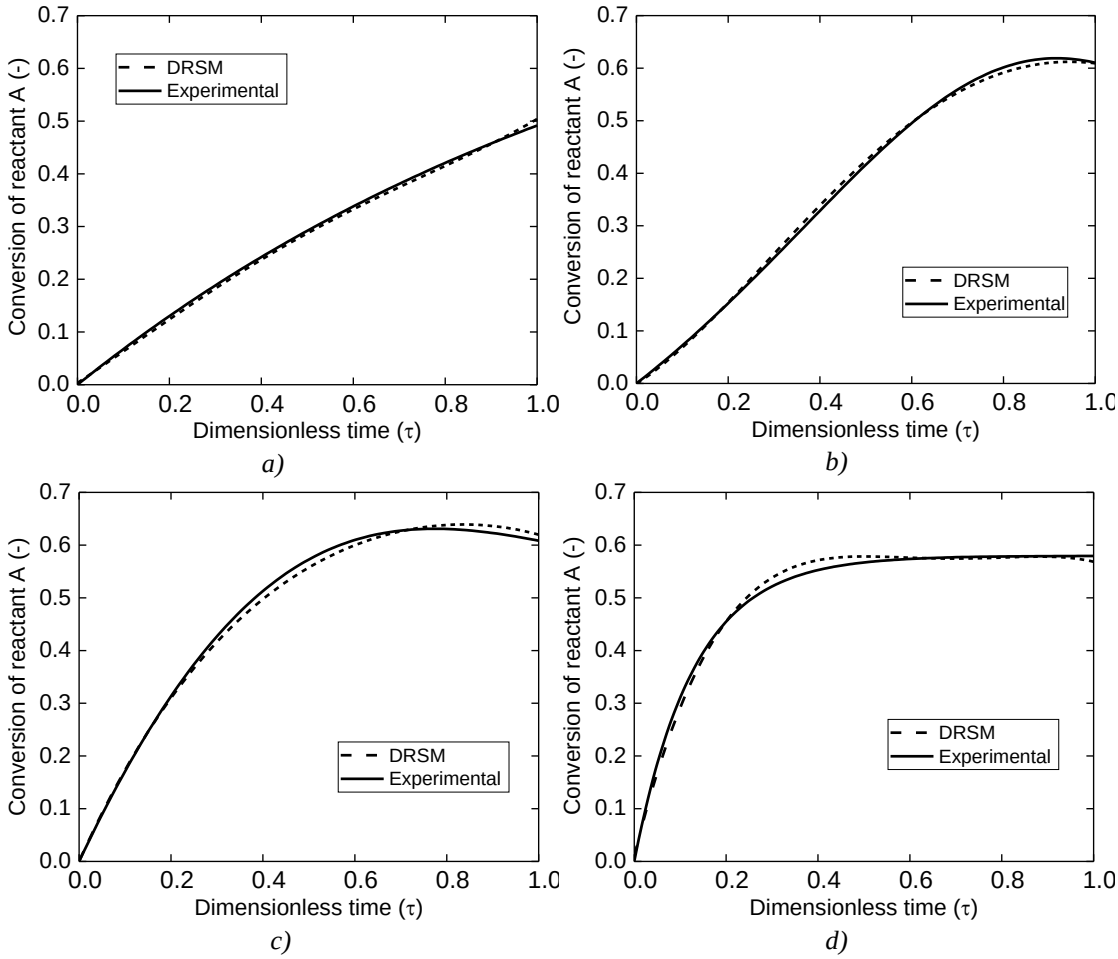


Figure 3.3 *In these figures the experimental data are compared to the DRSM predictions within the experimental design a) x_1=-1, x_2=0, b) x_1=0, x_2=1, c) x_1=0.5, x_2=0.5 and d) x_1=1, x_2=0.*

Once the accuracy of DRSM has been demonstrated, it is possible to use it as an alternative method for the optimization of the process.

3.4.1 Optimization using dynamic response surface model

DRSM, defined by Equation (3.8) and coefficients in Table 3.9, can be used for the optimization of the process conditions, as it describes both the time and the temperature dependence of reactant A conversion. Globally, DRSM defines a series of profiles that are shown in the Figure 3.4.

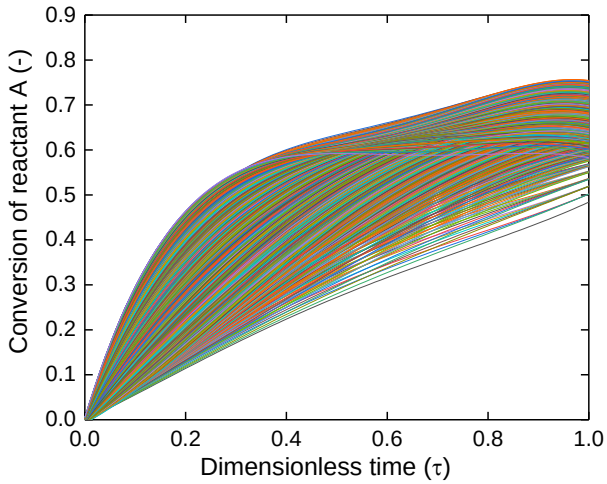
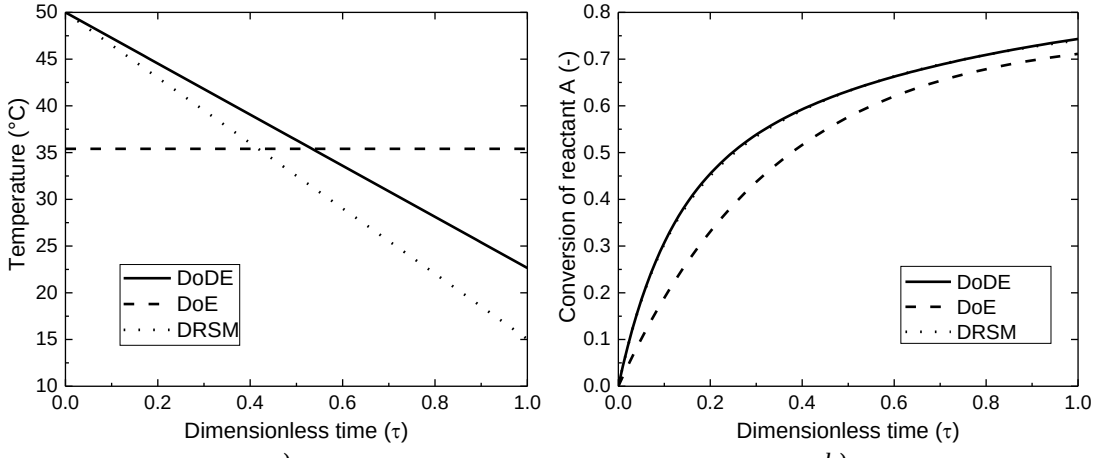


Figure 3.4 *Profiles of the reactant A conversion along time obtained using the DRSM in (3.8)*

One of the criteria that can be used for the definition of the optimum profile is to determine the value of the dynamic subfactors x_1 and x_2, and the related temperature profile, that defines the conversion profile, which maximizes the integral of the DRSM, from $\tau = 0$ to $\tau = 1$, according to Equation (1.33). The combination of dynamic subfactors that meets this condition is: $x_{1,OPT}$= 0.0013; $x_{2,OPT}$= -0.9987. The optimum profile calculated in this way, both for the profile of temperature and the profile of the conversion of the reactant A, can be compared with those obtained with the DoDE and DoE.

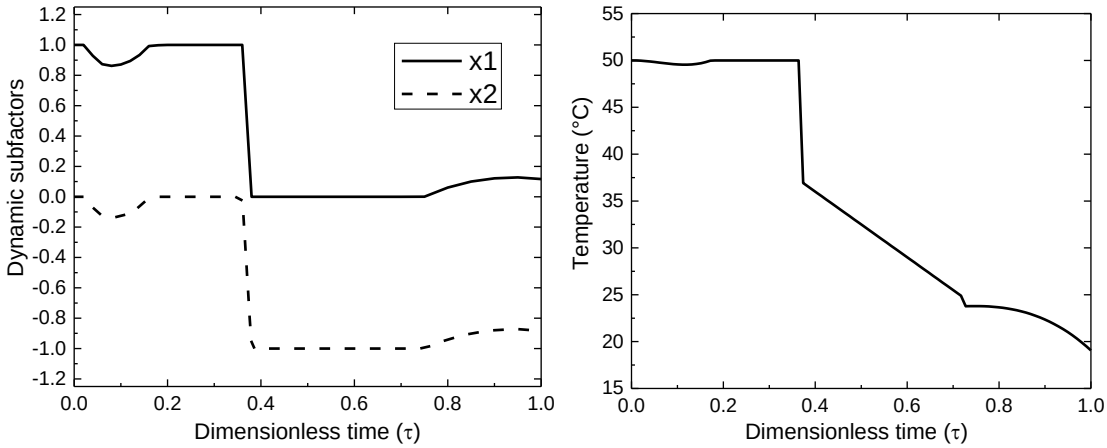


a) *b)*

Figure 3.5 *Comparison among the optimum calculated using DoE, DoDE and by maximizing the integral of the DRSM: a) reactor temperature; b) conversion of reactant A.*

Even if the temperature profiles calculated through the optimization procedure in DoDE and DRSM are different, the output profiles of the process are almost the same. Therefore, in this case the use of DRSM for the optimization does not provide a higher conversion of reactant A.

However, the profiles in Figure 3.4 suggests that there is not a combination of subfactors x_1 and x_2 that guarantees a profile of conversion that is higher than the others in each time instant. For this reason, the method (1.34) in Chapter 1 is considered to obtain the values of x_1 and x_2 that maximize the value of the $y_{DRSM}(\tau)$ in each time instant of the batch. Using DRSM, the values of x_1 and x_2 that maximize the response are calculated for the entire duration of the batch. The profiles of the dynamic subfactors and the related temperature profile are shown in Figure 3.6.



a) *b)*

Figure 3.6 *Optimal profiles of: a) the dynamic subfactors and b) temperature calculated maximizing the DRSM in a series of time instants.*

The corresponding profile of the reactant A conversion is shown in Figure 3.7. In this case, the predicted profile, as the one defined by DRSM, is compared to the simulated one.

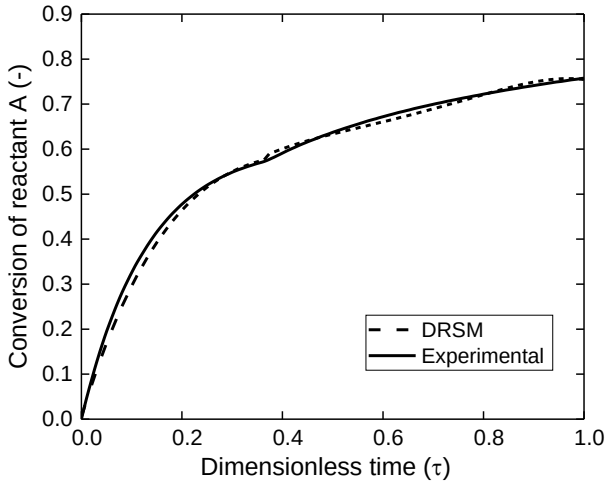


Figure 3.7 *Comparison between the approximated profiles of the concentration of reactant A according to the optimal temperature profile in Figure 3.6 and the real one.*

The difference between the profiles in Figure 3.7 is defined by the value of the *RSME* calculated according to Equation (1.35). In this case *RSME* is equal to 0.0127 that is lower than the value of σ=0.02. This means that the difference is not significant since it is lower than the measurement error. It is possible to represent the optimum simulated profile in Figure 3.4 to verify if the time-variation of the dynamic subfactors leads to define a profile of the process output that satisfy the request to be higher than all the other in the entire duration of the batch.

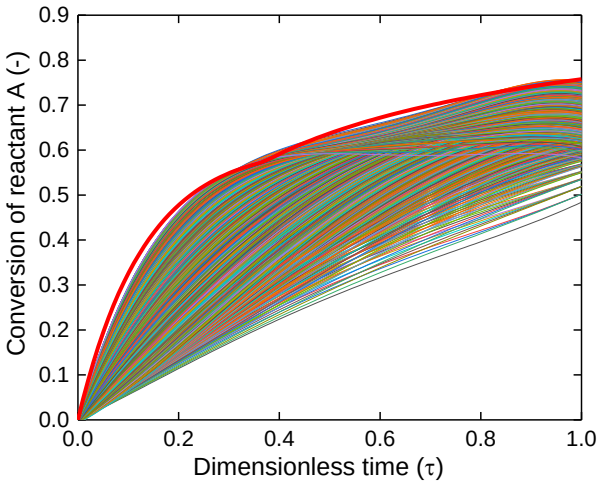


Figure 3.8 *Profiles of the conversion of reactant A calculated using DRSM and reported in Figure 3.4 are compared with the optimal real profile shown in Figure 3.7.*

The optimum profile in Figure 3.8, obtained varying the dynamic subfactors along the batch does not satisfy the optimality criterion described above. In fact, the optimal profile of the dynamic subfactors calculated using (1.34) would require an instant variation of the reactant temperature in τ= 0.4. Since this is not possible, it can be concluded that there is no profile of the dynamic subfactors variation and of the related temperature that allows providing a reactant A conversion profile that is higher than the others in the entire duration of the batch.

3.5 Conclusions regarding the optimization of a batch reactor

In the conclusions, it is possible to summarize the results obtained in order to define the operating optimum of the batch process. The following considerations are possible:

- The application of the DoDE methodology allows defining a time-varying profile of temperature that leads to determine a higher conversion of reactant A than the one obtainable performing a DoE, namely, considering constant temperature.

- The use of the DRSM as a means to optimize the process conditions allows to define a temperature profile different from the one defined by DoDE. However, this variation does not influence in a significant way neither the reactant A conversion profile nor its end-point value.

- The variation of the dynamic subfactors along the process defines a profile of the reactant A conversion that do not satisfy the optimization criterion considered

The operating optimum is the one defined by a linear profile of reactor temperature that is described by $x_{1,OPT} = 0.2187$ and $x_{2,OPT}$ = -0.7813 and provides a final conversion of reactant A equal y =74.32%.

Chapter 4

Optimization of a penicillin fermentation process

In this chapter, the optimization procedure defined in Chapter 1 is applied to case study of a penicillin fermentation process that is described in Section 2.3. By applying DoE and DoDE, it is possible to define the profile of an input variable that optimizes the value of the output. The construction of a DRSM, instead, allows studying the dynamic behavior of the output variable along the process.

4.1 General considerations

In this case study, the goal is to define the operating conditions that guarantee the maximum penicillin productivity at the end of the batch. The output variables considered are both the penicillin concentration (gr/L) and the grams of penicillin produced. The input variables that have to be optimized are the substrate income flowrate and initial biomass concentration. The latter is a static factor with a nominal value of 1.5 gmol/L that varies between 1 gmol/L and 2 gmol/L. The initial biomass concentration, indicated with $\bar{b}$, can be parametrized through this equation:

$$\bar{b} = 1.5 + 0.5w_1 \tag{4.1}$$

with:

$$-1 \leq w_1 \leq +1\,. \tag{4.2}$$

For what concern the dynamic factor, namely, the substrate income flowrate, an operating constraint on the maximum volume capacity of the reactor $V(t_b) = 10$ L is imposed. For the initial value of the reactor volume $V(0)$=7 L, the reference value of the substrate inflow $u_{0,s}(t)$ should satisfy the following constrain:

$$V(0) + \int_0^{t_b} u_{0,s}(t)dt = V(t_b)\,. \tag{4.3}$$

If the batch time: t_b is fixed to 130 hr, Equation (4.3) can be rewritten as:

$$\int_0^1 u_{0,s}(t)dt = \frac{3}{130}. \tag{4.4}$$

Considering a simple linear dependence of $u_{0,s}$ on the dimensionless time $\tau = t/t_b$ and imposing the following final condition $u_{0,s}(1) = 0$, the following equation is obtained:

$$u_{0,s}(\tau) = \frac{6}{130}(1-\tau). \tag{4.5}$$

The $\Delta u_s(\tau)$ is defined in the same way in Equation (4.4):

$$\Delta u_s(\tau) = \frac{6}{130}(1-\tau), \tag{4.6}$$

so that all possible feeding profiles between 0 and $2u_{0,s}(\tau)$ can be used. The dynamic behaviour of the substrate income flowrate can be described by the following general equation:

$$u_s(\tau) = u_{0,s}(\tau) + \Delta u_s(\tau)x(\tau), \tag{4.7}$$

with:

$$-1 \le x(\tau) \le 1. \tag{4.8}$$

The $x(\tau)$ is the codified version of the dynamic factor $u_s(\tau)$. The $u_s(\tau)$ must satisfy the following equation to fill the reactor but not overfill.

$$V(0) + \int_0^{t_b} u_s(t)dt = V(t_b). \tag{4.9}$$

According to (4.9), it is required that:

$$\int_0^1 u_s(t)dt = \int_0^1 u_{0,s}(\tau) + \Delta u_s(\tau)x(\tau) = \frac{3}{13} \tag{4.10}$$

The $x(\tau)$ is defined using a linear combination of Legendre polynomials and coefficients that are called dynamic subfactors.

The general expression is the following:

$$x(\tau) = x_1 P_0(\tau) + x_2 P_1(\tau) + x_3 P_2(\tau) \tag{4.11}$$

In the case study, a DoE should be performed by considering the dynamic factor as a static one. The variation of the substrate incoming flowrate should be defined using only one dynamic subfactor x_1. Therefore, DoE could be made by the combinations of values of x_1 that describes the substrate incoming flowrate and w_1 that defines the initial biomass concentration. However, this is not possible because, once fixed the batch time, there is only one value of x_1 that satisfies the constraint on the total volume in Equation (4.10). A DoDE must be performed.

4.2 Optimization of a penicillin fermentation process using DoDE methodology

The objective is to define an empirical model that describes the relationship between the inputs, which are the substrate income flowrate and the initial concentration of biomass and the process output, which can be represented either by the final penicillin concentration (gr/L) or by the end-point penicillin production (gr).

4.2.1 Selection of input factors

The initial biomass concentration ($\bar{b}$) is a static factor represented by w_1 as in Equation (4.2). The substrate income flowrate is a dynamic factor, whose variation is described by three dynamic subfactors x_1, x_2 and x_3. The m-th codified profile $x_m(\tau)$ of the substrate income flowrate is defined by the equation:

$$x_m(\tau) = x_{1,m} P_0 + x_{2,m} P_1 + x_{3,m} P_2. \tag{4.12}$$

The number of independent subfactors can be reduced considering two constraints. The first one imposes that each profile $x_m(\tau)$ goes to 0 at the end of the batch. This means that the combinations of the dynamic subfactors obey the following equation:

$$x_{1,m} + x_{2,m} + x_{3,m} = 0. \tag{4.13}$$

One of the three subfactors can be rewritten as a combination of the others:

$$x_{3,m} = -(x_{1,m} + x_{2,m}) \tag{4.14}$$

The second constraint involves Equation (4.10), which can be modified as:

$$\int_0^1 \left[\frac{6}{130}(1-\tau)+\frac{6}{130}(1-\tau)(x_{1,m}P_0+x_{2,m}P_1+x_{3,m}P_2)\right]=\frac{3}{130} \quad (4.15)$$

Using the (4.15), the dynamic subfactor $x_{2,m}$ can be rewritten as a combination of $x_{1,m}$ in this way:

$$x_{2,m}=3x_{1,m} \quad (4.16)$$

Equations (4.14) and (4.16) allow reducing the number of independent factors from four to two, which are w_1 and x_1.
The constraint (4.10) defines the variability range of the independent subfactor x_1 that is forced to vary between [-0.25, +0.25].

4.2.2 DoDE and experimentation

If the objective is to define a full quadratic RSM, at least 10 experiments are required to estimate the linear, the interaction and the quadratic effects of two independent factors on the output. A 3-level FF design with two factors w_1 and x_1, is performed, considering the range of variation defined previously. The design requires also 4 repeated experiments in the central point, useful to assess the normal variability of the process. The experiments are described in Table 4.1.

Table 4.1 *Three level full factorial design with two factors x_1 and w_1 the represent the variation of the substrate income flowrate and of initial biomass concentration.*

Run	Coded variables			
	w_1	x_1	x_2	x_3
1	1	-0.125	-0.375	0.5
2	1	0.125	0.375	-0.5
3	1	0	0	0
4	-1	-0.125	-0.375	0.5
5	-1	0.125	0.375	-0.5
6	-1	0	0	0
7	0	-0.125	-0.375	0.5
8	0	0.125	0.375	-0.5
9	0	0	0	0
10	0	0	0	0
11	0	0	0	0
12	0	0	0	0
13	0	0	0	0

The factor's values in Table 4.1 characterize thirteen experiments. The experimental data in terms of grams of penicillin produced and of penicillin concentration at the end of the batch are reported in Table 4.2.

Table 4.2 *Experimental data in terms of grams and concentration of the penicillin that are obtained at the end of the batch process according to the experimental campaign described in Table 4.1*

Run	Simulated production (gr)	Measured production (gr)	Simulated concentration (gmol/L)	Measured concentration (gmol/L)
1	36.6132	36.5901	3.6614	3.6651
2	77.5159	77.6183	7.7528	7.7604
3	56.9233	57.8223	5.6923	5.6880
4	31.3725	31.5649	3.1373	3.1513
5	68.0308	68.8202	6.8032	6.7908
6	49.5667	49.6694	4.9567	4.9362
7	34.3659	34.4023	3.4367	3.4414
8	73.6708	73.8966	7.3677	7.3847
9	53.8512	53.7091	5.3851	5.3793
10	53.8512	54.1855	5.3851	5.4066
11	53.8512	53.8087	5.3851	5.3442
12	53.8512	53.4813	5.3851	5.3562
13	53.8512	54.0857	5.3851	5.3024

The profiles of the concentration of the penicillin inside the reactor are shown in Figure 4.1, since they will be useful to define the DRSM in the next steps.

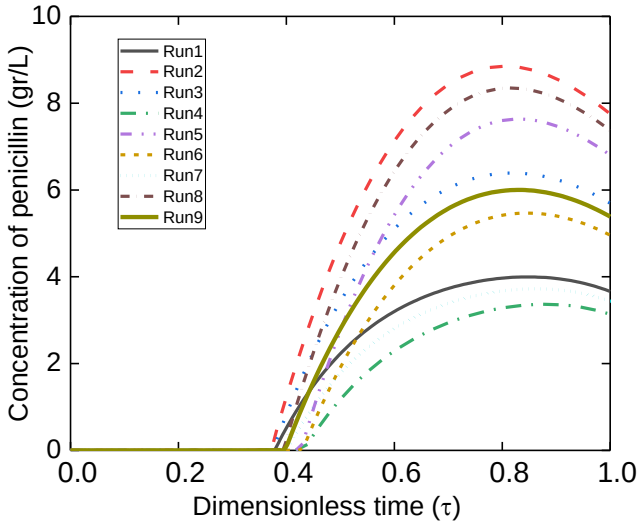


Figure 4.1 *Profiles of the penicillin concentration obtained by performing a 3-level full factorial design in the case of DoDE.*

4.2.3 RSM and optimization

Using the experimental data reported in last column of Table 4.2, the RSM that describes the relationship between the response variable y at the end of the process and the input factors (w_1 and x_1) can be defined. This is reported in Equation (4.17).

$$\hat{y} = b_0 + b_1 w_1 + b_2 x_1 + b_{12} w_1 x_1 + b_{11} {w_1}^2 + b_{22} {x_1}^2. \quad (4.17)$$

The response variable considered is the penicillin concentration. The estimated coefficients of Equation (4.17) are reported in Table 4.3.

Table 4.3 *Coefficients of response surface model in Equation (4.17) that define the relationship between input factors: substrate income flowrate and initial biomass concentration and the end-point penicillin concentration.*

Coefficients					
b_0	b_1	b_2	b_{12}	b_{11}	b_{22}
5.36±0.0342	0.3725±0.034	15.57 ±0.269	0.9116±0.33	-0.058 ± 0.049	2.75±3.172

The R^2-adj for the above regression is 1. The maximum predicted productivity is 7.77 gr/L, which corresponds to the values of $w_{1,OPT} = 1$, $x_{1,OPT}$ =0.125, $x_{2,OPT}$ =0.375 and $x_{3,OPT}$ =-0.5. The maximum concentration of the penicillin at the end of the batch is 7.75 gr/L that corresponds to the result of the second experiment in Table 4.2. In Figure 4.2 the optimum profile of both the income flowrate of substrate and the concentration of penicillin are shown.

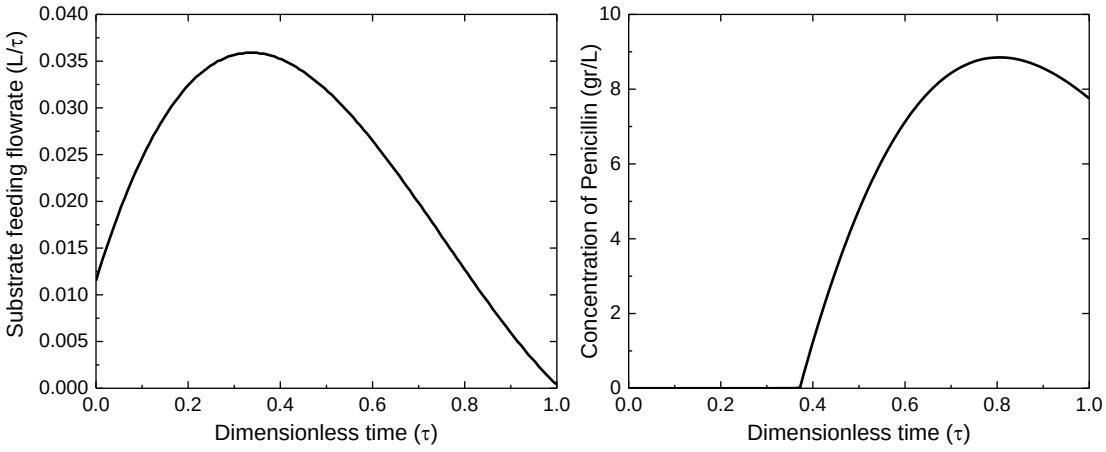


Figure 4.2 *Optimum profiles of a) the substrate feeding flowrate and b) the concentration of penicillin obtained with the DoDE.*

A DRSM is developed considering the profiles of the penicillin concentration obtained by performing a DoDE.

4.3 Definition of a dynamic response surface model

For the definition of the DRSM, the data of the penicillin concentration collected by performing a DoDE as the one in Subsection 4.3.1 are used. As it is possible to see in Figure 4.3, the profiles of the penicillin concentration show a discontinuity.

This makes the approximation more difficult. It has been decided to collect at least 14 points from each profile and use the first thirteen Legendre polynomials. The accuracy of the DRSM is defined through the calculation of $\widehat{SS}_{un}$ using Equation (1.26).

Table 4.4 *Values of $\widehat{SS}_{un}$ for dynamic response surface model with variable K and R calculated in the case of penicillin fermentation process*

R	K=Number of experiments for each profile							
	14	**15**	**16**	**17**	**18**	**19**	**20**	**21**
13	7.61E-04	3.87E-04	2.95E-04	5.78E-04	9.23E-04	9.41E-04	8.34E-04	8.40E-04
14		4.44E-04	3.01E-04	5.66E-04	6.57E-04	7.51E-04	6.80E-04	6.45E-04
15			3.06E-04	4.49E-04	6.62E-04	6.68E-04	8.86E-04	6.10E-04
16				4.48E-04	5.46E-04	6.72E-04	5.08E-04	3.74E-04
17					6.22E-04	5.65E-04	3.64E-04	4.54E-04
18						2.41E-04	8.35E-04	4.79E-04
19							2.67E-04	6.11E-04
20								2.73E-04

By looking to the results, it is possible to observe that the value of $\widehat{SS}_{un}$, cannot be reduced below 10^{-4} also using a large number of polynomials. This is due to the presence of a discontinuity in the experimental profiles of the penicillin concentration. However, the ratio F_0 and the *p*-value have to be calculated using the (1.29) in order to certify that the model adequately represents the non-random variability of the data. The values of *p*-value for each combination of *R* and *K* are reported in Table 4.5.

Table 4.5 *Values of p-values for dynamic response surface model with a variable K and R calculated in the case of penicillin fermentation process*

R	K=Number of experiments for each profile							
	14	**15**	**16**	**17**	**18**	**19**	**20**	**21**
13	1	0.615	0.89	1	1	1	1	1
14		1	1	1	1	1	1	1
15			1	0.98	1	1	1	1
16				1	1	1	0.99	0.99
17					1	1	0.92	0.99
18						0.82	1	1
19							0.83	1
20								0.87

An acceptable DRSM model is the one obtained with *K*=20 and *R*=19. The general equation of the DRSM is reported in Equation (4.18).

$$y_{DRSM}(\tau) = \sum_{r=1}^{R} g_{0,r}P_{(r-1)} + \left\{\sum_{r=1}^{R} g_{1,r}P_{(r-1)}\right\}w_1 + \left\{\sum_{r=1}^{R} g_{2,r}P_{(r-1)}\right\}x_1 + \left\{\sum_{r=1}^{R} g_{12,r}P_{(r-1)}\right\}w_1x_1 + \left\{\sum_{r=1}^{R} g_{11,r}P_{(r-1)}\right\}{w_1}^2 + \left\{\sum_{r=1}^{R} g_{22,r}P_{(r-1)}\right\}{x_1}^2 \quad (4.18)$$

The coefficients $\{g_{q,r}\}$ of the DRSM are calculated via stepwise regression that turns only the parameters that are significant. These are reported in Table 4.6.

Table 4.6 *Coefficients of dynamic response surface model in Equation (4.18) that define the relationship between the inputs and the time-variant output*

Coefficients				
g0,1	g0,2	g0,3	g0,4	g0,5
2.736	4.214±0.031	0.129±0.029	-2.096±0.030	-0.518±0.041
g0,6	g0,7	g0,8	g0,10	g0,11
1.103±0.036	-0.181±0.048	-0.608±0.040	0.318±0.043	-0.247±0.044
g0,12	g0,13	g0,17	g1,1	g1,2
-0.271±0.051	0.379±0.054	0.359±0.072	0.365±0.015	0.321±0.027
g1,3	g1,4	g1,5	g1,7	g1,8
-0.308±0.040	-0.300±0.045	0.309±0.056	-0.271±0.063	-0.114±0.062
g1,9	g1,10	g1,11	g1,12	g1,15
0.231±0.061	-0.147±0.059	-0.252±0.056	0.224±0.060	-0.121±0.102
g1,16	g1,18	g1,19	g2,1	g2,2
0.364±0.084	-0.151±0.083	0.243±0.111	8.471±0.111	12.815±0.187
g2,4	g2,5	g2,6	g2,8	g2,10
-7.028±0.281	-1.725±0.369	3.539±0.347	-1.790±0.394	0.946±0.417
g2,11	g2,12	g2,13	g2,17	g12,1
-0.698±0.431	-0.704±0.478	0.936±0.472	0.802±0.472	0.943±0.132
g12,2	g12,3	g12,4	g12,5	g12,7
0.904±0.215	-0.875±0.284	-0.491±0.291	0.772±0.367	-0.843±0.422
g12,9	g22,1	g22,2		
0.548±0.445	2.606±1.448	2.790±2.332		

The error SS_{un} that defines the accuracy of DRSM in the entire profile is calculated using (1.27). In this case, the experimental profiles of the concentration of penicillin are compared to the profiles approximated using the DRSM. The value of SS_{un} is 0.0018 that is higher than the $\widehat{SS}_{un}$. This result confirms the difficulty to approximate the profiles of the penicillin concentration along the batch. In Figure 4.3, where the DRSM predictions are plotted against the experimental profiles, it is possible to notice an oscillatory behaviour of the profiles predicted in the first part of the batch.

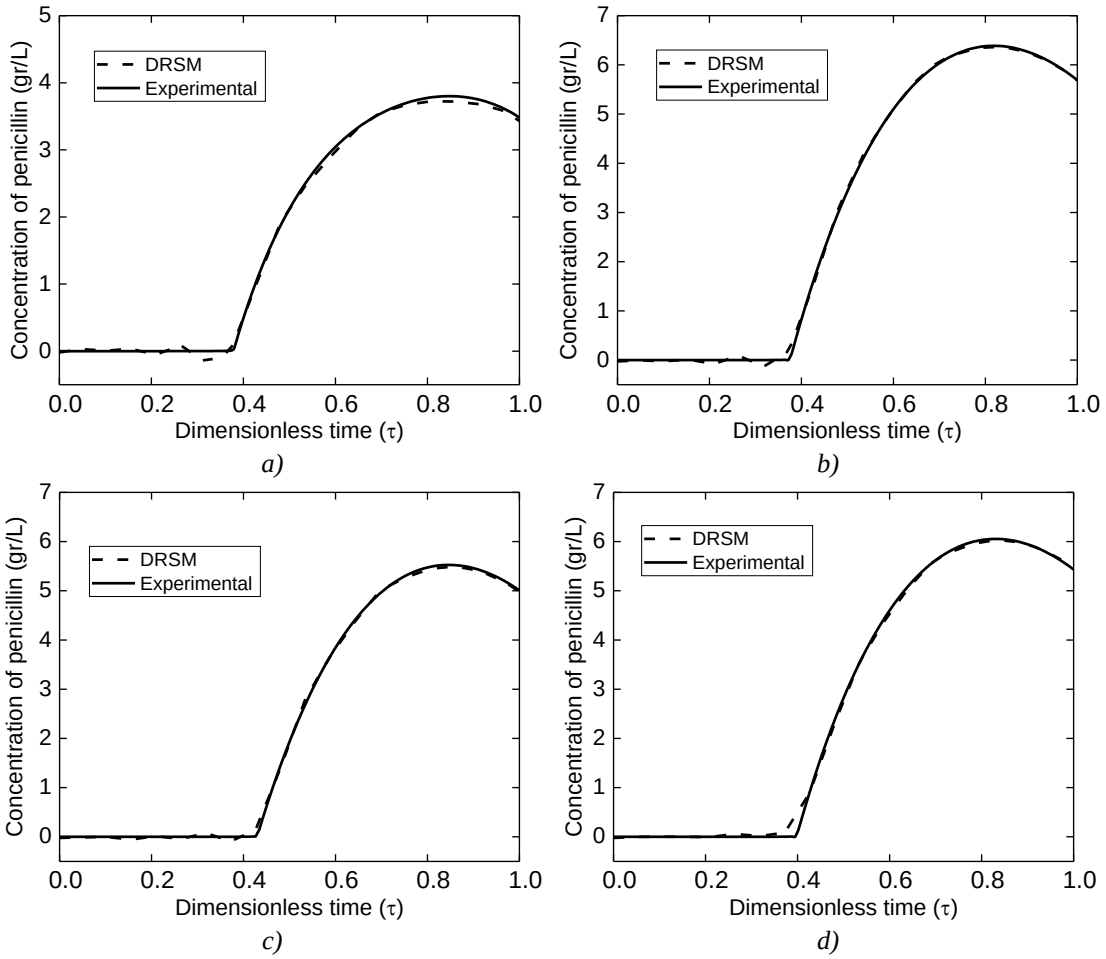


Figure 4.3 *In these figures the experimental profiles are compared to the DRSM predictions of penicillin concentration profiles within the experimental design in the case of a) w_1=1 and x_1= -0.125, b) w_1=1 and x_1=-0.125, c) w_1=-1 and x_1=0 and d) w_1=0 and x_1=0*

4.3.3.1 Optimization using dynamic response surface model

Using the DRSM calculated in Equation (4.18), it is possible to represent the profiles of the penicillin concentration inside the reactor for different combinations of w_1 and x_1. The profiles are shown in Figure 4.4.

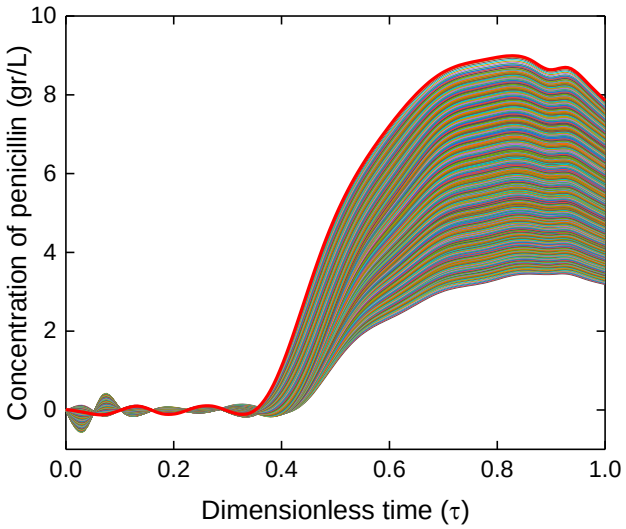


Figure 4.4 *Profiles of the Penicillin concentration along time obtained using the DRSM (4.28)*

In this case, the optimization procedure followed in Subsection 3.4.1 is not necessary. In fact, looking to the Figure 4.4, it is possible to notice that one of the profiles defined by the DRSM satisfies all the optimal criteria required. This profile is the same calculated through the DoDE (Subsection 4.2.3) and corresponds to the values of $w_{1,OPT} = 1$, $x_{1,OPT}$ =0.125, $x_{2,OPT}$ =0.375 and $x_{3,OPT}$ =-0.5. The maximum concentration of the penicillin at the end of the batch is 7.75 gr/L. Since its trend is always higher than the others, it is not necessary to vary the values of the dynamic subfactors along the process.

4.4 Optimization of a penicillin fermentation process with variable batch duration

The profiles of the response variable suggest the possibility to study the influence of the batch time on the calculation of the operating optimum. To do this, the batch time is considered a static factor whose variation is described by using the codified factor w_2 as in the following equation:

$$t_b = 130 + 30w_2 \tag{4.19}$$

with:

$$-1 \leq w_2 \leq +1 \tag{4.20}$$

In the application of DoE and DoDE, the static factor w_2 is combined with the dynamic factor $x(t)$ that describes the substrate income flowrate and the static factor w_1, which defines the initial biomass concentration.

4.4.1 Optimization of a penicillin fermentation process with variable batch duration using DoE

The calculation of the optimal batch duration requires the definition of an empirical model that described the linear, interaction and quadratic effects of the inputs factor on the output factor, which is the penicillin concentration at the end of the batch.

4.4.1.1 Selection of input factors

The DoE considers all the input factors as static. This means that the dynamic factor $x(t)$, that describes the variation of the substrate income flowrate, is parametrized using the first subfactor x_1. This allows defining only linear profiles of the substrate income flowrate. The subfactor x_1 is indicated as a static variable using w_3. The number of independent factors is decreased considering the constraint on the volume in Equation (4.15). This is rewritten as:

$$(130 + 30w_2)\int_0^1 \left[\frac{6}{130}(1-\tau) + \frac{6}{130}(1-\tau)(w_3 P_0)\right] = 3 \ . \tag{4.21}$$

It is possible to express w_2, which represents the batch duration, as a combination of w_3 in this way:

$$w_2 = \frac{-4.33w_3}{(1+w_3)} \ . \tag{4.22}$$

Since w_2 is bounded through Equation (4.20), the independent factors, w_1 and w_3, must vary according to the following inequalities:

$$-0.18 \leq w_3 \leq 0.30 \, , \tag{4.23}$$

$$-1 \ \leq w_1 \leq 1. \tag{4.24}$$

4.4.1.2 DoE and experimentation

The RSM that has to be defined is made by two independent factors (w_1 and w_3), which describe respectively the initial biomass concentration and the linear variation of the substrate income flowrate. If the model aims to explore the linear, the interaction and the quadratic effects of the input factors, at least 10 parameters have to be estimated. A 3-levels FFD is considered the best choice since it allows a complete collection of information.

Table 4.7 *Three level full factorial design that defines the combination of the three factors w_1, w_2, and w_3 that describe, respectively the variation of initial biomass concentration, of the batch duration and of the linear profile of the substrate income flowrate.*

Run	Coded Variables		
	w_1	w_2	w_3
1	-1	-0.999	0.30
2	-1	-0.245	0.06
3	-1	0.950	-0.18
4	-1	0.950	-0.18
5	0	-0.999	0.30
6	0	-0.245	0.06
7	0	-0.245	0.06
8	0	0.950	-0.18
9	1	-0.999	0.30
10	1	-0.999	0.30
11	1	-0.245	0.06
12	1	0.950	-0.18

The factor's values in Table 4.7 characterize twelve experiments, from which it is possible to collect the experimental data, in terms of grams of penicillin produced and of end-point penicillin concentration, that are reported in Table 4.8.

Table 4.8 *Experimental data in terms of grams and concentration of the penicillin obtained at the end of the batch process according to the experimental campaign described in Table 4.7.*

Run	Simulated production (gr)	Measured production (gr)	Simulated concentration (gr/L)	Measured concentration (gr/L)
1	45.8550	45.6586	4.5814	4.5689
2	47.2141	47.1224	4.7213	4.7207
3	53.8546	54.1316	5.3858	5.3643
4	53.8546	53.7901	5.3858	5.3720
5	41.2238	41.1307	4.1220	4.0914
6	51.8995	51.6454	5.1898	5.1764
7	51.8995	51.6054	5.1898	5.1719
8	56.8294	56.8357	5.6833	5.6998
9	45.5766	45.8136	4.5572	4.5904
10	45.5766	45.8572	4.5572	4.5808
11	55.2672	55.4219	5.5266	5.5555
12	58.9352	58.8441	5.8939	5.9139

The profiles of the penicillin concentration defined in DoDE are shown in Figure 4.5.

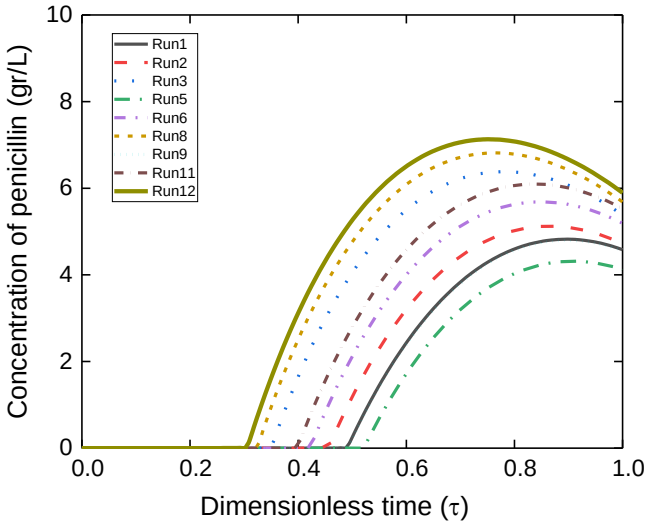


Figure 4.5 *Profiles of the penicillin concentration obtained by performing a 3-level full factorial design with three factors in the case of DoE*

4.3.1.3 RSM and optimization

Starting from the experimental data reported in Table 4.8, the following RSM can be defined:

$$\hat{y} = b_0 + b_1 w_1 + b_2 w_3 + b_{12} w_1 w_3 \tag{4.25}$$

As in Subsection 4.2, only the data regarding the penicillin concentration is taken into account. The estimated coefficients of the RSM in Equation (4.25) are in Table 4.9.

Table 4.9 *Coefficients of response surface model in Equation (4.25) that define the influence of the initial biomass concentration and of the substrate feeding flowrate on the end-point penicillin concentration considering a variable batch duration*

Coefficients			
b_0	b_1	b_2	b_{12}
5.24±0.1256	0.2557±0.1535	-2.587±0.6146	-0.570±0.7071

The RSM calculated does not consider the quadratic dependence of the input factors; this is due to the fact that the used function "*regress*" allows to determine the significance of the terms of which the RSM is composed. In this case, the quadratic effects of the factors do not influence in a significant way the final penicillin concentration. The R^2-adj for the above regression is 0.92. The optimum values of the input variables are those corresponding to the maximum predicted value of the end-point penicillin concentration. This is equal to 6.0712 gr/L and corresponds to the values of $w_{1,OPT}$ =1, $w_{2,OPT}$ = 0.950 and $w_{3,OPT}$= -0.18.

The penicillin concentration obtained by carrying out the process with the suggested optimal conditions concentration, is 5.89 gr/L.
The optimal batch time is equal to 158.5 hr but this result must be confirmed performing DoDE.

4.4.2 Optimization of a penicillin fermentation process with variable batch duration using DoDE

In this case, an empirical model that describes the influence of the input factors (both dynamic and static) must be calculated. The desired RSM considering both the linear and the non-linear terms (e.g. interactions and quadratic).

4.4.2.1 Selection of input factors

By applying DoDE, it has been possible to refer to the article Georgakis (2013) so that to compare the results obtained. In this case, the profiles of the substrate income flowrate are defined by the three dynamic subfactors $\{x_1, x_2, x_3\}$ as in Equation (4.11). Two static factors have to be considered, which are the initial biomass concentration w_1 and the batch duration w_2.
The number of independent factors (five), is reduced by imposing two constraints as in Section 4.2. The first one imposes that each profile $x_m(\tau)$ tends to 0 at the end of the batch, so one of the dynamic subfactors can be rewritten as a linear combination of the others, using Equation (4.14). The second constraint is on the reactor volume and is reported below:

$$(130 + 30w_2)\int_0^1 \left[\frac{6}{130}(1-\tau) + \frac{6}{130}(1-\tau)x_m(\tau)\right] = 3 \ . \tag{4.26}$$

This equation imposes the following relationship between the factor w_2 that describes the batch duration and the dynamic subfactors x_1 and x_2.

$$w_2 = -4.33\left(\frac{3x_1 - x_2}{3x_1 - x_2 + 3}\right) \ . \tag{4.27}$$

Equation (4.20) forces the w_2 to vary inside the range [-1, +1]. This bounds the variability range of the dynamic subfactors.

4.4.2.2 DoDE and experimentation

The independent factors are three: w_1, x_1 and x_2. A full quadratic RSM, as the one required, is composed by 10 parameters.

The experimental plan, made by the combinations of the three independent variables $\{w_1, x_1, x_2\}$, is built according to the d-optimal criterion. The latter provided the definition of sixteen experiments that are characterized in Table 4.10.

Table 4.10 *D-optimal design that defines the combinations of w_1, w_2, x_1, x_2 and x_3 in DoDE applied to the case study of penicillin fermentation process with variable batch duration.*

Run	Coded Variables				
	w_1	w_2	x_1	x_2	x_3
1	0.33	0.99	-0.1	0.26	-0.16
2	0.33	-0.62	0	-0.5	0.5
3	-1	1.04	-0.27	-0.23	0.5
4	-1	-0.99	0.13	-0.5	0.37
5	1	-0.3	0.18	0.32	-0.5
6	-1	-0.14	0.03	-0.01	-0.02
7	0	-1	0.35	0.15	-0.5
8	-1	0.99	-0.02	0.5	-0.48
9	1	-0.3	0.18	0.32	-0.5
10	1	1.04	-0.27	-0.23	0.5
11	0	1.04	-0.27	-0.23	0.5
12	-1	-0.99	0.13	-0.5	0.37
13	-0.33	0.03	0.12	0.38	-0.5
14	1	1.04	-0.27	-0.23	0.5
15	0	-0.63	0.14	-0.09	-0.05
16	1	-1	0.24	-0.18	-0.06

The substrate income profiles, defined using the value in Table 4.10 are shown in Figure 4.6.

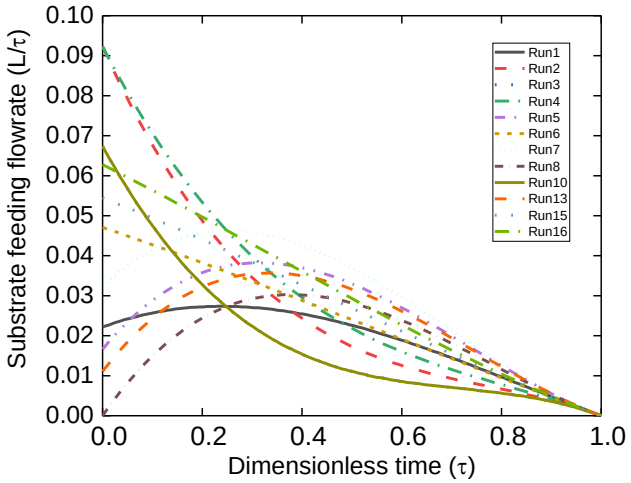


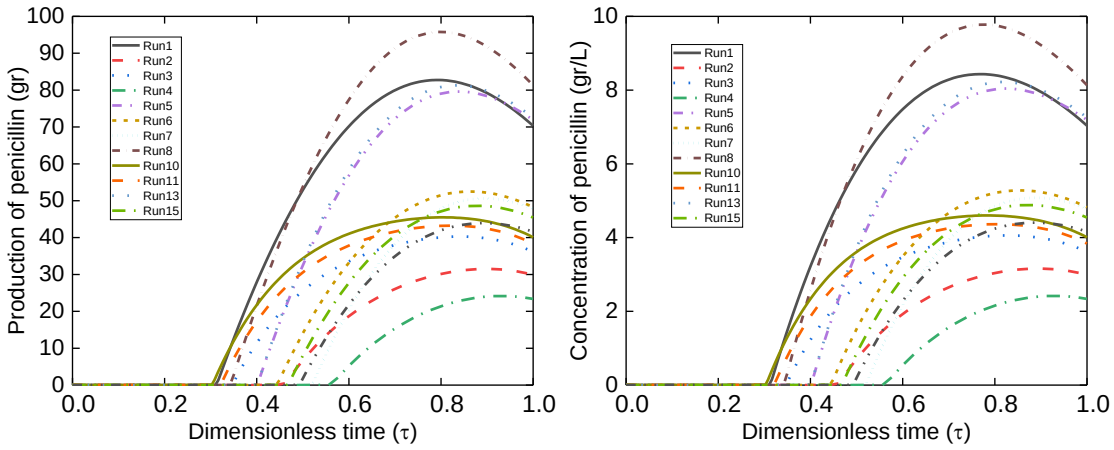
Figure 4.6. *Profiles of the substrate feeding flowrate obtained by performing a D-optimal design in the case of DoDE*

The experimental data, in terms of grams of penicillin produced and of final penicillin concentration, collected from the experiments in Table 4.10, are reported in Table 4.11.

Table 4.11 *Experimental data in terms of grams and concentration of the penicillin obtained at the end of the batch process according to the experimental campaign described in Table 4.10.*

Run	Simulated production (gr)	Measured production (gr)	Simulated concentration (gr/L)	Measured concentration (gr/L)
1	70.3113	69.7830	7.0334	6.9989
2	29.8981	29.8670	2.9901	2.9958
3	36.1361	35.8642	3.6136	3.6195
4	23.3673	23.5788	2.3364	2.3515
5	71.8808	71.8388	7.1905	7.2300
6	48.2309	48.5266	4.8231	4.8389
7	47.8810	47.9995	4.7881	4.7760
8	81.2945	81.6638	8.1336	8.1143
9	71.8808	71.4660	7.1905	7.1167
10	40.0795	40.3315	4.0078	3.9988
11	38.3801	38.6588	3.8380	3.8083
12	23.3673	23.1247	2.3364	2.3473
13	72.5724	72.5083	7.2573	7.2900
14	40.0795	39.9488	4.0078	4.0106
15	45.4576	45.4932	4.5458	4.5372
16	41.6087	41.4318	4.1608	4.1637

In Figure 4.7, the profiles of both the penicillin concentration and its productivity along the batch are shown.



a) *b)*

Figure 4.7 *Profiles of a) penicillin production and b) penicillin concentration related to the experimental design in Table 4.11.*

4.4.2.3 RSM and optimization

With the experimental data reported in the second and forth columns of Table 4.11, it is possible to define a RSM with the following formulation:

$$\hat{y} = b_0 + b_1 w_1 + b_2 x_1 + b_3 x_2 + b_{12} w_1 x_1 + b_{13} w_1 x_2 + b_{23} x_1 x_2 + b_{11} {w_1}^2 + b_{22} {x_1}^2 + b_{33} {x_2}^2 \quad (4.28)$$

The RSM, in this case, is defined both for the penicillin concentration and for the penicillin produced in order to compare the results with those in Georgakis (2013). The estimated coefficients of Equation (4.28), in the case in which the response variable $\hat{y}$ is the end-point penicillin concentration, are reported in Table 4.12.

Table 4.12 *Coefficients of the response surface in Equation (4.28) that describes the influence of inputs on the final penicillin concentration*

Coefficients				
b_0	b_1	b_2	b_3	b_{12}
5.312± 0.0935	0.3367 ±0.0472	-1.1092 ±0.2415	5.5398 ± 0.1485	0.4704 ± 0.2397
b_{13}	b_{23}	b_{11}	b_{22}	b_{33}
0.0622 ± 0.1426	-0.885 ± 1.2207	-0.0013 ±0.074	-7.6426 ±1.4163	1.5342 ±0.4657

In Table 4.12, the optimum values of the input variables, calculated according to the procedure in Sub-subsection (1.1.5.1), are reported.

Table 4.13 *Optimum value of the input factors calculated maximizing the RSM in Equation (4.28), with the coefficients in Table (4.12)*

	Concentration	**Coded variables**				
	(gr/L)	w_1	w_2	x_1	x_2	x_3
Calculated	8.8512	1	1	-0.021	0.5	-0.479
Simulated	8.7311					

The RSM can be defined also for the penicillin production. The coefficients, calculated using a least squares algorithm for the estimation of the coefficients in the multiple linear regression empirical model, are reported in Table 4.14

Table 4.14 *Coefficients of the response surface in Equation (4.28) that describes the influence of inputs on the end-point penicillin production*

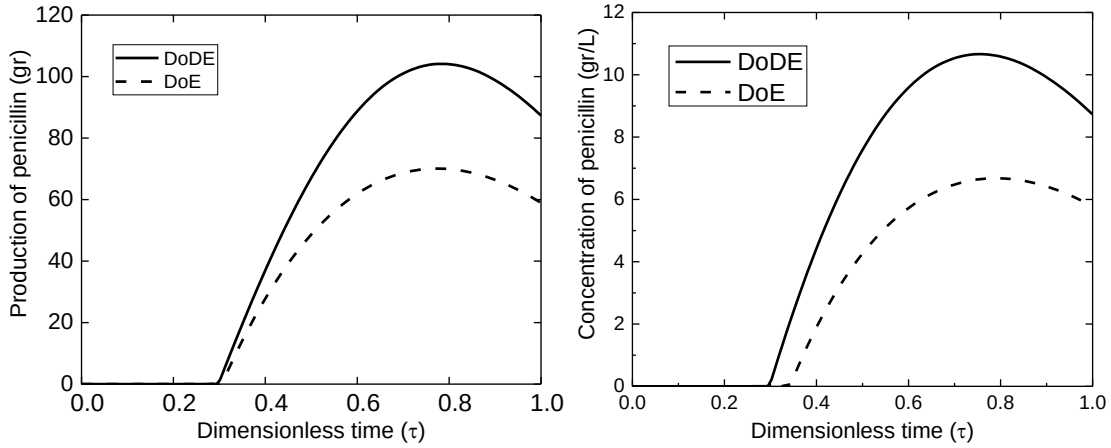
Coefficients				
b_0	b_1	b_2	b_3	b_{12}
53.08± 0.7297	3.291 ±0.3677	-11.156 ±1.8835	55.523 ± 1.1584	4.3765 ± 1.8696
b_{13}	b_{23}	b_{11}	b_{22}	b_{33}
0.2685 ± 1.1123	-8.8513 ± 9.5202	-0.0412 ±0.5773	-74.359 ±11.045	15.404 ±3.632

In Table 4.15, the optimum values of the input variables are reported.

Table 4.15 *Optimum value of the input factor calculated maximizing the RSM in Equation (4.28) with the coefficients in Table (4.14)*

	Production	**Coded variables**				
	(gr)	w_1	w_2	x_1	x_2	x_3
Calculated	88.2786	1	1	-0.021	0.5	-0.479
Simulated	87.3122					

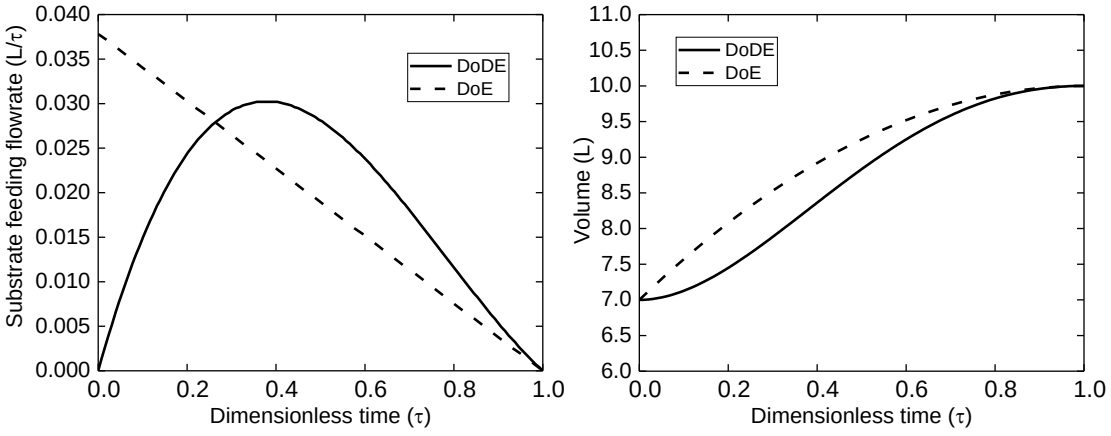
In Figure 4.8, the profiles of the penicillin concentration and of the penicillin produced, obtained by carrying out the process at the suggested optimal conditions, are shown. The profiles obtained by performing a DoDE are compared to those obtained by DoE.



a) *b)*

Figure 4.8. *Comparison between the optimum profile calculated in DoDE and DoE of a) the penicillin produced and b) the penicillin concentration*

In Figure 4.9, the profiles of the volume and of the substrate feeding flowrate, obtained by carrying out the process at the suggested optimal conditions, are shown, both in the case of DoE and DoDE.



a) *b)*

Figure 4.9. *Comparison between optimum profile in the case of DoDE and DoE of a) the substrate feeding flowrate, b) the volume*

Looking to Figure 4.8, it is possible to observe that the optimum conditions provided by DoDE leads to an improvement in terms of penicillin productivity, with respect to those provided by DoE. This means that the optimal profile of the substrate income flowrate is a non-linear one.

The results obtained are compared to those in Georgakis (2013). The maximum value of the final penicillin concentration calculated is significantly higher than the one in Georgakis (2013). The aim of the procedure in this section was to understand if the calculation of the optimal batch duration brings an improvement in the maximum value of the penicillin concentration. In Figure 4.10*b*, the penicillin production obtained by applying the optimization procedure in the case of fixed batch time t_b=130 hours is compared to that calculated considering the variation of batch duration (e.g. an optimum batch duration equal to t_b=160 hr).

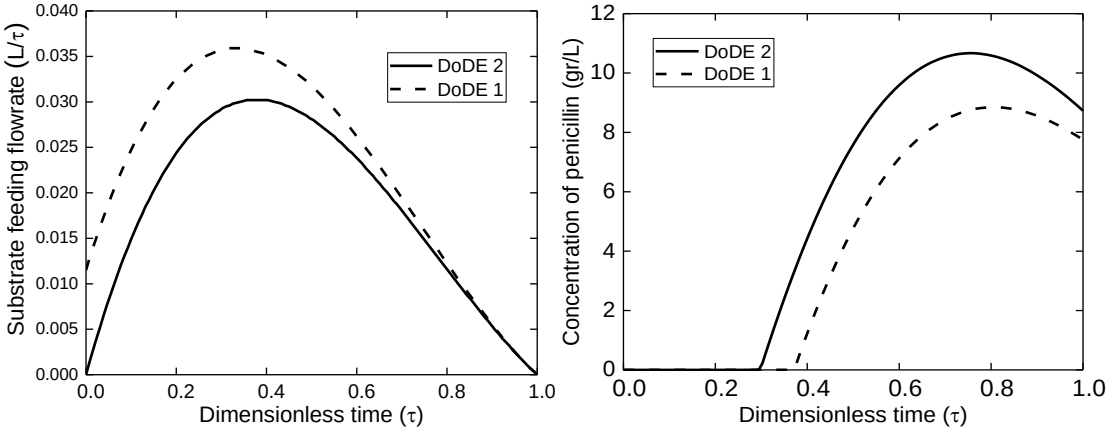


Figure 4.10 *Comparison between the optimum profile calculated performing a DoDE with variable batch time (continuous line) and the one calculated performing DoDE with fixed batch time (dashed line) of: a) the substrate feeding flowrate and b) the penicillin concentration.*

The increase of the batch duration affects the final value of the penicillin concentration significantly.

4.4.3 Definition of a dynamic response surface model

For the definition of the DRSM, the data of the penicillin concentration collected by performing a DoDE as the one in Subsection 4.4.2 are used. Respect to the case developed in Section 4.3, the DRSM is composed by three independent factors. For this reason, it is expected that the approximation will be more complicated than the one performed above. In fact, it has been decided to collect at least 18 experimental point from each profile. The accuracy of the DRSM is defined through the calculation of $\widehat{SS}_{un}$ using Equation (1.26).

Table 4.16 *Values of $\widehat{SS}_{un}$ for dynamic response surface model with variable K and R calculated in the case of penicillin fermentation process with variable batch duration*

R	K=Number of experiments for each profile						
	18	**19**	**20**	**21**	**22**	**23**	**24**
17	8.60E-04	5.70E-04	6.32E-04	6.94E-04	9.64E-04	8.42E-04	6.07E-04
18		5.71E-04	8.13E-04	0.0013	5.10E-04	7.49E-04	5.30E-04
19			7.45E-04	0.001	5.46E-04	4.21E-04	7.18E-04
20				6.94E-04	5.54E-04	0.0012	6.36E-04
21					0.0011	8.21E-04	7.97E-04
22						8.13E-04	6.77E-04
23							7.17E-04

By looking to the results, it is possible to observe that the value of $\widehat{SS}_{un}$, cannot be reduced below 10^{-4} also using many polynomials.
However, the ratio F_0 and the *p*-value have to be calculated using the (1.29) in order to certify that the model adequately represents the non-random variability of the data. The *p*-value for each combination of *R* and *K* are reported in Table 4.17.

Table 4.17 *Values of p-values for dynamic response surface model with a variable K and R calculated in the case of penicillin fermentation process with variable batch duration*

R	K=Number of experiments for each profile						
	18	**19**	**20**	**21**	**22**	**23**	**24**
17	1	0.99	1	1	1	0.98	1
18		1	1	1	1	1	0.98
19			1	1	0.88	0.78	1
20				1	1	1	0.99
21					1	1	1
22						1	1
23							1

An acceptable DRSM model is the one obtained with *K*=23 and *R*=19. The general equation of the DRSM is reported in Equation (4.29).

$$y_{DRSM}(\tau) = \sum_{r=1}^{R} g_{0,r}P_{(r-1)} + \left\{\sum_{r=1}^{R} g_{1,r}P_{(r-1)}\right\}w_1 + \left\{\sum_{r=1}^{R} g_{2,r}P_{(r-1)}\right\}x_1 + \left\{\sum_{r=1}^{R} g_{3,r}P_{(r-1)}\right\}x_2 \quad (4.29)$$

$$+\left\{\sum_{r=1}^{R} g_{12,r}P_{(r-1)}\right\}w_1x_1 + \left\{\sum_{r=1}^{R} g_{13,r}P_{(r-1)}\right\}w_1x_2 + \left\{\sum_{r=1}^{R} g_{23,r}P_{(r-1)}\right\}x_1x_2$$

$$+\left\{\sum_{r=1}^{R} g_{11,r}P_{(r-1)}\right\}{w_1}^2 + \left\{\sum_{r=1}^{R} g_{22,r}P_{(r-1)}\right\}{x_1}^2 + \left\{\sum_{r=1}^{R} g_{33,r}P_{(r-1)}\right\}{x_2}^2$$

The coefficients $\{g_{q,r}\}$ of the DRSM are calculated via stepwise regression that turns only the parameters that are significant. These are reported in Table 4.18.

Table 4.18 *Coefficients of dynamic response surface model in Equation (4.29) that define the relationship between the inputs and the time-variant output*

Coefficients				
g0,1	g0,2	g0,3	g0,4	g0,5
2.732	4.089±0.0469	0.214±0.0582	-2.020±0.0621	-0.429±0.0677
g0,6	g0,7	g0,8	g0,10	g0,12
0.956±0.0518	0.118±0.0767	-0.523±0.0602	0.368±0.0968	-0.229±0.0981
g1,1	g1,2	g1,3	g1,4	g1,5
0.337±0.0153	0.317±0.0257	-0.267±0.0331	-0.190±0.0348	0.211±0.0421
g1,6	g1,7	g1,9	g2,1	g2,2
0.064±0.0407	-0.153±0.0495	0.063±0.0504	-3.366±0.0721	-1.830±0.1226
g2,3	g2,4	g2,5	g2,6	g2,7
4.893±0.1587	1.604±0.1440	-4.384±0.2097	-0.382±0.2151	3.042±0.2576
g2,8	g2,9	g2,10	g2,11	g2,12
-0.822±0.2553	-2.389±0.2581	1.054±0.2815	0.726±0.2284	-1.293±0.2740
g2,14	g2,17	g3,1	g3,2	g3,3
0.978±0.3473	0.812±0.4256	4.055±0.0442	4.817±0.0762	-1.845±0.0958
g3,4	g3,5	g3,6	g3,7	g3,8
-2.950±0.0939	1.166±0.1243	1.103±0.1442	-1.195±0.1497	-0.254±0.1826
g3,9	g3,10	g3,11	g3,12	g3,13
0.931±0.1423	-0.273±0.1963	-0.606±0.1485	0.404±0.1704	0.274±0.1673
g3,14	g3,16	g12,1	g12,2	g12,3
-0.330±0.2184	0.264±0.2763	0.157±0.0798	0.582±0.1380	0.351±0.1650
g12,4	g12,5	g12,6	g12,8	g12,10
-0.661±0.1932	-0.293±0.1968	0.697±0.2356	-0.813±0.2643	0.520±0.2786
g13,1	g13,2	g13,3	g13,5	g13,7
0.196±0.0480	0.089±0.0758	-0.345±0.1082	0.199±0.1482	-0.457±0.1883
g13,11	g13,13	g13,17	g23,1	g23,2
-0.331±0.1910	0.289±0.1907	0.518±0.2848	-4.927±0.3851	-1.716±0.6504
g23,3	g23,5	g23,6	g23,7	g23,8
8.699±0.898	-4.973±1.2497	3.508±1.173	6.742±1.5562	-6.063±1.3964
g23,10	g23,12	g23,13	g23,14	g23,15
5.826±1.5460	-5.653±1.5215	2.210±1.2628	3.375±1.8874	-4.904±1.6154
g23,19	g11,1	g11,2	g11,3	g22,1
-3.564±1.8510	-0.077±0.0230	-0.060±0.0372	0.053±0.0456	-1.297±0.4329
g22,2	g22,3	g22,4	g22,5	g22,6
-6.107±0.7376	-5.970±0.9719	5.903±0.8529	4.024±1.3071	-7.598±1.2398
g22,7	g22,8	g22,10	g22,12	g22,13
-5.746±1.5733	8.760±1.4524	-8.357±1.7282	7.359±1.7318	-2.518±1.1145
g22,14	g22,15	g22,19	g33,1	g33,2
-3.413±1.5396	4.594±1.3629	3.478±1.5761	2.175±0.1412	1.950±0.2428
g33,3	g33,4	g33,5	g33,7	g33,9
-2.000±0.3029	-1.497±0.3512	1.720±0.3843	-1.786±0.4476	1.247±0.3218
g33,10	g33,11	g33,12	g33,13	g33,16
-1.076±0.5587	-1.279±0.3205	0.870±0.5260	0.873±0.3763	0.795±0.5469

In Figure 4.11, where the DRSM predictions are plotted against the experimental profiles it is possible to notice an oscillatory behaviour of the profiles predicted in the first part of the batch.

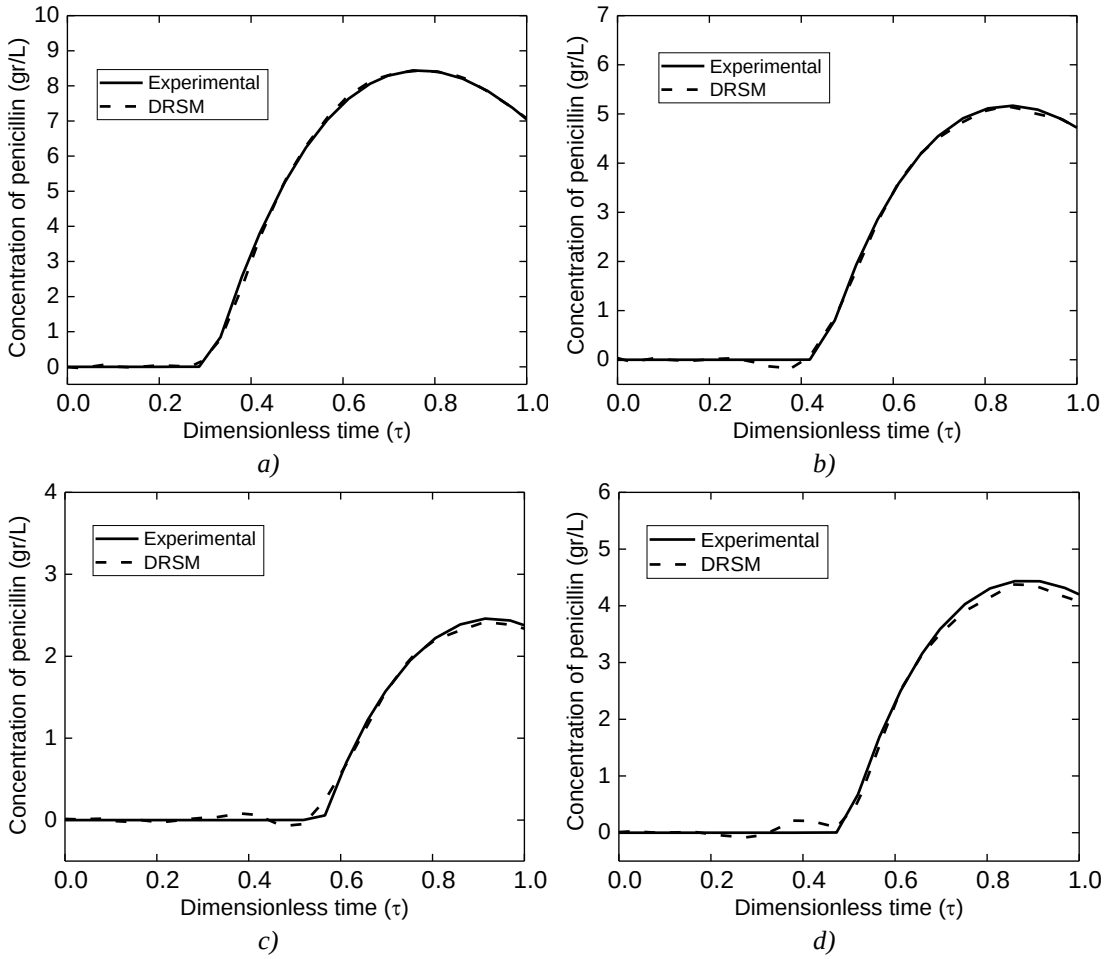


Figure 4.11 *In these figures the experimental profiles are compared to the DRSM predictions of penicillin concentration profiles within the experimental design in the case of a)* w_1*=0.33,* x_1*=-0.1 and* x_2*=0.26 b)* w_1*=-1,* x_1*=0.03 and* x_2*=-0.01 c)* w_1*=-1,* x_1*=0.13 and* x_2*=-0.5 d)* w_1*=1,* x_1*=0.25 and* x_2*=-0.18*

4.4.3.1 Optimization using dynamic response surface model

Using the DRSM calculated in Equation (4.29), it is possible to represent the profiles of the penicillin concentration inside the reactor for different combinations of w_1, x_1 and x_2. The profiles are shown in Figure 4.12.

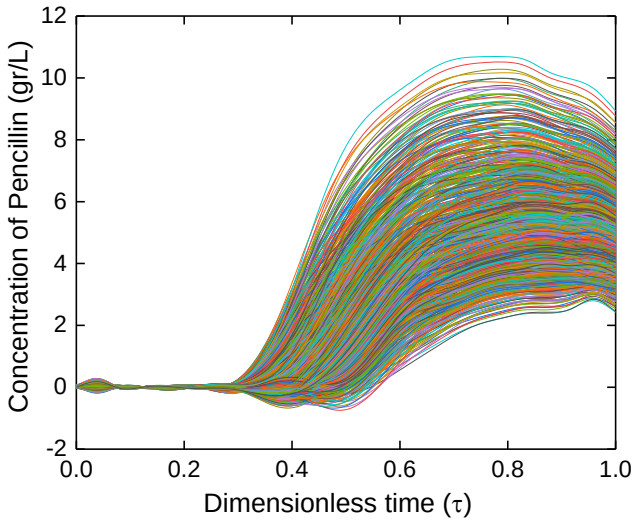
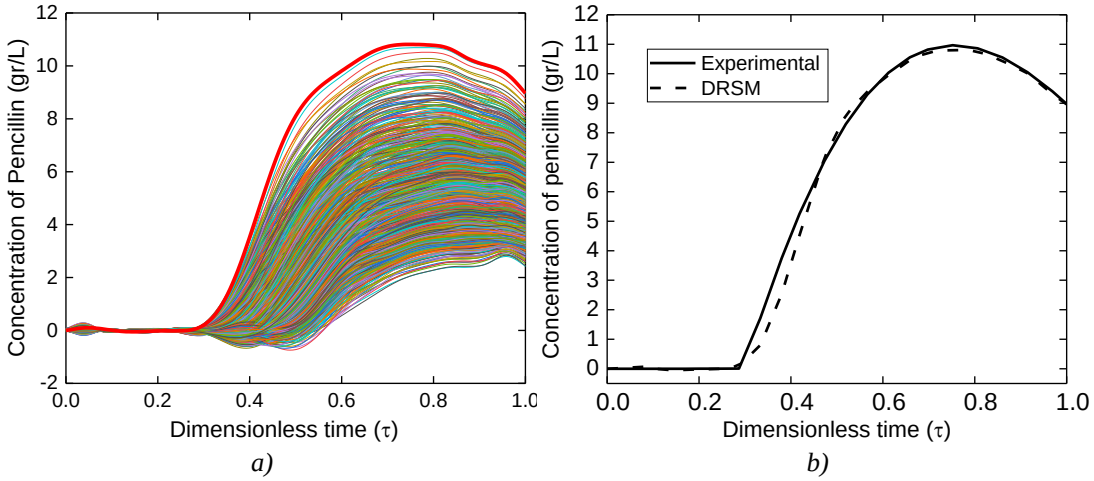


Figure 4.12 *Profiles of the penicillin concentration obtained using the DRSM in (4.29)*

One of the criteria that can be used for the definition of the optimum profile is to determine the value of the factors: w_1, x_1 and x_2, and the related temperature profile, that defines the conversion profile, which maximizes the integral of the DRSM, from τ = 0 to τ = 1, according to Equation (1.33). The combination of dynamic subfactors that meets this condition is: $w_{1,OPT}$= 1, $x_{1,OPT}$= -0.021; $x_{2,OPT}$= 0.5. The factors calculated are the same obtained in DoDE. By plotting this profile together with the profiles in Figure 4.12, it is possible to see that this satisfies also to the criterion in (1.34) since it is higher than all the others during the whole process.



a) *b)*

Figure 4.13. *Comparison between the optimum predicted profile of the penicillin concentration and a) the profiles obtained using the DRSM in (4.29) and b) the optimum simulated profile of the penicillin concentration*

The difference between the profiles in Figure 4.13*b* is defined by the value of the *RSME* calculated according to Equation (1.35). In this case *RSME* is equal to 0.36 that is higher than the value of σ=0.02.

This means that the difference between the predicted profile and the simulated one is higher than the measurement error added to the experimental data.

4.5 Conclusions regarding the optimization of the penicillin fermentation process

The obtained results can be summarized in order to compare the methodologies applied and to define the operating optimum of the penicillin fermentation process. The following considerations, regarding both the case of fixed batch duration and the variable one, are possible:

- In the case of fixed batch time (t_b=130 hours), DOE cannot be performed, and the comparison is made between DoDE results and DRSM. The calculation of the operating optimum using the DRSM does not bring any improvements with respect to DoDE, neither of penicillin concentration profile, nor of its end-point value. Therefore, the optimum is defined by the following factors values: $w_{1,OPT} = 1$, $x_{1,OPT}$ =0.125, $x_{2,OPT}$ =0.375 and $x_{3,OPT}$ =-0.5. The related final penicillin concentration is 7.75 gr/L.
- Considering the variability of the batch duration, both the DoE and DoDE have been applied. The DoDE leads to determine a higher value of the penicillin concentration at the end of the batch compared to the one that can be obtained using DoE. The operating optimum is characterized by the following values of the factors: $w_{1,OPT} = 1$, $w_{2,OPT} = 1$, $x_{1,OPT}$ = -0.021, $x_{2,OPT}$ =0.5 and $x_{3,OPT}$ =-0.479. The optimal batch duration is equal to t_b=160 hours. The corresponding final penicillin concentration is 8.73 gr/L. This result is higher than the one obtained by DoDE with fixed batch duration.

Conclusions

The objective of this thesis has been to discuss a methodology that allows defining the optimum operating conditions of a dynamic batch or semi-batch process. The methodology uses the data-driven approach commonly used in the case of time-invariant input variables, (namely, the design of experiments; DoE) and a new methodology (called design of dynamic experiments; DoDE) that allows considering the dynamic behavior of the input variables. The application of these two methodologies allows to obtain a large amount of information by minimizing the number of experiments to be performed. The data collected, in terms of process output, have to ensure the definition of a response surface model (RSM) that describes the effects of the factors on the output end-point value. A larger number of experimental measurements is required for the construction of a dynamic response surface model (DRSM), which is a data-driven model that captures the relationship between process input and time-resolved output. The RSM has been used to define the profile of the manipulated variables that optimize the final value of the process output. Instead, through the use of a DRSM, it is possible to optimize the process output profile. In the latter case, two example criteria have been considered. The first one aims to maximize the integral of the profile of the output variable while the second criterion requests to define the profile of the response variable that is greater than all the others during the whole process.

The complete methodology, made by both DoE and DoDE, and the related RSM and DRSM, has been applied to three case studies. Results indicate that the application of DoDE in batch and semi-batch processes can lead to the definition of better values of the process output than those achievable by DoE. This result shows how the time variation of the input variables can affect significantly the product quality.

For what concerns the construction of a DRSM, the examples considered have highlighted some critical aspects. Firstly, the output variable must be measurable during the entire batch. Furthermore, the number of experimental data, required for the definition of a DRSM, quickly increases as the number of input factors increases or as the process output trend shows a discontinuity. In the latter occurrence, the DRSM calculated may be inaccurate. Considering the results regarding the three case studies, it can be concluded that the use of the DRSM approach does not allow to improve significantly the output end-point value if compared to DoDE.

However, the DRSM methodology may be very useful to determine the optimum profile of the process output.

List of symbols

Acronyms

CCD	=	Central Composite Design
DODE	=	Design of Dynamic Experiments
DOE	=	Design of Experiment
DOD	=	D-Optimal Design
DRSM	=	Dynamic Response Surface Model
FFD	=	Full Factorial
RSM	=	Response Surface Model
RSME	=	Root Mean Square Error
SSE	=	Sum of Squares Error
SSR	=	Sum of Squares Regression

Symbols

$[A]$	=	Concentration of A (gmol/L)
$[B]$	=	Concentration of B (gmol/L)
B_T	=	Total amount of B fed (gmol)
b	=	Biomass concentration (gr_b/L)
$\bar{b}$	=	Initial biomass concentration (gr_b/L)
b_q	=	Coefficients of a Response Surface Model
$[C]$	=	Concentration of C (gmol/L)
$[D]$	=	Concentration of D (gmol/L)
D	=	Total number of the significance parameters of DRSM
d	=	Total number of parameters of RSM

$[E]$	=	Concentration of E (gmol/L)
E_A	=	Activation energy of direct reaction J/mol
E_B	=	Activation energy of inverse reaction J/mol
$g_{q,r}$	=	Coefficients of a Dynamic Response Surface Model
H	=	Total number of time instants in which the DRSM is maximized
I	=	Total number of variation levels of a static factor
J	=	Total numbers of static factors
K	=	Total number of experimental points for each profile used in DRSM
k_A	=	Kinetic constant of direct reaction (L/gmol/h)
k_B	=	Kinetic constant of inverse reaction (L/gmol /h)
k_{A0}	=	Pre-exponential factor of direct reaction (h^{-1})
k_{B0}	=	Pre-exponential factor of inverse reaction (h^{-1})
k_1	=	Kinetic constant of reaction 1 (L/gmol/ h)
k_2	=	Kinetic constant of reaction 2 (L/gmol/ h)
k_3	=	Kinetic constant of reaction 3 (h^{-1})
k_b	=	Saturation parameter for biomass growth (gr_b/gr_s)
k_p	=	Saturation parameter for penicillin production (gr_s/L)
k_{in}	=	Inhibition parameter for penicillin production (gr_s/L)
k_d	=	Penicillin degradation (h^{-1})
k_m	=	Saturation parameter for maintenance consumption (gr_s/L)
M	=	Total number of variation profiles of a dynamic factor
m_s	=	Maintenance consumption rate (gr_s/gr_b/h)
N	=	Total number of dynamic subfactors related to a dynamic factor
p	=	Penicillin concentration (gr_p/L)
P_0	=	Shifted Legendre polynomial of zero degree
P_1	=	Shifted Legendre polynomial of first degree
P_2	=	Shifted Legendre polynomial of second degree
R	=	Total number of shifted Legendre polynomials used in DRSM

r	=	Rate of reaction (gmol/L/h)
r_1	=	Rate of reaction (gmol/L/h)
r_2	=	Rate of reaction (gmol/L/h)
r_3	=	Rate of reaction (gmol/L/h)
R_g	=	Gas Constant (J/mol/K)
s	=	Substrate concentration (gr_s/L)
s_f	=	Feed concentration of substrate (gr_s/L)
t	=	Time (h)
t_b	=	Batch time (h)
T	=	Temperature (K)
$\boldsymbol{T}$	=	Set of temperature profiles (°C)
u_B	=	Feeding flowrate of B (gmol/h)
u_s	=	Substrate feeding flowrate (L/h)
V	=	Volume (L)
W	=	Static factor
w	=	Codified static factor
Y_{XS}	=	Yield factor, substrate to biomass (gr_b /gr_s)
Y_{PS}	=	Yield factor, substrate to product (gr_p /$gr_{s)}$
y_{exp}	=	Perturbed value of the response variable
y	=	Simulated value of the response variable
$\hat{y}$	=	Response Surface Model
y_{DRSM}	=	Dynamic Response Surface Model
$\boldsymbol{X}$	=	Set of dynamic factor profiles
$\boldsymbol{x}$	=	Set of codified dynamic factor profiles
X	=	Dynamic factor
x	=	Codified dynamic factor
x_1	=	Dynamic subfactor of factor x
x_2	=	Dynamic subfactor of factor x
x_3	=	Dynamic subfactor of factor x
$\mathbf{Z}$	=	Matrix of set of candidate samples

z	=	Submatrix of **Z** defined according to D-optimal criterion

Greek letters

μ_{max}	=	Maximum specific biomass growth rate (h^{-1})
ρ_{max}	=	Maximum specific production rate ($gr_p/gr_{b/}\,h$)
τ	=	Dimensionless time

Subscript

h	=	*h*-th time instants in which the DRSM is maximized
i	=	*i*-th variation level of a static factor
j	=	*j*-th static factor
m	=	*m*-th profile of a dynamic factor
n	=	*n*-th dynamic subfactor of a dynamic factor
q	=	Index of variation of the parameters according to the used model in RSM and DRSM

Appendix A

Optimization of a semi-batch reactor

In this appendix, the optimization procedure defined in Chapter 1 is applied to case study of a semi-batch reactor that is described in Section 2.2. By applying DoE and DoDE, it is possible to define the profile of an input variable that optimizes the value of the output. The construction of a DRSM, instead, allows studying the dynamic behavior of the output variable along the process.

A.1 Optimization of a semi-batch reactor using DoE methodology

The optimization procedure, described in Chapter 1, is applied in order to define an empirical model that describes the relationship between the feeding flowrate of B, which is the input factor, and a process output that is the concentration of product C. The model that has to be defined described both the linear and the quadratic effects of the input on the concentration of product C.

A.1.1 Selection of input factors

The feeding flowrate of B is the input variable that must be manipulated to define the optimum value of the concentration of product C, that is the output of the batch. The input variable has an intrinsically dynamic behaviour that tends to 0 at the end of the process and it is indicated with $u_B(\tau)$. The general equation, that describes the variation of the dynamic factor u_B in the dimensionless time, is the following:

$$u_B(\tau) = u_{0,B}(\tau) + \Delta u_B(\tau) x(\tau) \tag{A.1}$$

with:

$$u_{0,B}(\tau) = 30(1-\tau) \ ; \tag{A.2}$$

$$\Delta u_B(\tau) = 20(1-\tau) \ ; \tag{A.3}$$

$$x(\tau) = x_1 P_0(\tau) \ . \tag{A.4}$$

The profile $x(\tau)$ is the codified version of the dynamic factor $u_B(\tau)$, described by the combination of the dynamic subfactor x_1 with the first Legendre polynomial $P_0(\tau)$. Equations (A.2) and (A.3) define the range of variation of the factor $u_B(\tau)$: [0, 50]. The initial value $u_B(0)$ varies between 10 gmol/L and 50 gmol/L.
An operating constraint is introduced to consider the total amount of B to be fed along the process. It is defined using the equation (A.5).

$$B_T = \int_0^1 \left\{u_{0,B}(\tau) + \Delta u_B(\tau)x(\tau)\right\} d\tau \tag{A.5}$$

The reference value for B_T is set to 15 gmol while its variation is bounded between 10 gmol and 20 gmol. The total amount of B fed is parametrized by introducing a decision variable a as follows:

$$B_T = 15 + 5a \tag{A.6}$$

with the constraint $-1 \leq a \leq +1$.

In this case, in which DoE is going to applied, the variation of the feeding flowrate of B is described using only the subfactor x_1. This means that the feeding flowrate of B varies linearly with time. Therefore, x_1 can be considered as a static factor and rewritten as w_1. Equation (A.1) is simplified as follows:

$$u_B(\tau) = 30 + 20w_1 - (30 + 20w_1)\tau \tag{A.7}$$

The range of variation of the factor w_1 is bounded according to the operating constraint on the total amount of B fed in Equations (A.5) in this way:

$$-0.5 \leq w_1 \leq +0.5. \tag{A.8}$$

A.1.2 DoE and experimentation

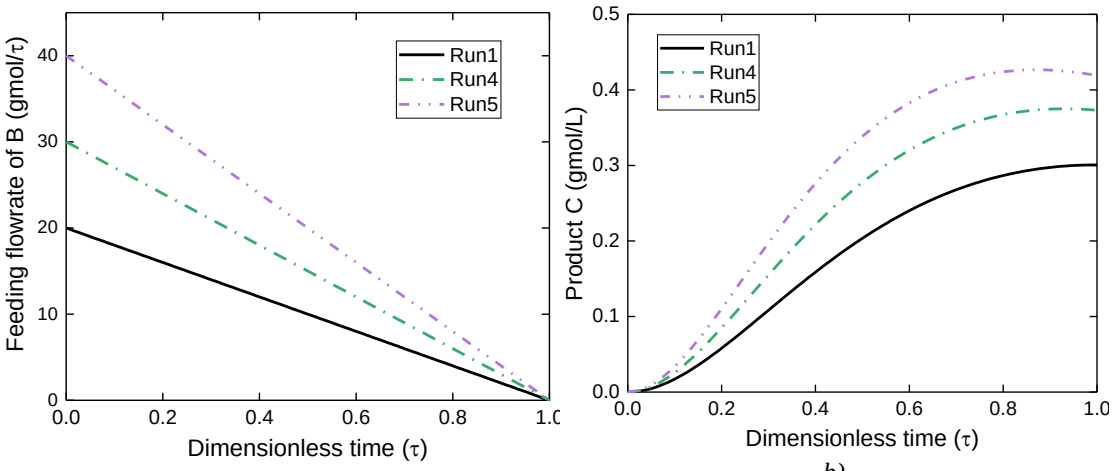
Different profiles of $u_B(\tau)$ are described by varying w_1 according to the design strategy chosen to characterize the experimental campaign. The latter depends on the empirical model that it has to be defined to describe the influence between the input variable, w_1 and the process output y, which is the product C concentration. Since the desired RSM has to include both the linear and the quadratic term, the experimental design defines at least three experiments, in which the input factor assumes respectively, the highest, the intermediate and the lowest value of its variability range.

Two repeated experiments in the central point of the experimental space are added. The five experiments, according to the variation range defined in (A.8), are described in Table A.1.

Table A.1 *Concentration of product C by applying the DOE methodology*

	Run1	**Run2**	**Run3**	**Run4**	**Run5**
$u_B(0)$ **(gmol/time)**	20	30	30	30	40
Coded variable w_1	-0.5	0	0	0	0.5
Simulated product C	0.3007	0.3732	0.3732	0.3732	0.4190
Measured product C	0.3015	0.3748	0.3757	0.3686	0.4187

The profiles of the feeding flowrate of B, according to the experiments in Table A.1, are shown in Figure A.1, with the related value of the process output, which is the concentration of product C.



a) *b)*

Figure A.1 *Profiles of a) the feeding flowrate of B and b) the concentration of C obtained performing a DoE described in Table A.1*

A.1.3 RSM and optimization

The RSM that has to be obtained and the relates the input factor w_1 with the response variable $\hat{y}$, namely, the concentration of product C at the end of the batch is the following:

$$\hat{y} = b_0 + b_1 w_1 + b_2 w_1^2 \tag{A.9}$$

The coefficients $\{b_0, b_1, b_2\}$ that define the RSM in (A.9) are calculated using a least square estimation method and are reported in Table A.2.

Table A.2 *Coefficients of the response surface model in Equation (A.9) that defines the quadratic relationship between the feeding flowrate of B and the concentration of product C*

Coefficients		
b_0	b_1	b_2
0.373 ±0.0097	0.1172± 0.0273	-0.0517 ± 0.0613

The R^2 adj is 0.9957. It is possible to define the value of the factor w_1 that guarantees the highest productivity. Since the trend of the productivity is a growing one, it reaches its maximum value at the upper limit of the variation range of factor w_1. The maximum predicted value is $\hat{y}$ = 0.4187 gmol/L, that is related to $w_{1,OPT}$= 0.5.
The confidence intervals of the parameters define a range of variation of the predicted maximum that is [0.3812 gmol/L, 0.4573 gmol/L]. Since the maximum real concentration of B is y =0.4190 gmol/L, the empirical model defined in Equation (A.9) is able to describe the process optimum.

A.2 Optimization of a semi-batch reactor using DoDE methodology

In this case, the objective is to define a RSM that consider both linear and non-linear effect (e.g. interactions and quadratic) of the input variable, which is the feeding flowrate of B, on the product C concentration.
Differently from the DoE described in Section (A.1), the feeding flowrate of B varies non- linearly with time.

A.2.1 Selection of input factors

The time dependence of the feeding flowrate of B can be non-linear if three dynamic subfactors: x_1, x_2 and x_3, are used. Equation (A.4), which describes the codified variation of feeding flowrate of B: $u_B(\tau)$, can be rewritten in this way:

$$x(\tau) = x_1 P_0(\tau) + x_2 P_1(\tau) + x_3 P_2(\tau)\,, \tag{A.10}$$

with

$$-1 \leq x(\tau) \leq +1 \tag{A.11}$$

where the $P_1(\tau)$ and $P_2(\tau)$ are, respectively, the first and second order Legendre polynomials. The subfactors combinations, that will be used for the characterization of the *M* profiles of feeding flowrate of B, have to satisfy the inequality in (A.12).

$$-1 \leq x_{1,m} \pm x_{2,m} \pm x_{3,m} \leq +1\ . \tag{A.12}$$

Since the experimental region shrinks to zero at the end of the batch, the constraint $x_m(1) = 0$, which yields $x_{1,m} + x_{2,m} + x_{3,m} = 0$, must be imposed. In this way, the number of independent subfactors is reduced from three to two independent subfactors $\{x_2, x_3\}$ since $x_{1,m}$ can be rewritten in this way:

$$x_{1,m} = -(x_{2,m} + x_{3,m}) \tag{A.13}$$

The constraint (A.11) defines the following variation range of the subfactors:

$$-0.5 \leq x_2 \leq +0.5 \tag{A.14}$$

$$-0.5 \leq x_3 \leq +0.5 \tag{A.15}$$

$$-0.5 \leq x_{2,m} + x_{3,m} \leq +0.5 \tag{A.16}$$

The operating constraint in Equation (A.5) has to be considered. In particular, it is necessary to consider the parameter a, which parametrizes the total amount of B fed in the reactor, considering that it depends on the value of x_2 and x_3 as follows:

$$a = -2\left(\frac{4x_2}{3} + x_3\right) \tag{A.17}$$

Since the value of a has to be always inside the range [-1, +1], the following constraint on the values of the dynamic subfactors has to be imposed:

$$-1.5 \leq 4x_2 + 3x_3 \leq +1.5 \tag{A.18}$$

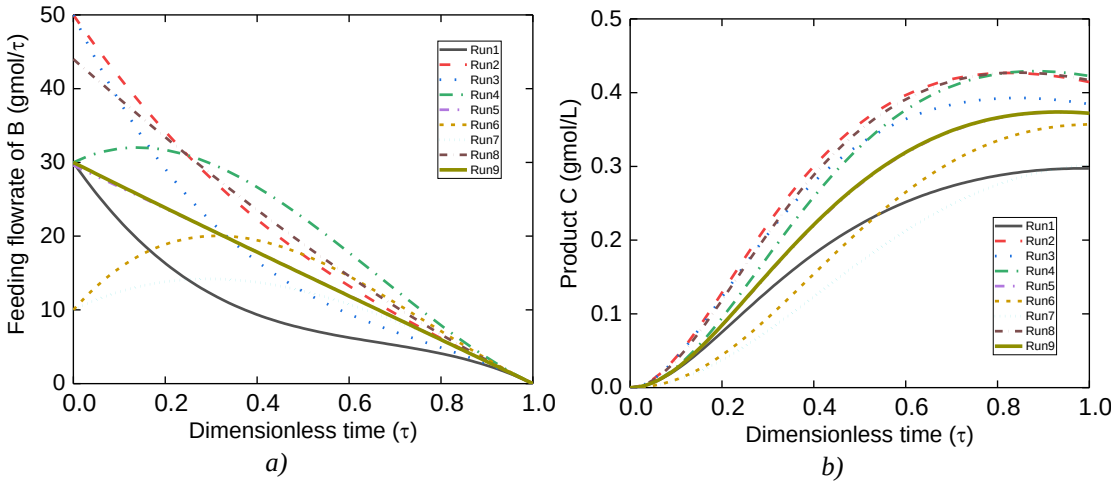
A.2.2 DoDE and experimentation

Since there are two independent subfactors x_2 and x_3, a full quadratic RSM requires the estimation of six parameters. The experiments are characterized by different combinations of the two dynamic subfactors that satisfy the constraints in Subsection A.2.1. A d-optimal design is the design strategy commonly used in the case in which the design space is bounded by constraints. To estimate a quadratic RSM thirteen experiments are required, of which 4 are replicates. In Table A.3, the combinations of the dynamic subfactors are reported. In the last column of the same table the results of the experiments, in terms of product C concentration at the end of the batch are reported.

Table A.3 *Feeding flowrate of B profiles defined according to a D-optimal design applied to three dynamic subfactors and the related value of the product C concentration at the end of the batch.*

Run	Coded Variables				Simulated product C (gmol/L)	Measured product C (gmol/L)
	x_1	x_2	x_3	a		
1	-0.5	0	0.5	-1	0.2972	0.2947
2	0.33	-0.5	0.17	0.993	0.4144	0.4144
3	0	-0.5	0.5	0.333	0.3844	0.3859
4	0.5	0	-0.5	1	0.4225	0.4230
5	-0.01	0.01	0	-0.027	0.3717	0.3682
6	0	0.5	-0.5	-0.333	0.3570	0.3560
7	-0.33	0.5	-0.17	-0.993	0.3019	0.3015
8	0.38	-0.35	-0.03	0.993	0.4168	0.4206
9	-0.01	0	0.01	-0.020	0.3720	0.3737
10	-0.01	0	0.01	-0.020	0.3720	0.3719
11	-0.01	0	0.01	-0.020	0.3720	0.3744
12	-0.01	0	0.01	-0.020	0.3720	0.3701
13	-0.01	0	0.01	-0.020	0.3720	0.3714

The profiles of the feeding flowrate of B described by the dynamic subfactors in Table A.3 and the respective variation of the response, namely, the concentration of product C are shown in Figure A.2.



a) *b)*

Figure A.2 *Profiles of: a) the feeding flowrate of B and b) the concentration of product C obtained performing a D-optimal design as the one in Table A.3.*

A.2.3 RSM and optimization

The RSM that describes the influence of the non-linear profile of the feeding flowrate of B on the value of the product C concentration at the end of the batch has the following formulation:

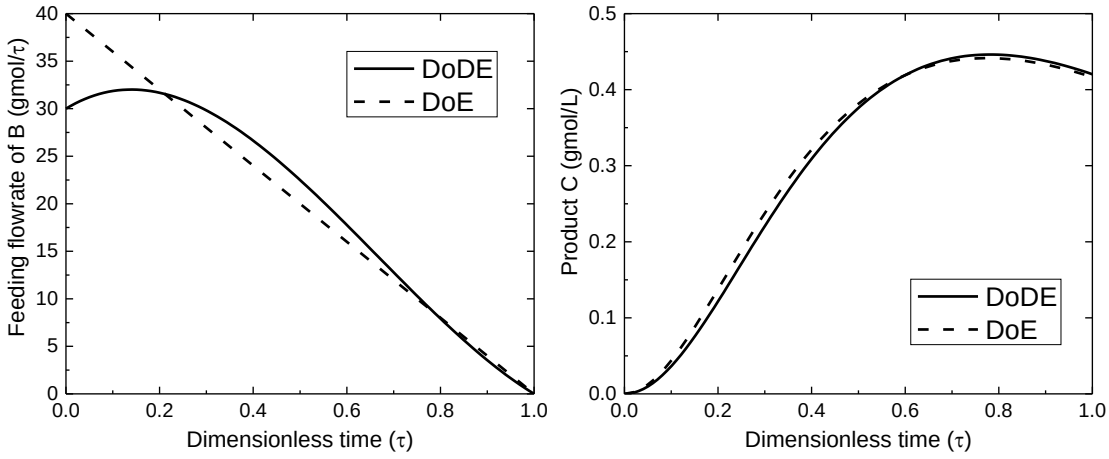
$$\hat{y} = b_0 + b_1 x_2 + b_2 x_3 + b_{12} x_2 x_3 + b_{11} {x_2}^2 + b_{22} {x_3}^2 \,. \qquad \text{(A.19)}$$

The coefficients, calculated using a least square algorithm for the estimation of the coefficients in the multiple linear regression empirical model, are reported in Table A.4.

Table A.4 *Coefficients of response surface model in Equation (A.19) that defines the full quadratic relationship between two independent subfactors and the value of product C concentration*

Coefficients					
b_0	b_1	b_2	b_{12}	b_{11}	b_{22}
0.3731±0.02	0.1585±0.0062	-0.128±0.0063	-0.146 ±0.0352	-0.1 ±0.024	-0.056±0.0166

The R^2-adj for the above regression is 1. DoDE has been applied in order to define the profile of feeding flowrate of B that provides the maximum concentration of product C at the end of the batch. The maximum predicted value of the product C, is 0.4235 gmol/L, which corresponds to a feeding profile defined by $x_{1,OPT}$= 0.4999, $x_{2,OPT}$=0.0001 and $x_{3,OPT}$= -0.4999. The maximum real conversion is 0.4225 gmol/L that is inside the range of variability [0.4025 gmol/L, 0.4445 gmol/L] calculated considering the coefficients in Table A.4 at the ends of their confidence interval. This result of the optimization can be compared with the one obtained performing the DoE in Figure A.3.



a) *b)*

Figure A.3 *Comparison among the optimum calculated using DoE and the optimum obtained by DoDE: a) feeding flowrate of B and b) product C concentration.*

In Figure A.3, it is possible to notice that the difference between the optimum concentration obtained by DoDE and the one obtained by DoE is not significant as it is less than the value of the measurement error of 1% imposed in the experimental campaign.

A.3 Definition of the optimal batch duration

Since the profiles of the product C concentration show a parabolic trend with time, it might be useful to define the optimal batch time. In this Section, DoE and DoDE are applied considering the batch time as a static factor. This case studies has been developed in Troup and Georgakis (2013).
The batch time t_b is parametrized through the decision variable w_2 as follow:

$$t_b = 1 + 0.5w_2 \tag{A.20}$$

where w_2 is the codified factor that describes the variation of the batch duration.
The w_2 is bounded by the following constraint:

$$-1 \leq w_2 \leq +1 \tag{A.21}$$

According to Equations (A.20) and (A.21), the batch time varies between 0.5 h and 1.5 h. In the application of DoE and DoDE, the static factor w_2 is combined with the dynamic factor $x(t)$, which describes the feeding flowrate of B.

A.3.1 Optimization of a semi-batch reactor with variable batch duration using DoE methodology

The calculation of the optimal batch duration requires the definition of an empirical model that described the linear, interaction and quadratic effects of the inputs factor on the output factor, which is the product C concentration at the end of the batch.

A.3.1.1 Selection of input factors

In DoE, two factors are considered, one in the static factor w_2 that describes the variable batch duration, the other is the factor w_1. The latter defines the linear variation of the feeding flowrate of B as defined in Subsection A.1.1. The variation range of the two factors are in Equations (A.8) and (A.21).

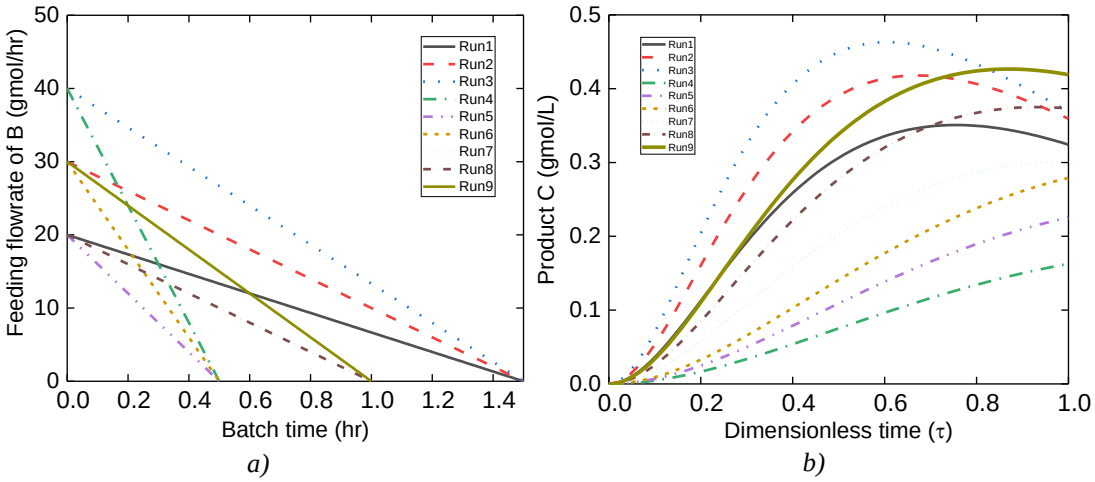
A.3.1.2 DoE and experimentation

The definition of a full quadratic RSM with two factors w_1 and w_2 requires a complete collection of information that can be provided by a three-levels FFD. It is made by twelve experiments among which three are replicas. The CCD cannot be performed since the two factors vary among different levels. The combinations of the factors and the related value of the product C concentration of at the end of the batch are reported in Table A.5.

Table A.5 *Full factorial design applied to the case of two input factors: feeding flowrate and batch time duration and the that vary among three levels and the related value of the product C concentration at the end of the batch*

Run	Coded Variables		Simulated product C (gmol/L)	Measured product C (gmol/L)
	w_2	w_1		
1	1	-0.5	0.3242	0.3258
2	1	0	0.3593	0.3590
3	1	0.5	0.3687	0.3692
4	-1	-0.5	0.1627	0.1632
5	-1	0	0.2256	0.2255
6	-1	0.5	0.2792	0.2799
7	0	-0.5	0.3007	0.2988
8	0	0	0.3732	0.3753
9	0	0.5	0.4190	0.4170
10	0	0	0.3732	0.3698
11	0	0.5	0.4190	0.4219
12	-1	-0.5	0.1627	0.1627

The profiles of the feeding flowrate of B and of the product C concentration, according to design plan in Table A.5, are shown in Figure A.4.



a) *b)*

Figure A.4 *Profiles of: a) the feeding flowrate of B and b) the concentration of product C obtained by performing a 3-levels FFD in the case of DoE*

A.3.1.3 RSM and optimization

The effects of the two factors w_1 and w_2 on the product C concentration (y) can be described by the following RSM:

$$\hat{y} = b_0 + b_1 w_2 + b_2 w_1 + b_{12} w_2 w_1 + b_{11} {w_2}^2 + b_{22} {w_1}^2. \quad \text{(A.22)}$$

Using the experimental data in Table A.5, the coefficients of Equation (A.22) are estimated and reported in Table A.6.

Table A.6 *Coefficients of response surface model of Equation (A.22) that defines the quadratic effects of the two factors on the product C concentration*

Coefficients					
b_0	b_1	b_2	b_{12}	b_{11}	b_{22}
0.372± 0.015	0.0631± 0.010	0.093± 0.0192	-0.0351± 0.0242	-0.08± 0.016	-0.035 ± 0.0647

The R^2-adj for the above regression is 0.99. Equation (A.22) allows calculating the maximum predicted value of the concentration of product C that is equal to 0.4172 gmol/L. This corresponds to the values of $w_{1,OPT}$=0.5 and $w_{2,OPT}$ =0.2854. The optimum value that described the feeding flowrate of B is the same as the one obtained in the case of fixed batch time in Section A.1. However, the optimal batch time is t_b=1.14 h which is 14% longer than the fixed duration. The variability range of the RSM is [0.3747, 0.4697], which includes the real maximum value of product C that is y =0.4168 gmol/L. This value is lower than the one calculated in DoE with fixed batch length, but the difference is smaller than the measurement error equal to 0.005. This means that, either keeping the batch time at 1 hour, or increasing it to 1.14 hours, the maximum value of the concentration of C obtainable does not change.

A.3.2 Optimization of a semi-batch reactor with variable batch duration using DoDE methodology

In this subsection the empirical model that describes the linear, interaction and quadratic effects of the input factors on the output factor, which is the product C concentration at the end of the batch, is defined. Differently from DoE, the feeding flowrate profiles considered are non-linear.

A.3.2.1 Selection of input factors

In this case, three dynamic subfactors $\{x_1, x_2, x_3\}$ are used to characterize the non-linear profiles of the feeding flowrate of B. These are unified to the static factor w_2 that describes the variation of the batch duration according to Equation (A.20). According to the proceeding in Subsection A.2.1, only two of the three dynamic subfactors are independent, namely, (x_2, x_3).
The variability range of the two dynamic subfactors are the same of Equations (A.14)-(A.18).

A.3.2.2 DoDE and experimentation

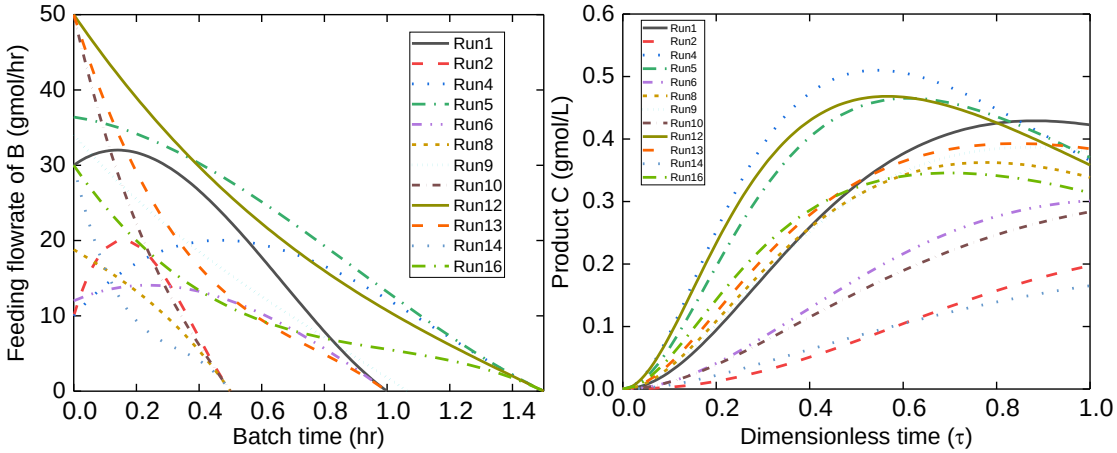
A design strategy is applied to define the combinations of the two dynamic subfactors x_2, x_3 and the static factor w_2. To obtain a full quadratic RSM with three independent factors, at least 10 experiments are required. In this case, a D-optimal design is

performed, which define sixteen experiments among which three are replicates. These are suggested in Table A.7.

Table A.7 *D-optimal design applied to the case of three input factors: feeding flowrate and batch time duration and the related value of the product C concentration at the end of the batch*

Run	Coded Variables					Simulated product C (gmol/L)	Measured product C (gmol/L)
	w_2	x_1	x_2	x_3	a		
1	0	0.5	0	-0.5	1	0.4225	0.4264
2	-1	0	0.5	-0.5	-0.333	0.1973	0.1983
3	-1	0	0.5	-0.5	-0.333	0.1973	0.1982
4	1	0	0.5	-0.5	-0.333	0.3654	0.3650
5	1	0.44	-0.16	-0.28	0.987	0.3703	0.3706
6	0	-0.35	0.45	-0.10	-1	0.3010	0.3014
7	-1	0.01	0	-0.01	0.020	0.2267	0.2268
8	1	-0.32	0.28	0.04	-0.827	0.3386	0.3372
9	0.13	0	-0.1	0.1	0.067	0.3803	0.3796
10	-1	0.33	-0.5	0.17	0.993	0.2835	0.2833
11	-1	0.33	-0.5	0.17	0.993	0.2835	0.2828
12	1	0.33	-0.5	0.17	0.993	0.3583	0.3598
13	0	0	-0.5	0.5	0.333	0.3844	0.3892
14	-1	-0.5	0	0.5	-1	0.1652	0.1641
15	-1	-0.5	0	0.5	-1	0.1652	0.1653
16	1	-0.5	0	0.5	-1	0.3139	0.3116

The profiles of both the feeding flowrate of B defined by the values of factors in Table A.7 and the concentration of the product C are reported in Figure A.5.



a) *b)*

Figure A.5 *Profiles of a) the feeding flowrate of B and b) the concentration of product C obtained performing a d-optimal design described in Table A.7.*

A.3.2.3 RSM and optimization

The RSM that has to be obtained is the following:

$$\hat{y} = b_0 + b_1 w_2 + b_2 x_2 + b_3 x_3 + b_{12} w_2 x_2 + b_{13} w_2 x_3 + b_{23} x_2 x_3 + b_{11} {w_2}^2 + b_{22} {x_2}^2 + b_{33} {x_3}^2 \quad \text{(A.23)}$$

The RSM describes the influence of the batch duration w_2 and the feeding flowrate of B, described by x_2 and x_3 on the final product C concentration. The coefficients of Equation (A.23) in Table A.8, are calculated using a least square algorithm.

Table A.8 *Coefficients of response surface model related to Equation (A.23) that defines the influence of the batch duration and of the non-linear profile of feeding flowrate of B on the product C concentration*

Coefficients				
b_0	b_1	b_2	b_3	b_{12}
0.3662± 0.0106	0.0642 ±0.0072	-0.1276 ±0.0172	-0.101 ± 0.0154	0.068 ± 0.0184
b_{13}	b_{23}	b_{11}	b_{22}	b_{33}
0.0264 ± 0.0216	-0.1243 ± 0.069	-0.0764 ±0.0099	-0.4453 ±0.2015	0.037 ±0.1506

Equation (A.23) allows defining the optimum combination of the values of the dynamic and static factors as the one that provides the higher concentration of C at the end of the batch. The optimum profile is described by the values of subfactors x_1,x_2, x_3 and w_2 that are respectively 0.5, 0, -0.5 and 0.334. The maximum concentration predicted is 0.4346 gmol/L. The optimum value of the batch time is equal to 1.16 hr. In these conditions, the concentration of product C is equal to 0.4204 gmol/L.
The difference between this value and the one calculated in the case of fixed batch time is not significant. This means that the variation of the batch time from 1 hour to 1.16 hours is not necessary and the lowest one can be considered the optimal.

A.4 Definition of a dynamic response surface model

The data collected during DoDE in terms of variation of the concentration of product C, are used for the construction of DRSM. It is considered to use the results obtained in the DoDE with fixed batch. (Figure A.2*b*). DRSM has the aim to approximate the relationship between the feeding flowrate of B and the concentration of product C considering also the time variation of the latter. The construction of DRSM consists in the definition of the Legendre polynomial degree *R-1* that approximates the *KM* experimental points. Using the *M* profiles performed through DoDE, *K* measurements of the output variable are collected from each experiment. A series of DRSM,

characterized by different values of *R* and *K*, has been taken into account. The lowest $\widehat{SS}_{un}$ value, identify the most appropriate *R* and *K*. In Table A.9, the values of $\widehat{SS}_{un}$ that characterize the different DRSM, are reported.

Table A.9 *Values of* $\widehat{SS}_{un}$ *related to dynamic response surface model with variable K and R calculated in the case of semi-batch*

R	K=Number of experiments for each profile							
	3	**4**	**5**	**6**	**7**	**8**	**9**	**10**
2	0.0443	0.0304	0.0298	0.0293	0.029	0.0286	0.0285	0.0258
3		0.0071	0.0051	0.0052	0.005	0.0053	0.0045	0.004
4			0.0011	0.0013	0.0022	0.0019	0.0019	0.0015
5				1.40E-04	4.39E-04	3.78E-04	1.92E-04	4.06E-04
6					4.48E-04	4.03E-04	1.44E-04	1.95E-04
7						5.39E-04	3.63E-05	5.98E-05
8							2.23E-04	2.83E-04
9								3.41E-04

In Table A.9, it is possible to see that for *K*=8, 9 and 10 there are several values of *R* (= 4, 5, 6, 7, 8) that are characterized by the lowest values of $\widehat{SS}_{un}$, and thus are candidates for appropriate DRSM. The ratio F_0 and the *p*-value are calculated using Equation (1.29) to verify that the model adequately represents the non-random variability of the data. If the *p*-value is ≤ 0.95 the *K* time-resolved measurements demonstrate to be sufficient. The *p*-value for each combination of *R* and *K* are reported in Table A.10.

Table A.10 *P-values related to dynamic response surface model with variable K and R calculated in the case of semi-batch*

R	K=Number of experiments for each profile							
	3	**4**	**5**	**6**	**7**	**8**	**9**	**10**
2	1	1	1	1	1	1	1	1
3		1	1	1	1	1	1	1
4			1	1	1	1	1	1
5				1.32E-04	0.3706	0.0281	1.64E-07	2.76E-04
6					0.24	0.0174	1.18E-07	4.81E-07
7						0.336	5.86E-23	5.746E-06
8							0.0021	1.62E-04
9								4.53E-04

An acceptable DRSM is obtained with *K*=10 and using a sixth order polynomial. The DRSM, made by 42 coefficients calculated via stepwise regression, is reported in Equation (A.24).

$$
\begin{aligned}
y_{DRSM}(\tau) = {} & g_{0,1}P_0 + g_{0,2}P_1 + g_{0,3}P_2 + g_{0,4}P_3 + g_{0,5}P_4 + g_{0,6}P_5 + g_{0,7}P_6 \\
& + \left(g_{1,1}P_0 + g_{1,2}P_1 + g_{1,3}P_2 + g_{1,4}P_3 + g_{1,5}P_4 + g_{1,6}P_5 + g_{1,7}P_6\right)x_2 \\
& + \left(g_{2,1}P_0 + g_{2,2}P_1 + g_{2,3}P_2 + g_{2,4}P_3 + g_{2,5}P_4 + g_{2,6}P_5 + g_{2,7}P_6\right)x_3 \\
& + \left(g_{12,1}P_0 + g_{12,2}P_1 + g_{12,3}P_2 + g_{12,4}P_3 + g_{12,5}P_4 + g_{12,6}P_5 + g_{12,7}P_6\right)x_2x_3 \\
& + \left(g_{11,1}P_0 + g_{11,2}P_1 + g_{11,3}P_2 + g_{11,4}P_3 + g_{11,5}P_4 + g_{11,6}P_5 + g_{11,7}P_6\right){x_2}^2 \\
& + \left(g_{22,1}P_0 + g_{22,2}P_1 + g_{22,3}P_2 + g_{22,4}P_3 + g_{22,5}P_4 + g_{22,6}P_5 + g_{22,7}P_6\right){x_3}^2
\end{aligned}
\tag{A.24}
$$

The coefficients of the DRSM in (A.24) are reported in Table (A.11).

Table A.11 *Coefficients of DRSM in Equation (A.24) that defines the influence of feeding flowrate of B and batch duration on the product C concentration*

Coefficients				
g0,1	g0,2	g0,3	g0,4	g0,5
0.2331	0.2131±0.0002	-0.0676±0.002	-0.0248 ±0.0026	0.0183 ±0.0024
g0,6	g0,7	g1,1	g1,2	g1,3
-0.0056 ±0.0028	0	-0.1589 ±0.0039	-0.0815 ±0.0065	0.1094 ±0.008
g1,4	g1,5	g1,6	g1,7	g2,1
0	-0.0245 ±0.0086	0.0099 ±0.0070	0	-0.0828 ±0.004
g2,2	g2,3	g2,4	g2,5	g2,6
-0.0846 ±0.0063	0.0317 ±0.0081	0.0260 ±0.0067	-0.0101 ±0.0088	0
g2,7	g12,1	g12,2	g12,3	g12,3
0	0	-0.0170 ±0.0163	0	0

In Table A.11, it is possible to notice that the stepwise regression has eliminated the coefficients that are not significant. In particular, all the parameters that describe the quadratic dependence of the dynamic subfactor have been deleted, this means that the quadratic influence of the two subfactors on the profile of the reactant A conversion is not significant. As explained in Subsection 1.1.4.2, it is possible to calculate the distance of the predicted profiles from the experimental ones using Equation (1.27). The value of SS_{un} is 1.2113e-04. The DRSM predictions are plotted against the experimental data in Figure A.6.

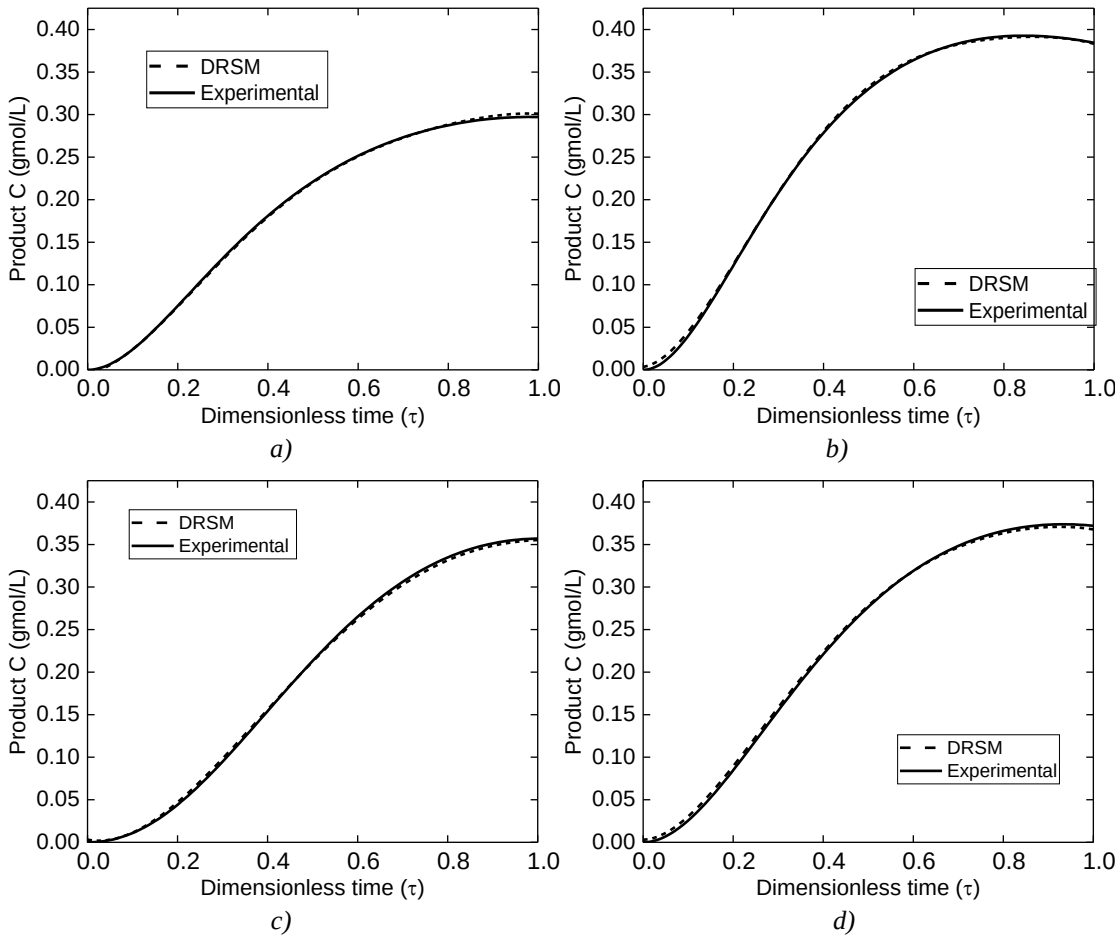


Figure A.6 *In these figures the experimental profiles are compared to the DRSM predictions of concentration of C profiles within the experimental design in the case of a) x_1=-0.5, x_2=0 and x_3=0.5, b) x_1=0, x_2=-0.5 and x_3=0.5, c) x_1=0, x_2=0.5 and x_3=-0.5 and d) x_1=-0.01, x_2=0 and x_3=0.01.*

By comparing the DRSM calculated to the one in Klebanov and Georgakis (2016), it is possible to see how the choice is very similar, since in the article the DRSM is approximate with a fifth order polynomial, instead of sixth.

A.4.1 Optimization using dynamic response surface model

Using the DRSM, it is possible to represent the profiles of the concentration of product C for different combinations of the independent subfactors x_2 and x_3. The profiles are shown in Figure A.7.

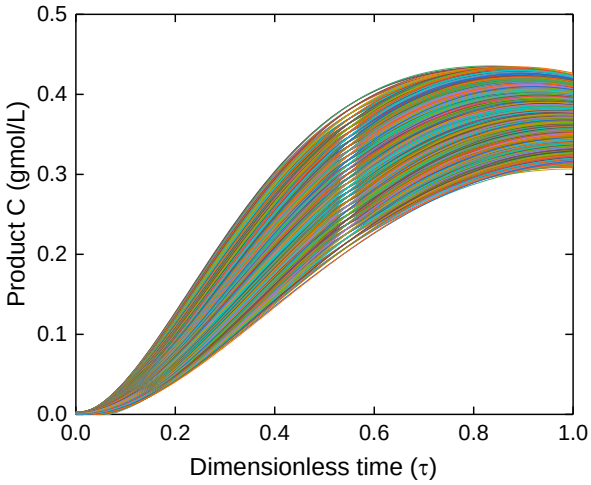
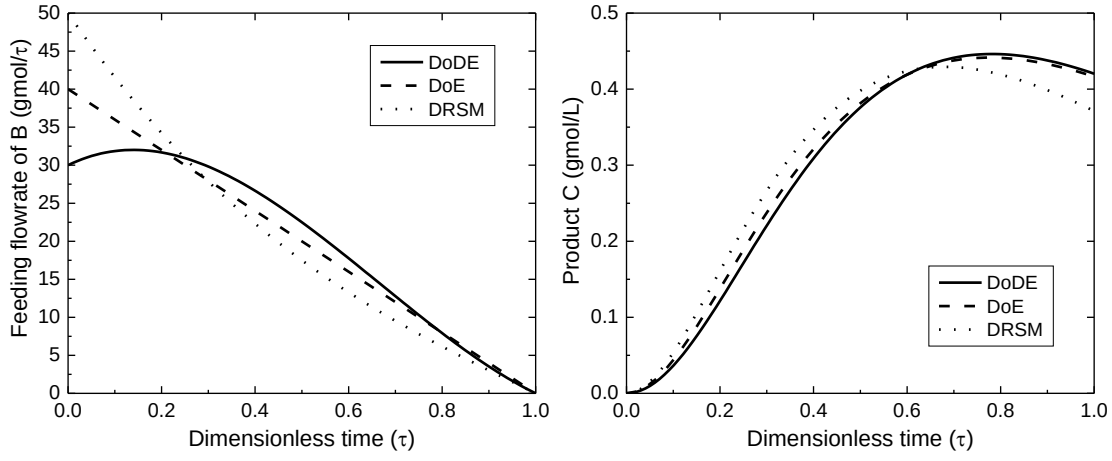


Figure A.7 *Profiles of product C concentration along time obtained using the DRSM in (A.24)*

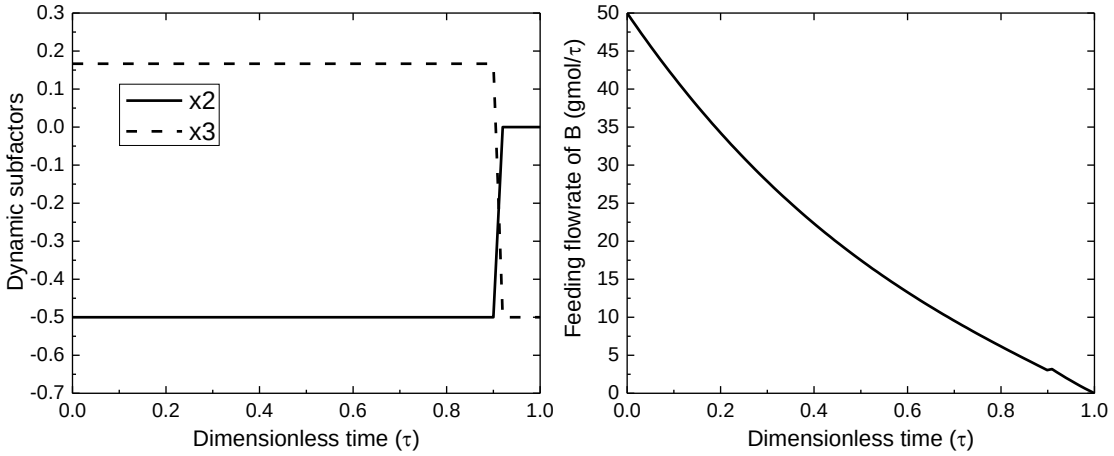
The DRSM allows defining the values of the dynamic subfactors x_2 and x_3 that determine the profile of the concentration of product C whose integral, in the dimensionless time τ, is greater than all the others. The values of the subfactors, calculated using (1.33), are x_2 =-0.5 and x_3= +0.1667. The input profile, defined in this way, is different from those obtained using DoE and DoDE. In Figure (A.8), the optimum profiles of the feeding flowrate of B and the related profiles of the concentration of product C, obtained using the three different methods, are compared. In this case, the optimum profile calculated using DRSM is different from the one defined by DoDE, except in the last part of the process. In fact, the end-point value of the profile of the product concentration is the same.



a) *b)*

Figure A.8 *Comparison among the* **o***ptimum calculated using DoE, DoDE and by maximizing the integral of the DRSM: a) feeding flowrate of B and b) concentration of product C*

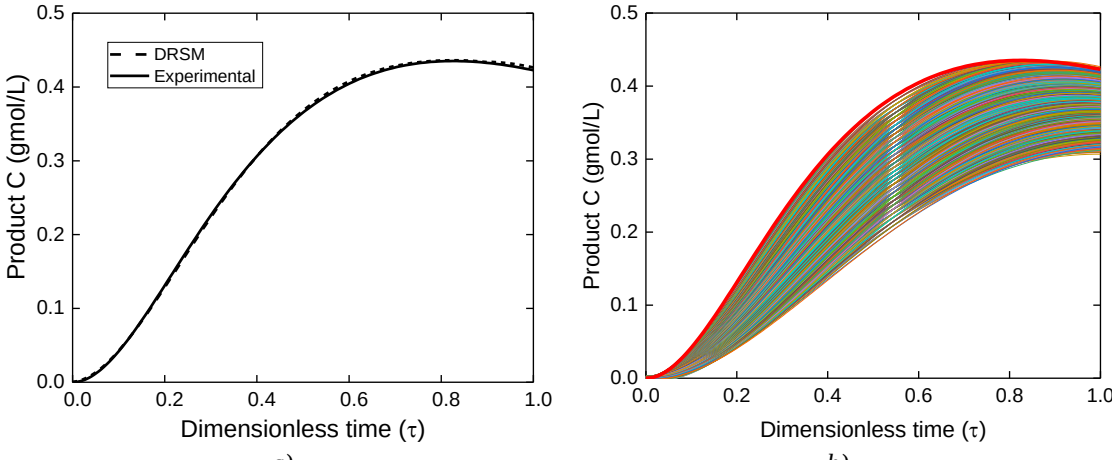
However, the profiles in Figure A.7 suggest that to increase the value of the concentration of product C, the variation of the values of subfactors x_2 and x_3 along the process must be considered. In fact, there is no profile that shows, in each point, a higher trend than all the others. Equation (1.34) allows calculating the combinations of values x_2 and x_3 that maximize the concentration of C in a series of time laps. The profiles of the dynamic subfactors and the related feeding flowrate of B are reported in Figure A.9.



a) *b)*

Figure A.9 *Optimal profile of a) the dynamic subfactors and b) the feeding flowrate of B calculated maximizing the DRSM in a series of time instants.*

The concentration profile of product C, in case of the variation of the feeding flowrate of B as in Figure A.9*b*, is reported in Figure A.10*a*.



a) *b)*

Figure A.10 *Comparison between the real profile of the product C concentration according to the optimal feeding flowrate of B in Figure A.9b and a) the predicted profile and b) the profiles of the product C concentration calculated using the DRSM*

In Figure A.10*b*, the predicted concentration profile, related to the input profile in Figure A.9*b*, and the simulated one are compared.

The *RSMe,* calculated using (1.35), is 0.0057, which is smaller than the measurements error equal to 0.02. This means that the DRSM is accurate in describing the output profile in the case of dynamic behavior of the dynamic subfactors. Furthermore, the simulated profile, in red in Figure A.10*b*, corresponds to the request to maximize the output variable profile in the whole process.

A.5 Conclusions regarding the optimization of a semi-batch reactor

The results obtained by applying the optimization procedure, both in the case of fixed batch duration and in the variable one, are summarized below.

- In the case of fixed batch time ($t_b = 1$ hour), DoDE and DoE leads approximately to the same result in term of maximum concentration of product C at the end of the batch. This means that either varying the feeding flowrate of B linearly or non-linearly the optimum value of the process output is the same.
- The use of the DRSM to maximize the integral value of the profile of product C concentration leads to a lower value at the end of the batch if compared to DoDE or DoE. The time variation of the dynamic subfactors, instead, allows defining a profile that is higher than all the others in the whole process.
- In the case of variable batch duration, DoDE leads to determine a higher value of the product C concentration at the end of the batch if compared to the one achievable using DoE. However, the optimum value of the process output calculated is not better than the one defined in the case of fixed batch time.

The operating optimum is characterized by the following values of the factors: $x_{1,OPT} = 0.4999$, $x_{2,OPT} = 0.0001$ and $x_{3,OPT} = -0.4999$, for a batch time $t_b = 1$ hour. The corresponding final product C concentration is 0.4225 gmol/L.

References

Bajpai, R. K.; Reuss, M. (1980) A Mechanistic Model for Penicillin Production. J. *Chem. Technol. Biotechnol*, **30**(6), 332−344.

Box, G., Draper,N.R. (1987) Empirical Model-Building and Response Surface; John Wiley Sons, Incorporated: New York; p 688

Chatterjee, S., and Hadi A. S. (1986) Influential Observations, High Leverage Points, and Outliers in Linear Regression. *Statistical Science*., **1**, 379–416.

Draper, N. R., and Smith. H. (1998). *Applied Regression Analysis*. Hoboken, NJ: Wiley-Interscience, pp. 307–312.

Fischer A.(1935), The Design of Experiments, 9th Edition; Hafner Publishing Company, NJ

Georgakis, C. (2013) Design of Dynamic Experiments: A Data-Driven Methodology for the Optimization of Time-Varying Processes. *Ind. Eng. Chem. Res*., **52**, 12369−12382

Klebanov, N. and Georgakis, C. (2016) Dynamic Response Surface Models: A Data-Driven Approach for the Analysis of Time-Varying Process Outputs. *Ind. Eng. Chem. Res*. **55**, 4022-4034

Myers, R. H.; Montgomery, C.C. Anderson-Cook, C.M. (2009) *Response Surface methodology: process and product optimization using designed experiments*, 3rd ed.; Wiley: Hoboken, N,J., p xiii, 680

Montgomery, D. (2013) *Design and Analysis of Experiments*, 8th Edition; Wiley: Hoboken, NJ

Riascos, C. A. M.; Pinto, J. M. (2004) Optimal control of bioreactors: a simultaneous approach for complex systems. *Chem. Eng. J*. 2004, **99**, 23−34.

Trutna, L.; Spagon, P.M.; Del Castillo, E.; Moore T.; Hartley, S.; Hurwitz, A. (2012) E-Handbook of Statistical Methods. Retrieved from https://www.itl.nist.gov/div898/handbook/index.htm. Last checked in 30th March 2018

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