UNIVERSITÀ 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

AN OPTIMISATION FRAMEWORK TO DEFINE MAXIMUM PARAMETRIC UNCERTAINTY GUARANTEEING ACCEPTABLE MODEL FIDELITY

Relatore: Prof. Fabrizio Bezzo
Correlatrice: Ing. Geremia Margherita

Laureando: GIULIO CISCO

ANNO ACCADEMICO 2022-2023

Abstract

Quantitative models are useful tools to support and accelerate the development of pharmaceutical manufacturing processes. The assessment of the predictive capability of such models is fundamental to enhance their usage in a systematic way. This is particularly important in a highly regulated sector such as the pharmaceutical industry, where model uncertainty evaluation may have to be quantified and filed if model results are to be included as part of a regulatory submission. In this work, an optimization framework has been developed, which defines the maximum allowable parametric uncertainty capable of guaranteeing a satisfactory prediction of both key performance indicators and critical quality attributes, which we will generically call key indicators. The proposed approach aims at quantifying the impact of model parameters with respect to a key indicator of interest, and assesses whether a statistically satisfactory estimation for all parameters is always necessary in order to meet the required prediction fidelity. The validity of the methodology is examined through the implementation on a direct compression process. Results demonstrate the possibility to implement a procedure where parametric uncertainty is explicitly optimised in order to target a pre-set fidelity performance of model predictions.

Riassunto esteso

Nell'industria farmaceutica è sempre più diffuso l'utilizzo di modelli matematici al fine di descrivere le unità di processo che caratterizzano il settore. Tali modelli ampliano le conoscenze in merito ai sistemi analizzati e consentono di accelerarne ed ottimizzarne lo sviluppo facendo fronte all'elevato consumo di risorse e tempo legato alle campagne sperimentali. Tuttavia, il completo innesto di modelli quantitativi nell'ambiente farmaceutico è ad oggi limitato dalla mancanza di una procedura rigorosa che permetta di quantificare l'effettiva capacità di tali modelli di riprodurre in modo fedele la realtà. Un possibile approccio è stato recentemente proposto da Geremia *et al.* (2023) dove, attraverso l'utilizzo di una metodologia basata su tecniche statistiche multivariate, si stabilisce in quali circostanze si giunga alla predizione delle variabili di interesse entro un certo range di tolleranza prestabilito. Questo avviene non solo evidenziando l'incertezza riguardante la stima dei parametri, ma anche identificando quali siano i parametri di maggior influenza rispetto alle variabili analizzate. Tuttavia, l'approccio proposto in Geremia *et al.* (2023) prevede una linearizzazione del modello matematico e questo potrebbe rappresentare un limite all'efficacia del metodo. Per tale motivo, in questo lavoro si propone lo sviluppo e l'applicazione di una nuova metodologia in grado di mantenere fede alla forma originale del modello. L'interesse è sempre quello di stabilire quanto la predizione di una variabile dipenda dai parametri e dalla loro incertezza, servendosi però di tecniche diverse da quelle utilizzate in Geremia *et al.* (2023) per il conseguimento dell'obiettivo. La metodologia è stata interamente sviluppata utilizzando il software gPROMS® (Siemens Process Systems Engineering) sfruttando sia la possibilità del programma di rappresentare modelli dinamici, sia la vasta gamma di strumenti che questo mette a disposizione per l'analisi di sistemi. Il metodo sviluppato prevede una preliminare analisi di sensitività globale del modello al fine di classificare l'importanza dei parametri nei confronti della variabile di interesse. In seguito, cercando di migliorare la precisione con cui vengono stimati i parametri più impattanti, viene realizzata una pianificazione degli esperimenti basata su modello, dal quale deriva poi la possibilità di simulare un esperimento in silico che verrà utilizzato per una successiva stima parametrica. Infine, viene compiuta un'ottimizzazione sulle incertezze dei parametri stimati andando a determinare l'incertezza massima consentita per ciascun parametro al fine di mantenere la variabile di interesse all'interno di un intervallo desiderato.

Per validare la metodologia descritta, questa è stata testata sullo stesso processo di compressione diretta utilizzato in Geremia *et al.* (2023), favorendo così un successivo confronto tra i due metodi. Il sistema si compone di un'unità operativa di compressione delle polveri, un'unità di disintegrazione delle compresse e un'unità di dissoluzione *in vitro* dei prodotti della disintegrazione; i modelli matematici di queste fasi del processo vantano

rispettivamente sette, cinque e due parametri oggetto di stima. I risultati ottenuti per le unità di disintegrazione e di dissoluzione mostrano un'affinità tra le due metodologie; diversamente, per l'unità di compressione, il raggiungimento di un'accettabile fiducia del modello attraverso l'approccio proposto in questo progetto richiede tre esperimenti aggiuntivi rispetto al metodo basato su tecniche statistiche multivariate. La ragione è legata al fatto che l'approccio proposto non tiene conto del fatto che l'incertezza raggiunta nella stima di alcuni parametri potrebbe essere inferiore al limite massimo consentito e quindi compensare l'effetto di incertezza di altri parametri per i quali potrebbe essere accettabile un valore di incertezza superiore rispetto a quello ottenuto dall'ottimizzatore. Per far fronte a questo problema la metodologia è stata affinata attraverso un metodo adattativo che tiene conto dell'effettiva incertezza parametrica ottenuta dopo ciascuna stima. I risultati conseguiti a seguito dell'implementazione di questa metodologia si dimostrano soddisfacenti: per tutti e tre i modelli del processo a compressione diretta il numero di esperimenti necessario al raggiungimento della condizione di stop è pari a quello ottenuto attraverso il metodo proposto in Geremia *et al.* (2023). Questo suggerisce, inoltre, che almeno per il caso studio trattato, la linearizzazione del modello matematico non costituisce un limite.

Table of contents

List of symbols

Acronyms

API	=	active pharmaceutical ingredient
CI	=	confidence interval
CQA	=	critical quality attribute
DAE	=	differential algebraic equations
DC	=	direct compression
DoE	=	design of experiments
GSA	=	global sensitivity analysis
KI	=	key indicator
KPI	=	key performance indicator
KS	=	knowledge space
LV	=	latent variable
LSA	=	local sensitivity analysis
MBDoE	=	model-based design of experiments
MC	=	Monte Carlo
PC	=	principal component
PCA	=	principal component analysis
PLS	=	partial least squares
R&D	=	research and development
USP	=	Unites States Pharmacopeia

Greek letters

β	=	sensitivity of formulation to lubrication $[-]$
γ	=	lubrication rate constant [dm^{-1}]
δ_j	=	acceptable tolerance in the model prediction for $K_{M,j}$
$\boldsymbol{\delta}$	=	vector of acceptable tolerances δ_j
$\acute{\delta}$	=	Dirac delta function
ε_i	=	uncertainty of parameter θ_i
$\boldsymbol{\varepsilon}$	=	vector of parameters uncertainties ε_i
ε_i^h	=	value of ε_i at scenario h
$\varepsilon_{i,max}$	=	maximum uncertainty of parameter θ_i performing the optimization on N scenarios
$\boldsymbol{\varepsilon_{max}}$	=	vector of maximum uncertainties $\varepsilon_{i,max}$
$\varepsilon_{i,max,b}$	=	maximum uncertainty of parameter θ_i performing the optimization on 2^n boundary scenarios
$\boldsymbol{\varepsilon_{max,b}}$	=	vector of maximum uncertainties $\varepsilon_{i,max,b}$

$\left(\varepsilon_{i,max,b}\right)^{(k)}$	=	maximum uncertainty of parameter θ_i performing the optimization on 2^n boundary scenarios at iteration k
ε_s	=	average porosity of the swollen product $[-]$
ϵ_{ijk}	=	measurement error
$\dot{\epsilon}$	=	erosion rate [m s^{-1}]
θ_i	=	ith model parameter
$\boldsymbol{\theta}$	=	set of model parameters θ_i
$\theta_i^{i.g.}$	=	initial guess for parameter θ_i
θ_i^h	=	value of θ_i at scenario h
$\boldsymbol{\theta}^h$	=	set of parameters at scenario h
θ_i^0	=	ith estimated model paramenter
$\boldsymbol{\theta^0}$	=	set of estimated model parameters θ_i^0
$\boldsymbol{\Theta}$	=	matrix of parameters combinations for PCA/PLS models
λ	=	swelling rate [s^{-1}]
μ	=	liquid viscosity [Pa s]
ν_i^h	=	variable used both to consider different uncertainty signs and to characterize all the scenarios parameters combinations
$\boldsymbol{\nu}^h$	=	vector of ν_i at each scenario h
ξ_i	=	ith relative uncertainty with respect to θ_i^0
$\boldsymbol{\xi}$	=	vector of relative uncertainties ξ_i
$\xi_{i,max}$	=	ith relative maximum uncertainty with respect to θ_i^0
$\boldsymbol{\xi_{max}}$	=	vector of maximum relative uncertainties $\xi_{i,max}$
ρ_p	=	density of particles [kg m^{-3}]
σ_{ijk}	=	measurement standard deviation
$\tilde{\sigma}_1^{rs}$	=	rsth element of the inverse of the estimate of the residuals' variance-covariance matrix
$\boldsymbol{\Sigma}_y$	=	variance-covariance matrix of measurements errors
τ	=	total stress [MPa]
τ_{or}	=	average tablet tortuosity $[-]$
$\boldsymbol{\varphi}$	=	experiment design vector
ϕ	=	shape factor of particles $[-]$
Ψ	=	Maximum Likelihood function

Latin letters

a_1	=	extended Kushner parameter (1) [MPa]
a_2	=	extended Kushner parameter (2) $[-]$
a_{sf}	=	Kawakita model parameter (1) $[-]$
A_t	=	tablet surface area [m^2]
b_1	=	extended Kushner parameter (3) $[-]$
b_2	=	extended Kushner parameter (4) $[-]$
B_{API}	=	rate of release of API [s^{-1}]

b_{sf}	=	Kawakita model parameter (2) [MPa^{-1}]
$\boldsymbol{C}$	=	matrix of ν_i^j for the yeast fermentation problem
C_2	=	Peppas and Colombo parameter (1) [MPa]
C_3	=	Peppas and Colombo parameter (2) [MPa]
c_{API}	=	bulk concentration of API [kg m^{-3}]
c_{sat}	=	saturation concentration of API [kg m^{-3}]
d_h	=	tablet hydraulic diameter [m]
E	=	elastic constant (1) [−]
F_L	=	tablet fracture load [MN]
G_0	=	elastic constant (2) [MPa]
H_{coat}	=	thickness of the coating layer [m]
K	=	extent of lubrication [m]
$\boldsymbol{\mathcal{K}}$	=	response matrix for PLS model
k_{API}	=	mass transfer coefficient of API [m s^{-1}]
$\boldsymbol{K}_M$	=	vector of $K_{M,j}$
K_{M_1}	=	key indicator of tablet press model (i.e., TS) [MPa]
K_{M_2}	=	key indicator of disintegration model (i.e., t_{dis}) [s]
K_{M_3}	=	key indicator of dissolution model (i.e., LC) [%]
$K_{M,j}$	=	key indicator of the model
$\overline{K_{M,J}}$	=	target value of $K_{M,j}$
l	=	particle size [m]
$l_{0,API}$	=	particle size at the beginning of the process [m]
LC	=	label content [%]
$\widehat{LC}$	=	label content predicted by the model [%]
LC_{min}	=	minimum label content value for K_{M_3} acceptability [%]
LC_{max}	=	maximum label content value for K_{M_3} acceptability [%]
m	=	number of parameter combinations via MC simulations
M_1	=	tablet press model
M_2	=	disintegration model
M_3	=	in vitro dissolution model
M_t	=	tablet mass [kg]
$M_{t,0}$	=	initial tablet mass [kg]
n	=	number of parameters in the model
N	=	total number of scenarios for the optimization problem resolution
n_{API}	=	order of dissolution of API [−]
N_{API}	=	number of dissolving API particles [−]
N^{exp}	=	total number of measurements taken during all the experiments
n_s	=	swelling parameter [−]
N_{KI}	=	number of model key indicators
NE	=	number of experiments performed
NM_{ij}	=	number of measurements of the jth variable in the ith experiment

NV_i	=	number of variables measured in the ith experiment
Obj	=	objective function of the optimization problem
P	=	compaction pressure [MPa]
p_c	=	capillary pressure [Pa]
P_d	=	water penetration depth [m]
q	=	number of responses for $\boldsymbol{V}(\boldsymbol{\theta}, \boldsymbol{\varphi})$ definition
$R_{API,l}$	=	particle dissolution coefficient [m^2 s^{-1}]
s_f	=	solid fraction $[-]$
S_i	=	first-order index of Sobol analysis
$S_{i,j}^{\lambda}$	=	similarity factor $[-]$
S_p	=	shape factor of pores $[-]$
S_r	=	matrix of partial derivatives of the rth equation in the model
$S_{T,i}$	=	total-effect index of Sobol analysis
t	=	time [s]
t_{dis}	=	disintegration time [s]
$\widehat{t_{dis}}$	=	tensile strength predicted by the model [s]
$t_{dis,min}$	=	minimum disintegration time value for K_{M_2} acceptability [s]
$t_{dis,max}$	=	maximum disintegration time value for K_{M_2} acceptability [s]
t^{exp}	=	experimental duration
$\boldsymbol{t_{sp}}$	=	vector of sampling times
$T_{t/2}$	=	half tablet thickness [m]
TS	=	tensile strength [MPa]
$\widehat{TS}$	=	tensile strength predicted by the model [MPa]
TS_0	=	tensile strength at zero porosity [MPa]
TS_{min}	=	minimum tensile strength value for K_{M_1} acceptability [MPa]
TS_{max}	=	maximum tensile strength value for K_{M_1} acceptability [MPa]
$\boldsymbol{u}(t)$	=	vector of time-varying model control variables
u_1	=	dilution factor [h^{-1}]
u_2	=	substrate concentration in the feed [g/L]
$V(y_j)$	=	variance of the jth model output
$V(\boldsymbol{\theta}, \boldsymbol{\varphi})$	=	parameter variance-covariance matrix
$V_{1,2,\ldots,k}$	=	kth-order contribution of factor i to the variance of the output
V_c	=	coating volume [m^3]
V_i	=	first-order contribution of factor i to the variance of the output
$V_{i,l}$	=	second-order contribution of factor i to the variance of the output
V_m	=	liquid volume in the vessel [m^3]
V_{θ_i}	=	variance based first order effect for factor θ_i
$V_{\theta_{\sim i}}$	=	variance based first order effect for factor $\theta_{\sim i}$
w_l	=	liquid content in the tablet $[-]$
$\boldsymbol{x}(t)$	=	vector of time-dependent state variables
x_1	=	biomass concentration [g/L]

x_1^0	=	initial biomass concentration [g/L]
x_2	=	substrate concentration [g/L]
x_{API}	=	mass fraction of API $[-]$
$\boldsymbol{y}(t)$	=	vector of measured outputs
$\hat{\boldsymbol{y}}(t)$	=	vector of outputs predicted by the model
$\boldsymbol{y}_0$	=	vector of state variables initial conditions
z_{ijk}	=	kth predicted value of variable j in experiment i
$\tilde{z}_{ijk}$	=	measured value of z_{ijk}

Introduction

In the pharmaceutical environment, there is a significant need to accelerate the development of industrial processes in order to maintain competitiveness of companies. Quantitative models have been increasingly employed to reduce time-demanding and resource-intensive experimental campaigns that are typical of this industrial sector. However, the use of modeling for pharmaceutical process development and optimization is not as widespread as it could be. This is mainly due to stakeholders lack of confidence in the prediction capability of a quantitative model to accurately represent both the real process and the key indicators of interest. The objective of this study is to address this issue by proposing an approach that allows to ensure an acceptable level of model fidelity.

The Thesis is organised as follows.

In Chapter 1 the context of modeling in the pharmaceutical environment is presented, underlining the difficulty related to its application due to the stakeholders reluctance. After discussing the methodology described in Geremia *et al.* (2023), which represents the term of comparison for this work, motivations and objectives associated to the definition of a more rigorous methodology are discussed.

Chapter 2 describes the mathematical techniques adopted in the definition of the algorithm (i.e., sensitivity analysis, model based design of experiments, parameters estimation, and optimization) focusing on the formulation of model, constraints, and objective function of the optimization problem. Later, a motivating example is discussed in detail.

The methodology validity is then assessed in Chapter 3, where a direct compression process is used as a case study. The same process was also employed for the implementation of the systematic approach discussed in Geremia *et al.* (2023). In this way, a comparison of the results obtained with both methodologies is performed assessing whether the attainment of acceptable model fidelity is limited by one of the two methods.

Chapter 4 underlines the limits of the optimization procedure suggesting a possible improvement. An enhanced algorithm is presented and applied to the same direct compression process in Chapter 3. Results obtained are commented and compared again with those achieved through the procedure discussed in Geremia *et al.* (2023).

Some final remarks conclude the work.

Chapter 1

Challenges in pharmaceutical manufacturing: modeling support to process development

This chapter summarizes current challenges and trends in modeling for pharmaceutical manufacturing development, focusing on the requirements to attain a pre-set fidelity in model prediction. Later, a model evaluation framework capable to achieve this goal, proposed by Geremia *et al.* (2023), is briefly described. Finally, the motivations and objectives related to the formulation of a new methodology are presented.

1.1. Modeling for pharmaceutical manufacturing

Pharmaceutical manufacturing research and development (R&D) has traditionally relied on design of experiments (DoE) approaches to investigate the impact of process input parameters (i.e., operating variables, material properties) on product manufacturing and/or quality requirements. The experimental campaigns are generally very long and resource-intensive (e.g., consuming a lot of active pharmaceutical ingredient (API), which is expensive during process development). Therefore, the use of quantitative models in pharmaceutical manufacturing has recently gained momentum to accelerate process development and optimization, as well as enhancing process understanding.

Although modeling in pharmaceutical process environment has been demonstrated to be beneficial in saving substantial time and resources (Destro and Barolo, 2022), there is still a lack of confidence in the systematical use of quantitative models. One of the main obstacles to their implementation is due to the reluctance from stakeholders towards models prediction capability with respect to key performance indicators (KPIs), and critical quality attributes (CQAs), which we will generically call key indicators (KIs) (Braakman *et al.*, 2022). Moreover, many pharmaceutical processes involve very complex phenomena, and these may not be easily captured by first-principles models; this is particularly true for the manufacturing of small molecules and biopharmaceuticals. Thus, high in-house expertise and resources are still required (Polak et al., 2023).

Pharmaceutical manufacturing processes typically comprise of many unit operations. Test units can be also modeled to assess whether a pharmaceutical product meets the required specifications.

These mathematical models are generally defined by a set of differential algebraic equations

(DAEs) of the form:

$$\begin{cases} \boldsymbol{f}(\boldsymbol{x}(t),\ \dot{\boldsymbol{x}}(t), (\boldsymbol{\theta}), \boldsymbol{u}(t), t) = 0 \\ \hat{\boldsymbol{y}} = \boldsymbol{g}\big(\boldsymbol{x}(t)\big) \\ \boldsymbol{K}_M = \boldsymbol{h}\big(\boldsymbol{x}(t)\big) \end{cases}, \tag{1.1}$$

where $\boldsymbol{x}(t)$ is the vector of time-dependent state variables, $\boldsymbol{u}(t)$ are the model control variables, $\boldsymbol{\theta}$ is the set of unknown model parameters to be estimated, $\hat{\boldsymbol{y}}(t)$ is the vector of outputs predicted by the model (which in general may differ from the measured ones $\boldsymbol{y}(t)$), and $\boldsymbol{K}_M$ are the KIs on which constraints are imposed. These models can be used to describe both operating and test units, and to predict the values of the KIs. Therefore, quantitative mathematical models have the capability of explaining the phenomena taking place throughout the manufacturing process; in this way rational design decisions are supported. In recent years, there have been numerous instances of quantitative models application to the pharmaceutical industry, where more than one single unit operation is present. Just to give some references, Bano et al. (2022) guided the development of an industrial dry granulation process for immediate release tablets. Yang *et al.* (2022) design a recombinant adeno-associated virus drug manufacturing process. In Diab *et al.* (2022), mathematical models were used to find the optimal process setpoints to optimize the manufacturing process of API production.

To increase the acceptance of quantitative models in pharmaceutical manufacturing development, it is necessary to employ standardized model evaluation methods for the assessment of model prediction fidelity (Zineh, 2019). Whenever a reliable mathematical model is available, its predictive capability with respect to the KIs strongly depends on the precision of model parameters estimates.

In this context, the following questions need to be addressed:

- since accounting for uncertainty in model predictions is fundamental, how can we quantitatively assess the model fidelity, so as to ensure an assigned confidence in the prediction of the model KIs?
- should all model parameters be estimated in a statistically satisfactory way, or is it sufficient to focus on just a subset of them? What is the parametric precision that must be attained to satisfy the prediction requirements for the KIs?

The objective of this work is to answer the questions above, and to aid the systematic use of mathematical models within pharmaceutical manufacturing.

1.2. Methodology based on multi-variate statistical methods

The aim is to enhance the usage of quantitative models within a pharmaceutical manufacturing environment. Geremia *et al.* (2023) proposed a systematic procedure,

combining model based design of experiments (MBDoE) with new methods based on data analytics to assess model fidelity and to support practitioners in model usage for pharmaceutical development.

The methodology consists of four sequential steps:

1. model identifiability and parameters ranking;
2. impact of parameter uncertainty on the fidelity of model predictions;
3. MBDoE and experimentation;
4. parameter estimation.

The overall procedure, depicted in Figure 1.1, is iteratively repeated until satisfaction of the stop criterion, which is attained when all predictions of model KIs are within the desired tolerance.

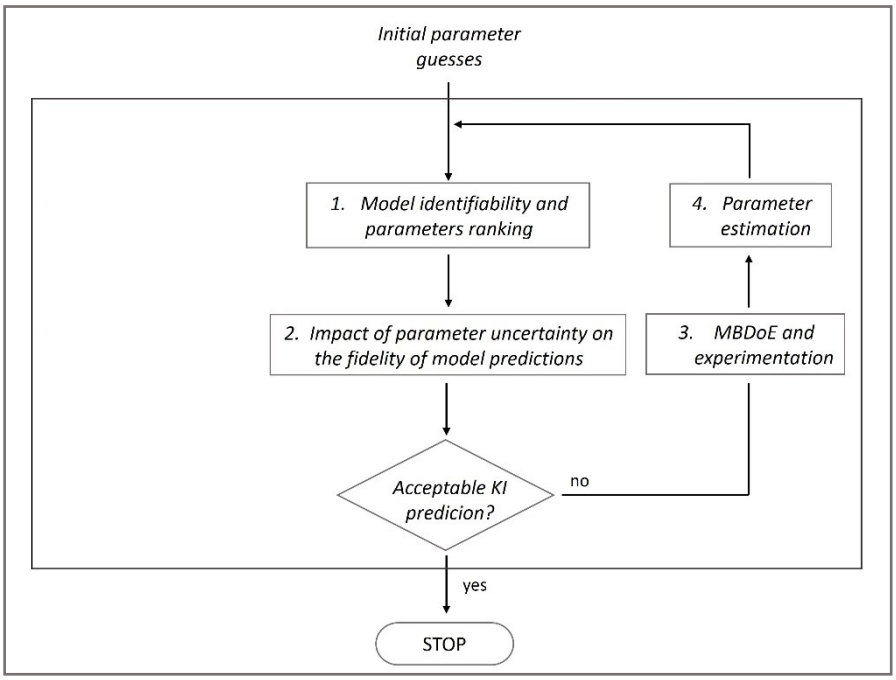

Figure 1.1. *Schematic of the general framework proposed by Geremia et al. (2023) to quantify the model parameter impact on the prediction fidelity of model KIs.*

The techniques adopted to implement the first and the second steps are briefly described in the following, while detailed descriptions of MBDoE and parameter estimation are presented in §2.1.2 and §2.1.3 respectively.

1.2.1. Model identifiability and parameters ranking

The objective is to characterize the parameter impact on the prediction of the selected KIs, so that the most influential parameters can be identified. Once it has been verified that all parameters can be estimated (Miao et al, 2011), the impact of model parameters on the KIs prediction is *ranked* with respect to each KI of interest. This is relevant in order to prioritize

the subsequent experimental effort (step 3 in Figure 1.1) for parameter estimation towards the most influential parameters (Saltelli et al., 2008).

Different techniques, such as local and global sensitivity analysis (presented in §2.1.1), can be used to evaluate the influence of parameter estimates on selected KIs; Geremia *et al.* (2023) proposed an alternative approach based on PCA (Montgomery, 2013).

The objective is to evaluate the relationship between the fidelity of a particular $K_{M,j}$ and n model parameters that are affected by uncertainty. Therefore, a matrix $\boldsymbol{\Theta}$ $[m \times n]$ of m combinations of n model parameters is built via Monte Carlo (MC) simulations according to the current ranges of parametric uncertainty, and the values of $K_{M,j}$ for each parameters combination are computed exploiting the original model. At this point, PCA is used to summarize the information of $\boldsymbol{\Theta}$ by projecting the model parameters onto a coordinate system of independent variables called principal components (PCs). The influence and correlation of each model parameter is established based on the length of the projected loadings along the PCs, and mutual positions in the latent space; the greater the θ_i loading, the more influential that parameter is in the prediction of the $K_{M,j}$ of interest. To avoid the scaling effect of different orders of magnitude of the model parameters, $\boldsymbol{\Theta}$ is autoscaled, i.e., data are mean-centered and scaled to unit variance (Wise and Gallagher, 2006).

If compared to any traditional sensitivity analysis, one relevant advantage of the PCA methodology is the possibility to assess how variations in control variables affect the interaction between parameters and their relative ranking by comparing different PCA models (i.e., models obtained with different fixed operating conditions); this can be evaluated through the similarity factor, $S_{i,j}^{\lambda}$, whose mathematical definition is discussed in Gunther *et al.* (2009).

Therefore, PCA represents an efficient method capable to define the correlation among parameters, their influence on $K_{M,j}$ of interest, and their dependence on process manipulated variables.

1.2.2. Impact of parameter uncertainty on the fidelity of model prediction

The second step of the algorithm presented in Figure 1.1 aims at quantifying how uncertainty in parameter values impacts on the prediction fidelity of the model $\boldsymbol{K}_M$, assessing whether or not the attained precision on the $\boldsymbol{K}_M$ is acceptable. In order to assist the interpretation of these results, a partial least-squares (PLS) regression model is exploited to quantify the impact of each model parameter to the selected $K_{M,j}$.

Let $\boldsymbol{\Theta}$ $[m \times n]$ be the same matrix defined in §1.2.1, and $\boldsymbol{\mathcal{K}}$ $[m \times N_{KI}]$ the correspondent response matrix of m combinations of N_{KI} $K_{M,j}$ computed via the original model. The stop criterion (Figure 1.1) is reached when an acceptable $K_{M,j}$ is achieved for all parameter combinations. This is evaluated through a PLS regression model (Geladi and Kowalski, 1986), where both the regressor $\boldsymbol{\Theta}$ and the response variables $\boldsymbol{\mathcal{K}}$ are projected onto a common

latent space; since it is advisable to capture most of the variance in the input and output data, the selection of the significant latent variables (LVs) to build the model is typically executed employing a cross-validation procedure (Wold et al., 2001). In the latent space, the confidence limits are represented by a hyper-ellipsoid, which outlines the boundary of the knowledge space (KS) (MacGregor and Bruwer, 2008); only points falling within this region can be used for inference with the PLS model. As a matter of example, Figure 1.2 displays the projections of the calibration samples onto the score space as light blue circles, and the KS boundary as a green ellipse.

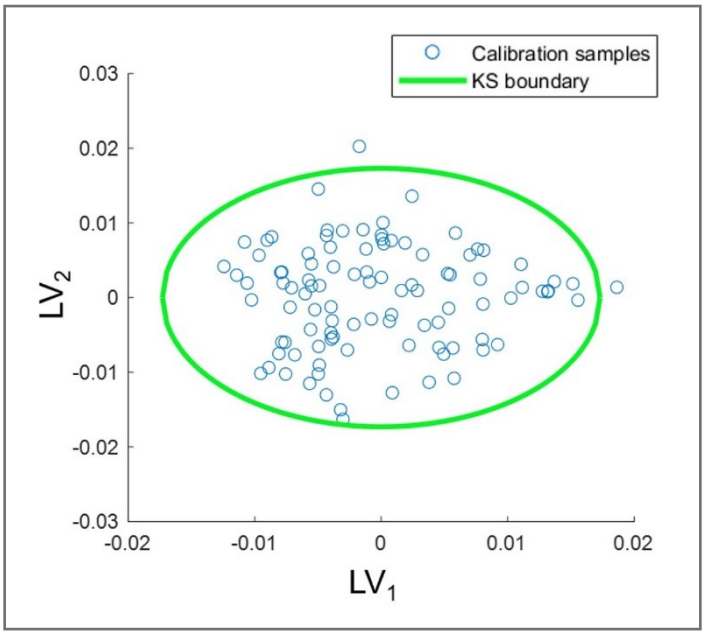

Figure 1.2. *Illustrative example of PLS: projection of the calibration samples (light blue circles) onto the score space and KS boundary (green ellipse).*

Once the PLS model is defined, it can be inverted (Jaeckle and MacGregor, 2000), as shown in Figure 1.3, to determine the set of parameter combinations which guarantee that the prediction of the $K_{M,j}$ of interest is within the range of desired tolerance.

Bound values for the response prediction define the $K_{M,j}$ acceptability region; if at least one projection falls outside the $K_{M,j}$ acceptability region, as shown in Figure 1.3a, it is necessary to conduct further experiments to enhance the parameter precision. The stop criterion is reached only when all uncertainties fall inside the $K_{M,j}$ acceptability region (Figure 1.3b).

Therefore, the proposed methodology can be exploited to ensure pre-set requirements on parameters in the prediction of model KIs, establishing which are the most influential parameters and focusing only on their statistically acceptable estimation. One main advantage consists in the graphical representation of both parameter uncertainty and KIs in a common latent space, allowing for an easier interpretation of results.

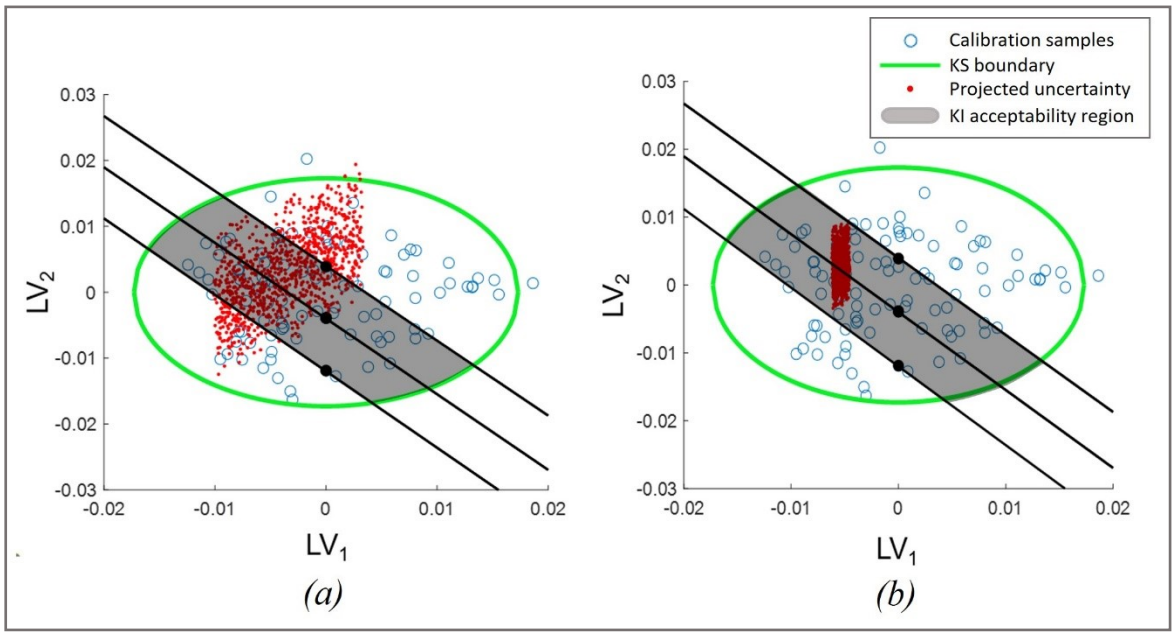

Figure 1.3. *Illustrative example of PLS inverted model: (a) Case in which some projections of parameter uncertainty (red points) fall outside the $K_{M,j}$ acceptability region (grey area). (b) Case in which all projections of parameter uncertainty (red points) fall inside the $K_{M,j}$ acceptability region (grey area)*

The procedure has been efficiently implemented on a direct compression process, leading to satisfactory findings for the modeling development in the pharmaceutical industry.

1.3. Motivations and objectives

The methodology proposed by Geremia *et al.* (2023) represents an efficient tool capable to assess whether or not an acceptable model fidelity is ensured and can be exploited for model evaluation. Nevertheless, it is based on a linear PLS model, i.e., an approximation of the rigorous model; this can represent a limit in the evaluation of model fidelity.

The aim of this work is to find an alternative approach for the quantification of the parameters impact toward the prediction fidelity of model KIs, while overcoming the issue above. Therefore, a procedure will be proposed that relies on the original model (i.e., without any linearization). This methodology is computed trying to specify the maximum parametric uncertainty capable to guarantee acceptable model fidelity. To reach this goal, an optimization framework is presented in the following chapter.

Chapter 2

Methodology

This chapter firstly aims at providing a thorough overview of techniques and software tools adopted in this work. Later, the interest is focused on both the framework and application of a new method developed capable to ensure an acceptable fidelity of models analysed, whose general definition is the one presented in Equation (1.1) in §1.1. At the end, an example of the procedure's implementation is presented.

2.1. Software and mathematical tools

The project development requires a dynamic process modelling and simulation software. In this study, Siemens Process System Enterprise's advanced process modelling software gPROMS® Model Builder v. 7.0. has been adopted. Along with the capability to perform dynamic simulation and optimization, the power in advanced model parameter estimation and design of experiments should be mentioned.

2.1.1. Sensitivity analysis

Each physical model able to describe a specific process unit is characterised by factors on which responses depend (i.e., operating conditions/ model parameters). In this work, the first objective consists in assessing which model parameters have the greatest influence towards output variability.

Sensitivity analysis provides metrics to rank the importance of parameters with respect to an output of interest: the larger the values of the metrics, the more sensitive the model response with respect to parameter changes (Saltelli *et al.*, 2008). As a result, it is possible to underline those parameters to focus on during the subsequent parameter estimation activity (§2.1.3). Clearly, influential parameters have an important role towards the prediction of the output of interest, and so an accurate estimate is required on them; quite the opposite, non-influential parameters can be fixed to nominal values. This reduction in the number of free parameters provides both greater confidence in model output predictions and a basis for model reduction (Braakman *et al.*, 2022).

Generally, sensitivity analysis can be classified into two main branches:

- local sensitivity analysis (LSA): is a one-at-time method that considers single parameter perturbations around nominal value and often quantifies sensitivity by means of partial derivatives;

- global sensitivity analysis (GSA): focuses on output uncertainty over the entire input space, and typically resorts to sampling-based approaches.

During this project GSA is adopted, that is generally recommended when very little preliminary information on parameter values is available (Saltelli *et al.*, 2008). In particular, among all types of global sensitivity analysis techniques, the Sobol method is adopted. This is a variance-based methodology based on Saltelli (2002) in which the variance of the ith model output is decomposed as follow:

$$V(y_j) = \sum_{i=1}^{k} V_i + \sum_{i=1}^{k} \sum_{l=i+1}^{k} V_{i,l} + \cdots + V_{1,2,\ldots,k} \quad , \tag{2.1}$$

where V_i expresses the first-order contribution of factor i to the variance of the output, while $V_{i,l}$ and $V_{1,2,\ldots,k}$ are the variance contribution due to second- and kth-order interactions. According to Sobol (1993, 2001), there are two types of variance-based sensitivity indices: a first-order (S_i) index and a total-effect ($S_{T,i}$) index. The former represents the main effect of each parameter on the variance of the output and it is defined as:

$$S_i = \frac{V_{\theta_i}(E_{\boldsymbol{\theta}\sim i}(y_j|\theta_i))}{V(y_j)} \quad , \tag{2.2}$$

where $V_{\theta_i}(E_{\boldsymbol{\theta}\sim i}(y_j|\theta_i))$ indicates a variance based first order effect for a generic factor θ_i. If a first-order index is large, then the corresponding parameter is influential on the output. The latter captures both the variance that can be attributed to a particular parameter, and all of its interactions with other parameters:

$$S_{T,i} = 1 - \frac{V_{\boldsymbol{\theta}\sim i}(E_{\theta_i}(y_j|\boldsymbol{\theta}_{\sim i}))}{V(y_j)} \quad , \tag{2.3}$$

where $V_{\boldsymbol{\theta}\sim i}(E_{\theta_i}(y_j|\theta_{\sim i}))$ is the first order effect of $\boldsymbol{\theta}_{\sim i}$, so that $V(y_j)$ minus $V_{\boldsymbol{\theta}\sim i}(E_{\theta_i}(y_j|\theta_{\sim i}))$ must give the contribution of all terms in the variance decomposition which do include θ_i. If a total-effect index is small, then the corresponding parameter is non-influential, and could be fixed during subsequent analysis. Total-effect index should be always higher than the first-order one.

Formulae used to estimate the sensitivity indices are those proposed in Saltelli *et al.* (2010). Within gPROMS® the sensitivity analysis is performed through the `Global System Analysis` entity that allows investigating the system behaviour. Variability of selected outputs is computed after specifying the range of variability of each input of interest.

2.1.2. Model based design of experiments

Model based design of experiments (MBDoE) techniques are fundamental for rapidly designing experiments containing the most informative data for use in estimation of model parameters (Galvanin *et al.*, 2007). They are effective mathematical tools, which, starting from the knowledge of the system under study, bring to the maximization of the information content from the designed experiment.

MBDoE is the first step of a relevant procedure that is composed of three sequential key activities:

1. design of the optimal experiment, possibly based on previous knowledge of the process;
2. execution of the new experiments: in this context they have been simulated in silico and so in a process based on an ideal reference parameter set. Random errors are, then, added to the output of interest, based on the sensor's standard deviation;
3. parameter estimation through both historical and new designed experimental set; discussed in §2.1.3.

The procedure above must be iteratively repeated, until the achievement of acceptable parameter estimation results (i.e., sufficient fidelity in model predictions and/or statistically satisfactory estimates) (Figure 2.1).

The mathematical description of the MBDoE fundamentals is briefly presented, recalling what is exposed by Franceschini and Macchietto (2008). All experiment design variables are collected in a single design vector, $\boldsymbol{\varphi}$:

$$\boldsymbol{\varphi} = \boldsymbol{\varphi}(\boldsymbol{y^0}, \boldsymbol{u}(t), \boldsymbol{w}, \boldsymbol{t_{sp}}, t^{exp}) \quad , \tag{2.4}$$

where $\boldsymbol{y^0}$ is a set of $\boldsymbol{x}$ and $\dot{\boldsymbol{x}}$ sufficient to establish the initial conditions of the system, $\boldsymbol{t_{sp}}$ is the vector of sampling times, and t^{exp} is the experiment duration. The objective is to reduce the expected model parameter uncertainty region by acting on $\boldsymbol{\varphi}$, and this is mathematically intuitive translated into a decrease of the elements of the parameter variance-covariance matrix, $\boldsymbol{V}(\boldsymbol{\theta}, \boldsymbol{\varphi})$, defined according to Asprey and Macchietto (2002):

$$\boldsymbol{V}(\boldsymbol{\theta}, \boldsymbol{\varphi}) = \left[\sum_{r=1}^{q}\sum_{s=1}^{q} \tilde{\sigma}_1^{rs} \boldsymbol{S}_r^T \boldsymbol{S}_s\right]^{-1} \quad , \tag{2.5}$$

where q is the number of responses, $\tilde{\sigma}_1^{rs}$ is the rsth element of the inverse of the estimate of the residuals' variance-covariance matrix. $\boldsymbol{S}_r$ is the matrix of partial derivatives of the rth equation in the model with respect to the parameters $\boldsymbol{\theta}$ calculated at the $n+1$ experimental points:

$$S_r = \frac{\partial \hat{\boldsymbol{y}}(\boldsymbol{\varphi}, \boldsymbol{\theta}, t)}{\partial \boldsymbol{\theta}} \quad . \tag{2.6}$$

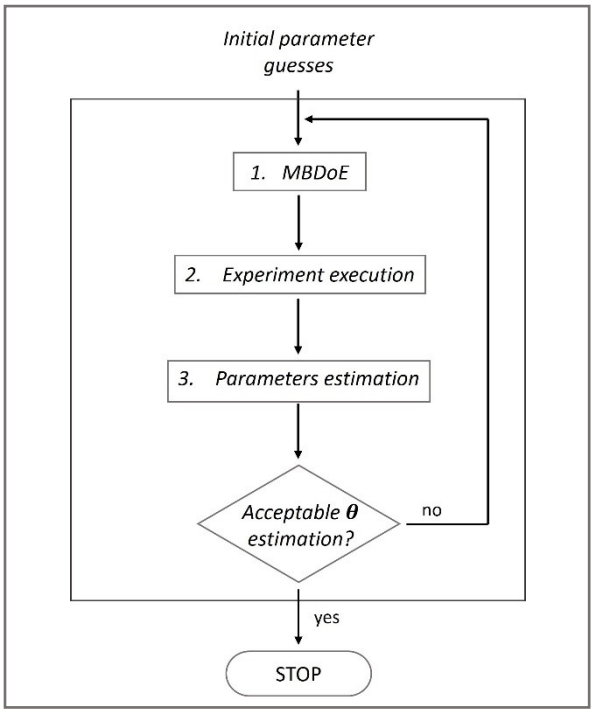

Figure 2.1. *Schematic of MBDoE iterative procedure: (1) model-based design, (2) experiment execution, (3) parameter estimation.*

In order to compare the magnitude of different matrices, various real-valued functions have been suggested as metrics. Three common criteria are:

- A-Optimality: minimise the trace of $\boldsymbol{V}(\boldsymbol{\theta}, \boldsymbol{\varphi})$, and thus minimises the dimensions of the enclosing box around the joint confidence region;
- E-Optimality: minimise the largest eigenvalue of $\boldsymbol{V}(\boldsymbol{\theta}, \boldsymbol{\varphi})$, and thus minimises the size of the major axis of the joint confidence region;
- D-Optimality: minimise the determinant of $\boldsymbol{V}(\boldsymbol{\theta}, \boldsymbol{\varphi})$, and thus minimises the volume of the joint confidence region.

The meaning of each criterion is geometrically presented in Figure 2.2.

From the algorithm perspective, the implementation of the MBDoE approach in gPROMS® occurs within the `Experiment Design` entity, that requires the construction of the `Experiment` entity, specifically of its …`to be designed` section. This tool allows to specify the full details of an experiment that has to be set. Therefore, the user can indicate the nature of the experiment itself (dynamic or steady-state) and provide guesses for the initial conditions, the sensor and the values of manipulated variables that later are to be optimally determined by the MBDoE algorithm. Moreover, it is also possible to add some endpoint or interior-point constraints that need to be placed on the experiment that will be designed.

Once the new test is specified, it is executed and later represented through the … `performed` section of the `Experiment` entity. This requires the specification of all the experiment conditions: the type (dynamic or steady-state), the initial conditions of the system, the values

of the time-invariant, piecewise constant and piecewise linear variables that are adjusted during the experiment, and finally the measured data and sensor used with also the possibility to describe the measurement variance model.

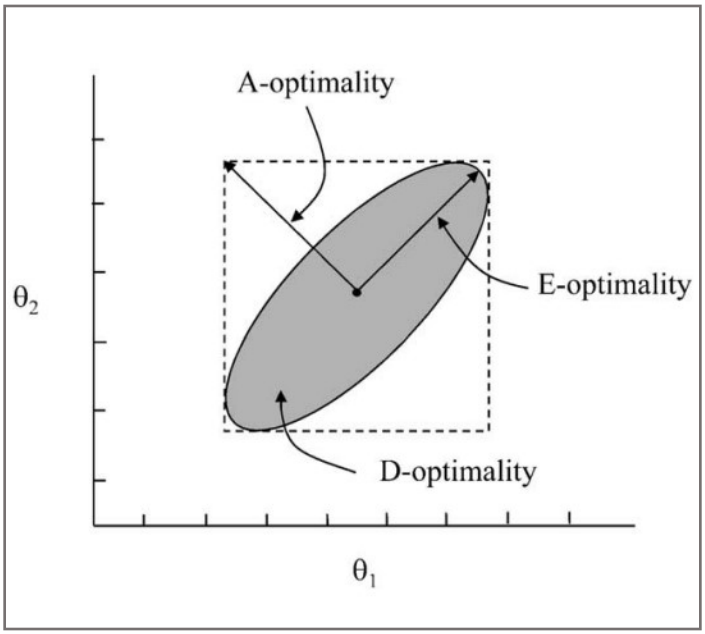

Figure 2.2. *Geometric interpretation of the various design criteria.*

In order to increase the maximum information obtainable from a designed experiment, the `Experiment Design` calculations should take into account any previously performed experiments. This may result in different regions of the operating space being explored from the point of view of yielding more knowledge of the model.

2.1.3. Estimation of parameters

A detailed process model is constructed from equations describing the physical and chemical phenomena that take place in the system and these equations usually involve parameters that must be calibrated to make the model predictions match reality.

The software supports the parameter estimation through the `Model Validation` entity. Parameter estimation in gPROMS® is based on the Maximum Likelihood formulation which provides simultaneous estimation of parameters within the process physical model.

When solving a Maximum Likelihood parameter estimation problem, gPROMS® attempts to determine values for the uncertain physical model parameters able to maximise the probability that the mathematical model will predict the measurement values obtained from the experiments. Assuming independent, normally distributed measurement errors, ϵ_{ijk}, with zero means and standard deviations, σ_{ijk}, this Maximum Likelihood goal can be captured through the following objective function (Bard, 1974):

$$\Psi = \left(\frac{N^{exp}}{2}\right) \cdot \ln(2\pi) + \left(\frac{1}{2}\right) \cdot min_{\theta} \sum_{i-1}^{NE} \sum_{j=1}^{NV_i} \sum_{k=2}^{NM_{ij}} \left[\ln\left(\sigma_{ijk}^2\right) + (\tilde{z}_{ijk} - z_{ijk})^2 / \sigma_{ijk}^2\right] \quad ,(2.7)$$

where N^{exp} is the total number of measurements taken during all the experiments, NE is the number of experiments performed, NV_i is the number of variables measured in the ith experiment, NM_{ij} is the number of measurements of the jth variable in the ith experiment, and $\tilde{z}_{ijk}$ is the measured value of the kth predicted one of variable j in experiment i, z_{ijk}. Once the estimation is performed, not only new values of parameters are achieved but also statistical information about both their confidence intervals (CIs) and t-value are available. If the 95% t-value of a parameter is lower than the reference one, that parameter is not estimated precisely: available data from experiments may not be sufficient and new ones are needed.

As will be presented in §2.2.3, above all the knowledge attained from the `Model Validation` entity, the interest is focused on the 95% CIs: within the iterations of the new procedure described in this project, these are compared with the results of an optimization problem.

2.1.4. Optimization

Once a process is defined with all its features, we focus on the achievement of some specific purposes. More and more in industry there is the will to attain desired conditions in order to keep companies at a competitive level, and this is translated into an optimization of the processes. Therefore, an objective function is declared and this one must be satisfied through the manipulation of selected variables.

From the algorithm point of view, gPROMS® supports an `Optimization` entity capable to optimize both the steady-state and the dynamic behaviour of a continuous or batch process. After the objective function definition, the user can choose either to maximise or to minimise it through the manipulation of time-invariant, piecewise constant, or piecewise linear variables, and whatever their nature initial guesses, lower bounds and upper bounds must be set. Considering the set of DAEs presented in Equation (1.1) in §1.1, a general optimization problem follows:

$$\begin{cases} \min\limits_{\boldsymbol{u}(t),\boldsymbol{\theta}} \boldsymbol{x}(t) \\ \boldsymbol{u}^{min} \leq \boldsymbol{u}(t) \leq \boldsymbol{u}^{max} \\ \boldsymbol{\theta}^{min} \leq \boldsymbol{\theta} \leq \boldsymbol{\theta}^{max} \end{cases} . \quad (2.8)$$

Moreover, since the calculations refer to process units, it might be required to impose a target on the outputs. Therefore, the `Optimization` entity also allows the specification of end-point equality, end-point inequality and interior path constraints. At the end, when all the decisions are made, the role of the optimization is to provide the values of the manipulated

variables of interest in order to reach the goal imposed on the objective function while satisfying all the constraints.

2.2. Formulation of the optimization problem

The objective of the research is to provide maximum uncertainties allowable on parameters ensuring a certain pre-set model fidelity. This goal is reached guaranteeing that one or more key indicators (KIs), are maintained inside a desired tolerance range at one single point in time or during the entire process duration. Since the interest is focused on the values of time-independent parameters, treated as manipulated variables, assuring some constraints on other factors, the task can be translated into an optimization problem defined through the `Optimization` entity within gPROMS®.

2.2.1. General structure

The optimization is characterised by its own model, constraints, and objective function.
The model is the one of the process under investigation, generally described by Equation (1.1) in §1.1. However, for the project purposes, the model structure of the optimization problem requires a different definition of parameters, leading to:

$$\begin{cases} \boldsymbol{f}(\boldsymbol{x}(t),\ \dot{\boldsymbol{x}}(t), (\boldsymbol{\theta^0} + \boldsymbol{\varepsilon}), \boldsymbol{u}(t), t) = 0 \\ \hat{\boldsymbol{y}} = \boldsymbol{g}\big(\boldsymbol{x}(t)\big) \\ \boldsymbol{K}_M = \boldsymbol{h}\big(\boldsymbol{x}(t)\big) \end{cases} , \tag{2.9}$$

where $\boldsymbol{\theta^0}$ are the estimated parameters, and $\boldsymbol{\varepsilon}$ is the vector of uncertainties applied to those parameters. Each uncertainty, ε_i, can be either positive or negative.
Once the model structure is defined, the interest is focused on the desired constraints:

$$(\boldsymbol{K}_M - \overline{\boldsymbol{K}_M})^2 \leq \boldsymbol{\delta}^2 \quad , \tag{2.10}$$

where $\boldsymbol{\delta}$ is the vector of acceptable tolerances in the model prediction for $\boldsymbol{K}_M$, and $\overline{\boldsymbol{K}_M}$ are the target values imposed for $\boldsymbol{K}_M$. Therefore, if for example in a reactor a temperature, $\bar{T}$, equal to 350°C is needed, and the acceptable tolerance, δ, is set at 10%, from condition (2.10) the measured T is constrained inside the range [315; 385]°C.
Finally, the problem structure requires to delineate an objective function. The aim of this one is to find maximum uncertainties that can be applied simultaneously to all parameters ensuring that constraint (2.10) is fulfilled. Thus, it is defined as:

$$Obj = max \prod_{i=1}^{n} \xi_i \quad , \tag{2.11}$$

where n is the number of parameters in the model, and ξ_i is the relative uncertainty with respect to θ_i^0. Specifically:

$$\xi_i = \varepsilon_i / \theta_i^0 \quad . \tag{2.12}$$

A good practice is to avoid a too low Obj, because it can provide numerical issues. We observed that if Obj order of magnitude is at least 1, problems are reduced. Therefore, since ξ_i is imposed lower than 0.500, it may be reasonable to multiply the Obj to 10^{α}, where α should be high enough to reach the goal.

2.2.2. Code framework

In order to reach the optimization scope, the code is built to calculate the maximum parameters uncertainty that ensures the specified constraints. Since ε_i can be either positive or negative, the possible parameter combinations refer to those obtainable considering all different $\boldsymbol{\varepsilon}$. Easily, the number of total combinations available is equal to 2^n, where 2 represents the two possible signs ($+$ or $-$), i.e., the high and low bounds of each parameter. As a matter of example, the case with two parameters, θ_1 and θ_2, is presented:

Table 2.1. *Example: all* $(\boldsymbol{\theta}^0 + \boldsymbol{\varepsilon})$ *combinations for a model with two parameters.*

Combinations	$\boldsymbol{\theta_1 = \theta_1^0 + \varepsilon_1}$	$\boldsymbol{\theta_2 = \theta_2^0 + \varepsilon_2}$
1^{st}	$\varepsilon_1 > 0$	$\varepsilon_2 > 0$
2^{nd}	$\varepsilon_1 > 0$	$\varepsilon_2 < 0$
3^{rd}	$\varepsilon_1 < 0$	$\varepsilon_2 > 0$
4^{th}	$\varepsilon_1 < 0$	$\varepsilon_2 < 0$

The interest is focused on the research of the maximum uncertainties allowable, $\boldsymbol{\varepsilon_{max}}$, and so for at least one $\boldsymbol{\theta}$ it is expected that Equation (2.10) is at its equality condition. For clarity purposes, an example is presented on Figure 2.3, where $K_{M,j}$ fidelity is assumed to be dependent on a single θ_i; $\varepsilon_{i,max,b}$ is the maximum uncertainty allowing to fulfil condition (2.10).

This is attained within gPROMS® with the use of a composite model (i.e., a model that contains one or more other model entities as a sub-model). In particular, only one sub-model is used for the problem resolution, leading to a total of two models:

- lower-level model: different scenarios are considered, each one characterised by the same set of equations but different variables values;
- higher-level model: collects the various scenarios results.

The number of scenarios evaluated at this point is equal to the number of total parameters combinations, 2^n. Once this is specified, each single θ_i at each single scenario h, with $1 \leq h \leq 2^n$, must be defined:

$$\theta_i^h = (1 + \xi_i \cdot \nu_i^h) \cdot \theta_i^0 \quad , \tag{2.13}$$

where θ_i^h is the value of parameter i at scenario h, and ν_i^h is a variable used both to consider different uncertainty signs and to characterize all the scenarios parameters combinations, $\boldsymbol{\theta}^h$. Variable ν_i^h is defined as follow:

$$for\ 1 \leq i \leq n\ and\ 1 \leq h \leq 2^n \quad \nu_i^h = \{-1; +1\} \quad s.t. \quad \boldsymbol{\nu}^h \neq \boldsymbol{\nu}^{k \neq h} \quad , \tag{2.14}$$

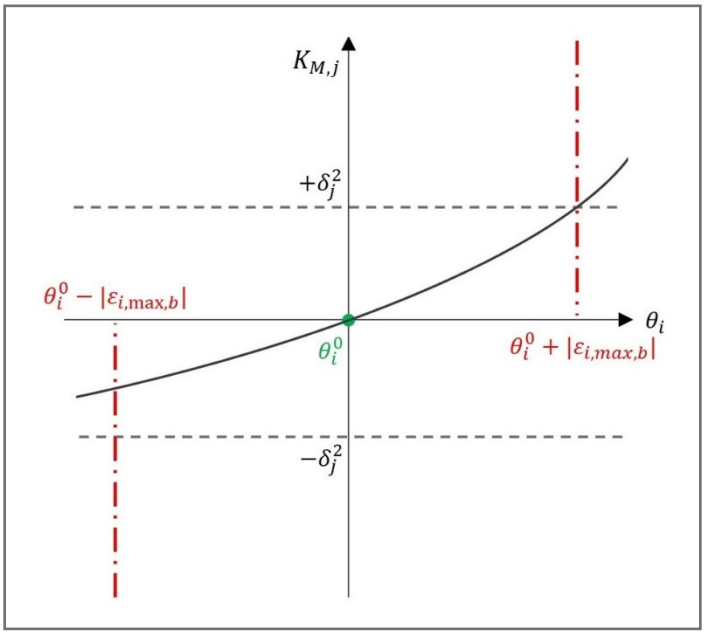

Figure 2.3. *Maximum uncertainty allowable, $\varepsilon_{i,max,b}$, on a parameter, θ_i, in order to respect the constraint on the key indicator variable, $K_{M,j}$, when a monotonic dependence between θ_i and $K_{M,j}$ is assumed.*

The number of scenarios evaluated at this point is equal to the number of total parameters combinations, 2^n. Once this is specified, each single θ_i at each single scenario h, with $1 \leq h \leq 2^n$, must be defined:

$$\theta_i^h = (1 + \xi_i \cdot \nu_i^h) \cdot \theta_i^0 \quad , \tag{2.13}$$

where θ_i^h is the value of parameter i at scenario h, and ν_i^h is a variable used both to consider different uncertainty signs and to characterize all the scenarios parameters combinations, $\boldsymbol{\theta}^h$. Variable ν_i^h is defined as follow:

$$for\ 1 \leq i \leq n\ and\ 1 \leq h \leq 2^n \quad \nu_i^h = \{-1; +1\} \quad s.t. \quad \boldsymbol{\nu}^h \neq \boldsymbol{\nu}^{k \neq h} \quad , \tag{2.14}$$

In this way, parameter uncertainty at each scenario, ε_i^h, is identified as follows:

$$\varepsilon_i^h = \xi_i \cdot \nu_i^h \cdot \theta_i^0 \quad , \tag{2.15}$$

while the general ε_i:

$$\varepsilon_i = \xi_i \cdot \theta_i^0 \quad . \tag{2.16}$$

The optimization is directly performed on ξ_i, considering all parameters concurrently. Note that when handling complex models, converge issues may arise; therefore, a more detailed procedure should be adopted:

1. optimize one or two parameters maintaining the others fixed;
2. use the results attained both as initial guesses and upper bounds; add one or two more parameters on the optimization procedure: a maximum of four parameters are simultaneously considered at this point;
3. repeat 2 until the final step is reached: all the parameters are evaluated simultaneously.

Analysing the entire set of parameters, maximum uncertainties of interest are achieved:

$$\boldsymbol{\varepsilon_{max} = \xi_{max} \cdot \theta^0} \quad . \tag{2.17}$$

However, in Figure 2.3 a monotonic dependence is assumed between θ_i and $K_{M,j}$ and this is not always expected. If a non-monotonic dependence occurs, as in the case of Figure 2.4, outcomes obtained with this optimization framework, $\varepsilon_{i,max,b}$, are misleading since it provides only boundary values of uncertainty.

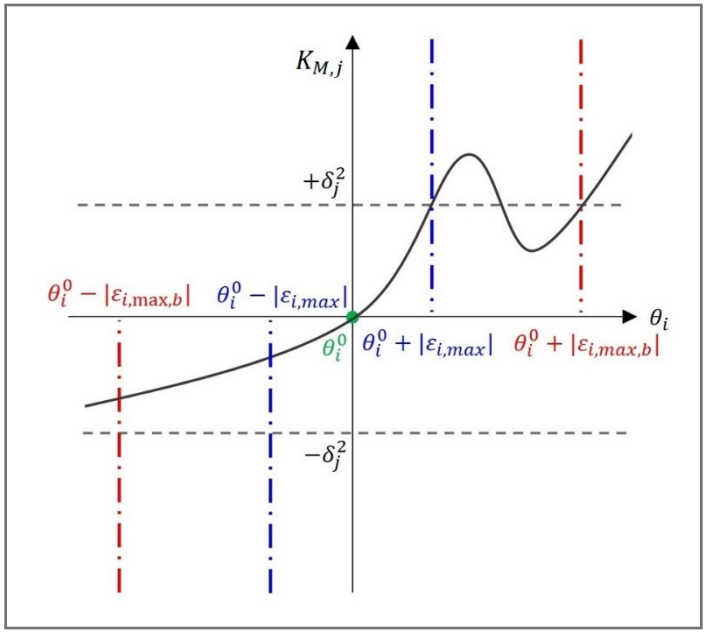

Figure 2.4. *Maximum uncertainty allowable, $\varepsilon_{i,max}$, on a parameter, θ_i, in order to respect the constraint on the key indicator variable, $K_{M,j}$, when a non-monotonic dependence between θ_i and $K_{M,j}$ is assumed.*

To solve this issue, additional scenarios are introduced where $0 \leq \left|\varepsilon_i^h\right| \leq \left|\varepsilon_{i,max}\right|$. Constraints on $\boldsymbol{K}_M$ must be ensured not only for the firsts 2^n scenarios, but for all scenarios, N. In this way, the attained uncertainties guarantee that the acceptable tolerances on $\boldsymbol{K}_M$ are respected for all combinations of parameters with uncertainties lower than $\boldsymbol{\varepsilon_{max}}$ (but not for those with higher ones).

The code structure on gPROMS® should adapt accordingly. The hierarchical sub-model decomposition is maintained, but on the lower-level model additional stochastic scenarios are evaluated. The definition for θ_i^h is as in Equation (2.13), where the following definition is adopted ν_i^h:

$$for\ 1 \leq i \leq n \quad \nu_i^h = \begin{cases} \{-1;+1\} \ \ s.t. \ \ \boldsymbol{\nu}^h \neq \boldsymbol{\nu}^{k \neq h} & for\ \ 1 \leq h \leq 2^n \\ uniform(-1,+1) & for\ \ 2^n < h \leq N \end{cases}, \tag{2.18}$$

where $uniform(-1,+1)$ indicates a uniform distribution of values in the range $[-1;+1]$, and N is the total number of scenarios. Since ξ_i is the same for all scenarios and ν_i^h is smaller in absolute terms for these new systems with respect to the first 2^n, situations with parameters higher than $\boldsymbol{\theta^0}$ and lower than $\boldsymbol{\theta^0} + |\boldsymbol{\varepsilon_{max}}|$ are analysed, guaranteeing that all $\boldsymbol{K}_M$ constraints are respected. Thus, if there are conditions for ε_i smaller than $\varepsilon_{i,max,b}$ where δ_j is no more ensured, another $\varepsilon_{i,max}$ distinct from $\varepsilon_{i,max,b}$ is provided by the optimization problem (Figure 2.4).

This second code structure is significantly more computational demanding: the more the scenarios evaluated, the more the time requested. Therefore, as discussed §2.2.3, it is exploited only as at a final level.

2.2.3. Procedure implementation

Once the optimization framework is defined, we must assess whether the current knowledge on parameter uncertainty is sufficient. Therefore, we need to compare optimization outcomes with parameter estimates, and build an iterative scheme capable to reach an acceptable model fidelity. The total procedure comprises a maximum of four steps:

1. MBDoE and experimentation;
2. parameter estimation;
3. optimization problem on 2^n scenarios;
4. optimization problem on N scenarios.

For each iteration a new experiment is performed to get both more precise and accurate parameters values. Specifically, the MBDoE is attained focusing on the most influential parameters specified by a preliminary Sobol analysis. Later, the optimization is executed on the estimates achieved at the same iteration, but this is done evaluating only the firsts 2^n boundaries scenarios and not the stochastic ones. Uncertainties given by the optimization are

later compared with the half of the 95% CIs, (95%CIs)/2, provided by the parameter estimation in order to assess whether the estimates precision is sufficient to reach the acceptable model fidelity. Even if for a single parameter a situation as the one presented in Figure 2.5 occurs, where $\varepsilon_{i,max,b}$ is lower than the confidence interval, another iteration is needed starting again with a new MBDoE. These iterations are repeated until all maximum uncertainties allowable on parameters are larger than (95%CIs)/2, as shown in Figure 2.6.

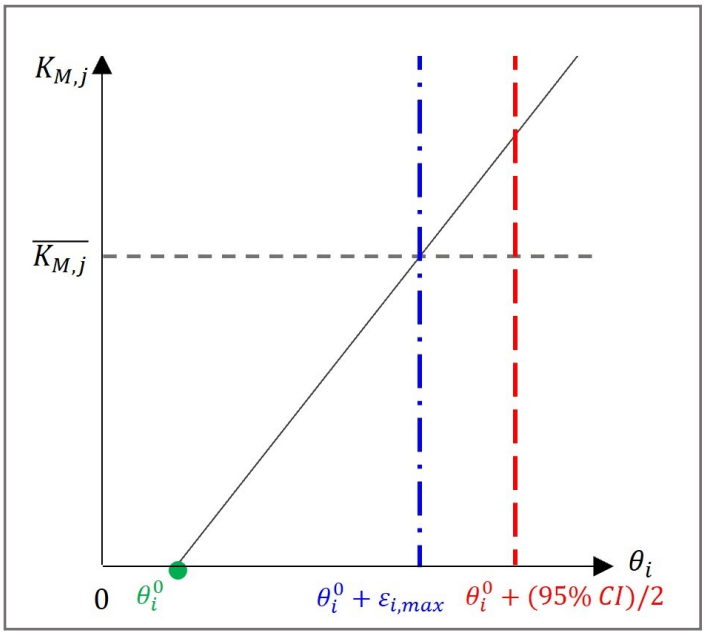

Figure 2.5. *Comparison between maximum uncertainty, $\varepsilon_{i,max}$, allowable on a parameter, θ_i, and the half of its 95% confidence interval: situation in which the former is lower than the latter.*

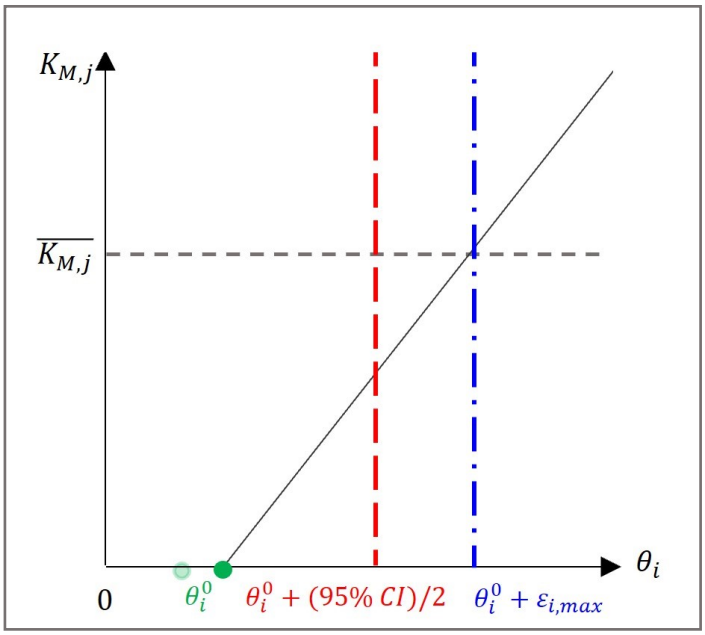

Figure 2.6. *Comparison between maximum uncertainty, $\varepsilon_{i,max}$, allowable on a parameter, θ_i, and the half of its 95% confidence interval: situation in which the former is higher than the latter.*

If a parameter is characterized by a $\varepsilon_{i,max,b}$ higher than the 50% of its estimated value, it is assumed that has very low influence on $\boldsymbol{K}_M$ and so it is omitted during the evaluation of the stop condition. Therefore, an upper bound for $\xi_{i,max}$ equal to 0.500 is set for those parameters.

At this point, the optimization with stochastic scenarios must be performed to ensure that $\boldsymbol{\varepsilon_{max}}$ found is acceptable. Two possible situations may occur:

1. $\boldsymbol{\varepsilon_{max}}$ obtained with the optimization on 2^n scenarios is satisfactory;
2. $\boldsymbol{\varepsilon_{max}}$ obtained with the optimization on 2^n scenarios is not satisfactory:
    a. the new $\boldsymbol{\varepsilon_{max}}$ achieved is larger than (95%CIs)/2;
    b. the new $\boldsymbol{\varepsilon_{max}}$ achieved is smaller than (95%CIs)/2.

The stop condition with the definitive $\boldsymbol{\varepsilon_{max}}$ is reached only if 1 or 2.a occurs, otherwise another iteration is needed.

The overall scheme of iterations is summarised in Figure 2.7.

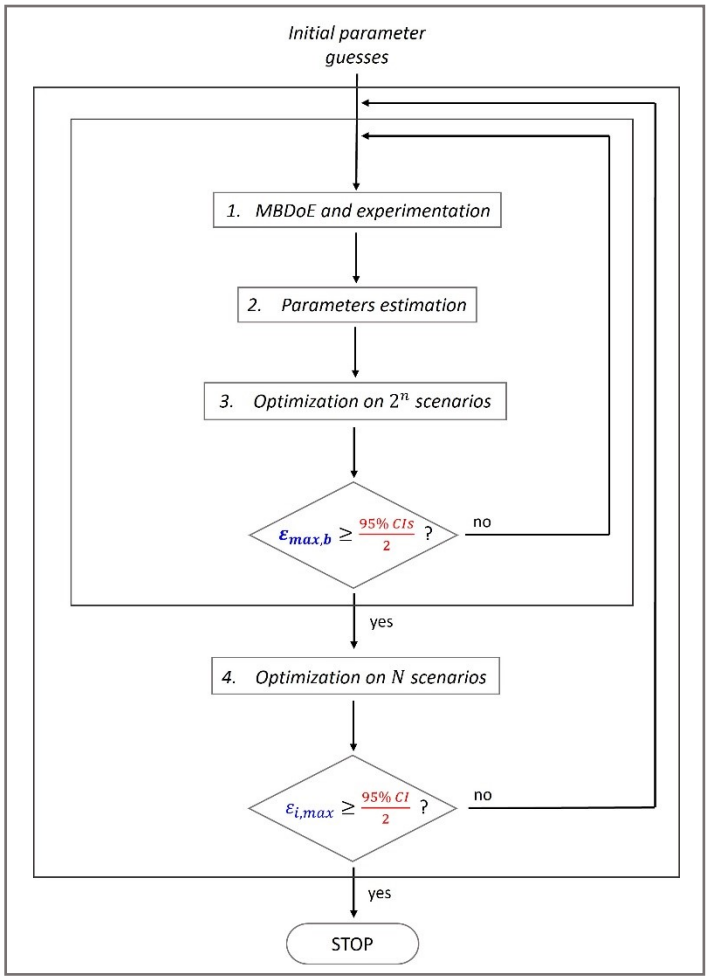

Figure 2.7. *Complete scheme of iterations: after guesses for parameters are assumed, the loop containing MBDoE, parameters estimation, and optimization is computed until the desired outcomes are achieved; later an optimization on N scenarios occurs.*

Since $\boldsymbol{\varepsilon_{max}}$ provides a situation in which acceptable tolerance on $\boldsymbol{K_M}$ is guaranteed, the state highlighted in Figure 2.6 ensures that the estimation on parameters is precise enough with respect to the same constraints on $\boldsymbol{K_M}$. Thus, the last iteration certifies a state where a satisfactory model prediction is achieved.

2.3. Application example

The methodology discussed in the previous section is applied to a biomass fermentation process for baker's yeast growth.

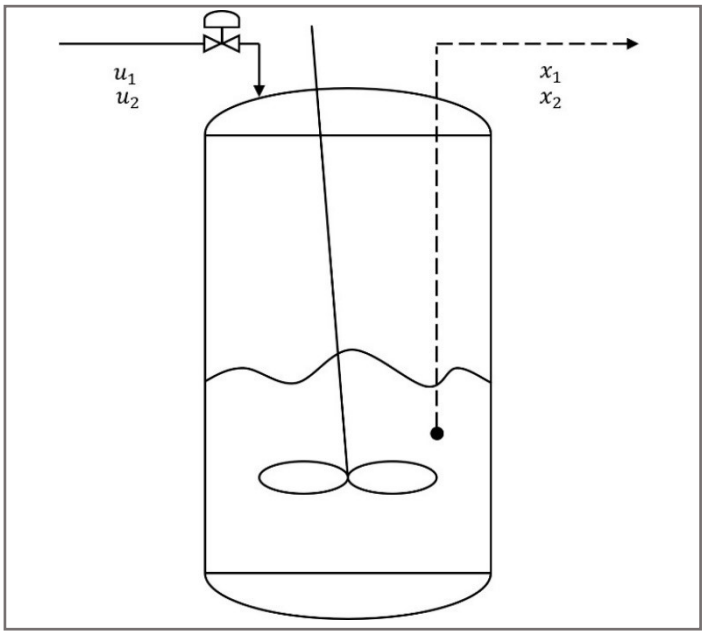

Figure 2.8. *Process scheme of a baker's yeast fermentation (adapted from Asprey and Macchietto (2002).*

2.3.1. Problem description

Assuming Monod-type kinetics for biomass growth and substrate consumption, the system is described by the set of DAEs presented in Galvanin *et al.* (2007):

$$\begin{cases} dx_1/dt = (r - u_1 - \theta_4) \cdot x_1 \\ dx_2/dt = -(r \cdot x_1/\theta_3) + u_1(u_2 - x_2) \\ r = (\theta_1 \cdot x_2)/(\theta_2 + x_2) \end{cases}, \tag{2.19}$$

where x_1 is the biomass concentration [g/L], x_2 is the substrate concentration [g/L], u_1 is the dilution factor [h^{-1}], and u_2 is the substrate concentration in the feed [g/L]. The experimental conditions that characterize a particular experiment are the initial biomass concentration $x_1^0 = [1.000; 10.00]$ g/L, the dilution factor $u_1 = [5.000 \times 10^{-2}; 0.200]$ h^{-1}, and the substrate

concentration in the feed $u_2 = [5.000; 35.00]$ g/L. The initial substrate concentration x_2^0 is set to 0 g/L and cannot be manipulated for experimental design purposes. Both x_1 and x_2 can be measured during the experiment, and so $\boldsymbol{x}(t) = \widehat{\boldsymbol{y}}(t)$. For the optimization purposes only x_1 is selected as key indicator. Thus, the definition of the optimization problem can be formulated according to Equation (2.9):

$$\begin{cases} dx_1/dt = \left(r - u_1 - (\theta_4^0 + \varepsilon_4)\right) \cdot x_1 \\ dx_2/dt = -(r \cdot x_1/(\theta_3^0 + \varepsilon_3)) + u_1(u_2 - x_2) \\ r = \left((\theta_1^0 + \varepsilon_1) \cdot x_2\right)/\left((\theta_2^0 + \varepsilon_2) + x_2\right) \\ x_1(0) = 2.453 \\ x_2(0) = 0 \\ \widehat{\boldsymbol{x}} = \boldsymbol{g}\left(\boldsymbol{x}(t)\right) \\ K_M = \widehat{x_1} \end{cases} . \tag{2.20}$$

Constraints on K_M are set as follows:

$$(K_M - \overline{K_M})^2 \leq (10\% \cdot \overline{K_M})^2 \quad , \tag{2.21}$$

where $\overline{K_M}$ is identified as x_1 at assumed process operating conditions but with "true" nominal values of parameters (Galvanin *et al.*, 2007). Thus, the true process is presented in Table 2.2,

Table 2.2. *Process settings at nominal conditions.*

Variable/Parameter	Units	Nominal
t	h	48.00
u_1	h^{-1}	0.190
u_2	g/L	34.79
θ_1	-	0.310
θ_2	-	0.180
θ_3	-	0.550
θ_4	-	5.000×10^{-2}

where t [h] is the time horizon of the process. At $t = 48$ h, $x_1 = \overline{K_M} = 14.89$ g/L, so:

Table 2.3. *Constraints on* x_1.

$x_{1,min}$	13.40 g/L
$x_{1,max}$	16.37 g/L

2.3.2. Problem resolution

The optimization problem (Equation (2.20)) requires initial guesses for model parameters. Average values within ranges given by Asprey and Macchietto (2002) are considered:

$$\theta_{1,2,3} = [5 \times 10^{-2};\ 0.980] \Rightarrow \theta_{1,2,3}^{i.g.} = 0.515 \quad , \tag{2.22}$$

$$\theta_4 = [1 \times 10^{-2};\ 0.980] \Rightarrow \theta_4^{i.g.} = 0.495 \quad . \tag{2.23}$$

Based on this, a MBDoE is performed: the experiment is a dynamic one with initial guess for t at 48 h and u_1 and u_2 settled as piecewise constant variables for ten time intervals. Global sensitivity analysis is carried out before the MBDoE (Table 2.4).

Table 2.4. *Sobol's sensitivity indices for parameters with respect to* x_1.

	$\boldsymbol{\theta_1}$	$\boldsymbol{\theta_2}$	$\boldsymbol{\theta_3}$	$\boldsymbol{\theta_4}$
$\boldsymbol{S_i}$	0.246	3.300×10^{-3}	0.565	0.104
$\boldsymbol{S_{T,i}}$	0.322	1.300×10^{-2}	0.623	0.164

As already underlined in §2.1.1, the index representing the first-order effect of parameters on the output should be lower with respect to the total-effect one. However, Table 2.4 shows that this condition is not verified for θ_2. The reason may be found in a strong correlation between parameters, while Sobol analysis considers them as independent factors (Song *et al.*, 2016). Therefore, for the specific case-study under evaluation, all parameters are considered during the first MBDoE. Later, the new experiment achieved is applied to the true process, the one defined by Table 2.2 conditions, in order to obtain a set of experimental data. A normally distributed noise, with a mean of zero and defined by the covariance matrix $\sum_y$, is added to the original dataset to reproduce the experimental error:

$$\sum_y = \begin{bmatrix} 1 \times 10^{-2} & 0 \\ 0 & 5 \times 10^{-2} \end{bmatrix} \quad , \tag{2.24}$$

Values attained are then used to perform a first estimation of parameters:

Table 2.5. *Estimation of model parameters: first iteration.*

	Units	**Value**	**95% CI**	**95% t-value**
$\boldsymbol{\theta_1^0}$	-	0.325	0.153	2.12
$\boldsymbol{\theta_2^0}$	-	0.326	0.386	0.844
$\boldsymbol{\theta_3^0}$	-	0.553	1.402×10^{-2}	39.43
$\boldsymbol{\theta_4^0}$	-	5.088×10^{-2}	2.706×10^{-3}	18.80
Reference t-value (95%)				1.660

Once values of $\boldsymbol{\theta^0}$ are acquired, these are used inside Equation (2.20). At this point, since values of parameters different with respect to the nominal ones are obtained, an optimization on the process manipulated variables must be performed ensuring that the desired $\overline{K_M}$ is achieved. The settings of the model after the first iteration are presented in Table 2.6.

Table 2.6. *Process settings after the first estimation of parameters.*

Variable/Parameter	Units	Value
t	h	48.00
u_1	h^{-1}	0.193
u_2	g/L	35.00
θ_1	-	0.325
θ_2	-	0.326
θ_3	-	0.553
θ_4	-	5.088×10^{-2}

Results attained in Table 2.5 are compared with 95% reference t-value: 1.660. In situations, as for θ_2, where the 95% t-value of the parameters is lower than the reference one, the parameters are not estimated precisely.

Thus, the procedure discussed in §2.2.2 is computed on estimated parameters obtained from the first iteration. Since the model is characterized by four parameters, we analyze a total of 16 possible combinations. For each scenario, the θ_i^h description is the one presented at Equation (2.13). Each $\boldsymbol{\theta}^h$ definition differs from the others for the $\boldsymbol{\nu}^h$ specification. All the $\boldsymbol{\nu}^h$ can be collected in a single matrix $\boldsymbol{C}$ to get a better interpretation of the different scenario. Therefore, for the specific case under analysis, $\boldsymbol{C}$ is structured as:

$$\boldsymbol{C} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & -1 & 1 \\ -1 & -1 & 1 & -1 \\ -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 \end{bmatrix} . \tag{2.25}$$

Once all the combinations are defined, the optimization is performed on ξ_i (Table 2.7).

Table 2.7. *Maximum relative percentage amounts, $\xi_{i,max}$, obtained for different parameters at the first iteration.*

	$\boldsymbol{\theta_1}$	$\boldsymbol{\theta_2}$	$\boldsymbol{\theta_3}$	$\boldsymbol{\theta_4}$
$\boldsymbol{\xi_{i,max}}$	7.109×10^{-2}	0.500*	4.546×10^{-2}	0.152

Consequently, following Equation (2.17), maximum uncertainties are obtained (Table 2.8).

Table 2.8. *Maximum uncertainties, $\varepsilon_{i,max,b}$, obtained for different parameters at the first iteration.*

	$\boldsymbol{\theta_1}$	$\boldsymbol{\theta_2}$	$\boldsymbol{\theta_3}$	$\boldsymbol{\theta_4}$
$\boldsymbol{\varepsilon_{i,max,b}}$	2.311×10^{-2}	0.163*	2.513×10^{-2}	7.752×10^{-3}

As expressed in §2.2.3, since $\xi_{2,max}$ (Table 2.7) is higher than 0.500, and so $\varepsilon_{2,max,b}$ is larger than the 50% of θ_2 estimated, this parameter is not considered during the final step of iteration. Table 2.9 reports the comparison between $\boldsymbol{\varepsilon_{max,b}}$ and (95%CIs)/2 of the first estimation (see Table 2.5).

Table 2.9. *First estimation-optimization iteration: uncertainties comparison.*

	$\boldsymbol{\theta_1}$	$\boldsymbol{\theta_3}$	$\boldsymbol{\theta_4}$
$\boldsymbol{\varepsilon_{i,max,b}}$	2.311×10^{-2}	2.513×10^{-2}	7.752×10^{-3}
(95%CI)/2	7.666×10^{-2}	7.011×10^{-3}	1.353×10^{-3}

The stop criterion is attained if all (95%CIs)/2 are lower than the maximum uncertainties allowable given by the optimization procedure. Clearly, this is not achieved after the first iteration for θ_1. Starting from the subsequent iteration, MBDoE is performed on θ_1 only. The desired outcome is achieved after a total of three iterations. The final estimation of model parameters is reported in Table 2.10.

Table 2.10. *Estimation of model parameters: last iteration.*

	Units	**Value**	**95% CI**	**95% t-value**
$\boldsymbol{\theta_1^0}$	-	0.327	5.826×10^{-3}	56.10
$\boldsymbol{\theta_2^0}$	-	0.307	4.668×10^{-2}	6.57
$\boldsymbol{\theta_3^0}$	-	0.557	7.125×10^{-3}	78.21
$\boldsymbol{\theta_4^0}$	-	5.194×10^{-2}	1.704×10^{-3}	30.49
Reference t-value (95%)				1.649

Again, new u_1 and u_2 are defined in order to keep x_1 at specification (Table 2.11).

Table 2.11. *Process settings after the last estimation of parameters.*

Variable/Parameter	**Units**	**Value**
t	h	48.00
u_1	h^{-1}	0.190
u_2	g/L	34.88
θ_1	-	0.327
θ_2	-	0.307
θ_3	-	0.557
θ_4	-	5.194×10^{-2}

Later, another optimization on ξ_i is carried out (Table 2.12).

Table 2.12. *Maximum relative percentage amounts, $\xi_{i,max}$, obtained for different parameters at the last iteration.*

	$\boldsymbol{\theta_1}$	$\boldsymbol{\theta_2}$	$\boldsymbol{\theta_3}$	$\boldsymbol{\theta_4}$
$\boldsymbol{\xi_{i,max}}$	8.359×10^{-2}	0.500*	4.577×10^{-2}	0.155

Results are stated in the form of maximum uncertainties (Table 2.13).

Table 2.13. *Maximum uncertainties, $\varepsilon_{i,max,b}$, obtained for different parameters at the last iteration.*

	$\boldsymbol{\theta_1}$	$\boldsymbol{\theta_2}$	$\boldsymbol{\theta_3}$	$\boldsymbol{\theta_4}$
$\boldsymbol{\varepsilon_{i,max,b}}$	2.732×10^{-2}	0.153*	2.55×10^{-2}	8.074×10^{-3}

Again, $\xi_{2,max}$ is higher than 0.500, so it is not included in the $\boldsymbol{\varepsilon_{max,b}}$ comparison to (95%CIs)/2 of the last estimation (Table 2.14).

Table 2.14. *Last estimation-optimization iteration: uncertainties comparison.*

	$\boldsymbol{\theta_1}$	$\boldsymbol{\theta_3}$	$\boldsymbol{\theta_4}$
$\boldsymbol{\varepsilon_{i,max,b}}$	2.732×10^{-2}	2.550×10^{-2}	8.074×10^{-3}
(95%CI)/2	2.913×10^{-3}	3.562×10^{-3}	8.518×10^{-4}

Since all (95%CIs)/2 are lower than the maximum uncertainties allowable given by the optimization procedure, the estimation might be acceptable. However, this condition must be verified, and this is done avoiding the possibility of situations like that of Figure 2.4 to occur. Therefore, a process simulation with more scenarios (e.g., 10^4) is considered. For all scenarios, the θ_i^h description is the one presented in Equation (2.13), where:

- ξ_i is the one outlined in Table 2.12 and is fixed for all the scenarios.
- v_i^h is the one presented in Equation (2.18).
- θ_i^0 is the one specified in Table 2.10 and is fixed for all the scenarios.

Since among all the scenarios $x_{1,min}$ and $x_{1,max}$ do not exceed the constraints imposed on x_1, a situation as the one presented in Figure 2.3 occurs. Thus, uncertainties highlighted in Table 2.13 are the maximum allowable on parameters in order to both ensure the acceptable tolerance on K_M and guarantee an adequate parameters prediction.

Moreover, another consideration might be added by focusing on Table 2.12: the lower the $\xi_{i,max}$, the smaller the uncertainty allowable on θ_i, and consequentially the more influent the parameter is on the K_M under investigation. The impact of model parameters towards the prediction of x_1 can be presented in the following decreasing order: θ_3, θ_1, θ_4 and θ_2. It is

important to notice how this result accurately reflects the scale obtained considering the total-effect indices, $S_{T,i}$, underlined by the Sobol analysis in Table 2.4

Chapter 3

Methodology application on direct compression process models

This chapter focuses on the assessment of model fidelity for pharmaceutical tableting manufacturing and quality. This is achieved through the application of the overall methodology described in §2.2 to direct compression (DC) process models.

3.1. Direct compression process models

In the pharmaceutical field, direct compression (DC) is a manufacturing process aimed at producing oral solid dose tablets, which benefits from using only a small number of unit operations that are continuously enhanced in order to streamline specific requests.
In this study, we are not considering blend or content uniformity or tablet weight variability. Only three models are evaluated (Figure 3.1):

- the tablet press unit operation (M_1);
- the tablet disintegration test unit (M_2);
- the *in vitro* dissolution test unit (M_3).

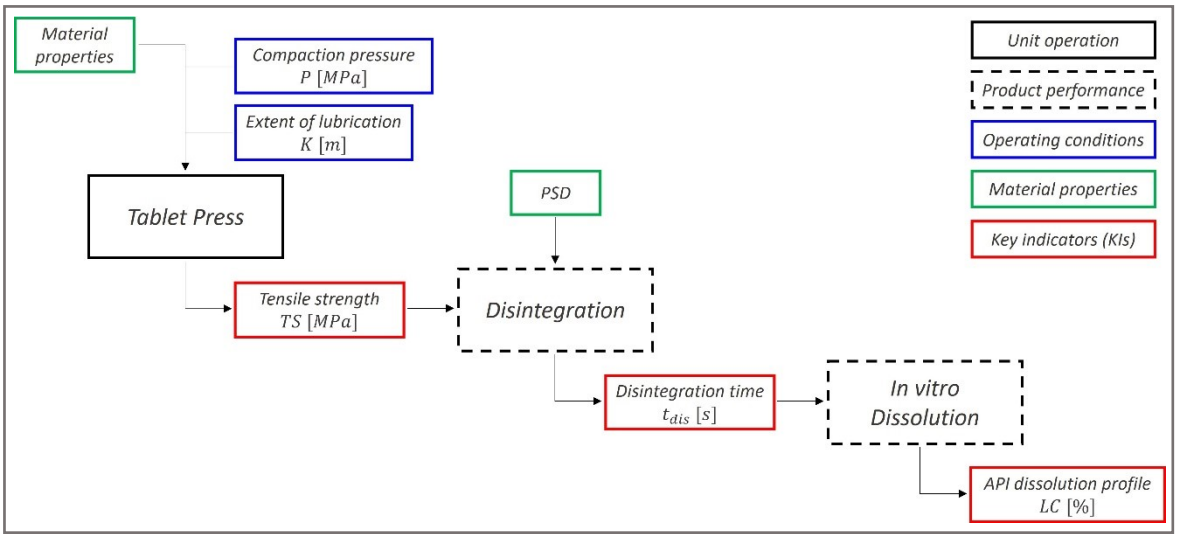

Figure 3.1. *Direct compression (DC) process configuration.*

Moreover, the tablet press model is the only model concerning a real unit operation in the manufacturing process; however, the methodology we propose is general and thus the results are not affected by omission of the other previous unit operations in a typical DC line. The

other models represent test units for the assessment of product quality, and require information from the tablet press model, i.e., the lubrication extent attained in the upstream powder blending, and the compaction pressure exerted by the press. In this example, each model outputs a KI, i.e., the tensile strength from the tablet press unit operation, the disintegration time from the tablet disintegration test unit, and the API dissolution profile from the *in vitro* dissolution test unit.

3.1.1. Direct compression process

The first step of the DC process, as shown in Figure 3.1, is a tablet press unit that leads to the compaction of tablets. Dies are filled with powder, which is compressed through rigid punches: these bring to a plastic deformation of the powder increasing the contact surface between particles and so cohesive forces capable to form a dense compact. Finally, the product is ejected from the die. The quality and behaviour of the drug product greatly rely on the extent of lubrication, which should be carefully controlled during development and scale-up. Lubrication is applied to enhance the pharmaceutical tablets ejection process by diminishing the wall friction between the tablet and the die walls, resulting in a reduced ejection force. It also improves powder flowability, diminishing the possibility of powder adhering to metal surfaces during tablet compression, decreases tablet hardness and tensile strength, expands disintegration time and reduces dissolution rate (Nassar *et al.*, 2021).
Therefore, the compaction pressure, P [MPa], and the extent of lubrication, K [m], must be set and controlled to ensure the product requirements (i.e., P and K are the two manipulated variables of the process). The output of interest is the tensile strength, TS [MPa], which is the first key indicator variable, K_{M_1}, of the process.
Pharmaceutical products need to fulfil strict regulations. Therefore, additional test units are considered (i.e., disintegration and dissolution test units). Particles initially disintegrate into granules, and then into primary particles: we assume the second mechanism to be very fast, and so negligible, with respect to the first one. The stages of tablet disintegration are complex and only partially understood (Markl *et al.*, 2017). From a simplified perspective, the first stage of tablet disintegration is water penetration, whose rate depends on both wettability and porosity of the tablet. For tablets with high drug loading and highly soluble API, water penetration is typically negligible, and tablet disintegration mainly occurs at the solid-liquid interface according to a purely erosion-driven mechanism. On the other hand, if the rate of water penetration is not negligible and water is able to penetrate into the tablet matrix, a swelling step initiates leading to the enlargement of particles and the creation of intra-tablet stresses that can cause the tablet break-up. The swelling rate generally depends on the given formulation, while the level of stress required to break up the tablet is related to the tablet tensile strength. For some formulations, an intermediate situation where the disintegration of the tablet occurs mostly with an erosion-driven mechanism, but swelling is not negligible

since the rate of water penetration into the tablet is relatively fast, may arise. In this scenario, swelling modifies the "apparent" volume of the tablet and increases the solid-liquid interface subject to the erosion mechanism (Bano *et al.*, 2022).

As presented in Figure 3.1, the tensile strength measured after the tablet press unit is an input for the disintegration model. The output of interest is the time required for disintegration from tablet to granule, t_{dis} [s], which corresponds to the second key indicator variable, K_{M_2}, of the process.

Dissolution testing, generally called *in vitro* dissolution (or release) testing, has emerged as a very important tool in the generic pharmaceutical industry. It is very widely used in formulation development and approval of generic solid oral dosage forms, in monitoring the manufacturing process and as a quality control test. Moreover, it is utilized to predict the *in vivo* performance of certain products. In developing a dissolution test for a generic product, investigators should consider the official methods and standards published in the United States Pharmacopeia (USP). The USP describes different dissolution apparatuses and techniques, which can be used to develop an appropriate dissolution method based on the drug product characteristics (Anand *et al.*, 2011). We refer to USP <701> dissolution test specifications. After the tablets disintegration into granules and then into primary particles, the rate of release and dissolution of each element of the formulation is analysed.

Previous information on tablet press and disintegration are used in this final stage where the API dissolved, expressed as percentage of the label content, LC [%], represents the key indicator variable, K_{M_3}.

3.1.2. Modular approach

Different workflows are possible in order to apply the optimization procedure presented in this study. Within this chapter, the optimization technique on the direct compression process models is implemented using a modular approach.

In a modular approach, the $\boldsymbol{K}_M$ of all units are targeted sequentially and the model parameters are estimated on a sub-system basis. Practically, the goal is to obtain a reliable prediction of $\boldsymbol{K}_M$ following the process layout. For instance, let us consider the process in Figure 3.1. Initially, the focus is on the tablet press unit operation, where the $K_{M,j}$ of interest is the tablet tensile strength (i.e., hardness). Once the tablet press model parameters are precise enough to guarantee sufficient tensile strength prediction fidelity, attention turns to the disintegration model. After achieving the required precision for the relevant parameters, the interest is focused on the API dissolution profile from the *in vitro* dissolution test unit. Since experiments are typically designed and implemented for one unit at a time, a modular approach can be convenient to organize the experimental campaign. On the other hand, it is assumed that the parametric precision attained on that unit will be sufficient to reach the $K_{M,j}$ fidelity required in subsequent units.

3.2. Tablet press unit operation

The tablet press is the first unit of the DC process on which the optimization procedure defined in §2.2 is computed.
The model is characterized by seven parameters: a_1 [MPa], a_2 [-], a_{sf} [-], b_1 [-], b_2 [-], b_{sf} [MPa^{-1}], and γ [dm^{-1}].

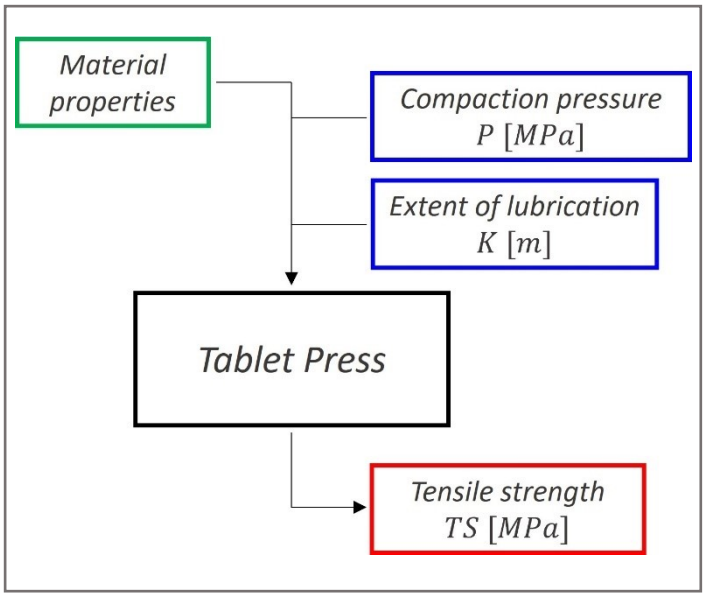

Figure 3.2. *Tablet press unit configuration.*

The schematic representation for the tablet press unit is shown in Figure 3.2.

3.2.1. Model for the tablet press unit operation

Although they do not capture the complex stress state and inhomogeneity that may occur during compaction, the compression/compaction process has been described by multiple semi-empirical models relating both pressure to solid fraction for compressibility and solid fraction to tensile strength for compactability (Nassar *et al.*, 2021).
Tablet compressibility measures the variation in tablet porosity due to the pressure exerted by the tablet press; different models have been proposed in the literature to quantify this relationship. In this study, the approach proposed by Kawakita is used:

$$\begin{cases} s_f = \left[a_{sf} \cdot (1 + b_{sf} \cdot P)\right] / (1 + a_{sf} \cdot b_{sf} \cdot P) \\ \varepsilon_t = 1 - s_f \end{cases}, \tag{3.1}$$

where s_f is the tablet solid fraction [-], and ε_t is the tablet porosity.
To account for the significance of lubrication, it is necessary to introduce a new variable, β [-], defined as the sensitivity of formulation to lubrication itself:

$$\beta = a_2 \cdot \left(1 - s_f\right) + b_2 \quad . \tag{3.2}$$

Finally, tablet compactability is expressed through the Kushner equations, capable to describe the effect of lubrication on the tensile strength:

$$\begin{cases} TS = TS_0 \cdot \left[(1-\beta) + \beta \cdot e^{(-\gamma \cdot 10 \cdot K)}\right] \\ TS_0 = a_1 \cdot e^{[b_1 \cdot (1 - s_f)]} \end{cases} . \tag{3.3}$$

For the optimization purposes, TS is selected as key indicator variable: K_{M_1}. Thus, the definition of the optimization problem can be formulated according to Equation (2.9):

$$\begin{cases} s_f = \left[\left(a_{sf} + \varepsilon_3\right) \cdot \left(1 + \left(b_{sf} + \varepsilon_6\right) \cdot P\right)\right] / \left[1 + \left(a_{sf} + \varepsilon_3\right) \cdot \left(b_{sf} + \varepsilon_6\right) \cdot P\right] \\ \beta = \left(a_2 + \varepsilon_2\right) \cdot \left(1 - s_f\right) + \left(b_2 + \varepsilon_5\right) \\ TS = TS_0 \cdot \left[(1-\beta) + \beta \cdot e^{(-(\gamma + \varepsilon_7) \cdot 10 \cdot K)}\right] \\ TS_0 = \left(a_1 + \varepsilon_1\right) \cdot e^{[(b_1 + \varepsilon_4) \cdot (1 - s_f)]} \\ \widehat{TS} = \boldsymbol{g}\left(\boldsymbol{x}(t)\right) \\ K_{M_1} = \widehat{TS} \end{cases} . \tag{3.4}$$

Constraints on K_{M_1} are set as follows:

$$\left(K_{M_1} - \overline{K_{M_1}}\right)^2 \leq \left(10\% \cdot \overline{K_{M_1}}\right)^2 \quad , \tag{3.5}$$

where $\overline{K_{M_1}}$ is identified as TS at assumed process operating conditions but with "true" nominal values of parameters selected from Nassar *et al.* (2021), as reported in Table 3.1.

Table 3.1. *Tablet press settings at nominal conditions.*

Variable/Parameter	Units	Nominal
K	m	99
P	MPa	214
a_1	MPa	11.04
a_2	-	1.091
a_{sf}	-	0.463
b_1	-	-8.202
b_2	-	0.326
b_{sf}	MPa^{-1}	2.460×10^{-2}
γ	dm^{-1}	1.211×10^{-3}

These conditions, which correspond to the ones of the true process, lead to $TS = \overline{K_{M_1}} =$ 2.000 MPa. Therefore, following Equation (3.5), bounds obtained for the tensile strength show $TS_{min} = 1.800$ MPa and $TS_{max} = 2.200$ MPa.

3.2.2. Quantification of parameters uncertainty for the tablet press unit operation

The optimization problem (Equation (3.4)) requires initial guesses for model parameters. These are randomly chosen inside a $\pm 50\%$ interval with respect to their nominal values (Table 3.2).

Table 3.2. *Initial guess,* $\theta_i^{i.g.}$*, for tablet press model parameters.*

	a_1	a_2	a_{sf}	b_1	b_2	b_{sf}	γ
$\theta_i^{i.g.}$	14.81	1.433	0.394	-6.287	0.242	1.710×10^{-2}	7.368×10^{-4}

The units of measurements are the ones in Table 3.1. Based on this, a MBDoE is performed. The tablet press is a steady-state model: at each experiment corresponds a single point of TS. However, since seven parameters need estimation from experimental measurements, the first parameters estimation requires at least seven points to supply CIs. Therefore, initially MBDoE is applied to design seven experiments to perform where different values are highlighted for the two time-invariant manipulated variables: the applied stress, P, and the extent of lubrication, K.

Global sensitivity analysis is carried out before the MBDoE; results are reported in Table 3.3, where the most influential model parameters are in boldface.

Table 3.3. *Sobol's sensitivity indices for parameters with respect to* TS*.*

	a_1	a_2	a_{sf}	b_1	b_2	b_{sf}	γ
S_i	3.309×10^{-2}	9.319×10^{-3}	**0.589**	**0.156**	7.126×10^{-2}	$\mathbf{8.124 \times 10^{-2}}$	8.805×10^{-3}
$S_{T,i}$	4.471×10^{-2}	1.004×10^{-2}	**0.630**	**0.177**	9.631×10^{-2}	$\mathbf{9.297 \times 10^{-2}}$	9.628×10^{-3}

Higher influence of a_{sf}, b_1, and b_{sf} is highlighted with respect to other parameters. In particular b_{sf} is selected instead of b_2 based on additional analyses discussed in Geremia *et al.* (2023). Therefore, for the specific case-study under evaluation, a_{sf}, b_1, and b_{sf} are considered during the first MBDoE. Later, manipulated variables obtained are applied to the true process, the one defined by Table 3.1 conditions, in order to obtain a set of experimental data; values attained are then used to perform a first estimation of parameters (Table 3.4).

Table 3.4. *Estimation of model parameters: first iteration.*

	Units	Value	95% CI	95% t-value
a_1^0	MPa	11.93	0.210	56.95
a_2^0	-	1.449	9.535×10^{-2}	15.19
a_{sf}^0	-	0.413	4.230×10^{-3}	97.61
b_1^0	-	-6.235	0.105	59.66
b_2^0	-	0.205	1.758×10^{-2}	11.67
b_{sf}^0	MPa^{-1}	1.890×10^{-2}	3.891×10^{-4}	48.56
γ^0	dm^{-1}	7.365×10^{-4}	2.050×10^{-4}	3.593
Reference t-value (95%)				1.943

Once values of $\boldsymbol{\theta^0}$ are acquired, these are used inside Equation 3.4. At this point, since new parameter values are obtained, an optimization on the process manipulated variables must be performed ensuring that the desired $\overline{K_{M_1}}$, the one of the true process, is achieved. This leads to $K = 99$ m and $P = 192$ MPa.

Thus, the procedure discussed in §2.2.2 is computed on estimated parameters obtained from the first iteration. Since the model is characterized by seven parameters, we analyze a total of 128 (i.e., 2^7) possible combinations.

Once all the combinations are defined, the optimization is performed on ξ_i (Table 3.5),

Table 3.5. *Maximum relative percentage amounts, $\xi_{i,max}$, obtained for different parameters at the first iteration.*

	a_1	a_2	a_{sf}	b_1	b_2	b_{sf}	γ
$\xi_{i,max}$	1.354×10^{-2}	6.046×10^{-3}	6.31×10^{-3}	9.328×10^{-3}	9.833×10^{-2}	1.363×10^{-2}	5.536×10^{-2}

and maximum uncertainties are obtained (Table 3.6).

Table 3.6. *Maximum uncertainties, $\varepsilon_{i,max,b}$, obtained for different parameters at the first iteration.*

	a_1	a_2	a_{sf}	b_1	b_2	b_{sf}	γ
$\varepsilon_{i,max,b}$	0.162	8.761×10^{-2}	2.573×10^{-3}	5.816×10^{-2}	2.016×10^{-2}	2.576×10^{-4}	4.077×10^{-5}

Table 3.7 reports the comparison between $\boldsymbol{\varepsilon_{max,b}}$ and (95%CIs)/2 of the first estimation (see Table 3.4).

Table 3.7. *First estimation-optimization iteration: uncertainties comparison.*

	a_1	a_2	a_{sf}	b_1	b_2	b_{sf}	γ
$\varepsilon_{i,max,b}$	0.162	8.761×10^{-2}	2.573×10^{-3}	5.816×10^{-2}	2.016×10^{-2}	2.576×10^{-4}	4.077×10^{-5}
(95%CI)/2	0.105	4.768×10^{-2}	2.115×10^{-3}	5.226×10^{-2}	8.791×10^{-3}	1.946×10^{-4}	1.025×10^{-4}

The stop criterium is attained if all (95%CIs)/2 are lower than the maximum uncertainties given by the optimization procedure. As can be noticed in Table 3.7, this is not achieved for γ. Therefore, starting from the subsequent iteration, MBDoE is performed on γ only, designing a new optimal experiment. Every new experiment, together with the old ones, will be used to estimate all parameters.
The desired outcome is achieved after a total of five iterations. The final estimation of model parameters is reported in Table 3.8.

Table 3.8. *Estimation of model parameters: last iteration.*

	Units	Value	95% CI	95% t-value
a_1^0	MPa	11.04	0.133	82.75
a_2^0	-	1.078	8.810×10^{-2}	12.23
a_{sf}^0	-	0.456	3.332×10^{-3}	136.98
b_1^0	-	-8.097	0.108	74.68
b_2^0	-	0.326	1.023×10^{-2}	31.82
b_{sf}^0	MPa^{-1}	2.494×10^{-2}	3.773×10^{-4}	66.11
γ^0	dm^{-1}	1.210×10^{-3}	1.172×10^{-4}	10.3220
Reference t-value (95%)				1.812

Again, new K and P are defined in order to keep TS at specification, leading to $K = 99$ m and $P = 213$ MPa.
Later, another optimization on ξ_i is carried out (Table 3.9).

Table 3.9. *Maximum relative percentage amounts, $\xi_{i,max}$, obtained for different parameters at the last iteration.*

	a_1	a_2	a_{sf}	b_1	b_2	b_{sf}	γ
$\xi_{i,max}$	1.353×10^{-2}	8.117×10^{-2}	6.223×10^{-3}	1.060×10^{-2}	4.190×10^{-2}	1.376×10^{-2}	5.296×10^{-2}

Results are stated in the form of maximum uncertainties (Table 3.10).

Table 3.10. *Maximum uncertainties, $\varepsilon_{i,max,b}$, obtained for different parameters at the last iteration.*

	a_1	a_2	a_{sf}	b_1	b_2	b_{sf}	γ
$\varepsilon_{i,max,b}$	0.150	8.747×10^{-2}	2.841×10^{-3}	8.581×10^{-2}	1.364×10^{-2}	3.432×10^{-4}	6.408×10^{-5}

The comparison between $\boldsymbol{\varepsilon_{max,b}}$ and (95%CIs)/2 after the last estimation is reported in Table 3.11.

Table 3.11. *Last estimation-optimization iteration: uncertainties comparison.*

	a_1	a_2	a_{sf}	b_1	b_2	b_{sf}	γ
$\varepsilon_{i,max,b}$	0.150	8.747×10^{-2}	2.841×10^{-3}	8.581×10^{-2}	1.364×10^{-2}	3.432×10^{-4}	6.408×10^{-5}
(95%CI)/2	6.674×10^{-2}	4.405×10^{-2}	1.666×10^{-3}	5.421×10^{-2}	5.117×10^{-3}	1.886×10^{-4}	5.862×10^{-5}

At this point all the (95%CIs)/2 are lower than the acceptable maximum uncertainty for each parameter. However, according to the algorithm (Figure 2.7 in §2.2.3) an additional check is required by running the optimization on N scenarios.

Results obtained show that evaluating all the scenarios TS_{min} and TS_{max} do not exceed the constraints imposed on TS. Hence, the uncertainties presented in Table 3.10 are the maximum allowable in order to both ensure the acceptable tolerance on K_{M_1} and guarantee a faithful parameters prediction.

Moreover, focusing on Table 3.9, the lowest relative uncertainties are for a_{sf} and b_1. As stated in §2.3.2, the lower the $\xi_{i,max}$, the more influent the parameter θ_i on the KI under investigation, TS. Consequently, a_{sf} and b_1 represent the most influential parameters for K_{M_1}, and this reflects the outcomes obtained by the Sobol analysis in Table 3.3.

3.3. Disintegration test unit

The analysis of the DC process proceeds with the second unit: disintegration.

This model is characterized by five parameters: C_2 [MPa], C_3 [MPa], $\dot{\epsilon}$ [m s^{-1}], n_s [-], and S_p [-].

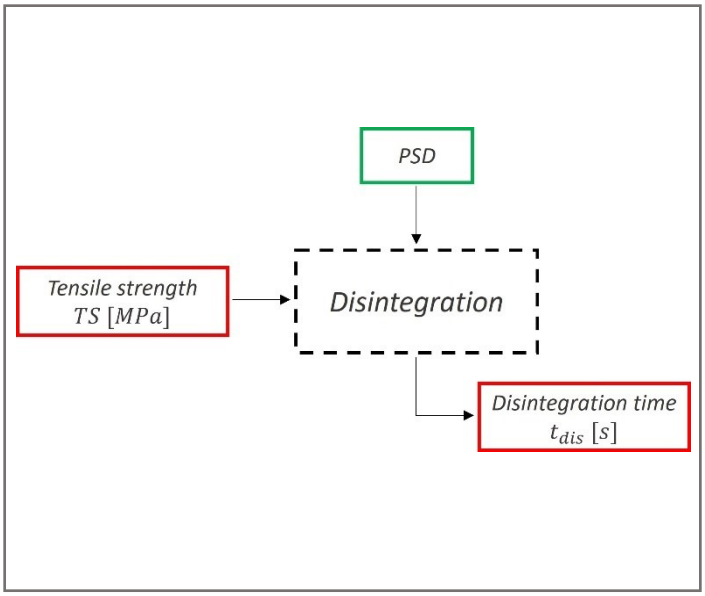

Figure 3.3. *Disintegration unit configuration.*

Figure 3.3 represents a schematic depiction of the disintegration test unit. The dependence on the tensile strength underlines a connection with the tablet press model. Therefore, following a modular approach, calculations on disintegration unit are performed considering the values of tablet press parameters already obtained (Table 3.8).

3.3.1. Model for the disintegration test unit

Both erosion and swelling are considered (Markl *et al.*, 2017). The first one is described as:

$$V_c = (H_{coat} - \dot{\epsilon} \cdot t) \cdot A_t \quad , \tag{3.6}$$

where V_c [m^3] is the coating volume varying with time t [s], H_{coat} [m] is the thickness of the coating layer, A_t [m^2] is the tablet surface area. The dynamic evolution of the penetration depth due to swelling is modelled as:

$$dP_d/dt = [P/(F_L/A_t)]^{n_s \cdot (T_{t/2} - P_d)/T_{t/2}} \cdot \left[\left({d_h}^2 \cdot \varepsilon_s\right)/\left(S_p \cdot {\tau_{or}}^2 \cdot \mu \cdot P_d\right)\right] \cdot p_c \quad , \tag{3.7}$$

where P_d [m] is the water penetration depth, F_L[MN] is the fracture load, d_h [m] is the tablet hydraulic diameter, τ_{or} [-] is the average tablet tortuosity, μ [Pa s] is the liquid viscosity, p_c [Pa] is the capillary pressure, $T_{t/2}$ [m] is the time-dependent half tablet thickness, and ε_s [-] represents the average porosity of the swollen product. The stress due to tablet expansion from swelling is defined according to Peppas and Colombo (1989):

$$\tau = -TS + C_2 \cdot w_l + C_3 \cdot \sqrt{w_l} \quad , \tag{3.8}$$

where τ [MPa] is the total stress, and w_l [-] is the liquid content in the tablet. From τ it is possible to compute ε_s to be included in dP_d/dt equation:

$$\tau = [G_0 \cdot exp[-(E \cdot \varepsilon_s)/(1 - \varepsilon_s)] \cdot \lambda \cdot t]/T_{t/2} \quad , \tag{3.9}$$

where G_0 [MPa] and E [-] are elastic constants, and λ [s^{-1}] is the swelling rate (Kuentz and Luenberger, 1998).

The disintegration time, t_{dis}, is defined as the time for which the tablet stops disintegrating (i.e., $dP_d/dt = 0$), and for the optimization purposes it is selected as the key indicator variable, K_{M_2}, for the model under investigation.

The optimization problem is formulated according to the structure of Equation (2.9):

$$\begin{cases} V_c = (H_{coat} - (\dot{\epsilon} + \varepsilon_3) \cdot t) \cdot A_t \\ \frac{dP_d}{dt} = \left[P/\left(\frac{F_L}{A_t}\right)\right]^{(n_s+\varepsilon_4)\cdot[(T_{t/2}-P_d)/T_{t/2}]} \cdot \left[\left(d_h{}^2 \cdot \varepsilon_s\right)/\left(\left(S_p + \varepsilon_5\right) \cdot \tau_{or}{}^2 \cdot \mu \cdot P_d\right)\right] \cdot p_c \\ \tau = -TS + (C_2 + \varepsilon_1) \cdot w_l + (C_3 + \varepsilon_2) \cdot \sqrt{w_l} \\ \tau = [G_0 \cdot exp[-(E \cdot \varepsilon_s)/(1 - \varepsilon_s)] \cdot \lambda \cdot t]/T_{t/2} \\ P_d(0) = 1.000 \times 10^{-6} \\ \widehat{t_{dis}} = \boldsymbol{g}(\boldsymbol{x}(t)) \\ K_{M_2} = \widehat{t_{dis}} \end{cases} . \quad (3.10)$$

Constraints on K_{M_2} are set as follows:

$$\left(K_{M_2} - \overline{K_{M_2}}\right)^2 \leq \left(25\% \cdot \overline{K_{M_2}}\right)^2 \quad , \quad (3.11)$$

where $\overline{K_{M_2}}$ is identified as t_{dis} at assumed process operating conditions but with "true" nominal values of parameters (Table 3.12).

Table 3.12. *Disintegration settings at nominal conditions.*

Parameter	Units	Nominal
C_2	MPa	1.000×10^2
C_3	MPa	1.000×10^2
$\dot{\varepsilon}$	m s^{-1}	1.000×10^{-3}
n_s	-	0.900
S_p	-	0.524

These conditions, which correspond to the ones of the true process, combined with those expressed in Table 3.1 for tablet press, lead to $t_{dis} = \overline{K_{M_2}} = 243$ s. Therefore, following Equation 3.11: $t_{dis,min} = 180$ s; $t_{dis,max} = 300$ s.

3.3.2. Quantification of parameters uncertainty for the disintegration test unit

As for the tablet press unit, the initial guesses for disintegration parameters are arbitrarily chosen inside a $\pm$50% interval with respect to their nominal values (Table 3.13).

Table 3.13. *Initial guess,* $\theta_i^{i.g.}$*, for disintegration model parameters.*

	C_2	C_3	$\dot{\epsilon}$	n_s	S_p
$\theta_i^{i.g.}$	63.00	1.410×10^2	1.300×10^{-3}	1.019	0.688

The units of measurements are the ones in Table 3.12. Taking this into account, a MBDoE is carried out. The disintegration is a dynamic model where, differently from the tablet press,

each experiment provides a trajectory of the key indicator variable of interest. Therefore, even if the model includes more than one parameter, a single experiment is sufficient to perform a first estimation. Since the dependence of the disintegration section to the tablet press, the values of manipulated variables provided by the MBDoE are those of previous unit: P and K.
Global sensitivity analysis is carried out before the MBDoE; results are reported in Table 3.14, where the most influential model parameters are in boldface.

Table 3.14. *Sobol's sensitivity indices for parameters with respect to* t_{dis}.

	C_2	C_3	$\dot{\epsilon}$	n_s	S_p
S_i	2.815×10^{-4}	2.209×10^{-4}	3.603×10^{-4}	**0.641**	$\mathbf{1.312 \times 10^{-2}}$
$S_{T,i}$	3.992×10^{-4}	3.731×10^{-4}	5.042×10^{-4}	**0.881**	**0.101**

Therefore, for the case-study under evaluation, n_s and S_p are considered during the first MBDoE. The time-invariant values of P and K resulting from the design of experiments are then applied to the true process, the one specified by parameters in Table 3.12. Later, based on the set of experimental data attained, a first estimation of parameters is performed (Table 3.15).

Table 3.15. *Estimation of model parameters: first iteration.*

	Units	Value	95% CI	95% t-value
C_2^0	MPa	52.23	7.475×10^{5}	6.987×10^{-5}
C_3^0	MPa	99.93	1.649×10^{5}	6.060×10^{-4}
$\dot{\epsilon}^0$	m s^{-1}	1.480×10^{-3}	0.373	3.969×10^{-3}
n_s^0	-	0.905	0.184	4.920
S_p^0	-	0.535	0.547	0.978
Reference t-value (95%)				1.683

Then, the optmization on 2^n boundary scenarios is computed. The model is characterized by five parameters, so a total of 32 (i.e., 2^5) possible combinations is evaluated.
Once all the combinations are defined, the optimization is performed on ξ_i (Table 3.16),

Table 3.16. *Maximum relative percentage amounts,* $\xi_{i,max}$*, obtained for different parameters at the first iteration.*

	C_2	C_3	$\dot{\epsilon}$	n_s	S_p
$\xi_{i,max}$	0.500*	0.500*	0.500*	1.855×10^{-2}	0.106

and maximum uncertainties are obtained (Table 3.17).

Table 3.17. *Maximum uncertainties, $\varepsilon_{i,max,b}$, obtained for different parameters at the first iteration.*

	C_2	C_3	$\dot{\epsilon}$	n_s	S_p
$\boldsymbol{\varepsilon_{i,max,b}}$	26.11*	49.97*	7.402×10^{-4}*	1.679×10^{-2}	5.686×10^{-2}

Table 3.17 shows that for C_2, C_3, and $\dot{\epsilon}$ the $\xi_{i,max}$ is higher than 0.500. Thus, as expressed in §2.2.3, since for these parameters the maximum uncertainty allowable, $\varepsilon_{i,max,b}$, is larger than the 50% of θ_i^0, they are not considered during the final step of iteration. Table 3.18 reports the comparison between $\boldsymbol{\varepsilon_{max,b}}$ and (95%CIs)/2 of the first estimation (see Table 3.15).

Table 3.18. *First estimation-optimization iteration: uncertainties comparison.*

	n_s	S_p
$\boldsymbol{\varepsilon_{i,max,b}}$	1.679×10^{-2}	5.686×10^{-2}
(95%CI)/2	9.199×10^{-2}	0.274

The first iteration is not sufficient to reach the stop condition: $\varepsilon_{i,max,b}$ is lower than (95%CI)/2 for both n_s and S_p. Therefore, also for the next iteration the MBDoE is performed considering n_s and S_p.

The desired outcome is achieved after a total of five iterations. The final estimation of model parameters is reported in Table 3.19.

Table 3.19. *Estimation of model parameters: last iteration.*

	Units	**Value**	**95% CI**	**95% t-value**
$\boldsymbol{C_2^0}$	MPa	88.72	3.895×10^{4}	2.278×10^{-3}
$\boldsymbol{C_3^0}$	MPa	99.17	2.174×10^{3}	4.561×10^{-2}
$\boldsymbol{\dot{\epsilon}^0}$	m s^{-1}	9.835×10^{-4}	6.964×10^{-2}	1.412×10^{-2}
$\boldsymbol{n_s^0}$	-	0.903	3.491×10^{-2}	25.86
$\boldsymbol{S_p^0}$	-	0.533	0.110	4.853
Reference t-value (95%)				1.652

Another optimization on ξ_i is carried out (Table 3.20).

Table 3.20. *Maximum relative percentage amounts, $\xi_{i,max}$, obtained for different parameters at the last iteration.*

	C_2	C_3	$\dot{\epsilon}$	n_s	S_p
$\boldsymbol{\xi_{i,max}}$	0.500*	0.500*	0.500*	2.001×10^{-2}	0.116

Results are stated in the form of maximum uncertainties (Table 3.21).

Table 3.21. *Maximum uncertainties, $\varepsilon_{i,max,b}$, obtained for different parameters at the last iteration.*

	C_2	C_3	$\dot{\epsilon}$	n_s	S_p
$\boldsymbol{\varepsilon_{i,max,b}}$	44.36*	49.58*	4.918×10^{-4}*	1.806×10^{-2}	6.178×10^{-2}

Again, $\xi_{i,max}$ is higher than 0.500 for C_2, C_3, and $\dot{\epsilon}$, so these parameters are not included in the $\boldsymbol{\varepsilon_{max,b}}$ comparison to (95%CIs)/2 of the last estimation (Table 3.22).

Table 3.22. *Last estimation-optimization iteration: uncertainties comparison.*

	n_s	S_p
$\boldsymbol{\varepsilon_{i,max,b}}$	1.806×10^{-2}	6.178×10^{-2}
(95%CI)/2	1.746×10^{-2}	5.495×10^{-2}

The outcomes achieved suggest an acceptable estimation: all (95%CIs)/2 are lower than the maximum uncertainties allowable given by the optimization procedure.
Nevertheless, this result should be verified through the optimization on N scenarios, which confirms that the findings $t_{dis,min}$ and $t_{dis,max}$ are fulfilled for all the scenarios, and so the stop condition is reached.
Furthermore, focusing on Table 3.20, the lowest relative uncertainties are for n_s and S_p. Therefore, as stated in §2.3.2, these parameters are the most influential on K_{M_2}; this result reflects the outcomes obtained by the Sobol analysis in Table 3.14.

3.4. In vitro dissolution test unit

Finally, the *in vitro* dissolution is the last unit evaluated for the DC process.
The model is characterized by two parameters: k_{API} [m s^{-1}], and n_{API} [-].
As for the dissolution model, also in this case there is a dependence on previous units. Therefore, following a modular approach, calculations on the *in vitro* dissolution unit are performed considering both tablet press and disintegration values of parameters as the ones achieved at their last iteration, respectively presented in Table 3.8 and Table 3.19.

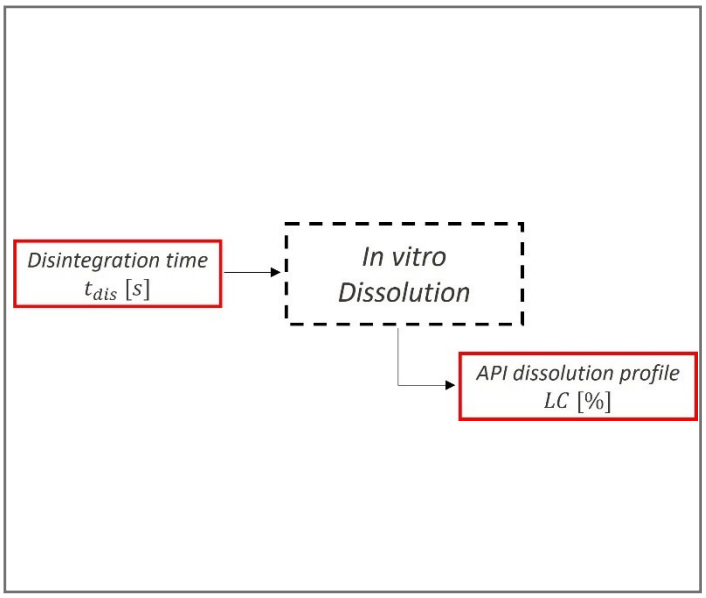

Figure 3.4. *Disintegration unit configuration.*

A schematic representation of the *in vitro* dissolution unit is shown in Figure 3.4.

3.4.1. Model for the in vitro dissolution test unit

The rate of release and dissolution of each element can be modelled using a population balance approach as proposed by Wilson *et al.* (2012). For this purpose, the individual components of the tablets are divided in three groups: API, soluble excipients, and insoluble excipients.

Focusing only on the API, the dynamic evolution of the number of particles, N_{API}, is modelled as:

$$\partial N_{API}/\partial t = B_{API} \cdot \acute{\delta} \cdot \left(l - l_{0,API}\right) + R_{API,l} \cdot (\partial N_{API}/\partial t) \quad , \tag{3.12}$$

where B_{API} [s^{-1}] is the rate of release of API from the tablet, l [m] is the particle size at given t, $l_{0,API}$ [m] is the initial particle size of API, R_{API} [m^2 s^{-1}] is the API dissolution coefficient, and $\acute{\delta}$ is the Dirac delta function. Following the assumptions discussed in Bano *et al.* (2022), B_{API} can be expressed as:

$$B_{API} = \rho_p^{-1} \cdot \left[x_{API}/\left(\phi \cdot l_{0,API}^3\right)\right] \cdot (\partial M_t/\partial t) \quad , \tag{3.13}$$

where x_{API} [-] is the mass fraction of the API, ρ_p [kg m^{-3}] is the particles density, ϕ [-] is the shape factor of particles ($\phi = \pi/6$ for spherical particles), and M_t [kg] is the tablet mass. The API dissolution coefficient, R_{API}, can be described as:

$$R_{API,l} = k_{API} \cdot (c_{sat} - c_{API})^{n_{API}} \quad , \tag{3.14}$$

where c_{sat} [kg m^{-3}] is the API saturation concentration, and c_{API} [kg m^{-3}] is the API bulk concentration. the liquid content in the tablet. Usually, the extent of dissolution is defined as the percentage of dissolved label content, LC [%]:

$$LC = 100 \cdot \left[(c_{API} \cdot V_m)/\left(x_{API} \cdot M_{t,0}\right)\right] \quad , \tag{3.15}$$

where V_m [m^3] is the liquid volume in the test vessel, and $M_{t,0}$ [kg] is the initial mass of the tablet.

For the optimization purposes, LC is selected as the *in vitro* dissolution model key indicator variable, K_{M_3}.

The optimization problem is formulated as follow:

$$\begin{cases} \partial N_{API}/\partial t = B_{API} \cdot \delta \cdot \left(l - l_{0,API}\right) + R_{API,l} \cdot (\partial N_{API}/\partial t) \\ B_{API} = \rho_p^{-1} \cdot \left[x_{API}/\left(\phi \cdot l_{0,API}^3\right)\right] \cdot (\partial M_t/\partial t) \\ R_{API,l} = (k_{API} + \varepsilon_1) \cdot (c_{sat} - c_{API})^{(n_{API}+\varepsilon_2)} \\ LC = 100 \cdot \left[(c_{API} \cdot V_m)/\left(x_{API} \cdot M_{t,0}\right)\right] \\ LC(0) = 0 \\ \widehat{LC} = \boldsymbol{g}(\boldsymbol{x}(t)) \\ K_{M_3} = \widehat{LC} \end{cases} . \tag{3.16}$$

Moreover, constraints on K_{M_3} are set:

$$\left(\overline{K_{M_3}} - 15\%\right) \leq K_{M_3} \leq \overline{K_{M_3}} \quad , \tag{3.17}$$

where $\overline{K_{M_3}}$ is identified as LC at assumed process operating conditions but with "true" nominal values of parameters (Table 3.23).

Table 3.23. *In vitro dissolution settings at nominal conditions.*

Parameter	Units	Nominal
k_{API}	m s^{-1}	2.300×10^{-12}
n_{API}	-	1.000

These conditions, which correspond to the ones of the true process combined with those expressed in Table 3.1 and in Table 3.12 for tablet press and disintegration unit respectively, lead to $LC = \overline{K_{M_3}} = 80$ %. Thus, following Equation (3.17): $LC_{min} = 65$ % and $LC_{max} = 80$ %.

3.4.2. Quantification of parameters uncertainty for the in vitro dissolution test unit

As for the tablet press unit, the initial guesses for disintegration parameters are arbitrarily chosen inside a $\pm 50\%$ interval with respect to their nominal values (Table 3.24).

Table 3.24. *Initial guess,* $\theta_i^{i.g.}$*, for disintegration model parameters.*

	k_{API}	n_{API}
$\theta_i^{i.g.}$	2.996×10^{-12}	0.762

The units of measurements are the ones in Table 3.23. Based on this, a MBDoE is executed. Similarly to disintegration, *in vitro* dissolution is described by a dynamic model where, differently from the tablet press, each experiment provides a trajectory of the key indicator variable of interest. Therefore, even if the model includes more than one parameter, a single experiment is sufficient to perform a first estimation. Since the dependence to the tablet press, the values of manipulated variables provided by the MBDoE are those of previous unit: P and K.

Global sensitivity analysis is carried out before the MBDoE; results are reported in Table 3.25, where the most influential model parameters are in boldface.

Table 3.25. *Sobol's sensitivity indices for parameters with respect to LC.*

	k_{API}	n_{API}
S_i	**0.275**	1.165×10^{-2}
$S_{T,i}$	**0.353**	1.333×10^{-2}

Therefore, the first MBDoE is performed considering only k_{API}. Later, the new experiment achieved is applied to the true process, the one defined by Table 3.23 conditions. Based on the set of experimental data attained, a first estimation of parameters is performed (Table 3.26).

Table 3.26. *Estimation of model parameters: first iteration.*

	Units	Value	95% CI	95% t-value
k_{API}^0	m s^{-1}	2.226×10^{-12}	9.432×10^{-16}	2.360×10^{3}
n_{API}^0	-	1.100	1.521×10^{-3}	9.538×10^{2}
Reference t-value (95%)				1.652

Then, the optmization on 2^n boundary scenarios is computed. The model is characterized by two parameters, so a total of 4 (i.e., 2^2) possible combinations is evaluated.

Once all the combinations are defined, the optimization is performed on ξ_i (Table 3.27),

Table 3.27. *Maximum relative percentage amounts, $\xi_{i,max}$, obtained for different parameters at the first iteration.*

	k_{API}	n_{API}
$\xi_{i,max}$	1.882×10^{-3}	4.950×10^{-3}

and maximum uncertainties are obtained (Table 3.28).

Table 3.28. *Maximum uncertainties, $\varepsilon_{i,max,b}$, obtained for different parameters at the first iteration.*

	k_{API}	n_{API}
$\varepsilon_{i,max,b}$	4.163×10^{-15}	5.446×10^{-3}

Table 3.29 reports the comparison between $\boldsymbol{\varepsilon_{max,b}}$ and (95%CIs)/2 of the first estimation (see Table 3.26).

Table 3.29. *First estimation-optimization iteration: uncertainties comparison.*

	k_{API}	n_{API}
$\varepsilon_{i,max,b}$	4.163×10^{-15}	5.446×10^{-3}
(95%CI)/2	4.716×10^{-16}	5.767×10^{-4}

Results obtained show that a single iteration is sufficient to reach a situation in which $\boldsymbol{\varepsilon_{max,b}}$ is greater than (95%CIs)/2 for all parameters.
Once again, this condition must be verified: the optimization is performed considering also stochastic scenarios.
The findings attained underline that LC_{min} and LC_{max} do not exceed the constraints imposed on LC. Thus, the uncertainties presented in Table 3.28 are the maximum allowable on parameters in order to both ensure the acceptable tolerance on K_{M_3} and guarantee a faithful parameters prediction.
Furthermore, focusing on Table 3.27, the lowest relative uncertainties is for k_{API}. Consequentially, as stated in §2.3.2, this is the most influential parameter on K_{M_3} and this reflects the outcomes obtained by the Sobol analysis in Table 3.25.

3.5. Comments on results

The application of the optimization procedure on DC process models provides satisfying results: for all the units a condition where $\boldsymbol{\varepsilon_{max}}$ is greater than (95%CIs)/2 is obtained. The number of experiments required to reach the stop condition for each unit is summarized in Table 3.30.

Table 3.30. *Optimization procedure: number of experiments required to reach the stop condition for each unit.*

Unit	Exp
Tablet press	11
Disintegration	5
Dissolution	1

As discussed in §1.2, methodologies based on multi-variate statistical methods represent another approach to ensure whether or not an acceptable model fidelity is achieved. Table 3.31 summarises the number of experiments required when the methodology proposed by Geremia *et al.* (2023) was implemented.

Table 3.31. *Methodologies based on multi-variate statistical methods: number of experiments required to reach the stop condition for each unit.*

Unit	Exp
Tablet press	8
Disintegration	5
Dissolution	1

It can be observed, that the results are identical for the disintegration and dissolution test units, while for the tablet press unit operation, the optimization procedure appears to be less efficient and requires three additional experiments.
In the following chapter, the reasons for such differences are discussed, and an improved optimization procedure is presented and tested in order to reduce the number of iterations needed to guarantee an acceptable model fidelity.

Chapter 4

Enhanced optimization procedure

Within this chapter an improvement of the optimization procedure defined in §2.2 is presented. Later, the implementation of this new technique to the DC process models is discussed.

4.1. Definition of the improved optimization procedure

The reason associated to the development of a new optimization approach is related to the strict conservatism of the former methodology; the enhanced optimization aims at the possibility of ensuring an acceptable model fidelity with a reduced number of iterations.

4.1.1. Previous optimization limits and possible solution

The application of the optimization procedure to the direct compression process models provides satisfying findings. However, outcomes achieved with methodologies based on multi-variate statistical methods (§3.5) suggest that, sometimes, the optimization technique may overestimate the number of required experiments to reach the stop condition, at least if compared to the methodology proposed in Geremia *et al.* (2023).
As a matter of example, we can refer to the second iteration of the tablet press model presented in Table 4.1; units of measurements of parameters are those defined during the tablet press unit definition (§3.2).

Table 4.1. *Second estimation-optimization iteration: uncertainties comparison.*

	a_1	a_2	a_{sf}	b_1	b_2	b_{sf}	γ
θ_i^0	11.09	1.088	0.455	-7.961	0.321	2.445×10^{-2}	1.202×10^{-3}
$\varepsilon_{i,max,b}$	0.150	8.225×10^{-2}	2.823×10^{-3}	8.420×10^{-2}	1.366×10^{-2}	3.356×10^{-4}	6.249×10^{-5}
(95%CI)/2	8.935×10^{-2}	5.523×10^{-2}	2.128×10^{-3}	6.650×10^{-2}	7.272×10^{-3}	2.395×10^{-4}	1.559×10^{-4}

Since $\varepsilon_{\gamma,max,b}$ is lower than its (95%CI)/2, the stop condition is not achieved at the second iteration and another one is needed.
However, if (95%CIs)/2 reported in Table 4.1 are assumed as $\boldsymbol{\varepsilon}$ inside the model of the optimization problem defined for the tablet press (Equation (3.4) in §3.2.1), constraints on K_{M_1} are fulfilled. This is a consequence of the precision in the estimation of some parameters:

for $\boldsymbol{\theta}$, except γ, the (95%CI)/2 is lower than $\varepsilon_{i,max}$, and this "compensate" for a larger uncertainty in γ.

In this way a limitation of the actual optimization procedure is highlighted: the current optimization calculates Obj as the maximum product of all ξ_i (Equation (2.11) in §2.2.1), but the solution may provide very high ξ_i for some parameters and very low ξ_i for others.

A possible solution to this problem may result in fixing an upper bound for the maximum uncertainty allowable on parameters. Thus, after the application of the optimization procedure, if for a parameter θ_i, $\varepsilon_{i,max} >$ (95%CI)/2, then (95%CI)/2 is set as upper bound for $\varepsilon_{i,max}$ at the next iteration. A graphical example is presented in Figure 4.1.

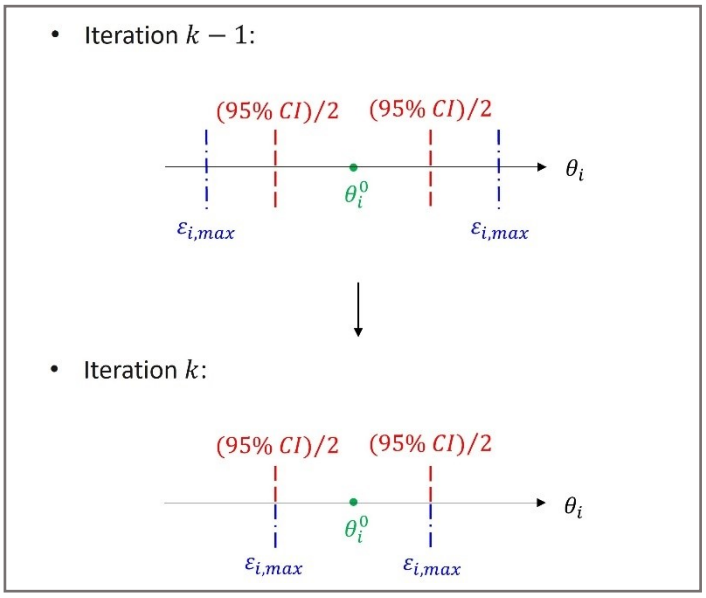

Figure 4.1. *Improved optimization framework.*

It is worth declaring that CIs are not always numerically stable: sometimes a parameter estimation may lead to 95%CIs higher than the ones achieved at the previous iteration. This is probably related to the occurrence of local minima where the estimation settles during its calculations. Therefore, a new condition is added to the improved optimization: if for a parameter θ_i, $\varepsilon_{i,max} >$ (95%CI)/2, then $\varepsilon_{i,max}$ at the next iteration is set at the maximum (95%CI)/2 between the one of the current and the previous estimation.

As a consequence of setting a limit to the maximum uncertainty for some parameters, a larger uncertainty may be achieved for the others. In this way, there is a greater possibility that less iterations are required to obtain $\varepsilon_{i,max,b} >$ (95%CI)/2 for all parameters.

4.1.2. New procedure implementation

The objective of the improved optimization framework is still the research of a condition where the current knowledge on parameter uncertainty is sufficient to reach an acceptable model fidelity. Therefore, as for the previous optimization, we need to compare the

uncertainties with the parameter estimates and build an iterative scheme capable to reach a stop condition. The procedure still comprises a maximum of four steps:

5. MBDoE and experimentation;
6. parameter estimation;
7. optimization problem on 2^n scenarios;
8. optimization problem on N scenarios.

Nevertheless, additional intermediate decisions are required, as illustrated in Figure 4.2, and commented on in the following.

Starting with the problem resolution, initial guesses for parameters are defined. Based on this, a new experiment is performed to get both more precise and accurate parameters values. In particular, the MBDoE is carried out by focusing only on the most influential parameters as specified by a preliminary Sobol analysis. The values of manipulated variables provided by the MBDoE are applied to the true process in order to obtain a set of experimental data produced in silico. Later, a first estimation of parameters is executed, and then an optimization on 2^n boundaries scenarios is performed. Uncertainties given by the optimization are compared with (95%CIs)/2 provided by the parameter estimation: even if for a single parameter $\varepsilon_{i,max,b}$ is lower than the confidence interval, another iteration is needed starting again with a new MBDoE.

The phases described for the first iteration are the same of the previous optimization problem; differences between the two methods may occur starting with the second iteration. Executing the optimization at a specific iteration $k > 1$, findings of iteration k-1 are analysed: if for a generic parameter, θ_i, $\left(\varepsilon_{i,max,b}\right)^{(k-1)}$ is greater than $((95\%CI)/2)^{(k-1)}$, then an upper bound is set for the uncertainty at iteration k, i.e., it is imposed that $\left(\varepsilon_{i,max,b}\right)^{(k)} = ((95\%CI)/2)^{(k-1)}$. However, as stated in §4.1.1, the CIs attained for $\boldsymbol{\theta^0}$ at the new iteration k, $((95\%CIs)/2)^{(k)}$, may be higher than the ones at the previous iteration, $((95\%CIs)/2)^{(k-1)}$. Therefore, following a more conservative approach, whenever $\left(\varepsilon_{i,max,b}\right)^{(k-1)} > ((95\%CI)/2)^{(k-1)}$, the maximum uncertainty at iteration k is defined as:

$$\left(\varepsilon_{i,max,b}\right)^{(k)} = max\ (\ ((95\%CI)/2)^{(k-1)}, ((95\%CI)/2)^{(k)}\ )\quad . \tag{4.1}$$

Once this new condition is imposed, the overall procedure is repeated, as for the original optimization procedure, until all maximum uncertainties allowable on parameters are larger than (95%CIs)/2. Moreover, also in this methodology, if a parameter is characterized by $\varepsilon_{i,max,b}$ higher than the 50% of its estimated value, it is assumed that it has very low influence on $\boldsymbol{K}_M$ and so it is omitted during the evaluation of the stop condition. Therefore, an upper bound for $\xi_{i,max}$ equal to 0.500 is set for those parameters.

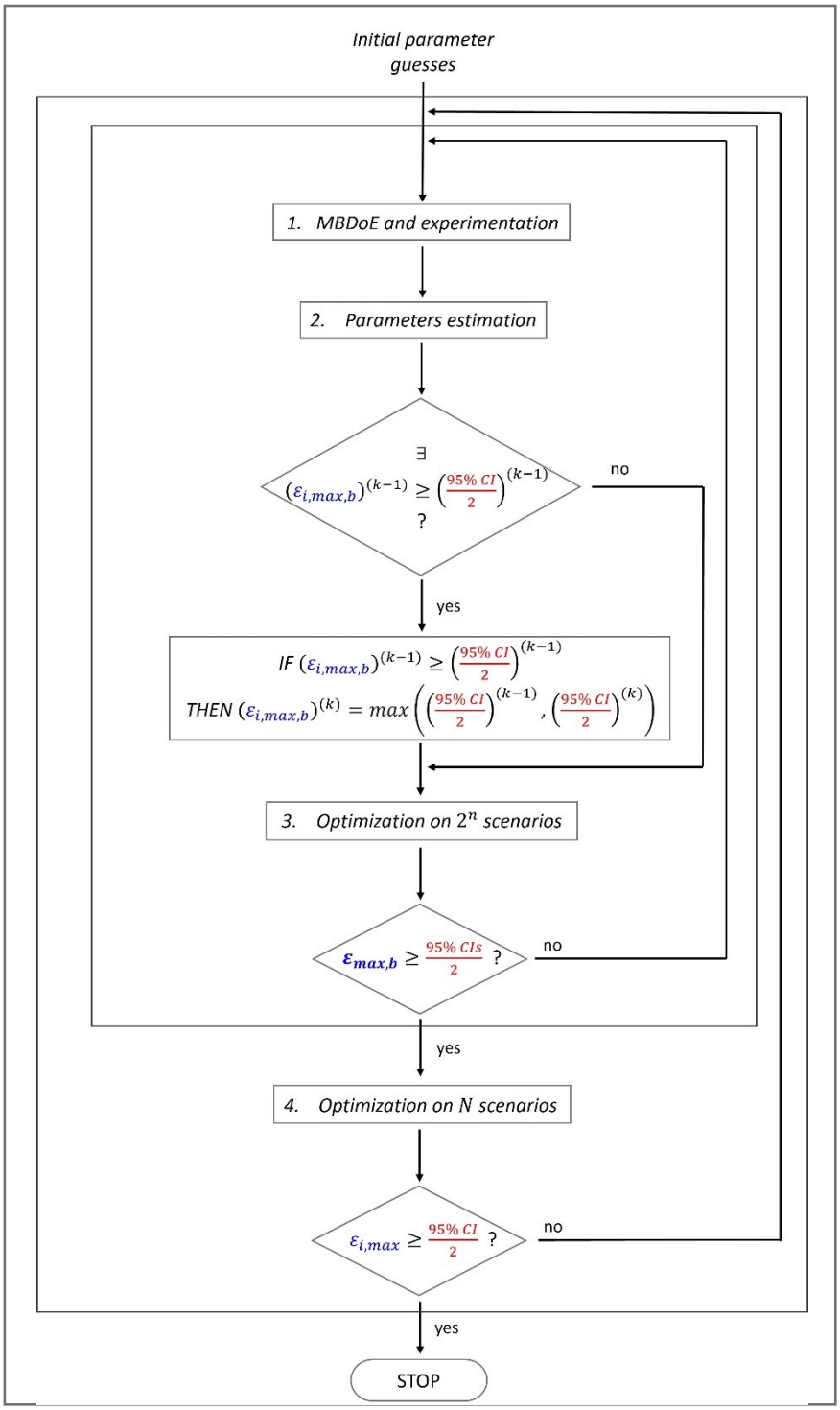

Figure 4.2. *Complete scheme of iterations for the improved optimization procedure: after guesses for parameters are assumed, the loop containing MBDoE, parameters estimation, and optimization is computed until the desired outcomes are achieved; later an optimization on N scenarios occurs.*

After $\boldsymbol{\varepsilon_{max,b}} >$ (95%CIs)/2 is obtained, the optimization with stochastic scenarios must be performed to ensure that $\boldsymbol{\varepsilon_{max,b}}$ found is acceptable. Two possible situations may occur:

3. $\boldsymbol{\varepsilon_{max,b}}$ obtained with the optimization on 2^n scenarios is satisfactory;
4. $\boldsymbol{\varepsilon_{max,b}}$ obtained with the optimization on 2^n scenarios is not satisfactory:

a. the new $\boldsymbol{\varepsilon_{max}}$ achieved is larger than (95%CIs)/2;
b. the new $\boldsymbol{\varepsilon_{max}}$ achieved is smaller than (95%CIs)/2.

The stop condition with the definitive $\boldsymbol{\varepsilon_{max}}$ is reached only if 1 or 2.a occurs, otherwise another iteration is needed.

The overall scheme of iterations is summarised in Figure 4.2, where it is possible to highlight that if at iteration k, condition $\left(\varepsilon_{i,max,b}\right)^{(k-1)} >$ ((95%CI)/2)$^{(k-1)}$ does not occur, then the iteration k for the improved optimization is equal to the one defined with the previous methodology (see Figure 2.7 in §2.2.3)

Since $\boldsymbol{\varepsilon_{max}}$ provides a situation in which acceptable tolerance on $\boldsymbol{K_M}$ is guaranteed, and $\boldsymbol{\varepsilon_{max}} >$ (95%CIs)/2, then the estimation on parameters is precise enough with respect to the same constraints on $\boldsymbol{K_M}$. Thus, the last iteration certifies a state where a satisfactory model prediction is achieved.

4.2. Application: DC process models

After the definition of the enhanced optimization procedure, this is applied to the DC process models to assess whether it would lead to the stop condition with a reduced number of iterations. Results for the tablet press unit operation and for the disintegration test unit are shown in the followings. The improved optimization cannot be tested on the *in vitro* dissolution model, since already only one iteration is needed.

4.2.1. Tablet press

As already stated in the previous section, the first iteration steps for the two methodologies, described in §2.2.3 and in §4.1.2 respectively, show no differences. For sake of clarity, the results that were obtained in §3.2.2, after the first iteration, are reported again in Table 4.2.

Table 4.2. *Improved optimization: first iteration results.*

	$\boldsymbol{a_1}$	$\boldsymbol{a_2}$	$\boldsymbol{a_{sf}}$	$\boldsymbol{b_1}$	$\boldsymbol{b_2}$	$\boldsymbol{b_{sf}}$	$\boldsymbol{\gamma}$
$\boldsymbol{\theta_i^0}$	11.93	1.449	0.413	-6.235	0.205	1.890×10^{-2}	7.365×10^{-4}
$\boldsymbol{\varepsilon_{i,max,b}}$	0.162	8.761×10^{-2}	2.573×10^{-3}	5.816×10^{-2}	2.016×10^{-2}	2.576×10^{-4}	4.077×10^{-5}
(95%CI)/2	0.105	4.768×10^{-2}	2.115×10^{-3}	5.226×10^{-2}	8.791×10^{-3}	1.946×10^{-4}	1.025×10^{-4}

Since $\varepsilon_{\gamma,max,b} <$ (95%CI)/2, then another iteration is required. Thus, a new MBDoE is executed trying to increase the precision on γ, since it is the only parameter for which the $\varepsilon_{i,max,b} >$ (95%CI)/2 condition is not satisfied; values achieved of manipulated variables are then applied to the true process (Table 3.1 in §3.2.1) leading to a new set of experimental data. Later, this is used together with the experimental dataset attained during the first iteration in order to perform a second estimation of parameters (Table 4.3).

Table 4.3. *Estimation of model parameters: second iteration.*

	Units	Value	95% CI	95% t-value
a_1^0	MPa	11.09	0.179	62.08
a_2^0	-	1.088	0.110	9,853
a_{sf}^0	-	0.455	4.256×10^{-3}	1.068×10^{2}
b_1^0	-	-7.961	0.133	59.89
b_2^0	-	0.321	1.454×10^{-2}	22.11
b_{sf}^0	MPa^{-1}	2.445×10^{-2}	4.791×10^{-4}	51.03
γ^0	dm^{-1}	1.202×10^{-3}	3.118×10^{-4}	3.856
Reference t-value (95%)				1.895

Once values of $\boldsymbol{\theta^0}$ are acquired, these are used inside the model of the optimization problem (Equation (3.4) in §3.2.1). Moreover, since new parameter values are obtained, an optimization on the process manipulated variables must be performed ensuring that the desired $\overline{K_{M_1}}$, the one of the true process, is achieved. This leads to $K = 99$ m and $P = 213$ MPa.

At this point, following the new optimization flowchart (Figure 4.2), we must assess if at the previous iteration for some parameter $\varepsilon_{i,max,b} >$ (95%CI)/2 occurs. This condition arises for all parameters except for γ, so an upper bound must be fixed for their maximum uncertainty. However, the limit has to be the maximum value between the (95%CI)/2 of the first and the second iteration. Therefore, outcomes achieved at the two iterations are grouped in Table 4.4, where the maximum (95%CI)/2 for each parameter is in boldface.

Table 4.4. *First and second iteration: (95%CI)/2 comparison; maximum (95%CI)/2 for each parameter is in boldface.*

	a_1	a_2	a_{sf}	b_1	b_2	b_{sf}
((95%CI)/2)¹	**0.105**	4.768×10^{-2}	2.115×10^{-3}	5.226×10^{-2}	$\mathbf{8.791 \times 10^{-3}}$	1.946×10^{-4}
((95%CI)/2)²	8.935×10^{-2}	$\mathbf{5.523 \times 10^{-2}}$	$\mathbf{2.128 \times 10^{-3}}$	$\mathbf{6.650 \times 10^{-2}}$	7.272×10^{-3}	$\mathbf{2.395 \times 10^{-4}}$

Once the largest (95%CIs)/2 are selected, these are set as $\boldsymbol{\varepsilon_{max,b}}$ for the optimization on 128 (i.e., 2^7) boundary scenarios during the second iteration; the resulting $\boldsymbol{\xi}$ are presented in Table 4.5.

Table 4.5. *Maximum relative percentage amounts, $\xi_{i,max}$, obtained for different parameters at the second iteration with the improved optimization procedure.*

	a_1	a_2	a_{sf}	b_1	b_2	b_{sf}	γ
$\xi_{i,max}$	9.446×10^{-3}	5.075×10^{-2}	4.680×10^{-3}	8.349×10^{-3}	2.734×10^{-2}	9.798×10^{-3}	0.135

Results are also presented in the form of maximum uncertainties (Table 4.6).

Table 4.6. *Maximum uncertainties, $\varepsilon_{i,max,b}$, obtained for different parameters at the second iteration with the improved optimization procedure.*

	a_1	a_2	a_{sf}	b_1	b_2	b_{sf}	γ
$\boldsymbol{\varepsilon_{i,max,b}}$	0.105	5.523×10^{-1}	2.128×10^{-3}	6.650×10^{-2}	8.791×10^{-3}	2.395×10^{-4}	1.623×10^{-4}

Finally, the comparison between $\boldsymbol{\varepsilon_{max,b}}$ and the (95%CIs)/2 must be performed considering the confidence intervals of the current iteration: $((95\%CI)/2)^{(2)}$. Outcomes are reported in Table 4.7.

Table 4.7. *Improved optimization: second iteration results.*

	a_1	a_2	a_{sf}	b_1	b_2	b_{sf}	γ
$\boldsymbol{\varepsilon_{i,max,b}}$	0.105	5.523×10^{-2}	2.128×10^{-3}	6.650×10^{-2}	8.791×10^{-3}	2.395×10^{-4}	1.623×10^{-4}
(95%CI)/2	8.935×10^{-2}	5.523×10^{-2}	2.128×10^{-3}	6.650×10^{-2}	7.272×10^{-3}	2.395×10^{-4}	1.559×10^{-4}

Results demonstrate that, by setting $\varepsilon_{i,max,b} =$ (95%CI)/2 for the parameters where a sufficient precision has been reached, the allowable uncertainty on γ can increase and become larger than its (95%CI)/2. Findings are confirmed after the evaluation of all N scenarios.

4.2.2. Disintegration

Following the new methodology, as for the tablet press, the most relevant outcomes of the first iteration for the disintegration test unit (Table 4.8) are presented grouping those obtained in Table 3.15 and Table 3.18 (§3.3.2) with the previous optimization procedure.

Table 4.8. *Improved optimization: first iteration results.*

	C_2	C_3	$\dot{\epsilon}$	n_s	S_p
$\boldsymbol{\theta_i^0}$	52.23	99.93	1.480×10^{-3}	0.905	0.535
$\boldsymbol{\varepsilon_{i,max,b}}$	-	-	-	1.679×10^{-2}	5.686×10^{-2}
(95%CI)/2	-	-	-	9.199×10^{-2}	0.274

Since for C_2, C_3, and $\dot{\epsilon}$ the $\xi_{i,max}$ is higher than 0.500 (Table 3.16 in §3.3.2), they are not considered during the $\boldsymbol{\varepsilon_{max,b}}$ and (95%CIs)/2 comparison. Results show that $\varepsilon_{i,max,b}$ is lower than (95%CI)/2 for both n_s and S_p; thus, another iteration is required. Since no parameter shows $\varepsilon_{i,max,b}$ greater than (95%CI)/2, the second iteration for the improved optimization is equal to the one of the previous technique. The first time that $\varepsilon_{i,max,b} >$ (95%CI)/2 occurs is in the fourth iteration, which is characterized by the parameters estimation presented in Table 4.9.

Table 4.9. *Estimation of model parameters: fourth iteration.*

	Units	Value	95% CI	95% t-value
C_2^0	MPa	90.31	5.334×10^{4}	1.693×10^{-3}
C_3^0	MPa	99.83	3.002×10^{3}	3.325×10^{-2}
$\dot{\epsilon}^0$	m s^{-1}	9.820×10^{-4}	8.079×10^{-2}	1.215×10^{-2}
n_s^0	-	0.903	3.651×10^{-2}	24.74
S_p^0	-	0.534	0.117	4.56
Reference t-value (95%)				1.654

The optmization on 2^n boundary scenarios is then computed on the estimates, leading to the ξ_i reported in Table 4.10,

Table 4.10. *Maximum relative percentage amounts, $\xi_{i,max}$, obtained for different parameters at the fourth iteration.*

	C_2	C_3	$\dot{\epsilon}$	n_s	S_p
$\xi_{i,max}$	0.500*	0.500*	0.500*	1.976×10^{-2}	0.113

and later to maximum uncertainties (Table 4.11).

Table 4.11. *Maximum uncertainties, $\varepsilon_{i,max,b}$, obtained for different parameters at the fourth iteration.*

	C_2	C_3	$\dot{\epsilon}$	n_s	S_p
$\varepsilon_{i,max,b}$	45.15*	49.92*	4.910×10^{-4}*	1.785×10^{-2}	6.036×10^{-2}

Again, Table 4.11 shows that for C_2, C_3, and $\dot{\epsilon}$ the $\xi_{i,max}$ is higher than 0.500, and so these are not considered during the $\varepsilon_{max,b}$ and (95%CIs)/2 comparison in Table 4.12.

Table 4.12. *Fourth estimation-optimization iteration: uncertainties comparison.*

	n_s	S_p
$\varepsilon_{i,max,b}$	1.785×10^{-2}	6.036×10^{-2}
(95%CI)/2	1.826×10^{-2}	5.848×10^{-2}

Differently from the findings attained at the first iteration (Table 4.8), $\varepsilon_{i,max,b}$ greater than its (95%CI)/2 occurs for S_p. Therefore, at the fifth iteration, the additional steps introduced with the algorithm discussed in §4.1.2 are applied: an upper bound will be set to $\varepsilon_{S_p,max,b}$ during the optimization on 2^n boundaries scenarios.

The estimates of parameters obtained at the fifth iteration are those presented in Table 3.19 (§3.3.2). At this point, the optimization must be performed fixing a limit to $\varepsilon_{S_p,max,b}$, and this is chosen as the maximum between $((95\%CI)/2)^{(4)}$ and $((95\%CI)/2)^{(5)}$; the value of interest is reported in boldface in Table 4.13.

Table 4.13. *Fourth and fifth iteration: (95%CI)/2 comparison; maximum (95%CI)/2 is in boldface.*

	S_p
((95%CI)/2)$^{(4)}$	**5.848×10^{-2}**
((95%CI)/2)$^{(5)}$	5.495×10^{-2}

Once the largest (95%CI)/2 is selected, this is set as $\varepsilon_{i,max,b}$ for S_p during the optimization on boundary scenarios: relative uncertainties achieved, $\boldsymbol{\xi}$, are presented in Table 4.14,

Table 4.14. *Maximum relative percentage amounts, $\xi_{i,max}$, obtained for different parameters at the fifth iteration with the improved optimization procedure.*

	C_2	C_3	$\dot{\epsilon}$	n_s	S_p
$\boldsymbol{\xi_{i,max}}$	0.500*	0.500*	0.500*	2.111×10^{-2}	0.110

and maximum uncertainties are reported in Table 4.15.

Table 4.15. *Maximum uncertainties, $\varepsilon_{i,max,b}$, obtained for different parameters at the fifth iteration with the improved optimization procedure.*

	C_2	C_3	$\dot{\epsilon}$	n_s	S_p
$\boldsymbol{\varepsilon_{i,max,b}}$	44.36*	49.58*	4.918×10^{-4}*	1.906×10^{-2}	5.848×10^{-2}

Finally, the comparison between $\boldsymbol{\varepsilon_{max,b}}$ and the (95%CIs)/2 must be performed considering the confidence intervals of the current iteration: ((95%CI)/2)$^{(4)}$. Results are reported in Table 4.16, excluding C_2, C_3, and $\dot{\epsilon}$ since their $\xi_{i,max}$ is higher than 0.500 (Table 4.14).

Table 4.16. *Improved optimization: fifth iteration results.*

	n_s	S_p
$\boldsymbol{\varepsilon_{i,max,b}}$	1.906×10^{-2}	5.848×10^{-2}
(95%CI)/2	1.746×10^{-2}	5.495×10^{-2}

Results show that all (95%CIs)/2 are lower than the maximum uncertainties allowable for the model parameters. In this case, the stop condition is reached at the same iteration obtained in §3.3.2 with the previous procedure. Thus, the application of the new optimization procedure to the disintegration test unit does not provide any additional improvement with respect to the previous optimization discussed in §2.2.3.

4.3. Comments on results

The application of the enhanced optimization procedure on DC process models provides satisfying results: for all the units a condition where $\boldsymbol{\varepsilon_{max}}$ is greater than (95%CIs)/2 is obtained. Additionally, for the tablet press unit operation, less iterations are needed with respect to the original optimization to achieve a condition where acceptable model fidelity in ensured. The number of experiments required to reach the stop condition for each unit is summarized in Table 4.17.

Table 4.17. *Enhanced optimization procedure: number of experiments required to reach the stop condition for each unit.*

Unit	Exp
Tablet press	8
Disintegration	5
Dissolution	1

The total number of experiments to be performed is identical for all units to those obtained in Table 3.31 (§3.5) implementing the methodology proposed by Geremia *et al.* (2023). We can conclude that the two approaches lead to comparable results to ensure the pre-set acceptable model fidelity.

Conclusions

The optimization problem developed in this study proved to be an effective tool for overcoming fidelity issues related to modeling in the pharmaceutical industry, specifically to address the following tasks:

- how to determine the most influential model parameters on the prediction of key indicators;
- assess whether the prediction of desired key indicators is ensured by analysing the parametric uncertainty.

Differently from other approaches, such as the one proposed in Geremia *et al.* (2023), the optimization method presented in this study strives to achieve the goal in a more rigorous way, by thoroughly examining the mathematical model without any linearization. The new procedure aims at establishing the maximum uncertainties allowable on model parameters to keep the key indicators within the desired range.

The new methodology was applied to the direct compression process analysed in Geremia *et al.* (2023), where three models are present: tablet press unit operation, disintegration test unit, and *in vitro* dissolution test unit. In this way, it was possible to compare the outcomes obtained from both methodologies. An acceptable model fidelity was ensured with the same number of required experiments for the disintegration and dissolution models, while three additional experiments were required for the tablet press model, when the approach proposed in this Thesis was used. The reason of this difference was attributed to the strict conservatism of the objective function definition within the optimization problem.

Therefore, the optimization framework was relaxed in order to take advantage of the precision already attained during the estimation of the model parameters. Results showed that in this way it was possible to use as many experiments as in Geremia *et al.* (2023) in order to target the required model fidelity towards the key indicators. Additionally, the optimization procedure indirectly suggested that the model linearization employed in the methodology based on multi-variate statistical methods does not restrict the ability to obtain reliable predictions for direct compression process models.

The actual optimization framework, similarly to the approach defined in Geremia *et al.* (2023), sets a limit equal to 50% to the maximum relative uncertainty evaluated for parameters. When a parameter meets this condition, it is assumed to have very low influence on the key indicator of interest and so it is excluded from the evaluation of the stop condition. This assumption may have an impact on the actual reliability a model may achieve, and should be further explored as future work. Furthermore, similarly to Geremia *et al.* (2023), the proposed optimisation procedure should be applied and tested on the overall systems model, where all unit models are accounted for simultaneously.

References

Anand, O., L. X. Yu, D. P. Conner and B. M. Davit (2011). Dissolution testing for generic drugs: an FDA perspective. *The AASP Journal*, **13**, pp. 328-335.

Asprey, S. P. and S. Macchietto (2002). Designing robust optimal dynamic experiments. *Journal of Process Control*, **12**, pp. 545-556.

Bano, G., R. M. Dhenge, S. Diab, D. J. Goodwin, L. Gorringe, M. Ahmed, R. Elkes and S. Zomer (2022). Streamlining the development of an industry dry granulation process for an immediate release tablet with systems modelling. *Chemical Engineering Research and Desing*, **178**, pp. 421-437.

Bard, Y. (1974). *Nonlinear Parameter Estimation.* Academic Press. New York (USA).

Braakman, S., P. Pathmanathan and H. Moore (2022). Evaluation framework for systems models. *CPT: Pharmacometrics & System Pharmacology*, **00**, pp. 1-26.

Destro, F. and M. Barolo (2022). A review on the modernization of pharmaceutical development and manufacturing – Trends, perspectives, and the role of mathematical modeling. *International Journal of Pharmaceutics*, **620**, 121715.

Diab, S., C. Christodoulou, G. Taylor and P. Rushworth (2022). Mathematical modeling and optimization to inform impurity control in an industrial active pharmaceutical ingredient manufacturing process. *Organic Process Research and Dev*elopment, **26**, pp. 2864-2881.

Franceschini, G. and S. Macchietto (2008). Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science*, **63**, pp. 4846-4872.

Galvanin, F., S. Macchietto and F. Bezzo (2007). Model-based design of parallel experiments. *Industrial & Engineering Chemistry Research*, **46**, pp. 871-882.

Geladi, P. and B. R. Kowalski (1996). Partial least-squares regression: a tutorial. *Analytica Chimica Acta*, **185**, pp. 1-17.

Geremia, M., S. Diab, C. Christodoulou, G. Bano, M. Barolo and F. Bezzo (2023). A general procedure for the evaluation of the prediction fidelity of pharmaceutical system models. *Chemical Engineering Science*. (submitted)

Gunther, J. C., J. Baclaski, D. E. Seborg and J. S. Conner (2009). Pattern matching in batch bioprocesses – comparisons across multiple products and operating conditions. *Computers and Chemical Engineering*, **33**, pp. 88-96.

Jaeckle, C. M. and J. F. MacGregor (2000). Industrial applications of product design through the inversion of latent variable models. *Chemometrics and Intelligent Laboratory Systems,* **50**, pp. 199-210.

MacGregor, J. F. and M. J. Bruwer (2008). A framework for the development of design and control spaces. *Journal of Pharmaceutical Innovation*, **3**, pp. 15-22.

Markl, D., S. Yassin, D. I. Wilson, D. J. Goodwin, A. Anderson and J. A. Zeitler (2017). Mathematical modelling of liquid transport in swelling pharmaceutical immediate release tablets. *International Journal of Pharmaceutics*, **526**, pp. 1-10.

Miao, H., Xia, X., Perelson, A. S., Wu, H. (2011). On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Rev*iew, **53**, pp. 3-39.

Montgomery, D. C. (2013). *Introduction to Statistical Quality Control*. John Wiley and Sons Ltd, Hoboken, 07030 (USA).

Nassar, J., B. Williams, C. Davies, K. Lief and R. Elkes (2021). Lubrication empirical model to predict tensile strength of directly compressed powder blends. *International Journal of Pharmaceutics*, **592**, 119980.

Peppas, N. A. and P. Colombo (1989). Development of disintegration forces during water penetration in porous pharmaceutical systems. *Journal of Controlled Release*, **10**, pp. 245-250.

Polak, J., M. Von Stosch, M. Sokolov, L. Piccioni, A. Streit, B. Schenkel and B. Guelat (2023). Hybrid modeling supported development of an industrial small-molecule flow chemistry process. *Computers and Chemical Engineering*, **170**, 108127.

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

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

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana and S. Tarantola (2008). *Global Sensitivity Analysis. The Primer*. John Wiley and Sons Ltd, West Sussex PO19 8SQ (England), pp. 11-33.

Sobol, I.M. (1993). Sensitivity estimates for nonlinear mathematical models, *MMCE*, **1**(4), pp. 407-414.

Sobol, I.M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, **55**, pp. 271-280.

Song, E., B. L. Nelson and J. Staum (2016). Shapley Effects for Global Sensitivity Analysis: Theory and Computation. *Society for Industrial and Applied Mathematics and American Statistical Association*, **4**, pp. 1060-1083.

Wilson, D., S. Wren and G. Reynolds (2012). Linkinf dissolution to disintegration in immediate release tablets using image analysis and a population balance modelling approach. *Pharm Res*, **29**, pp. 198-208.

Wise, B. M. and N. B. Gallagher (2006). The process chemometrics approach to process monitoring and fault detection. *Journal of Chemometrics*, **12**, pp. 301-321.

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

Yang, O., Y. Tao, M. Qadan and M. Ierapetritou (2022). Process design and comparison for batch and continuous manufacturing of recombinant adeno-associated virus. *Journal of Pharmaceutical Innovation*, **49**, 2215-2220.

Zineh, I. (2019). Quantitative systems pharmacology: a regulatory perspective on translation. *CPT Pharmacometrics & System Pharmacology,* **8**, 336-339.