Run-to-Run Optimization of Biochemical Batch Processes in the Presence of Model-Plant Mismatch

by

Rubin Hille

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Chemical Engineering

Waterloo, Ontario, Canada, 2018

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Prof. Christopher Swartz Chemical Engineering, McMaster University
Supervisor	Prof. Hector Budman Chemical Engineering, University of Waterloo
Internal Member	Prof. Ali Elkamel Chemical Engineering, University of Waterloo
Internal Member	Prof. Nasser Mohieddin Abukhdeir Chemical Engineering, University of Waterloo
Internal-external Member	Prof. Brian Ingalls Applied Mathematics, University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

For Chapter 3, the author implemented the documented methodologies, obtained the numerical results and completed the writing of the paper. Jasdeep Mandur and Hector Budman were involved in the conceptualization of the work, aided in the analysis and edited the manuscript. All authors read and approved the final manuscript.

Abstract

An increased demand for novel pharmaceuticals such as recombinant proteins with therapeutic potential has lead to significant advances in the operation of biotechnological processes. In general, biochemical processes are characterized by nonlinear behavior and a sensitivity to environmental conditions. Furthermore, due to their complex operation, exposure to contamination and the low volume of the obtained product, these processes are generally still frequently operated in batch or fed-batch reactors. The repetitive nature of batch processes motivates the use of previous experimental effort to improve the performance of future batch operations. In this way, a so-called run-to-run optimization can be performed where the measurements of the current batch-run are utilized to determine the input for the next experiment. To conduct this step in a systematic and reliable manner, fundamental process models can be used for prediction and optimization purposes. This way, it is possible to determine the input for the next iteration from the predicted optimum obtained by calibrating a model based on measurements from the current batch-run.

Fundamental models are typically derived from the underlying physical phenomena of the process. However, to make these models useful and tractable, it is common to make assumptions and simplifications during the model development. As a result, there often exists mismatch between the model and process under study. In the presence of model-plant mismatch, the set of model parameter estimates, which satisfy an identification objective, may not result in an accurate prediction of the gradients of the cost-function and constraints, which are essential for optimization. To still ensure convergence to the optimum, the method of simultaneous identification and

optimization aims at forcing the predicted gradients to match the measured gradients by adapting the model parameters. At the same time, a correction factor is introduced into the model output so that the previously achieved fitting accuracy can be maintained. This results in a set of model parameters that reconcile the objectives of identification and optimization in presence of model-plant mismatch. Although the method provides the potential for dealing with structural error in iterative optimization schemes, there exist several challenges that have to be addressed before it is applicable to more complex systems.

For example, when dealing with models containing a large number of parameters, it is unclear which parameters should be selected for calibration and adaptation. Since updating all available parameters is impractical due to estimability problems and over-fitting, there is a motivation for adapting only a subset of parameters. Furthermore, for this method to be more efficient under uncertainty, it is necessary to introduce additional robustness to uncertainty in initial conditions and gradient measurements. Finally, it is essential to develop experimental design criteria that will provide the user with more informative experiments to speed up convergence to the optimum and to calibrate the model with better accuracy.

Following the above, this work presents the following new contributions:

(i) An algorithmic approach to select a subset of parameters based on the sensitivities of the model outputs as well as of the cost function and constraint gradients.

(ii) A run-to-run optimization formulation that is robust to uncertainties in initial batch conditions based on polynomial chaos expansions that are used to quantify the uncertainty and to propagate it onto the optimization cost.

(iii) A modified parameter identification objective based on the minimization of the ratio of the sum of squared prediction errors to a parametric sensitivity measure to speed up convergence of the run-to-run procedure.

(iv) The use of uncertainty bounds on the predicted trajectories to ensure model accuracy while

solving the parameter identification problem described in item (iii) and to determine whether a model update is necessary at any given run.

(v) The use of a design of experiments approach within the run-to-run optimization procedure to optimally complement the cost gradient information that is already available from previous batch experiments.

The presented methods are shown to be efficient and facilitate the use of complex models for run-to-run optimization of batch processes. Several case studies of cell culture processes are presented to illustrate the improvements in robustness and performance. These case studies involve batch, fed-batch and perfusion operations. A part of this work has been developed in collaboration with an industrial partner whose main line of business is the development of perfusion growth media for mammalian cell culture operations.

Acknowledgements

I would like to sincerely thank my supervisor Professor Hector Budman for his continuous support, guidance and encouragement throughout the course of my research.

I also would like to express my gratitude to all my friends for the support during the span of this work. In particular, I want to thank my colleagues Ali Nikdel, Yuncheng Du, Ricardo Martinez-Villegas, Omar Santander, Vanessa Zavatti, Jasdeep Mandur, Manoj Mathew, Mariana Carvalho and Mehrez Agnaou for the engaging conversations during the last four years.

Finally, I want to thank the National Science and Engineering Research Council for funding my research.

Dedication

This dissertation is dedicated to my parents.

Table of Contents

List of Figures

List of Tables

1. Introduction

Biotechnological processes present a platform for the manufacturing of various products ranging from vaccines to therapeutic proteins with medical applications. Due to the low volume of the obtained products and characteristic behaviours such as the build up of toxic byproducts and the occurrence of contamination, bio-processes are still frequently operated in batch or fed-batch operations. The operation of batch processes presents a challenge compared to the largely continuous process operations predominant in the chemical industry due to the following distinctive features. Batch operations generally lack a steady-state as the desired product is obtained through various reactions starting from an initial starting material. As the quantities of interest undergo large changes in concentration between the initial and final time, these processes are usually more difficult to monitor and control and, mathematically, they exhibit a clear nonlinear behaviour. Furthermore, due to the high cost of experiments and limited availability of measurements, experimental data might only be scarcely available. Additional difficulties in the operation of batch process are presented by safety considerations and constraints on the manipulated variables.

Mathematical models consequently offer an opportunity to gain a better process understanding and to increase product yields. In terms of their characteristics, one can generally distinguish between black-box (data-driven) and first-principles (knowledge-driven) models. First-principle models require an intricate understanding of the process and therefore necessitate larger development time. Nonetheless, when compared to black-box models, fundamental models offer a larger domain of validity and are therefore particularly suited for optimization purposes. In addition, the repetitive

nature of batch processes allows for using past experiments to update model predictions and consequently enhance the operation of future batch-runs based on a progressively improved model. In this way, a so-called run-to-run optimization can be performed where information from previous batch-runs is utilized to update an existing model which determines the optimal input for the next experiment.

Due to their complex nature, run-to-run optimization therefore presents an attractive method when dealing with biotechnological processes. For cell culture processes, typical optimization goals may involve the maximization of productivity, while maintaining desired rates of certain metabolic reactions. Since this task requires knowledge of the metabolic pathways of the respective organism, it is advantageous to use this understanding to obtain a more accurate description of the process and enable a prediction over a broad range of operating conditions. On the other hand, due to the complexity of bio-organisms, it is common practice to describe biochemical systems with simplified models based on balances of intra and extra cellular components. Typical techniques involve the reduction of a model, derived from a metabolic flux analysis, to its most significant metabolic pathways (fluxes). The result of this reduction is a model which can only capture the significant interactions of the process and therefore leads to a certain amount of discrepancy between the model and process under study. This discrepancy, referred to as model-plant mismatch, presents a critical challenge when using models to optimize the process operation. In this case, methods have to be used which can deal with structural mismatch and drive the process reliably and effectively towards its optimal operating point.

Following these considerations, for achieving a successful process optimization, it is of utmost importance to employ a model capable of providing an accurate description of the given process optimality conditions. To provide for such accuracy, the model has to be calibrated first by fitting process output measurements. This task is often challenging due to uncertainty in initial parameter values, measurement noise as well as unknown disturbances entering the process. Thus, in addition

to a structural mismatch, stochastic sources of uncertainty reduce the accuracy of the model predictions. This is especially important in the context of batch-to-batch optimization since model-plant mismatch coupled with parametric uncertainty leads to conflicts between the objectives of model identification and optimization. In other words, the parameters that are identified to generate a satisfactory fitting between output measurements and model predictions may not lead to an accurate prediction of the cost-function and constraint gradients. Consequently, a model-driven optimization that is based on an imperfect model may result in a sub-optimal operating policy. To reach the process optimum despite the presence of model-plant mismatch, it is therefore vital to consider not just output measurements, but also available measurements of the gradients of the cost-function and constraints with respect to the decision variables.

If only process optimization is deemed to be of importance, it would be possible to utilize methods that maintain a model solely for an accurate prediction of the objective function. Such approaches for instance include direct search methods or empirical techniques based on experimental design methodologies. However, in the context of batch-to-batch optimization of bioprocesses, an accurate model is not only essential for optimization purposes, but often crucial for the following reasons: I - for predicting the outputs around the optimum in response to possible changes in operating conditions or disturbances. II - for calculating reference trajectories along the batch to be tracked under closed-loop control. III - to predict quantities of interest around the optimum that may not be directly available as measurements in a reasonable time. IV - to enforce constraints on variables of interest such as toxic species in bio-processes (e.g. ammonia and lactate).

Hence, the ultimate goal set for the current work is to drive the process to the optimum while identifying a model that can be effectively used for several purposes. To this end, the method of simultaneous identification and optimization provides a framework where the model parameters are first identified from available output measurements and subsequently adapted to match the measured cost-function and constraint gradients. At the same time, a correction term is introduced into

the model output to maintain an accurate prediction of outputs. However, there exist several limitations which prevents the method to be applied to models of larger biochemical systems with many parameters. For instance, it is unclear whether all or only a subset of parameters should be updated during each iteration of the algorithm. To reduce the effect of uncertainty on the performance, it is furthermore necessary to develop a more robust parameter estimation technique tailored for run-to-run optimization. This step also includes a model-update criterion to provide information about when to update which specific model-outputs in order to achieve an adequate model-fitting. Moreover, to improve the overall cost and efficiency of new experiments, it is beneficial to incorporate a design of experiments approach which takes the presence of structural mismatch into account.

To address these limitations, this work presents the following contributions:

1. Chapter 3 proposes a careful examination of parametric sensitivities. The analysis is performed not only with respect to the output sensitivities, but also with respect to the sensitivities of the gradients of the cost and constraint functions. A subsequent ranking and selection of suitable parameters is developed in order to improve the interplay between conflicting objectives of identification and optimization. This allows for the method to be applied to more complex biochemical models involving larger number of parameters.

2. In addition to dealing with structural model error, robustness to stochastic uncertainty in initial batch conditions is introduced in chapter 3 by an augmented optimization procedure using polynomial chaos expansions. This approach allows for the propagation of input uncertainties onto the objective, which ultimately enables the search for operating points that are robust and thus less susceptible to unknown disturbances.

3. In chapter 4, an approach involving set-based constraints on the model outputs is proposed to identify parameter values which are more valuable in terms of cost-function and constraint gradient sensitivities, allowing for an improved reconciliation of the objectives of identification and optimization. At the same time, the bounds on the model outputs enable

a model-update criterion, which avoids an unnecessary re-estimation of parameters for new batch runs.

4. In chapter 5, a design of experiments approach is implemented to determine operating points which increase parameter precision when correcting for errors in predicted gradients of cost-function and constraints. Simultaneously, cost-function information from previous batch runs is taken into account in a systematic way to reduce the effect of uncertainty in gradient measurements.

5. Finally, chapter 6 presents a model development to describe the cultivation of mammalian cells in perfusion systems. The model is subsequently utilized in a run-to-run optimization case study by applying the methods outlined above.

Overall, the developed methodologies greatly improve the performance of the simultaneous identification and optimization framework, not only in terms of speed of convergence to the process optimum, but also in the prediction capabilities around the optimum. The improvements are illustrated using several case studies of cell culture processes.

2. Background and Literature Review

For process optimization, both model-based and model-free methods have been proposed in the past. In the model-free or direct search method (Garcia & Morari, 1981), the path to the optimum is based on the measured gradients of the plant cost-function or it can be also sought by performing carefully planned experiments to improve the process performance (Box & Draper, 1969). On the other hand, in a model-based optimization the goal is to find the optimum of the process based on optimizing the objective function of a mathematical model. This task requires that a suitable model of the process under study is available.

2.1. Model Categories

In general, the types of models that can be utilized to seek for the process optimum can be classified as follows: i- data-driven models where a representation of the processes is solely derived from plant measurements but physical principles are not explicitly used, ii- fundamental models that explicitly consider the physical phenomena by using first principles to describe the underlying system and iii- hybrid models which consist of a combination of these two approaches.

2.1.1. Empirical Models

Empirical or black-box models are based on measurements obtained from experiments combined with a generic pre-determined model structure. As a result, the input-output relationship of the system under study can be determined by means of a mathematical representation (Box & Draper, 1987; Ljung, 1999). As no further knowledge about the plant is assumed, this approach generally requires an adequate number of experiments to be performed to achieve a satisfactory mapping. Model that are linear with respect to the process variables are unsuitable for approximating highly nonlinear process behavior and thus nonlinear empirical models may be needed. Nonetheless, empirical models present a cost-effective and reliable method to obtain a representation of the process. Data-driven modeling approaches for batch processes include design of experiments (DoE) approaches (Montgomery, 2012; Georgakis, 2013) as well as multivariate statistical techniques such as partial least squares (PLS) (Nomikos & MacGregor, 1995; Chen & Liu, 2002). However, the fact that empirical models are obtained through interpolation of experimental data implies that they are not necessarily well suited for extrapolation purposes (Bonvin et al., 2016). This is a major drawback if it is desired to optimize a plant for inputs outside the region of where experiments have been performed as it is often the case in run-to-run optimization procedures.

2.1.2. First-Principles Models

The primary goal of first-principles models is to utilize available knowledge about the fundamental physical phenomena that govern a given process to obtain an accurate mathematical description (Verma, 2014). In chemical engineering for example, it is common to use material and energy balances to derive relationships between input and output variables (Rodrigues & Minceva, 2005). Although the model parameters often have a physical meaning, it is common to make assumptions and simplifications such as lumping certain model parameters to reduce the model's complexity. As a result of these simplifications, it is still typically required to estimate the model parameters

from available experimental data (Bard, 1974). Although their development is substantially more extensive and costly when compared to black-box models, first-principle models offer the advantage of a broader range of validity (Walter & Kieffer, 2007) and are thus suitable for optimization purposes that often require extrapolation beyond the range of the available data used for model calibration (Yip & Marlin, 2004).

With respect to biochemical processes, fundamental models have also become more popular in the recent years. For example, several models of cell cultivation processes for the production of therapeutic proteins have been reported (Yahia et al., 2015). In many of these studies, fundamental dynamic metabolic models have been developed using the metabolic flux analysis (MFA) methodology (Naderi et al., 2011). To reduce the complexity of the entire reaction network, MFA has been utilized to determine metabolites and fluxes which are important for explaining the observed phenomena. This has led to metabolic models which incorporate understandings of the inner workings of the cells to provide a dynamic description of the process (Provost et al., 2006; Zamorano et al., 2013; Aghamohseni et al., 2014).

2.1.3. Hybrid Models

Hybrid models are based on the combination of first-principles models with data-driven approaches (Duarte & Saraiva, 2003; von Stosch et al., 2014). This approach deals with the fact that, typically, not all underlying phenomena of the process are known. In this case, a model can be augmented with data-driven elements such as neural networks (Oliveira, 2004) that are capable of representing the unknown parts of the model, e.g. reaction kinetics, by nonlinear relationships. However, unless these models are trained from sufficient data, they may also suffer from poor extrapolation abilities.

2.2. Parameter Estimation

Regardless of the type of models described in section 2.1, it is usually necessary to estimate unknown model parameters from experimental data. Following first principles, i.e. the mathematical description of processes based on conservation laws such as mass and energy balances, a process model can be generally expressed in terms of a system of ordinary differential equations (ODEs) as follows:

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\theta}, t) + \boldsymbol{\nu} \tag{2.1}$$

$$\boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}) + \boldsymbol{\eta} \tag{2.2}$$

where $\boldsymbol{x} \in \mathbb{R}^{n_x}$ are the state variables evaluated at sampling times t_i with $i \in \{1, \ldots, n_t\}$, $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$ the vector of unknown model parameters and $\boldsymbol{u} \in \mathbb{R}^{n_u}$ the constant vector of model inputs. $\boldsymbol{y} \in \mathbb{R}^{n_y}$ present the model outputs, $\boldsymbol{\nu} \in \mathbb{R}^{n_x}$ are uncertainties such as model-plant mismatch and process noise while $\boldsymbol{\eta} \in \mathbb{R}^{n_y}$ presents the measurement noise. The goal of a parameter estimation procedure is to find the parameter values which result in the best fit of model predictions to the experimental data. This task is typically accomplished by a minimization of the sum of squared errors (SSE) between the model prediction and process measurements as follows (Bard, 1974):

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n_t} \left\| \boldsymbol{y}_p(t_i) - \boldsymbol{y}(\boldsymbol{\theta}, t_i) \right\|^2 \\ &\text{s.t.} \quad (2.1) \text{ and } (2.2) \end{aligned} \tag{2.3}$$

where $\hat{\boldsymbol{\theta}}$ presents a particular set of parameter values that satisfies the parameter estimation objective. When dealing with nonlinear parameter estimation, it is often the case that only a subset of parameters can be identified from the given measurements. This lack of parameter observability is due to the fact that, typically, not all the states can be measured and the ones that can be measured

are generally corrupted by noise. Additionally, the parameter's effect on the model output can be highly correlated thus complicating their identification (McLean & McAuley, 2012). Also, the calibration of too many parameters to fit the available data may often lead to overfitting of noise thus resulting in poor prediction ability. For that reason it is often preferable to select an appropriate subset of parameters before performing a parameter estimation. To this end, it is first necessary to perform a sensitivity analysis as described below.

2.2.1. Sensitivity Analysis

A local sensitivity analysis considers sensitivities based on single parameter perturbations. The effect of an individual parameter perturbation on the model output is given as:

$$S_{\theta_i}^{y_j}(t_k) = \frac{\partial y_j}{\partial \theta_i}(\theta, t_k)\frac{\theta_i}{y_j(t_k)} \tag{2.4}$$

which is the derivative of the model output y_j with respect to the parameter θ_i scaled by the magnitude in the nominal parameter value and model output. To obtain an overall sensitivity of an output for different sampling times, the coefficients can be averaged up over all time points as follows:

$$S_{\theta_i}^{y_j} = \frac{1}{n_t}\sum_{k=1}^{n_t}\left|S_{\theta_i}^{y_j}(t_k)\right| \tag{2.5}$$

When it is desired to compare the effect of different parameters on all outputs, the quantities in (2.5) can be further averaged over all the outputs to obtain the following local sensitivity measure:

$$S_{\theta_i}^{y} = \frac{1}{n_y}\sum_{j=1}^{n_y} S_{\theta_i}^{y_j} \tag{2.6}$$

There are several techniques for computing the derivatives needed for the analysis. One method is

to approximate the derivative by means of a forward finite difference as follows:

$$f^{'}(x) \approx \frac{f(x+h) - f(x)}{h} \tag{2.7}$$

or alternatively by a central difference approximation:

$$f^{'}(x) \approx \frac{f(x+h) - f(x-h)}{2h} \tag{2.8}$$

2.2.2. Parameter Selection

For parameter selection within the context of a nonlinear regression problem, it is essential to determine the effect of each parameter on the model outputs as well as the correlations between the effects of different parameters. Ideally, one is interested in a set of parameters with very little correlation between the parameter's effect on the model output, as this would allow for a precise estimation. This information can be attained from the parameter covariance matrix $\boldsymbol{V}_\theta$, which generally has to be approximated by linearization. For nonlinear systems, it has been shown that the inverse of Fisher Information Matrix (FIM) provides a lower bound (Walter & Pronzato, 1990) for the covariance matrix. The FIM is defined as:

$$\boldsymbol{F} = \boldsymbol{S}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{S} \tag{2.9}$$

where $\boldsymbol{S}$ is the sensitivity matrix and $\boldsymbol{\Sigma}$ the measurement covariance matrix. Using the scaled sensitivities from (2.4), the matrix of output sensitivity functions $\boldsymbol{S}$ at sampling time t_k can be

defined as:

$$\boldsymbol{S}(t_k) = \begin{bmatrix} S_{\theta_1}^{y_1}(t_k) & S_{\theta_i}^{y_1}(t_k) & \cdots & S_{\theta_{n_\theta}}^{y_1}(t_k) \\ S_{\theta_1}^{y_j}(t_k) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ S_{\theta_1}^{y_{n_y}}(t_k) & \cdots & \cdots & S_{\theta_{n_\theta}}^{y_{n_y}}(t_k) \end{bmatrix} \tag{2.10}$$

which can be summed up over all time points to yield:

$$\boldsymbol{S} = \sum_{i=k}^{n_t} \boldsymbol{S}(t_k) \tag{2.11}$$

As the measurement noise is commonly assumed to be uncorrelated (white), the covariance matrix of the measurement noise $\boldsymbol{\Sigma}$ is typically given as:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_{n_y}^2 \end{bmatrix} \tag{2.12}$$

Several criteria have been proposed to determine a subset of parameters for a subsequent estimation (Walter & Pronzato, 1990). In Weijers & Vanrolleghem (1997), the modified E-criterion and the D-criterion have been applied to select the parameter subsets which yield the highest identifiability. In that regard, the E-criterion corresponds to the condition number of $\boldsymbol{F}$, whereas the D-criterion is based on the determinant of $\boldsymbol{F}$ (Atkinson & Donev, 1992). This is similar to the determinant measure applied to parameter subsets in Brun et al. (2002). Another approach to determine a group of identifiable parameters is by successively eliminating parameters which correspond to small eigenvalues of the $\boldsymbol{F}$ (Schittkowski, 2008). Many of these criteria correspond to geometric interpretations of the joint confidence region (Franceschini & Macchietto, 2008). For example, while the D-criterion describes the volume of the joint confidence region, the E-criterion relates to

the size of its major axis.

A method to address parameter correlations by ranking parameters according to their effect on the model output is the orthogonalization method proposed by Yao et al. (2003). In this method, the most influential parameter is determined by comparing the magnitude of the respective columns of the sensitivity matrix, given in (2.10). The parameter corresponding to the largest magnitude is deemed to be the most important. To address the correlation between parameters, the sensitivity matrix is successively adjusted to account for the effect of each parameter. For example, after computing the first significant parameter, the sensitivity vectors related to the remaining parameters are projected onto the space normal to the first sensitivity vector. This way, the "net influence" of the remaining parameters can be assessed. After the adjustment, the second most influential parameter can be determined by comparing again the magnitude of the remaining columns. This procedure can be continued until the column with the largest magnitude of the residual matrix, obtained after successive projections, is lower than a prescribed cut-off value. Using this approach it is possible to rank the model parameters based on their respective "net influence" on the model output. The orthogonalization method is a sequential method which optimizes the D-criterion at each step. However, due to its sequential nature, it is possible that the resulting parameter set might not be the most significant (Chu & Hahn, 2007).

Another important issue regarding parameter estimation for nonlinear models is the uncertainty in nominal parameter values. The particular choice of the nominal values can have a significant impact on the respective sensitivities and therefore may lead to different results. In order to account for the uncertainty present in the nominal parameter values and as well as in the model input, a global sensitivity analysis (Saltelli et al., 2008) can be performed over a range of operating conditions to assess the overall effect of parametric sensitivities .

2.2.3. Design of Experiments

Design of experiments (DoE) is an important aspect in improving parameter precision and model prediction (Atkinson & Donev, 1992; Franceschini & Macchietto, 2008). The goal of model-based experimental design is to obtain the next experimental conditions that minimize a measure of the parameter covariance matrix thus reducing the corresponding uncertainty in the estimated parameters. For example, the D-criterion, which minimizes the volume of the joint confidence region, corresponds to the maximization of the determinant of $\boldsymbol{V}_{\boldsymbol{\theta}}^{-1}$ (Franceschini & Macchietto, 2008), which, from (2.9), is given by:

$$\boldsymbol{V}_{\boldsymbol{\theta}}^{-1} = \boldsymbol{F} \tag{2.13}$$

This leads to the following objective function:

$$\psi(\hat{\boldsymbol{\theta}}, \boldsymbol{u}) = \det \left| \boldsymbol{V}_{\theta}^{-1} \right| = \det \left| \sum_{i=1}^{n_t} \boldsymbol{S}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{S}_i \right| \tag{2.14}$$

where $\boldsymbol{u}$ describes the input vector which is generally constrained by the design space. Using the objective in (2.14), it is possible to determine the input for the next batch run that is especially informative and would result in and increase in parameter precision:

$$\begin{aligned} \boldsymbol{u}_{DoE} \quad &= \quad \arg\min_{\boldsymbol{u}} -\psi(\hat{\boldsymbol{\theta}}, \boldsymbol{u}) \\ &\text{s.t.} \quad (2.1) \text{ and } (2.2) \\ &\quad\;\; \boldsymbol{u}^L \leq \boldsymbol{u} \leq \boldsymbol{u}^U \end{aligned} \tag{2.15}$$

where $\boldsymbol{u}_{DoE}$ describes the input for the input for the next experiment to reduce the parametric uncertainty. Using this approach, it is also possible to maximize the information content when multiple experiments are performed at the same time (Galvanin et al., 2007).

2.3. Run-to-Run Optimization

The repetitive nature of batch processes offers the advantage of making use of previous batch runs to improve the performance of future runs (Bonvin, 1998). In this fashion, a run-to-run optimization can be performed, where a model is updated after each batch-run to determine an optimal input for the next run. In the past, batch-to-batch optimization approaches have been mainly based on data-driven models such as neural networks (Dong et al., 1996), partial least squares (Camacho et al., 2007; Duran-Villalobos & Lennox, 2013) or hybrid modelling approaches (Doyle et al., 2003; Teixeira et al., 2006). In contrast to these data-driven modelling approaches and for the remainder of the work, we focus on the use of first-principles models that are updated and optimized in a run-to-run framework to ultimately reach an optimal set of operating conditions.

In general, a batch process can be seen as a static map between a vector of inputs $\boldsymbol{u}$ and outputs $\boldsymbol{y}$, i.e. $\boldsymbol{y} = \mathcal{M}(\boldsymbol{u})$ (Bonvin & Francois, 2017). It is also common for a run-end output (at the final time t_f of the batch) to be the quantity of interest to be optimized (maximized or minimized), such that $z = \boldsymbol{y}(t_f)$. The primary goal of a run-to-run optimization is to progressively minimize a cost-function ϕ through the course of numerous repetitive batch-runs to achieve a desired process improvement. A simple method that performs an iterative model update followed by an optimization step, referred to as the two-step approach, is described below.

2.3.1. Two-Step Method

In the two-step approach, reported in previous studies (Chen & Joseph, 1987; Ruppen et al., 1995; Marlin & Hrymak, 1997), an identification step, in which unknown parameters are identified, is followed by an optimization step, where the optimal operating point is determined. An illustration of the approach is given in figure 2.1.

In general, the first step is an identification step that involves the minimization of the difference

between the predicted and measured outputs, which can be expressed for batch run k as follows:

$$\begin{aligned} \boldsymbol{\theta}_k \quad &= \quad \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n_t} \left\| \boldsymbol{y}_{p,k}(t_i) - \boldsymbol{y}_k(\boldsymbol{\theta}, t_i) \right\|^2 \\ &\text{s.t.} \quad \dot{\boldsymbol{x}}_k = \boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{\theta}) \\ &\qquad \boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k) \\ &\qquad \boldsymbol{\theta} \in [\boldsymbol{\theta}_{lb}, \boldsymbol{\theta}_{ub}] \end{aligned} \tag{2.16}$$

where $\boldsymbol{y}_{p,k} \in \mathbb{R}^{n_t \times n_y}$ are the plant measurements and $\boldsymbol{y}_k \in \mathbb{R}^{n_t \times n_y}$ the model outputs. $\boldsymbol{u}_k \in \mathbb{R}^{n_u}$ are the decision variables defining the current operating point. $\boldsymbol{\theta}_k \in \mathbb{R}^{n_\theta}$ describes the particular set of parameter estimates which minimizes the identification objective. The bounds $\boldsymbol{\theta}_{lb} \in \mathbb{R}^{n_\theta}$ and $\boldsymbol{\theta}_{ub} \in \mathbb{R}^{n_\theta}$provide a permissible range of parameter values.

Once the model has been updated using the most recent output measurements, the second step involves finding the input for which the model predicts a minimum for the given cost-function. This optimization step is given by:

$$\begin{aligned} \boldsymbol{u}_{k+1} \quad &= \quad \arg\min_{\boldsymbol{u}} \phi(\boldsymbol{y}(\boldsymbol{\theta}_k), \boldsymbol{u}) \\ &\text{s.t.} \quad \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\theta}_k) \\ &\qquad \boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}) \\ &\qquad \boldsymbol{g}(\boldsymbol{y}(\boldsymbol{\theta}_k), \boldsymbol{u}) \leq \boldsymbol{0} \\ &\qquad \boldsymbol{u}^L \leq \boldsymbol{u} \leq \boldsymbol{u}^U \end{aligned} \tag{2.17}$$

The optimization is subject to constraints $\boldsymbol{g} \in \mathbb{R}^{n_g}$ and lower and upper bounds on the decision variables $\boldsymbol{u}^L$ and $\boldsymbol{u}^U$, both $\in \mathbb{R}^{n_u}$. According to the method, the two steps are successively repeated to account for changing process conditions and disturbances until convergence to the optimum is achieved. However, this two-step approach is only successful in determining the plant optimum

if there is no major discrepancy between the model and the plant. When such mismatch occurs, the two step approach outlined above must be modified to reach the actual optimum as explained below.

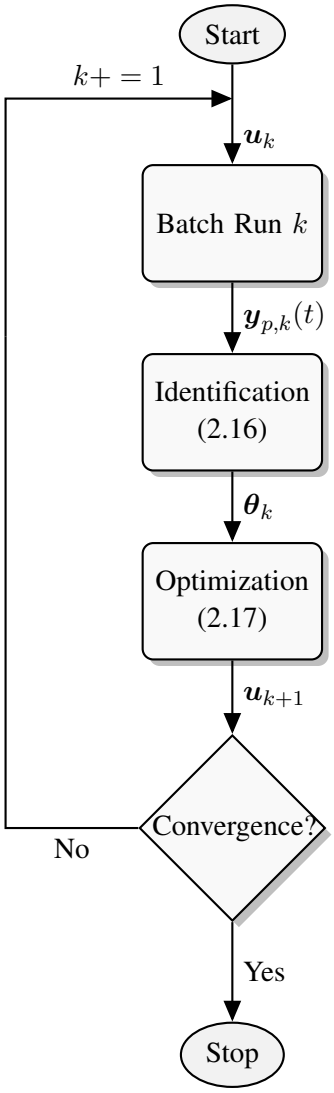

Figure 2.1.: Outline of the two-step procedure.

2.3.2. Model-Plant Mismatch

Structural mismatch between a fundamental model and process under study may arise due to simplifications and assumptions during the model development. Such assumptions generally arise due to lack of a priori knowledge about some phenomena occurring in the process, e.g. ignoring a particular metabolic reaction occurring in a biochemical system. Or it is intentionally introduced by the designer due to the need to reduce the complexity and to obtain a manageable model. Contrary to statistical uncertainty such as measurement noise, model-plant mismatch presents a systematic uncertainty and cannot be reduced by performing additional experiments. If such discrepancies exist, the model might converge to an optimum which does not coincide with that of the plant (Biegler et al., 1985; Agarwal, 1997). In addition, even for the parameters that are included in the model, it is necessary that sufficient excitation exists for estimating suitable parameter values. This leads to further problems in convergence due to a lack of synergy between the identification and optimization objective (Srinivasan & Bonvin, 2002). Specifically, the predicted gradients of the cost-function do not match those of the process so that it is impossible to satisfy the necessary conditions of optimality (NCO) of the process. Hence, the standard two step approach, described in the previous section, is therefore not practical in the presence of structural model-plant mismatch in combination with parametric uncertainty. Rather, it is necessary to either employ a model that is already shown to be adequate for optimization or to perform certain model corrections to specifically enforce model adequacy.

2.3.3. Model Adequacy

The first-order necessary conditions of optimality (NCOs) of the process are also known as the Karush-Kuhn-Tucker (KKT) conditions. Assuming that the objective and constraint functions in (2.17) and the real process are continuously differentiable at a set of optimal input conditions $\boldsymbol{u}^*$,

the KKT conditions of the process are given as:

$$\frac{\partial \phi}{\partial \boldsymbol{u}}(\boldsymbol{u}^*) + \boldsymbol{\mu}^T \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{u}}(\boldsymbol{u}^*) = 0 \tag{2.18}$$

$$\boldsymbol{\mu}^T \boldsymbol{g}(\boldsymbol{u}^*) = 0 \tag{2.19}$$

$$\boldsymbol{\mu} \geq \mathbf{0} \tag{2.20}$$

$$\boldsymbol{g}(\boldsymbol{u}^*) \leq \mathbf{0} \tag{2.21}$$

where $\boldsymbol{\mu}$ is the vector of KKT multipliers. The KKT conditions present a generalization of the method of Lagrange multipliers which presents a strategy of finding an optimum in the presence of equality constraints. (2.18) represents the stationary condition, i.e. the gradient of the objective function is canceled out by the gradients of the active constraints. This firstly requires that the feasibility of the primal problem (2.21) is satisfied. Furthermore, due to the addition of inequality constraints, which are not necessarily active at the optimum, the KKT conditions include the complementary slackness (2.19) and the dual feasibility (2.20).

In order to reach the actual process optimum by means of a model-based optimization it is necessary to determine if the model is suitable for the purpose of optimization. With the so called *model adequacy* requirements, Forbes et al. (1994) and Forbes & Marlin (1996) determined the ability of a model to accurately predict the process optimum. That is, given the optimal manipulated variable, the model should satisfy the KKT conditions at the plant optimum. For a combined identification and optimization problem, given the process optimum $\boldsymbol{u}^*$ and a set of parameters $\boldsymbol{\theta}^*$, the augmented

model adequacy criteria are given as follows:

$$\frac{\partial J_{ident}}{\partial \boldsymbol{\theta}}(\boldsymbol{y}(\boldsymbol{\theta}^*), \boldsymbol{u}^*) = \mathbf{0} \tag{2.22}$$

$$\nabla_r \phi(\boldsymbol{y}(\boldsymbol{\theta}^*), \boldsymbol{u}^*) = \mathbf{0} \tag{2.23}$$

$$\frac{\partial^2 J_{ident}}{\partial \boldsymbol{\theta}^2}(\boldsymbol{y}(\boldsymbol{\theta}^*), \boldsymbol{u}^*) \succ \mathbf{0} \ (positive\ definite) \tag{2.24}$$

$$\nabla_r^2 \phi(\boldsymbol{y}(\boldsymbol{\theta}^*), \boldsymbol{u}^*) \succ \mathbf{0} \ (positive\ definite) \tag{2.25}$$

where the identification and optimization objective are given by J_{ident} and ϕ respectively. The subscript r denotes the reduced gradient, i.e. the gradient in the direction where the active constraints are not affected. Point-wise adequacy is a necessary requirement for determining the true optimum in a model-based optimization framework. However, if model-plant mismatch is present and the optimum of the plant is unknown - which is generally the case - it is not possible to verify the adequacy of a model beforehand. Moreover, due to structural model-plant mismatch, a set of parameters satisfying the KKT conditions at the plant optimum might not exist. Therefore, for model-based optimization in the presence of model-plant mismatch, it is necessary to utilize a model which, following a suitable adaptation mechanism, will properly describe the process in the proximity of the optimum. To that purpose, the standard two-step approach outlined in the previous section must be modified to account for the differences between the model and process gradients of the constraints and the objective function as shown in the following section.

2.4. Optimization Under Model-Plant Mismatch

In the following, several approaches are presented that aim at enforcing model-adequacy by correcting for errors between predicted and measured cost-function and constraint gradients.

2.4.1. Modifier Adaptation

In the modified two-step algorithm of Roberts (1979) and Roberts & Williams (1981), a term is added to the objective function to account for the differences between the model and process objective function thus achieving the correct optimal process condition despite structural uncertainties. The method was subsequently extended in Brdys et al. (1986) to also account for differences in constraints. Additional correction terms have been proposed (Gao & Engell, 2005), so that the gradients of the cost function and constraints of the model are equal to those of the process. A similar approach referred to as the *integrated system optimization and parameter estimation (ISOPE) method* has been proposed in Brdys & Tatjewski (2005). In the works of Tatjewski (2002), the estimation step was replaced by introducing a linear term in the output which corrects for the difference between model prediction and measurements. In the subsequently developed *Modifier Adaptation* (MA) framework (Marchetti et al., 2009; Chachuat et al., 2009; Costello et al., 2016; Gao et al., 2016), the identification step is completely eliminated since for the purpose of converging to the optimum, it was only considered of importance to match the objective and the constraints as well as their respective gradients accurately. Accordingly, this fixed-model based method uses measurements and a mostly inaccurate process model to guide the process to an optimal operating point. The modified optimization problem can be stated as follows:

$$\begin{aligned} \boldsymbol{u}_{k+1} \quad = \quad & \arg\min_{u} \phi(\boldsymbol{y}(\boldsymbol{u},\boldsymbol{\theta}),\boldsymbol{u}) + \boldsymbol{\lambda}_{\phi k}^{T}\boldsymbol{u} \\ s.t. \quad & \dot{\boldsymbol{x}} = f(\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{u}) \\ & \boldsymbol{y} = h(\boldsymbol{x}) \\ & \boldsymbol{g}(\boldsymbol{y}(\boldsymbol{u},\boldsymbol{\theta}),\boldsymbol{u}) + \boldsymbol{\varepsilon}_{gk} + \boldsymbol{\lambda}_{gk}^{T}(\boldsymbol{u}-\boldsymbol{u}_k) \leq \boldsymbol{0} \end{aligned} \tag{2.26}$$

where the modifiers $\boldsymbol{\lambda}_{\phi}$ and $\boldsymbol{\lambda}_{g}$ are used to correct for the gradients of the objective and constraints. The modifier $\boldsymbol{\varepsilon}_{g}$ corrects for the mismatch between the predicted and measured constraint values.

At the k-th iteration the modifiers are calculated as follows:

$$\lambda_{\phi k_i} = \frac{\partial \phi_p}{\partial u_i}(\boldsymbol{y}_p(\boldsymbol{u}), \boldsymbol{u}) - \frac{\partial \phi}{\partial u_i}(\boldsymbol{y}(\boldsymbol{u}, \boldsymbol{\theta}), \boldsymbol{u}) \tag{2.27}$$

$$\lambda_{\phi k_{ij}} = \frac{\partial g_{j,p}}{\partial u_i}(\boldsymbol{y}_p(\boldsymbol{u}), \boldsymbol{u}) - \frac{\partial g_j}{\partial u_i}(\boldsymbol{y}(\boldsymbol{u}, \boldsymbol{\theta}), \boldsymbol{u}) \tag{2.28}$$

$$\varepsilon_{g k_j} = g_{j,p}(\boldsymbol{y}_p(\boldsymbol{u}), \boldsymbol{u}) - g_j(\boldsymbol{y}(\boldsymbol{u}, \boldsymbol{\theta}), \boldsymbol{u}) \tag{2.29}$$

It is important to notice that by disregarding the identification step, the model can no longer be reliably used for prediction purposes. However, as stated in chapter 1, the ability to do predictions is often very important for predicting the outputs around the optimum, reference trajectories for closed-loop control and to enforce constraints on variables of interest. Moreover, the MA method has been shown to be more susceptible to uncertainty in measurements of gradients thus requiring the use of an ad-hoc filter. If filtering is not applied the method results in excessive over corrections that may lead to oscillations converging towards the optimum. This may be especially problematic at the beginning of a batch-to-batch optimization procedure where enough data is not a priori available.

2.4.2. Direct Input Adaptation

Another class of methods involves transforming the iterative optimization problem into a feedback control problem to implicitly optimize the process. In this case, functions of the measured variables are to be controlled at pre-specified constant values in order to enforce optimal operation of the process (Morari et al., 1980). Skogestad (2000) proposed a method referred to as *self-optimizing control* where a linear combination of output variables is tracked in order to maintain close to optimal performance. However, since this tracking does not automatically result in an optimal operation, it was proposed (Francois et al., 2005) to select the NCO components as the controlled variables and thus enforcing optimality of the process. A similar approach is extremum-seeking

control (Ariyur & Krstic, 2003; Cougnon et al., 2011; Zhang et al., 2003), where the system states are driven towards desired set-points to maximize the process performance. However, in this class of methods, the model is no longer of importance and consequently it is not continuously updated. Moreover, these methods are particularly tailored for online or real-time optimization and are not suitable for optimizing batch processes (Bonvin & Francois, 2017).

2.4.3. Simultaneous Identification and Optimization

To address the problem of a combined identification and optimization under model-plant mismatch, Mandur & Budman (2015b) developed a framework where both objectives are satisfied simultaneously. In the work of Srinivasan & Bonvin (2002), the identification objective is modified by including a weighted optimization objective. This "modeling for optimization" paradigm results in a trade-off between identification and optimization. In contrast, Mandur & Budman (2015b) proposed to satisfy the identification objective first. The gradients of the objective are then matched to the plant gradients in order to satisfy the optimization objective to a specified extent. The accuracy of the identification step is maintained after adjusting the parameters to match the gradients by incorporating a linear correction term into the model output. The method was shown to provide a model which is accurate in each iteration up to a certain pre-specified error as well as good model-based filtering capabilities with respect to gradient uncertainty.

In the simultaneous identification and optimization framework proposed by Mandur & Budman (2015b), the identification step is performed in similar way as described in (2.16), where the dif-

ference between the prediction and measurements is minimized:

$$\begin{aligned}\boldsymbol{\theta}_k \quad &= \quad \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n_t} \left\| \boldsymbol{y}_{p,k}(t_i) - \boldsymbol{y}_k(\boldsymbol{\theta}, t_i) \right\|^2 \\ &\text{s.t.} \quad \dot{\boldsymbol{x}}_k = \boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{\theta}) \\ &\qquad \boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k) - \boldsymbol{c}_{k-1} \\ &\qquad \boldsymbol{\theta} \in [\boldsymbol{\theta}_{lb}, \boldsymbol{\theta}_{ub}] \end{aligned} \tag{2.30}$$

where the main difference to (2.16) is the correction term $\boldsymbol{c}_{k-1} \in \mathbb{R}^{n_t \times n_y}$, explained further below, which is subtracted from the model output to maintain the fitting accuracy when correcting for errors in the gradients.

Before a gradient correction can be performed, it is first of all necessary to obtain an estimate of the plant gradients. Here, we assume that measurements of the cost-function and constraints are available. One possibility to obtain a gradient estimate at input $\boldsymbol{u}_k$ is to run additional batch experiments with the perturbed inputs $\boldsymbol{u}_k + \boldsymbol{\delta}_{kj}$. Then, using a finite difference approach, the gradient estimates of the cost can be obtained as follows:

$$\nabla \phi_{p,j} = \frac{\phi(\boldsymbol{u}_k + \boldsymbol{\delta}_{kj}) - \phi(\boldsymbol{u}_k)}{\|\boldsymbol{\delta}_{kj}\|} \tag{2.31}$$

where, with $j \in \{1, \ldots, n_u\}$, the perturbations $\boldsymbol{\delta}_{kj}$ are implemented in each direction of the decision variables, i.e.:

$$\boldsymbol{\delta}_{kj} = \Delta u_j \boldsymbol{e}_j \tag{2.32}$$

where $\boldsymbol{e}_j$ presents the identity vector in the direction of variable j and Δu_j the corresponding perturbation step size. The constraint gradients can be obtained in the same way as shown in (2.31).

Following the gradient estimation, model adequacy is enforced by changing the previously obtained

parameter estimates by an amount $\Delta\boldsymbol{\theta}$ in the following way:

$$\begin{aligned}
\Delta\boldsymbol{\theta}_k \quad = \quad & \arg\min_{\Delta\boldsymbol{\theta}} \Bigg(\boldsymbol{w}_\phi^T \left| \nabla\boldsymbol{\phi}_p(\boldsymbol{u}_k) - \nabla\boldsymbol{\phi}(\boldsymbol{y}_k(\boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}), \boldsymbol{u}_k) \right| \\
& + \sum_{i=1}^{n_g} \boldsymbol{w}_{g,i}^T \left| \nabla\boldsymbol{g}_{p,i}(\boldsymbol{u}_k) - \nabla\boldsymbol{g}_i(\boldsymbol{y}_k(\boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}), \boldsymbol{u}_k) \right| \Bigg) \\
\text{s.t.} \quad & \dot{\boldsymbol{x}}_k = \boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}) \\
& \boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k) - \boldsymbol{c}_k \\
& \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta} \in [\boldsymbol{\theta}_{lb}, \boldsymbol{\theta}_{ub}] \\
& \left\| \boldsymbol{\epsilon}_k^T \right\|_\infty \leq \epsilon_{max}
\end{aligned} \tag{2.33}$$

With the gradient vectors of the cost-function and constraints given by $\nabla\boldsymbol{\phi} \in \mathbb{R}^{n_u}$ and $\nabla\boldsymbol{g}_i \in \mathbb{R}^{n_u}$ with $i \in \{1, \ldots, n_g\}$ respectively. The weights $\boldsymbol{w}_\phi \in \mathbb{R}^{n_u}$ and $\boldsymbol{w}_g \in \mathbb{R}^{n_u}$ are used to normalize the two gradient matching objectives. Furthermore, to maintain the same fitting accuracy as obtained in (2.30), a correction term $\boldsymbol{c}_k$ is introduced into the model output. This correction term is defined by a first-order Taylor series expansion and is updated in each iteration:

$$\boldsymbol{c}_k(t_i) = \boldsymbol{c}_{k-1}(t_i) + \boldsymbol{D}y_k(\boldsymbol{\theta}_k, t_i)\Delta\boldsymbol{\theta}_k \tag{2.34}$$

where the Jacobian of the model at sampling time t_i is given by $\boldsymbol{D}y_k(\boldsymbol{\theta}_k, t_i) \in \mathbb{R}^{n_y \times n_\theta}$. To avoid overfitting and reduce the effect of gradient uncertainty, the gradient correction step is constrained by an upper bound ϵ_{max} on the relative truncation error $\boldsymbol{\epsilon}_k$. The relative truncation error due to the linear correction is defined:

$$\boldsymbol{\epsilon}_k(t_i) = [\boldsymbol{y}_k(\boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}_k, t_i) - \boldsymbol{D}y_k(\boldsymbol{\theta}_k, t_i) - \boldsymbol{y}_k(\boldsymbol{\theta}_k, t_i)] \cdot [\mathrm{diag}\,(\boldsymbol{y}_k(\boldsymbol{\theta}_k, t_i))]^{-1} \tag{2.35}$$

The gradient correction step results in a set of updated parameter values:

$$\boldsymbol{\theta}'_k = \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}_k \tag{2.36}$$

Due to the correction term $\boldsymbol{c}_k$, the parameter values $\boldsymbol{\theta}'_k$ provide a simultaneous fitting of both model outputs and gradients of the cost-function and constraints. In this way, it is possible to reconcile the two objectives of identification and optimization despite the presence of model-plant mismatch. Finally, to determine the input for the next batch-run, a model-based optimization can be performed as follows:

$$\begin{aligned} \boldsymbol{u}_{k+1} \quad = \quad & \arg\min_{\boldsymbol{u}} \phi(\boldsymbol{y}(\boldsymbol{u}, \boldsymbol{\theta}'_k), \boldsymbol{u}) \\ \text{s.t.} \quad & \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\theta}'_k) \\ & \boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{c}_k \\ & \boldsymbol{g}(\boldsymbol{y}(\boldsymbol{u}, \boldsymbol{\theta}'_k), \boldsymbol{u}) \leq \boldsymbol{0} \\ & \boldsymbol{u}^L \leq \boldsymbol{u} \leq \boldsymbol{u}^U \end{aligned} \tag{2.37}$$

For clarification, the main steps of the algorithm are illustrated in figure 2.2.

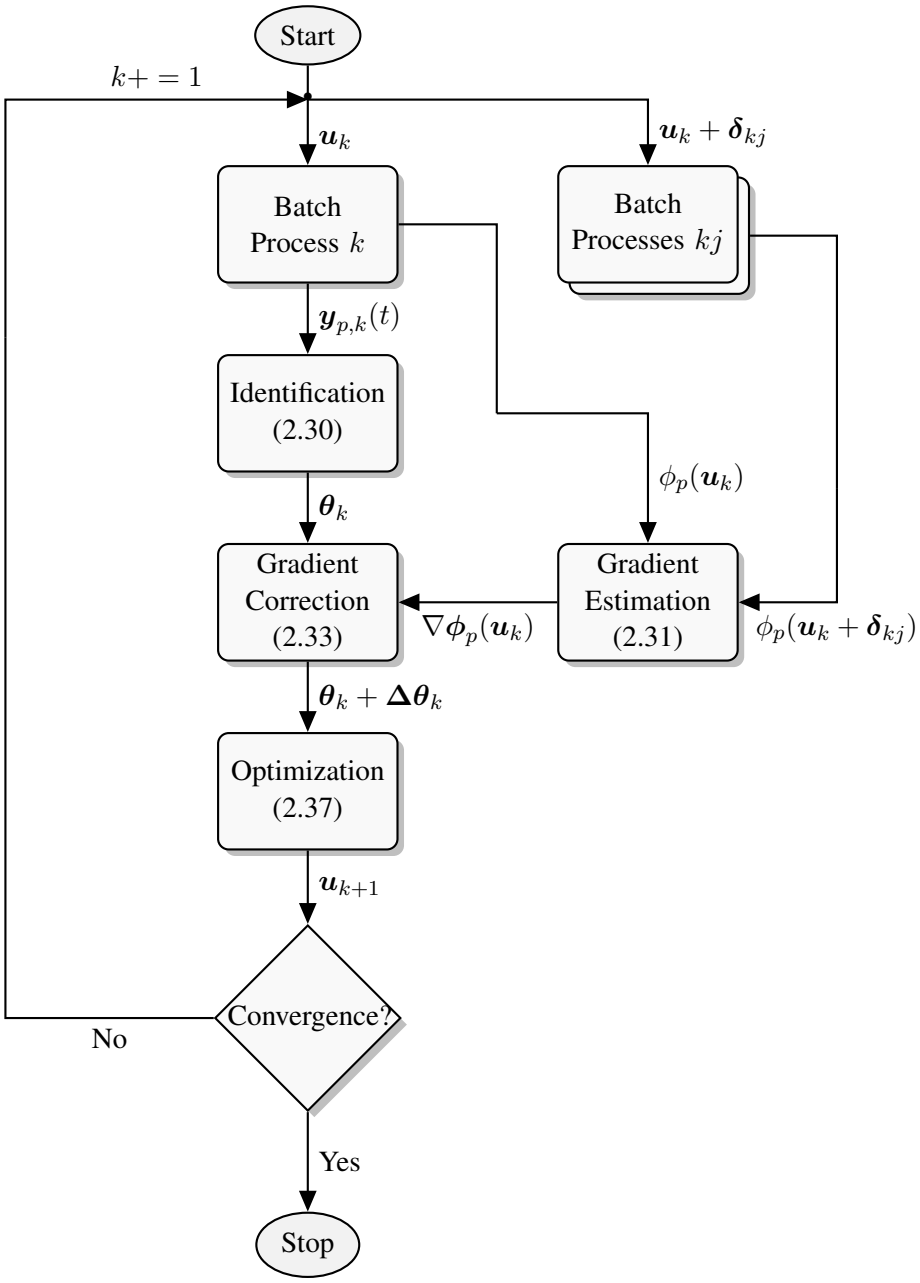

Figure 2.2.: Block diagram of the simultaneous identification and optimization methodology.

2.5. Robust Optimization

As explained in section 2.4, when performing optimization in the presence of model-plant mismatch it is essential to satisfy model adequacy conditions. In contrast, the primary goal of a robust optimization procedure is typically to optimize a worst case cost with respect to uncertainties in the model input - usually of stochastic nature - that affect the predicted optimum. Robustness is introduced by optimizing the nominal value or the mean, while minimizing the variation in the performance caused by various input uncertainties such as parametric uncertainty or uncertainty in initial batch conditions. To accomplish this, it is first of all necessary to propagate the effect of the input uncertainties onto the model output as described below.

2.5.1. Uncertainty Propagation

One of the main aspects of robust optimization is quantifying the effect of uncertainty on the objective function (cost) to be optimized. Common approaches approximate the objective function by a Taylor Series Expansion around the nominal parameter values (Ma et al., 1999). This method works well, if the objective function is linear or quadratic and the uncertainty in the parameters is reasonably small. However, this assumption are generally not met for most systems especially for those exhibiting highly nonlinear behavior. Moreover, the result is usually a worst-case solution, which may result in a very conservative or even infeasible solution.

Another approach for propagating uncertainty onto the cost utilizes Monte Carlo (MC) Sampling. While this method works well for nonlinear systems and a wide range of uncertainty descriptions, it is also computationally expensive, as it requires a large amount of simulated samples to correctly describe the stochastic effect of the inputs. This limits the usefulness of the Monte Carlo approach, especially if the analysis has to be performed in real time. Another disadvantage of the Monte Carlo method is that it does not provide a manageable representation of the process (Nagy &

Braatz, 2007). A more suitable method, which recently has received increasing attention, is based on Polynomial Chaos Expansions (PCEs).

2.5.2. Polynomial Chaos Expansions

The main advantage of PCEs is that any complex probability distribution can be described by the expansions and that the mean and variance of the resulting statistical distribution of the outputs or cost can be calculated analytically. Another benefit is that it requires less model simulations for the same accuracy as compared to conventional methods and thus it is computationally attractive. Recently, PCEs have been used for fault detection (Du et al., 2017), robustness analysis (Streif et al., 2016) and control (Paulson & Mesbah, 2017). This section provides an introduction to the theory of PCEs and the numerical methods which are necessary to compute the coefficients of the expansion. More information about the spectral approach of uncertainty quantification can be found in (Spanos & Ghanem, 1991; Xiu, 2010).

Given a probability space $(\Omega, \mathcal{F}, P)$, where Ω is a sample space, $\mathcal{F}$ a σ-algebra on Ω and P is a probability measure. With $\{\xi_i(\omega)\}_{i=1}^{\infty}$ we have a set of independent standard Gaussian random variables on Ω. Then we can represent any random variable $X : \Omega \to \mathbb{R}$, with finite variance, by a polynomial chaos expansion:

$$X(\omega) = a_0\Gamma_0 + \sum_{i_1=1}^{\infty} a_{i_1}\Gamma_1(\xi_{i_1}) + \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1} a_{i_1 i_2}\Gamma_2(\xi_{i_1}, \xi_{i_2})$$

$$+ \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1}\sum_{i_3=1}^{i_2} a_{i_1 i_2 i_3}\Gamma_3(\xi_{i_1}, \xi_{i_2}, \xi_{i_3}) + \cdots \tag{2.38}$$

where Γ_p is the Wiener Chaos (Wiener, 1938) of order p, ω is the random event and $a_{(.)}$ is the deterministic coefficient. With a one-to-one correspondence between the coefficients and functionals,

we can also rewrite the PCE in a more compact form (Spanos & Ghanem, 1991):

$$X(\omega) = \sum_{k=1}^{\infty} \hat{x}_k \Psi_k(\xi_1, \xi_2, \ldots) \tag{2.39}$$

For computational feasibility, one generally truncates the PCE in both order p and dimension n as follows:

$$X(\omega) = \sum_{k=1}^{P} \hat{x}_k \Psi_k(\xi_1, \xi_2, \ldots, \xi_n) \tag{2.40}$$

Thus, the number of terms in the PCE is given by:

$$P + 1 = \frac{(n+p)!}{n!p!} \tag{2.41}$$

Exponential convergence of this representation can be observed for random variables with the same density as that of $\boldsymbol{\xi}$. In the generalized polynomial chaos (gPC), developed by Xiu & Karniadakis (2002b), a class of orthogonal polynomials corresponds to the choice of distribution for the ξ_i (Najm, 2009). For the continuous case, Hermite polynomials correspond to a Gaussian distribution, Laguerre polynomial to a Laguerre distribution, Jacobi polynomials to a Beta distribution and Legendre polynomials to a Uniform distribution.

One important property of the PCE is that all basis functions are orthogonal to each other with respect to the probability distribution of $\boldsymbol{\xi}$. This orthogonality property leads to the following definition of the inner product:

$$\langle \Psi_i, \Psi_j \rangle = \int \Psi_i(\boldsymbol{\xi}) \Psi_j(\boldsymbol{\xi}) p(\boldsymbol{\xi}) d\boldsymbol{\xi} = \delta_{ij} \left\langle \Psi_i^2 \right\rangle \tag{2.42}$$

where $p(.)$ is the probability density of $\boldsymbol{\xi}$ and δ_{ij} the Kronecker delta. The orthogonality enables

us to evaluate the truncated PC representation by projecting it onto the corresponding PC basis:

$$\hat{x}_k = \frac{\langle X, \Psi_k \rangle}{\langle \Psi_k^2 \rangle} = \frac{1}{\langle \Psi_k^2 \rangle} \int X(\boldsymbol{\xi}) \Psi_k(\boldsymbol{\xi}) p(\boldsymbol{\xi}) d\boldsymbol{\xi} \tag{2.43}$$

The evaluation of the PCE coefficients can be either performed in an intrusive or non-intrusive way. In the intrusive approach, one can propagate the uncertainty through a model by using Galerkin projection to reformulate the governing equations into equations of the mode strength. The non-intrusive approach treats the original model as a black box and uses numerical evaluation of the PC modes of the model output. The resulting model for the non-intrusive approach is often referred to as a surrogate model since it is used to replace the original mathematical model.

Before these approaches will be explained in more detail, we first make the assumption that we have a model which produces the output $y = f(x)$, with the given input x (O'Hagan, 2013). If the input is a random variable then so is the output, thus we get $Y = f(X)$. Similar to (2.39), we can represent the output by a truncated PCE:

$$Y(\omega) = \sum_{j=1}^{P} \hat{y}_j \Psi_j(\boldsymbol{\xi}) \tag{2.44}$$

By using the model and PCE we can represent the uncertainty propagation in the following form:

$$\sum_{j=1}^{P} \hat{y}_j \Psi_j(\boldsymbol{\xi}) = f(\sum_{i=1}^{P} \hat{x}_i \Psi_i(\boldsymbol{\xi})) \tag{2.45}$$

2.5.2.1. Intrusive Approach

In the intrusive approach, we make use of the Galerkin projection for a reformulation of the governing equation. Applied to $y = f(x)$, we get:

$$\langle y, \Psi_k \rangle = \langle f(x), \Psi_k \rangle \tag{2.46}$$

for $k = 0, \ldots, p$. Similarly, for (2.45) we obtain the following:

$$\left\langle \sum_{j=1}^{P} \hat{y}_j \Psi_j, \Psi_k \right\rangle = \left\langle f(\sum_{i=1}^{P} \hat{x}_i \Psi_i), \Psi_k \right\rangle \tag{2.47}$$

For the simple case of $f(x) = \lambda x$ (Najm, 2009), with $\lambda = \sum_{i=1}^{P} \lambda_i \Psi_i$ (λ is also a random variable) and by employing orthogonality, we can rewrite (2.47) into:

$$\hat{y}_k = \sum_{i=1}^{P} \sum_{j=1}^{P} \lambda_i \hat{x}_j \frac{\langle \Psi_i \Psi_j \Psi_k \rangle}{\langle \Psi_k^2 \rangle}, \quad k = 0, 1, \ldots, P \tag{2.48}$$

The Galerkin method can be difficult to implement, as the reformulation has to be carried out each time a new model is used.

2.5.2.2. Non-Intrusive Approach

The non-intrusive approach has received much more attention in the literature than the intrusive approach. The main reason is that it is easier to implement as it does not afford a reformulation of the governing equations. In contrast to the intrusive method, the non-intrusive approach treats the model as a black-box. Here, one tries to solve the following version of the problem (O'Hagan, 2013):

$$\sum_{j=1}^{P} \hat{y}_j \Psi_j(\boldsymbol{\xi}) = f(\eta_p(\boldsymbol{\xi})) \tag{2.49}$$

where $\eta_p(\boldsymbol{\xi}) = \sum_{j=1}^{P} \hat{x}_j \Psi_j(\boldsymbol{\xi})$, using runs of the model for various samples of $\boldsymbol{\xi}$. With a n-dimensional basis $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_n)$ and the known PCEs, we follow the subsequent steps (Najm, 2009):

1. Based on the sampling strategy of interest, generate samples of $\boldsymbol{\xi}$,
2. From each sample $\boldsymbol{\xi}^i$, one can obtain $x^i = \sum_{j=1}^{P} \hat{x}_j \Psi_j(\boldsymbol{\xi}^i)$ and $y^i = f(x^i)$.

3. Then, from all the N samples, it is possible to compute the expectation of the Galerkin projection $\hat{y}_k = \frac{\langle Y, \Psi_k \rangle}{\langle \Psi_k^2 \rangle}, \quad \forall\, k \in \{0, 1, \ldots P\}$.

4. Finally, using the obtained $\hat{y}_k$ values, one computes $Y = \sum_{k=0}^{P} \hat{y}_k \Psi(\boldsymbol{\xi})$.

The computational burden in this approach originates from the computation of $y^i = f(x^i)$ for every x^i. For that reason, it is desired to achieve a given degree of accuracy by employing the least amount of samples. In this regard, random as well as deterministic sampling approaches can be used. Random sampling uses Monte Carlo evaluations and can make use of the structure of the integrand for more efficiency.

Deterministic sampling methods on the other hand present a reasonable alternative (Najm, 2009). They use numerical methods such as quadrature for the evaluation of projection integrals. These methods provide significant gains in efficiency over the random sampling methods, but only for low-dimensional systems where the dimensionality is related to the number of uncertain model parameters. If the systems are of higher dimensions, the number of quadrature points will rise exponentially ($(p+1)^n$), which renders these methods inefficient. In this regard, sparse-quadrature, Smolyak (Smoljak, 1963) or cubature methods can be implemented to tackle "the curse of dimensionality". In general, for problems with greater than two dimensions, Smolyak sparse grid approaches outperform tensor-product quadrature approaches (Eldred, 2009).

Another way to evaluate the projection integrals by means of a non-intrusive collocation approach is to evaluate the PCEs using a regression based on a selected set of points. Linear regression (also known as point collocation or stochastic response surfaces) uses a single linear least squares solution to solve the complete set of PCE coefficients. Although the method assures an accurate representation at the collocation points, it gives no possibility to control the error elsewhere.

3. Robust Batch-to-Batch Optimization in the Presence of Model-Plant Mismatch and Input Uncertainty

Overview

When performing a model-based optimization in the presence of model-plant mismatch, the set of model parameter estimates which satisfy an identification objective may not result in an accurate prediction of the gradients of the cost-function and constraints. To ensure convergence to the optimum, the predicted gradients must be forced to match the measured gradients by adapting the model parameters. Since updating all available parameters is impractical due to estimability problems and overfitting, there is a strong motivation for adapting a subset of parameters for updating the predicted outputs and gradients. This paper presents an approach to select a subset of parameters based on the sensitivities of the model outputs and of the cost function and constraint gradients. Furthermore, robustness to uncertainties in initial batch conditions is introduced using a robust formulation based on polynomial chaos expansions. The improvements in convergence to the process optimum and robustness are illustrated using a fed-batch bioprocess.

Adapted from Hille, R., Mandur, J., and Budman, H. (2017). Robust batch-to-batch optimization in the presence of model-plant mismatch and input uncertainty. *AIChE Journal*, 63, 2660-2670.

3.1. Introduction

In the chemical industry, batch processes are essential for the production of specialty products. For instance, common applications of batch and fed-batch operations are in the production of pharmaceutical products. In times of increasing competition, the chemical industry is faced with urgency to improve performance and minimize the cost of their operations. Because of the intricacy of these processes, mathematical models have become a fundamental part of the process industries' range of R&D activities. Beside their application to process design and control, a widespread use of process models is for performing model-based optimization by finding the minimum of a predicted cost function. In particular, batch-to-batch optimization has been used where the optimum is achieved in an iterative fashion after several batches.

This chapter deals with a model-based batch-to-batch optimization procedure that involves sequential identification and optimization steps performed for a series of batch experiments until an optimum is reached. In order to ensure convergence to the process optimum, the model must provide an accurate description of the given process in the neighbourhood of the optimum. However, when using mechanistic models (e.g. first-principles) for process optimization, structural mismatch between the model and the process is inevitable as simplifications and assumptions are often made to reduce the complexity of the model. In addition to this discrepancy, the calibration of a mechanistic model involves the identification of its parameters' values. This task is often challenging due to the presence of stochastic measurement noise and unknown input disturbances. Consequently, the structural and stochastic uncertainties reduce the accuracy of the calibrated model for predicting outputs over a large range of operating conditions. A necessary condition for model-based optimality is that the model used for optimization is able to predict the Karush-Kuhn-Tucker (KKT) conditions at the process optimum (Biegler et al., 1985), i.e. the model must accurately predict the gradients of the cost-function and constraints at the optimum (Forbes et al., 1994). To satisfy this condition in the presence of model-plant mismatch, it is essential to not only fit the model

predictions to measured process outputs but also to match the predicted gradients of cost-function and constraints to the measured ones. As a result of structural uncertainty, iterative optimization approaches that solely use repeated identification and optimization steps (Chen & Joseph, 1987; Marlin & Hrymak, 1997) may result in a sub-optimal operating policy or in a worst-case scenario, in violation of process constraints. This motivates the need for model-based optimization schemes which are robust to such mismatch.

To this end, a class of algorithms, referred to as Modifier Adaptation (Chachuat et al., 2009; Marchetti et al., 2010; Navia et al., 2015; Gao et al., 2016; Costello et al., 2016) has been reported based on earlier studies (Roberts, 1979; Brdys & Tatjewski, 2005; Gao & Engell, 2005). The main idea in these methods is to modify the cost-function and constraints to correct for the differences in predicted and measured gradients. The approach then guides the process to the optimal operating point by correcting for the mismatch iteratively through empirical modification of the cost function and constraints. Thus, the focus of these methods lies on correcting the cost function and constraints directly and does not explicitly update the model, which consequently cannot be used for prediction purposes. It is very often desirable to have a model that can provide accurate predictions around the optimum since such a model can be used for different objectives such as model predictions away from the optimum, developing soft-sensors and for calculating set-point trajectories for control. Then, the goal would be to find a set of parameter estimates, which can satisfy both, the identification and optimization objective. As this may not be possible due to structural mismatch, the problem of a combined identification and optimization has been addressed in (Srinivasan & Bonvin, 2002), where the identification objective is modified by including a weighted optimization objective. However, this approach results in a trade-off between the objectives of identification and optimization.

In order to reconcile the two objectives of identification and optimization in a systematic way, (Mandur & Budman, 2015b) proposed a simultaneous model identification and optimization frame-

work, where a separate gradient correction step is introduced after a conventional parameter identification step. By following this approach, the identification objective is satisfied first and subsequently the predicted gradients are matched to those of the process by adapting the model parameters. The accuracy of the identification step is maintained by incorporating a correction term into the model outputs and thus providing a set of parameter estimates, which can satisfy the identification and optimization objectives simultaneously. This method resulted in an accurate model in each iteration and also provided model-based filtering capability with an improved robustness to structural mismatch as well as to uncertainty in gradient measurements. Although in principle all model parameters could be updated and adapted for achieving matching of gradients as required in this latter procedure (Mandur & Budman, 2015b), this is often undesirable due to the sensitivity to noise, overfitting and computational effort. Furthermore, identifying all parameters may not be possible as often times, some of the states cannot be measured and, generally, the number of experiments is limited leading to insufficient excitation. In terms of model-fitting, it is typically desired to adjust the parameters that provide the largest effects on the model output. In that regard, there are several reported methods that are based on measures of the Fisher information matrix (FIM) and that have been applied to select the parameter subsets, which yield the highest parameter identifiability (Weijers & Vanrolleghem, 1997; Schittkowski, 2007). Methods related to the sensitivity matrix include a collinearity index (Brun et al., 2002), a principal component analysis (Degenring et al., 2004) and orthogonalization methods (Yao et al., 2003; Lund & Foss, 2008; Chu & Hahn, 2012).

An additional limitation of our earlier studies is that they only addressed model structure error but they did not address uncertainty in process inputs such as uncertainty in initial conditions. Obtaining a measure of the variance of the cost due to stochastic disturbances requires propagating their effect onto the outputs. A common approach is to use Monte Carlo (MC) methods in which the objective is calculated by sampling the uncertain inputs. However, to obtain an accurate estimate of the variance, MC methods are computationally expensive since they require a large number of in-

dividual model realizations. In recent years, Polynomial Chaos Expansions (PCEs) have been used as an efficient uncertainty propagation technique (Spanos & Ghanem, 1991; Xiu, 2010). Compared to MC, PCE offers the advantage that any commonly used probability density function can be propagated onto variables of interest and that the mean and variance can be calculated analytically (Nagy & Braatz, 2007). For these reasons, the PCE method has been recently applied to problems in control (Kim & Braatz, 2012), fault detection (Du et al., 2016) and robustness analysis (Mandur & Budman, 2014; Streif et al., 2016) and it is also used in this work for describing the effect of initial conditions on the outputs. Therefore, to address the limitations of our earlier studies (Mandur & Budman, 2015b), the current study presents three main novel contributions:

1. Development of a scheme for selecting a subset of parameters to be updated during each batch of the run-to-run procedure, whereas in the previous studies only a fixed parameter subset based on an a priori analysis has been used. The adaptive parameter selection is fully integrated into the iterative optimization framework, so that appropriate parameters are selected at each batch run. Considering that the sensitivities not only change between operating points, but also for changes in parameter values, it is shown that selecting an optimal set of parameters for each batch run can have a profound effect on the performance. The method is therefore not only adaptive in the sense of the operating point, but also takes into account the change in parameter values which also impacts the sensitivities of parameters due to correlation. Additional sampling of the sensitivities within the uncertain parameter space is performed in order to account for uncertainty in the previously estimated parameter values.

2. Selection of parameters based on the sensitivities of both the model outputs as well as of the gradients of the cost function and constraints. This is motivated by the fact that in contrast to parameter estimation problems where the goal is solely to fit the model predictions to a given set of data, the gradient correction step, required for model-based optimization, necessitates adapting parameters which more significantly affect the gradients of the cost function and

constraints. By using those sensitive parameters, it is shown that it is possible to match the predicted to measured gradients by smaller changes in the parameters thus increasing robustness. In addition, an orthogonalization technique (Yao et al., 2003) is incorporated to account for correlation among the parameters considered for updating.

3. Introduction of robustness into the simultaneous identification and optimization algorithm to account for both model structure error and uncertain inputs such as uncertain initial conditions. In contrast to previous studies of the co-authors, which only addressed robustness to model structure error, in this work we have also considered robustness to stochastic perturbations in the initial batch conditions. In that regard, we make use of PCE as an efficient uncertainty propagation technique in order to reduce computational effort which makes the methodology more practical and also potentially relevant for RTO applications.

The chapter is organized as follows: Section 3.2 reviews the parameter adaptation methodology. The proposed parameter selection algorithm is presented in section 3.3, whereas section 3.4 outlines the polynomial chaos-based uncertainty propagation approach. Finally, the proposed method and robustness analysis are illustrated by a case study of a penicillin fed-batch process in section 3.5 followed by conclusions in section 3.6.

3.2. Parameter Adaptation Methodology

The algorithm for the simultaneous identification and optimization (Mandur & Budman, 2015b) involves three steps that are separately reviewed below.

3.2.1. Parameter Estimation

A parameter identification is performed by minimizing the difference between model predictions and measurements collected at sampling times t_i with $i \in \{1, \ldots, n_t\}$ along a batch run:

$$\begin{aligned} \boldsymbol{\theta}_k \quad &= \quad \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n_t} \left\| \boldsymbol{y}_{p,k}(t_i) - \boldsymbol{y}_k(\boldsymbol{\theta}, t_i) \right\|^2 \\ &\text{s.t.} \quad \dot{\boldsymbol{x}}_k = \boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{\theta}) \\ &\qquad \boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k) - \boldsymbol{c}_{k-1} \\ &\qquad \boldsymbol{\theta} \in [\boldsymbol{\theta}_{lb}, \boldsymbol{\theta}_{ub}] \end{aligned} \tag{3.1}$$

where the subscript k denotes the batch index, $\boldsymbol{y}_{p,k} \in \mathbb{R}^{n_t \times n_y}$ the plant measurements, $\boldsymbol{x}_k \in \mathbb{R}^{n_t \times n_y}$ the states and $\boldsymbol{y}_k \in \mathbb{R}^{n_t \times n_y}$ the outputs of the model at the specified sampling times t_i. $\boldsymbol{u}_k \in \mathbb{R}^{n_u}$ is the vector of decision variables defining the current operating point at batch run k. $\boldsymbol{\theta}_k \in \mathbb{R}^{n_\theta}$ describes the particular set of parameter estimates, which minimizes the identification objective (3.1). $\boldsymbol{c}_{k-1} \in \mathbb{R}^{n_t \times n_y}$ is described below and presents a correction term necessary when correcting for errors in the cost-function and constraint gradients.

3.2.2. Gradient Correction

The parameter estimates $\boldsymbol{\theta}_k$, obtained in (3.1), lead to the following predicted gradients of the cost with respect to the decision variables at operating point $\boldsymbol{u}_k$:

$$\nabla \boldsymbol{\phi}(\boldsymbol{y}_k(\boldsymbol{\theta}_k), \boldsymbol{u}_k) = \frac{\partial \phi}{\partial \boldsymbol{u}}(\boldsymbol{y}_k(\boldsymbol{\theta}_k), \boldsymbol{u}_k) \tag{3.2}$$

where ϕ is the predicted cost-function and $\nabla \boldsymbol{\phi} \in \mathbb{R}^{n_u}$ the corresponding vector of gradients. Given

a set of process constraints $\boldsymbol{g} \in \mathbb{R}^{n_g}$, the predicted gradients of the constraints at $\boldsymbol{u}_k$ are as follows:

$$\nabla \boldsymbol{g}_j(\boldsymbol{y}_k(\boldsymbol{\theta}_k), \boldsymbol{u}_k) = \frac{\partial g_j}{\partial \boldsymbol{u}}(\boldsymbol{y}_k(\boldsymbol{\theta}_k), \boldsymbol{u}_k) \tag{3.3}$$

where $\boldsymbol{g}_j$ is a process constraint with $j \in \{1, \ldots, n_g\}$ and $\nabla \boldsymbol{g}_j \in \mathbb{R}^{n_u}$ is the respective constraint gradient vector. As mentioned in the introduction, a result of the discrepancy between the model and process is that the parameter estimates $\boldsymbol{\theta}_k$, which provide an accurate model-fit according to (3.1), do not necessarily lead to a correct prediction of the gradients. Convergence to the actual process optimum requires matching of the predicted to the measured cost function and constraint gradients, $\nabla \boldsymbol{\phi}_p$ and $\nabla \boldsymbol{g}_p$. This can be achieved by an additional optimization step where the originally obtained parameter estimates $\boldsymbol{\theta}_k$ are changed by an amount $\Delta\boldsymbol{\theta}$ calculated as follows:

$$\begin{aligned}
\Delta\boldsymbol{\theta}_k \quad = \quad & \arg\min_{\Delta\boldsymbol{\theta}} \Bigg(\boldsymbol{w}_\phi^T \left| \nabla \boldsymbol{\phi}_p(\boldsymbol{u}_k) - \nabla \boldsymbol{\phi}(\boldsymbol{y}_k(\boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}), \boldsymbol{u}_k) \right| \\
& + \sum_{j=1}^{n_g} \boldsymbol{w}_{g,j}^T \left| \nabla \boldsymbol{g}_{p,j}(\boldsymbol{u}_k) - \nabla \boldsymbol{g}_j(\boldsymbol{y}_k(\boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}), \boldsymbol{u}_k) \right| \Bigg) \\
\text{s.t.} \quad & \dot{\boldsymbol{x}}_k = \boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}) \\
& \boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k) - \boldsymbol{c}_k \\
& \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta} \in [\boldsymbol{\theta}_{lb}, \boldsymbol{\theta}_{ub}] \\
& \left\| \boldsymbol{\epsilon}_k^T \right\|_\infty \leq \epsilon_{max}
\end{aligned} \tag{3.4}$$

where the two gradient matching objectives are normalized by their respective weights $\boldsymbol{w}_\phi \in \mathbb{R}^{n_u}$ and $\boldsymbol{w}_{g,j} \in \mathbb{R}^{n_u}$. To maintain the same fitting accuracy that has been achieved in (3.1), it is necessary to introduce a correction term $\boldsymbol{c}_k \in \mathbb{R}^{n_t \times n_y}$ into the model output. The correction term is approximated by a first order Taylor series expansion and is updated as follows:

$$\boldsymbol{c}_k(t_i) = \boldsymbol{c}_{k-1}(t_i) + \boldsymbol{D} y_k(\boldsymbol{\theta}_k, t_i) \Delta\boldsymbol{\theta}_k \tag{3.5}$$

where $\boldsymbol{D}y_k(\boldsymbol{\theta}_k, t_i) \in \mathbb{R}^{n_y \times n_\theta}$ is the Jacobian of the model at sampling time t_i and $\boldsymbol{c}_{k-1}$ the correction term from the previous batch run. It should be noticed that approximating the correction term by the use of an expansion it is not strictly necessary but it is done for simplicity. Furthermore, the amount by which the parameter estimates are allowed to change is limited by a user selected upper bound ϵ_{max} on the relative truncation error, which is defined as the error introduced by the linear correction term in (3.5):

$$\boldsymbol{\epsilon}_k(t_i) = [\boldsymbol{y}_k(\boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}_k, t_i) - \boldsymbol{D}y_k(\boldsymbol{\theta}_k, t_i) - \boldsymbol{y}_k(\boldsymbol{\theta}_k, t_i)] \cdot [\mathrm{diag}\,(\boldsymbol{y}_k(\boldsymbol{\theta}_k, t_i))]^{-1} \tag{3.6}$$

where the $\mathrm{diag}(.)$ operator indicates a transformation of a vector into a diagonal matrix. The inequality in (3.4) yields a bound on the permitted change in parameter values and thus ensures the boundedness of the model after the gradient matching step. More importantly, it was shown that the specification of an upper bound on the relative truncation error provides robustness to uncertainty in the measured gradients of cost and constraints (Mandur & Budman, 2015b).

After performing the gradient correction, the adapted parameter values at batch run k are given by:

$$\boldsymbol{\theta}'_k = \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}_k \tag{3.7}$$

The overall idea behind the algorithm is that the parameter values in (3.7) satisfy the identification and gradient matching steps simultaneously and thus they reconcile the two objectives.

3.2.3. Robust Model-based Optimization

By using the updated parameter values given by (3.7), a model-based optimization is performed to determine the new optimal operating point. In order to introduce robustness with respect to uncertain initial conditions, the robust objective is formulated as a weighted sum of the expected

value and its variance:

$$\begin{aligned} \boldsymbol{u}_{k+1} \quad &= \quad \arg\min_{\boldsymbol{u}} \mathbf{E}\left[\phi\left(\boldsymbol{y}(\boldsymbol{\theta}_k'), \boldsymbol{u}\right)\right] + w\boldsymbol{Var}\left[\phi\left(\boldsymbol{y}(\boldsymbol{\theta}_k'), \boldsymbol{u}\right)\right] \\ &\text{s.t.} \quad \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\theta}_k') \\ &\qquad \boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{c}_k \\ &\qquad \boldsymbol{g}(\boldsymbol{y}(\boldsymbol{\theta}_k^{'}), \boldsymbol{u}) \leq \mathbf{0} \\ &\qquad \boldsymbol{u}^L \leq \boldsymbol{u} \leq \boldsymbol{u}^U \end{aligned} \tag{3.8}$$

where w represents a weight on the variance of the objective. It was shown in (Mandur & Budman, 2015b) that, upon convergence, the necessary conditions of optimality (NCOs) of the model are equal to those of the process at the plant optimum. In addition, to obtain a measure of the variance due to uncertainty in inputs, e.g. stochastic variability in initial conditions, it is necessary to propagate these probabilistic uncertainties into the model output. For that reason, PCE, a stochastic spectral method, is used in this work to quantify the effect of the input variability on the variation in the cost. When the variance term is ignored, the above formulation in (3.8) becomes a non-robust optimization.

3.2.4. Conditions of Convergence

For convergence of the parameter adaptation scheme to the neighbourhood of a local process optimum, it is required that the gradients of the cost and constraints can be matched to those of the process (Biegler et al., 1985). Matching of the process and model gradients has been referred to as model adequacy (Forbes et al., 1994), which, for a given set of model parameters $\boldsymbol{\theta}^*$, is satisfied if the following holds at the process optimum $\boldsymbol{u}^*$:

$$\nabla\boldsymbol{\phi}(\boldsymbol{y}(\boldsymbol{\theta}^*), \boldsymbol{u}^*) = \nabla\boldsymbol{\phi}_p(\boldsymbol{u}^*) \tag{3.9}$$

$$\nabla \boldsymbol{g}(\boldsymbol{y}(\boldsymbol{\theta}^*), \boldsymbol{u}^*) = \nabla \boldsymbol{g}_p(\boldsymbol{u}^*) \tag{3.10}$$

$$\nabla^2 \boldsymbol{\phi}(\boldsymbol{y}(\boldsymbol{\theta}^*), \boldsymbol{u}^*) > 0 \; (positive \; definite) \tag{3.11}$$

Following (3.4), this can be approximately achieved if a set of parameters $\boldsymbol{\theta}_k^{'}$ and a bound on the truncation error ϵ_{max} are such that the following amount of gradient correction can be satisfied in each iteration (Mandur & Budman, 2015b):

$$\left| \left(\nabla \boldsymbol{\phi}_p(\boldsymbol{u}_k) - \nabla \boldsymbol{\phi}(\boldsymbol{y}_k(\boldsymbol{\theta}_k'), \boldsymbol{u}_k) \right) + \sum_{j=1}^{n_g} \boldsymbol{w}_j^T \left(\nabla \boldsymbol{g}_{p,j}(\boldsymbol{u}_k) - \nabla \boldsymbol{g}_j(\boldsymbol{y}_k(\boldsymbol{\theta}_k'), \boldsymbol{u}_k) \right) \right| < \varepsilon \tag{3.12}$$

With the normalized weighting given by $\boldsymbol{w}_j = [\mathrm{diag}\,(\boldsymbol{w}_\phi)]^{-1} \boldsymbol{w}_{g,j}$. Convergence is then achieved when at the current input $\boldsymbol{u}_k$ the updated model predicts an optimum by satisfying the following stationary condition:

$$\nabla \boldsymbol{\phi}(\boldsymbol{y}_k(\boldsymbol{\theta}_k'), \boldsymbol{u}_k) + \sum_{j=1}^{n_g} \boldsymbol{\mu}_j^T \nabla \boldsymbol{g}_j(\boldsymbol{y}_k(\boldsymbol{\theta}_k'), \boldsymbol{u}_k) = 0 \tag{3.13}$$

Hence, combining (3.12) and (3.13) the following condition must hold for convergence:

$$\left| \nabla \boldsymbol{\phi}_p(\boldsymbol{u}_k) + \sum_{j=1}^{n_g} \boldsymbol{w}_j^T \nabla \boldsymbol{g}_{p,j}(\boldsymbol{u}_k) + \left(\boldsymbol{\mu}_j^T - \boldsymbol{w}_j^T \right) \nabla \boldsymbol{g}_j(\boldsymbol{y}_k(\boldsymbol{\theta}_k'), \boldsymbol{u}_k) \right| < \varepsilon \tag{3.14}$$

where convergence is achieved to the neighbourhood of the process optimum, which depends on the weighting ratio $\boldsymbol{w}_j$ and minimization tolerance ε. E.g. if $\boldsymbol{w}_j$ is such that $\boldsymbol{w}_j = \boldsymbol{\mu}_j$, then (3.14) simplifies to $\left| \nabla \boldsymbol{\phi}_p(\boldsymbol{u}_k) + \sum_{j=1}^{n_g} \boldsymbol{w}_j^T \nabla \boldsymbol{g}_{p,j}(\boldsymbol{u}_k) \right| < \varepsilon$. Thus, if the predicted optimum is also a KKT point of the actual process then we can set $\boldsymbol{u}^* = \boldsymbol{u}_k$ and $\boldsymbol{\theta}^* = \boldsymbol{\theta}_k^{'}$, so that the conditions of model adequacy (3.9) - (3.11) are approximately satisfied by the amount of gradient correction in (3.12). Condition (3.14) can be extended to the robust case by including also the variance in the condition. In the case that (3.14) does not hold for operating point $\boldsymbol{u}_k$ then the model-based

prediction of an optimum in (3.13) is not a KKT point of the actual process and the algorithm continues to iterate towards the process optimum.

3.3. Parameter Selection

In the presence of model-plant mismatch, it is necessary to update the model at new operating points reached during the batch-to-batch optimization procedure. This is mainly due to the fact that the model does not provide an accurate prediction over a wide range of operating conditions as the parameters have to compensate for structural mismatch and are affected by measurement noise. Although all parameters of the model could be updated simultaneously, this is very impractical for both computational reasons and because of the expected higher sensitivity to noise that may result from over-fitting noisy data. In addition, the measurements may not provide enough excitation in order to identify all parameters. Therefore, we propose to update only a sub-set of parameters that have large effects on the model outputs and the gradients of the cost function and constraints. To this end, a parametric sensitivity analysis is used to identify the model parameters that have most effect on these variables of interest.

3.3.1. Parametric Sensitivity for Outputs and Gradients

The scaled local sensitivity of a model-output $y_j \in \{y_1, \ldots, y_{n_y}\}$ with respect to a parameter $\theta_i \in \{\theta_1, \ldots, \theta_{n_\theta}\}$, at sample time $t_k \in \{t_1, \ldots, t_{n_t}\}$, at operating point $\boldsymbol{u}$ and nominal parameter values $\boldsymbol{\theta}$ is defined in the following as:

$$S_{y_j}^{\theta_i}(\boldsymbol{u}, \boldsymbol{\theta}, t_k) = \frac{\partial y_j}{\partial \theta_i}(\boldsymbol{u}, \boldsymbol{\theta}, t_k) \frac{\theta_i}{\overline{y}_j(\boldsymbol{u}, \boldsymbol{\theta})} \tag{3.15}$$

where the scaling is done with respect to the nominal parameter value θ_i and the average magnitude of the model output over all sampling times, which is given by:

$$\overline{y}_j\left(\boldsymbol{u},\boldsymbol{\theta}\right)=\frac{1}{n_t}\sum_{k=1}^{n_t}y_j\left(\boldsymbol{u},\boldsymbol{\theta},t_k\right) \tag{3.16}$$

Scaling can be also performed according to prior information on the permissible range in the parameter's magnitude. If no information is available, it is common to scale the parameters by their nominal value which corresponds to a 100 % possible change in their magnitude.

For a particular batch run, local sensitivities can be obtained for all outputs and parameters at a specific sampling time t_k, which results in the following output sensitivity matrix:

$$\boldsymbol{S}\left(\boldsymbol{u},\boldsymbol{\theta},t_k\right)=\begin{bmatrix} S_{y_1}^{\theta_1} & \cdots & S_{y_1}^{\theta_i} & \cdots & S_{y_1}^{\theta_{n_\theta}} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ S_{y_j}^{\theta_1} & \ddots & \ddots & \ddots & S_{y_j}^{\theta_{n_\theta}} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ S_{y_{n_y}}^{\theta_1} & \cdots & \cdots & \cdots & S_{y_{n_y}}^{\theta_{n_\theta}} \end{bmatrix} \tag{3.17}$$

The matrix in (3.17) can be computed for all sampling times during a batch to obtain:

$$\boldsymbol{S}\left(\boldsymbol{u},\boldsymbol{\theta}\right)=\begin{bmatrix} \boldsymbol{S}\left(\boldsymbol{u},\boldsymbol{\theta},t_1\right) \\ \vdots \\ \boldsymbol{S}\left(\boldsymbol{u},\boldsymbol{\theta},t_k\right) \\ \vdots \\ \boldsymbol{S}\left(\boldsymbol{u},\boldsymbol{\theta},t_{n_t}\right) \end{bmatrix} \tag{3.18}$$

Given the sensitivity matrix for all sampling times (3.18), it is then possible to obtain a lumped

metric of the overall effect of a parameter θ_i on all outputs as follows:

$$S_y^{\theta_i}(\boldsymbol{u},\boldsymbol{\theta}) = \sum_{j=1}^{n_y \times n_t} \left| \boldsymbol{S}(\boldsymbol{u},\boldsymbol{\theta})_{ji} \right| \tag{3.19}$$

Given the gradient of the cost function with respect to a particular decision variable by $\nabla\phi_l(\boldsymbol{u},\boldsymbol{\theta}) = \frac{\partial\phi}{\partial u_l}(\boldsymbol{u},\boldsymbol{\theta})$ with $l \in \{1,\ldots,n_u\}$. The scaled local sensitivity of the gradient of the cost-function with respect to parameter θ_i at operating point $\boldsymbol{u}$ and nominal parameter values $\boldsymbol{\theta}$ is given by:

$$S_{\nabla\phi_l}^{\theta_i}(\boldsymbol{u},\boldsymbol{\theta}) = \frac{\partial(\nabla\phi_l)}{\partial\theta_i}(\boldsymbol{u},\boldsymbol{\theta}) \frac{\theta_i}{\nabla\phi_l(\boldsymbol{u},\boldsymbol{\theta})} \tag{3.20}$$

where the scaling is with respect to the nominal gradient and parameter value. In order to obtain a lumped cost function gradient sensitivity measure, the sensitivities in (3.20) can be summed up to calculate a lumped sensitivity measure of cost gradients with respect to the parameter θ_i:

$$S_{\nabla\phi}^{\theta_i}(\boldsymbol{u},\boldsymbol{\theta}) = \sum_{l=1}^{n_u} \left| S_{\nabla\phi_l}^{\theta_i}(\boldsymbol{u},\boldsymbol{\theta}) \right| \tag{3.21}$$

Similarly, the local sensitivity of a constraint gradient with respect to a parameter θ_i is defined as:

$$S_{\nabla g_{kl}}^{\theta_i}(\boldsymbol{u},\boldsymbol{\theta}) = \frac{\partial(\nabla g_{kl})}{\partial\theta_i}(\boldsymbol{u},\boldsymbol{\theta}) \frac{\theta_i}{\nabla g_{kl}(\boldsymbol{u},\boldsymbol{\theta})} \tag{3.22}$$

With $k \in \{1,\ldots,n_g\}$. The lumped measure of the effect of parameter θ_i on the constraint gradients is then given by:

$$S_{\nabla g}^{\theta_i}(\boldsymbol{u},\boldsymbol{\theta}) = \sum_{k=1}^{n_g}\sum_{l=1}^{n_u} \left| S_{\nabla g_{kl}}^{\theta_i}(\boldsymbol{u},\boldsymbol{\theta}) \right| \tag{3.23}$$

3.3.2. Sampling of the Uncertain Parameter Space

The sensitivity measures described in the previous sections are local in the sense that they are calculated for a specific operating point $\boldsymbol{u}$ and a particular set of nominal parameter values $\boldsymbol{\theta}$. The

reason that a local sensitivity measure was chosen for this study as compared to a global measure (Saltelli et al., 2008) is that the identified model is required to be only locally correct since, due to the structural mismatch, it is successively updated in a batch-to-batch fashion. However, as the local sensitivities depend on a set of nominal parameter values, we must still account for parametric uncertainty, which results from model-plant mismatch and measurement noise. To this end, we calculate the sensitivity measures at different points from the uncertain parameter space based on the assumption of a multivariate parameter distribution. Given an available parameter co-variance matrix $\boldsymbol{V}_\theta$ and normally distributed measurement errors, a sample of the uncertain parameters can be calculated from the joint probability distribution:

$$\boldsymbol{\theta}_m \sim \mathcal{N}\left(\boldsymbol{\theta}^*, \boldsymbol{V}_\theta\right) \tag{3.24}$$

A lower bound of the parameter co-variance matrix $\boldsymbol{V}_\theta$ is provided by the inverse of the Fisher information matrix (Walter & Pronzato, 1990):

$$\boldsymbol{V}_\theta = \boldsymbol{F}^{-1} \tag{3.25}$$

where the Fisher information matrix $\boldsymbol{F}$ over all sample points is defined as follows:

$$\boldsymbol{F} = \sum_{k=1}^{n_t} \boldsymbol{S}\left(\boldsymbol{u}, \boldsymbol{\theta}, t_k\right)^T \boldsymbol{\Sigma}^{-1} \boldsymbol{S}\left(\boldsymbol{u}, \boldsymbol{\theta}, t_k\right) \tag{3.26}$$

where $\boldsymbol{S}\left(\boldsymbol{u}, \boldsymbol{\theta}, t_k\right)$ is the sensitivity matrix defined in (3.17) and $\boldsymbol{\Sigma}$ the co-variance matrix of the measurement noise, which is typically given as:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_{n_y}^2 \end{bmatrix} \tag{3.27}$$

As we are only interested in the mean of the sensitivities, a Latin hyper-cube sampling (LHS) (McKay et al., 1979) is used to obtain samples $\boldsymbol{\theta}_m$ from the multivariate distribution (3.24). LHS is preferred in this case over techniques like the Morris method (Morris, 1991) as we assumed an approximate distribution of the input parameters. Accordingly, if the number of samples is n_s, the mean of the sensitivities of the model output (3.19) are calculated as:

$$\bar{S}_y^{\theta_i} = \frac{1}{n_s} \sum_{m=1}^{n_s} S_y^{\theta_i} \left(\boldsymbol{u}, \boldsymbol{\theta}_m \right) \tag{3.28}$$

Similarly, the means of the sensitivity measures of the cost function (3.21) and constraint gradients (3.23) are given by:

$$\bar{S}_{\nabla\phi}^{\theta_i} = \frac{1}{n_s} \sum_{m=1}^{n_s} S_{\nabla\phi}^{\theta_i} \left(\boldsymbol{u}, \boldsymbol{\theta}_m \right) \tag{3.29}$$

$$\bar{S}_{\nabla g}^{\theta_i} = \frac{1}{n_s} \sum_{m=1}^{n_s} S_{\nabla g}^{\theta_i} \left(\boldsymbol{u}, \boldsymbol{\theta}_m \right) \tag{3.30}$$

Finally, in vector form, the averaged output and gradient sensitivity measures are given as follows:

$$\bar{\boldsymbol{S}}_y = \left[\bar{S}_y^{\theta_1}, \ldots, \bar{S}_y^{\theta_{n_\theta}} \right]^T \tag{3.31}$$

$$\bar{\boldsymbol{S}}_{\nabla\phi} = \left[\bar{S}_{\nabla\phi}^{\theta_1}, \ldots, \bar{S}_{\nabla\phi}^{\theta_{n_\theta}} \right]^T \tag{3.32}$$

$$\bar{\boldsymbol{S}}_{\nabla g} = \left[\bar{S}_{\nabla g}^{\theta_1}, \ldots, \bar{S}_{\nabla g}^{\theta_{n_\theta}} \right]^T \tag{3.33}$$

3.3.3. Parameter Ranking and Selection

As mentioned before, the goal of a parameter selection is to find the relevant parameters, which provide the largest effect on both model output and gradients. In order to obtain the overall effect of each parameter, the sensitivities in (3.31) – (3.33) can be combined into a single sensitivity vector. However, for an appropriate comparison of the different sensitivities it is first necessary to perform

the following normalization with respect to their average magnitude:

$$\hat{\boldsymbol{S}}_y = \frac{\bar{\boldsymbol{S}}_y}{\frac{1}{n_\theta}\sum_{i=1}^{n_\theta} \bar{S}_y^{\theta_i}} \tag{3.34}$$

$$\hat{\boldsymbol{S}}_{\nabla\phi} = \frac{\bar{\boldsymbol{S}}_{\nabla\phi}}{\frac{1}{n_\theta}\sum_{i=1}^{n_\theta} \bar{S}_{\nabla\phi}^{\theta_i}} \tag{3.35}$$

$$\hat{\boldsymbol{S}}_{\nabla g} = \frac{\bar{\boldsymbol{S}}_{\nabla g}}{\frac{1}{n_\theta}\sum_{i=1}^{n_\theta} \bar{S}_{\nabla g}^{\theta_i}} \tag{3.36}$$

Subsequently, an overall parametric sensitivity vector is defined as the sum of the normalized individual sensitivities as follows:

$$\hat{\boldsymbol{S}} = \hat{\boldsymbol{S}}_y + \hat{\boldsymbol{S}}_{\nabla\phi} + \hat{\boldsymbol{S}}_{\nabla g} \tag{3.37}$$

The resulting vector in (3.37) is used to provide a ranking of parameters in terms of their overall effect on model output and gradients of the cost function and constraints. However, correlation among model parameters has to be considered to correctly account for the effect of parameters on model outputs. As the correlation among the parameters' effects may affect their ranking, an orthogonalization method (Yao et al., 2003) is applied for this task as it impacts the output sensitivity vector $\hat{\boldsymbol{S}}_y$ appearing on the RHS of (3.37). The aforementioned orthogonalization method is applied to the sensitivity matrix in (3.18), which is iteratively adjusted to account for the effect of the selected parameters on the model output. Thus, during each iteration of the parameter selection procedure, we update the ranking vector $\hat{\boldsymbol{S}}^p$, where the superscript p indicates the number of selected parameters to be updated for each batch.

We initialize the selection procedure by setting $p = 0$ and follow the steps outlined below.

1. Using the ranking vector $\hat{\boldsymbol{S}}^p$, the most significant parameter at iteration p is obtained from:

$$i_{max} = \arg \max_{i \in \{1,\ldots,n_\theta\}} \hat{\boldsymbol{S}}_i^p \tag{3.38}$$

2. To account for correlation, the output sensitivity vectors related to the remaining parameters (i.e. that have not been yet selected) are projected onto the space normal to the selected sensitivity vector as follows (Yao et al., 2003):

$$\begin{aligned} s_i^{p+1} &= s_i^p - \frac{\left(s_i^p\right)^T s_{i_{max}}^p}{\left(s_{i_{max}}^p\right)^T s_{i_{max}}^p} s_{i_{max}}^p \quad \forall m \in \{1,\ldots,n_s\} \\ &\qquad \forall i \in \{1,\ldots,n_\theta\} \\ \text{s.t.} \quad & s_i^p = \boldsymbol{S}^p\left(\boldsymbol{u},\boldsymbol{\theta}_m\right)_{\cdot,i} \\ & s_{i_{max}}^p = \boldsymbol{S}^p\left(\boldsymbol{u},\boldsymbol{\theta}_m\right)_{\cdot,i_{max}} \end{aligned} \tag{3.39}$$

where s_i^p is a column of the sensitivity matrix (3.18) at iteration p and parametric sample vector $\boldsymbol{\theta}_m$. In this way, the "net influence" of the remaining parameters on the model output, which has not been covered by the already selected parameters can be assessed by the expression in (3.39).

3. A new output sensitivity vector $\hat{\boldsymbol{S}}^{p+1}$ is calculated from equations (3.19), (3.28) and (3.34). The sensitivity vectors of the cost function and constraint gradients are updated by eliminating the effect of the already selected parameters, i.e. by setting $\bar{S}_{\nabla\phi}^{\theta_{i_{max}}} = \bar{S}_{\nabla g}^{\theta_{i_{max}}} = 0$ to obtain $\hat{\boldsymbol{S}}_{\nabla\phi}^{p+1}$ and $\hat{\boldsymbol{S}}_{\nabla g}^{p+1}$ from (3.35) and (3.36) respectively. Finally, the updated ranking vector $\hat{\boldsymbol{S}}^{p+1}$ is obtained from (3.37).

4. Set the number of selected parameters to $p = p + 1$ and go back to step 1.

5. The selection procedure is continued until the remaining effect of a parameter on the model output, obtained after successive projections, is lower than a prescribed threshold value. The threshold is defined so that a minimum average effect of a parameter on a model output is related to the level of measurement noise as follows:

$$\max_{j\in\{1,\ldots,n_y\}} \left\{ \bar{S}_{y_j}^{\theta_{i_{max}}}\left(\boldsymbol{u},\boldsymbol{\theta}\right) \frac{\delta\theta_{i_{max}}}{\theta_{i_{max}}} \right\} \leq \varepsilon \tag{3.40}$$

where $\bar{S}_{y_j}^{\theta_{i_{max}}}(\boldsymbol{u},\boldsymbol{\theta}) = \frac{1}{n_t}\sum_{k=1}^{n_t}\left|S_{y_j}^{\theta_{i_{max}}}(\boldsymbol{u},\boldsymbol{\theta},t_k)\right|$ is the average effect of parameter $\theta_{i_{max}}$ on output y_j, ε is the level (standard deviation) of measurement noise and $\delta\theta_{i_{max}}$ is a user selected deviation in the parameter values, e.g. $\delta\theta_{i_{max}} = 0.1\theta_{i_{max}}$.

The complete proposed parameter selection procedure is illustrated in the flowchart in Figure 3.1 and is summarized as follows. After a batch run at operating point $\boldsymbol{u}_k$, the uncertain parameter space is sampled using LHS and the initial sensitivity vectors for the model output, cost function and constraint gradients are obtained. Then, the overall effects are assessed to select the most significant parameters according to the 5-step procedure outlined above. Subsequently, the model is updated based on the determined set of parameters by performing a parameter estimation (3.1) followed by the gradient correction (3.4). The final step is the computation of the next optimal batch run $\boldsymbol{u}_{k+1}$ by performing a model-based optimization according to (3.8), using the updated model.

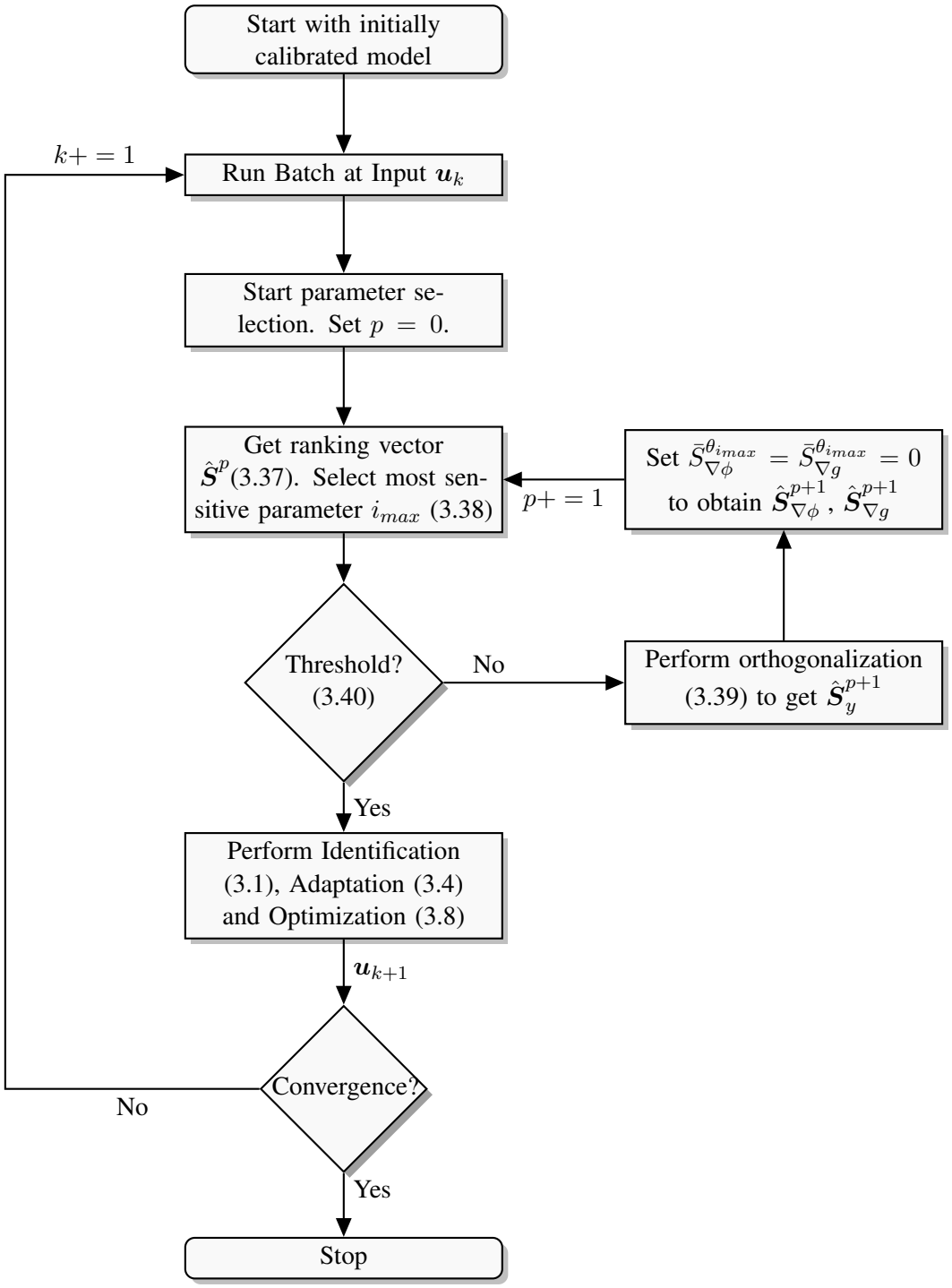

Figure 3.1.: Block diagram of the parameter selection procedure integrated into the simultaneous identification and optimization framework.

3.4. Uncertainty Propagation

To calculate the variance of the objective in (3.8), it is first necessary to characterize the uncertainty that is present in the model input. In general, the PCE method can be used for uncertainty propagation and quantification of parametric uncertainty or uncertainty in initial conditions (Nagy & Braatz, 2007; Mandur & Budman, 2014). In this work, we assume prior knowledge about the statistical properties of the initial batch conditions. A detailed background on PC expansions is given in (Spanos & Ghanem, 1991; Xiu, 2010) and a brief background is provided below for completeness.

Given a properly defined probability space $(\Omega, \mathcal{F}, P)$ and a set of independent random variables $\{\xi_i(\omega)\}_{i=1}^{\infty}$, PCE is used to express a random variable of interest X as a polynomial expansion of functions of standard random variables as follows:

$$X(\omega) = \sum_{k=0}^{N} \hat{x}_k \boldsymbol{\phi}_k(\boldsymbol{\xi}(\omega)) \tag{3.41}$$

where $\hat{x}_k$ are the PCE coefficients, $\boldsymbol{\phi}_k(\boldsymbol{\xi})$ the orthogonal basis functions of the standard random variables $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_n)$ and N is the total number of terms. The orthogonal basis functions have to be selected according to the choice of the statistical distribution of ξ. For example, Hermite polynomials are chosen as basis functions of normal distributions (Xiu & Karniadakis, 2002a). One key property of the PCE is that all basis functions are orthogonal to each other, which leads to the following definition of the inner product:

$$\left\langle \boldsymbol{\phi}_i, \boldsymbol{\phi}_j \right\rangle = \int \boldsymbol{\phi}_i(\boldsymbol{\xi})\, \boldsymbol{\phi}_j(\boldsymbol{\xi})\, W(\boldsymbol{\xi})\, d\boldsymbol{\xi} = \delta_{ij} \tag{3.42}$$

where δ_{ij} is the Kronecker delta and $W(\boldsymbol{\xi})$ a weighting function, which has the form of an n-dimensional Gaussian probability distribution. The orthogonality permits the evaluation of the PC

coefficients by a projection onto the corresponding basis function as follows:

$$\hat{x}_k = \frac{\langle X, \boldsymbol{\phi}_k \rangle}{\langle \boldsymbol{\phi}_k^2 \rangle} = \frac{\int X \boldsymbol{\phi}_k W(\boldsymbol{\xi})\, d\boldsymbol{\xi}}{\int \boldsymbol{\phi}_k^2 W(\boldsymbol{\xi})\, d\boldsymbol{\xi}} \tag{3.43}$$

To propagate the uncertainty in initial conditions onto the output, a mapping is first determined between the set of uncertain inputs X and the set of independent random variables $\boldsymbol{\xi}$ by constructing a PC expansion as shown in (3.41). Using the model equations, the output of interest Y can be expressed as a function of X and is therefore also a random variable:

$$Y = f(X) \tag{3.44}$$

Similar to the expression in (3.41), the output can also be expressed by a PC expansion as follows:

$$Y(\omega) = \sum_{k=0}^{N} \hat{y}_k \boldsymbol{\phi}_k(\boldsymbol{\xi}(\omega)) \tag{3.45}$$

So that for the PC coefficients of the output expansion we obtain:

$$\hat{y}_k = \frac{\langle Y, \boldsymbol{\phi}_k \rangle}{\langle \boldsymbol{\phi}_k^2 \rangle} = \frac{\int Y \boldsymbol{\phi}_k W(\boldsymbol{\xi})\, d\boldsymbol{\xi}}{\int \boldsymbol{\phi}_k^2 W(\boldsymbol{\xi})\, d\boldsymbol{\xi}} \tag{3.46}$$

By using a non-intrusive approach, for each value of $\boldsymbol{\xi}$ a value of the input can be generated and a value of the variable of interest Y can be calculated by using the relation in (3.44). In order to evaluate the integral in the numerator in (3.46), the model equations are solved at specific collocation points so that an approximation of the output distribution can be obtained. Finally, the mean and variance of the desired output Y that are required in the formulation in (3.8), are derived analytically from orthogonality as follows:

$$\mathbf{E}[Y] = \hat{y}_0 \tag{3.47}$$

$$\boldsymbol{Var}\left[Y\right] = \hat{y}_k^2 \left\langle \boldsymbol{\phi}_k^2 \right\rangle \tag{3.48}$$

3.5. Results and Discussion

The batch-to-batch optimization procedure with the proposed parameter selection and robustness analysis is illustrated in this section using a fed-batch penicillin process.

3.5.1. Case Study

A model describing a fed-batch penicillin process was proposed in (Birol et al., 2002). It was previously used as a case study in run-to-run optimization studies conducted in (Mandur & Budman, 2015b). The process is described by the following differential equations:

$$\frac{dX}{dt} = \left(\frac{\mu_X S X}{K_X X + S}\right) - \frac{X}{V}\frac{dV}{dt} \tag{3.49}$$

$$\frac{dP}{dt} = \left(\frac{\mu_P S X}{K_P + S + \frac{S^2}{K_I}}\right) - K_H P - \frac{P}{V}\frac{dV}{dt} \tag{3.50}$$

$$\begin{aligned}\frac{dS}{dt} = &-\left(\frac{1}{Y_{X/S}}\frac{\mu_X S X}{K_X X + S}\right) - \left(\frac{1}{Y_{P/S}}\frac{\mu_P S X}{K_P + S + \frac{S^2}{K_I}}\right) \\ &- m_X X + \frac{F s_f}{V} - \frac{S}{V}\frac{dV}{dt}\end{aligned} \tag{3.51}$$

$$\frac{dV}{dt} = F - V 6.226 \cdot 10^{-4} \tag{3.52}$$

where X is the biomass, P is the concentration of penicillin, S is the concentration of substrate and V the volume in the reactor. The constants are defined as follows: μ_X is the specific growth rate of biomass, μ_P is the specific rate of penicillin production, K_X and K_P are saturation constants, K_I is a substrate inhibition constant, K_H is a constant representing the rate of consumption of

penicillin by hydrolysis, $Y_{X/S}$ and $Y_{P/S}$ are the yields per unit mass of substrate for the biomass and penicillin respectively, m_X is the consumption rate of substrate for maintaining the biomass, F is the constant feed rate and s_f represents the concentration of substrate in the feed.

The process simulator is defined by the equations (3.49) – (3.52) and is used to generate *in silico* data of the process outputs and gradients. Additive Gaussian noise of $10\,\%$ of the average output values is assumed in the process outputs. Measurements of the gradients are estimated by running additional batch runs with a step change in the input in each direction of the decision variables and then taking the difference of the noisy measurement over the length of the step size. In order to reduce the amount of gradient uncertainty due to measurement noise we used a step size of $\Delta S_0 = 3\,\mathrm{g/l}$ for the non-robust case and a step size of $\Delta S_0 = 6\,\mathrm{g/l}$ in case of the presence of input disturbances. A step length of $\Delta F = 0.5\,\mathrm{l/h}$ has been used for flowrate gradient measurements.

In order to introduce model-plant mismatch, we introduce an intentional discrepancy between the model and the process. In this case, the hydrolysis term is assumed to be a priori unknown to the modeler and is therefore eliminated from the model used for optimization. This elimination reflects the lack of knowledge about the process and thus results in model structure error. Accordingly, the following rate of change of penicillin to be used in the model of the process instead of equation (3.50):

$$\frac{dP}{dt} = \left(\frac{\mu_P S X}{K_P + S + \frac{S^2}{K_I}} \right) - \frac{P}{V}\frac{dV}{dt} \tag{3.53}$$

In addition to model-plant mismatch, it is assumed that there are stochastic uncertainties in the initial concentration of biomass and substrate. For simplicity, we assume that we have prior knowledge that these uncertainties follow Gaussian distributions:

$$X_0 \sim \mathcal{N}\left(\bar{X}_0, \sigma_{X_0}\right) \tag{3.54}$$

$$S_0 \sim \mathcal{N}\left(\bar{S}_0, \sigma_{S_0}\right) \tag{3.55}$$

Biomass conc. $\left(\bar{X}_0\right)$	$0.1\,\mathrm{g/l}$
Biomass standard dev. (σ_{X_0})	$0.0007\,\mathrm{g/l}$
Substrate conc. (S_0)	$3\,\mathrm{g/l}$
Substrate standard dev. (σ_{S_0})	$1\,\mathrm{g/l}$
Product conc. (P_0)	$0\,\mathrm{g/l}$
Volume (V_0)	$100\,\mathrm{l}$
Input Feed (F)	$0.04\,\mathrm{l/h}$

Table 3.1.: Initial batch conditions.

The goal of the optimization is to maximize the amount of penicillin at the end of the batch while minimizing the variance due to uncertainty in the initial biomass and substrate concentration. The manipulated variables are the initial substrate concentration S_0 and the constant inlet feed rate F. The amount of feeding is constrained by an upper bound of $V_{max} = 120\,\mathrm{l}$ on the volume of the reactor. The robust optimization problem is stated as follows:

$$\begin{aligned} \min_{S_0,F} \quad & -\mathbf{E}\left[P\left(\boldsymbol{x}, X_0, S_0, F, \boldsymbol{\theta}, t_f\right)\right] + w\boldsymbol{Var}\left[P\left(\boldsymbol{x}, X_0, S_0, F, \boldsymbol{\theta}, t_f\right)\right] \\ \text{s.t.} \quad & (3.49) \text{ and } (3.51) - (3.53) \\ & V(\boldsymbol{x}, X_0, S_0, F, \boldsymbol{\theta}, t_f) \leq V_{max} \end{aligned} \tag{3.56}$$

The initial batch conditions for the simulator are given in Table 3.1. The values for the initial substrate concentration S_0 and the constant feed rate F are used for running the first batch and are updated according to the model-based prediction of the next optimal input. The model described by (3.49) and (3.51)-(3.53) contains eight potential parameters available for performing the model update and the gradient correction. The initial model parameter values are given in Table 3.2.

μ_X	K_X	μ_P	K_P	K_I	$Y_{X/s}$	$Y_{P/S}$	m_X
0.092	0.15	0.008	0.0002	0.1	0.45	0.9	0.014

Table 3.2.: Initial model parameter values.

3.5.2. Parameter Selection Results

In this section the effect of the proposed parameter selection procedure is investigated by comparing the proposed procedure to two cases:

1. An optimization where a fixed parameter subset based on a prior analysis is used.
2. An optimization where the parameters that are updated are chosen based on only output sensitivities.

For this case, the effect of the input uncertainty is omitted and a non-robust optimization is carried out by ignoring the variance term in (3.56). A separate analysis of the robust optimization case with respect to uncertainty in initial conditions is given in section 3.5.3.

3.5.2.1. Comparison to a Previous Subset

In an earlier application of the simultaneous identification and optimization algorithm (Mandur & Budman, 2014, 2015b), a fixed parameter sub-set containing K_I and K_X was chosen based on an a priori analysis for performing the parameter adaptation. In contrast, by using the proposed parameter selection scheme, it is found that the subset of parameters that need to be updated varies for different batches of the run-to-run optimization procedure. For instance, the subset of parameters that need to be adapted generally includes, depending on the iteration, different combinations of m_X, μ_P and μ_X. Based on the threshold defined in (3.40) , the ideal number of parameters that need to be adapted is three. However, for a fairer comparison to previous results where only two parameters were updated, the number of selected parameters in the current case study was limited to two.

In general, the upper bound on the truncation error ϵ_{max} in (3.4) should be selected so that the resulting model error is not larger than the level of measurement noise. In this way, the model fitting remains accurate enough to comply with constraints or other relevant prediction objectives.

Given the presence of $10\,\%$ measurement noise, the upper bound on the relative truncation error is therefore set to $\epsilon_{max} = 1\,\%$. Due to the constraint on the volume of the reactor, the optimal feed rate $F^* = 0.1728\,\mathrm{l/h}$ is achieved within one iteration. On the other hand, as a result of the structural mismatch, it takes up to 30 batch runs to obtain the optimal initial substrate concentration of $S_0^* = 54.72\,\mathrm{g/l}$. Figure 3.2 shows the results of the convergence of S_0 to the optimum for the proposed selection and the previously used fixed parameter sub-set. The optimization results are averaged over 10 different noise realizations and illustrated with 95% confidence intervals presented by the shaded areas.

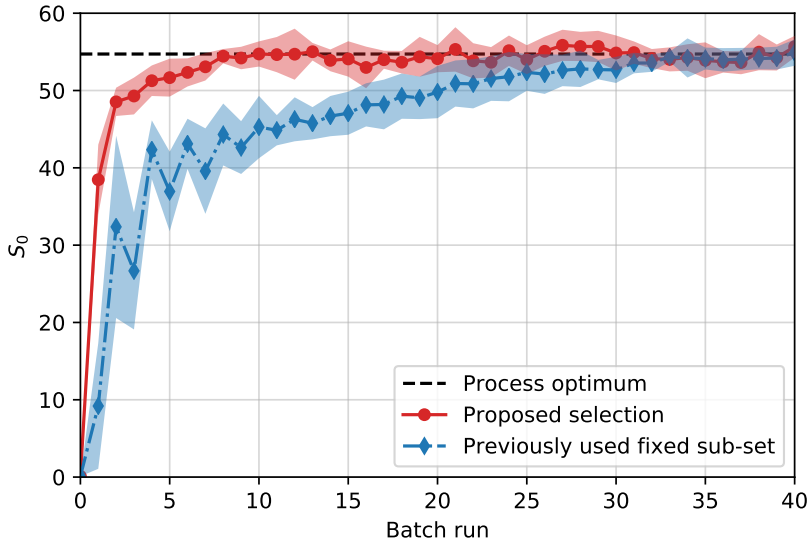

Figure 3.2.: Convergence of S_0 for proposed selection and previous parameter set for $\epsilon_{max} = 1\%$.

It is evident from Figure 3.2 that the use of a variable sub-set of model parameters as per the proposed selection leads to a very significant improvement in the speed of convergence to the process optimum. While the presented parameter selection is able to converge to the optimal initial substrate concentration on average within 8 iterations, more than 30 iterations are required for the fixed parameter set based on a prior analysis. This is mainly due to the fact that the selection of

appropriate parameters leads to a superior gradient correction and thus a better prediction of the objective function around the neighbourhood of the operating point at that iteration. Furthermore, as can be seen from the confidence intervals, the proposed selection leads to a reduction in the uncertainty in the prediction of the next optimal batch run.

3.5.2.2. Effect of Gradient Sensitivities

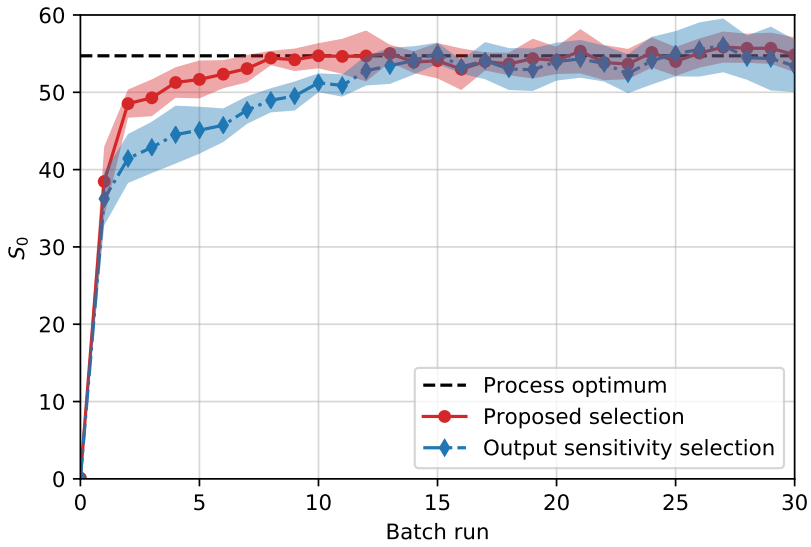

Figure 3.3.: Convergence of S_0 for proposed selection and output sensitivity selection for $\epsilon_{max} = 1\%$.

In this subsection we compare the proposed selection based on sensitivities of outputs, cost function and constraint gradients to a selection which is only based on output sensitivities. Figure 3.3 compares the results for the two selection methods for an upper bound on the truncation error of $\epsilon_{max} = 1\%$. As in the previous subsection, the optimization results are averaged over 10 different noise realizations and illustrated with 95% confidence intervals shown as shaded areas. By taking into account the parametric sensitivities of the gradients, it is possible to achieve a significantly faster convergence to the process optimum, especially during the initial iterations. The proposed

selection convergences on average within 8 iterations to the optimal substrate concentration, while it takes up to 14 iterations for a selection without gradient sensitivities. The difference in the selected parameters in each iteration is illustrated in figure 3.4, showing that the choice of the parameters' subset can change between iterations. Furthermore, the proposed parameter selection based on gradients and outputs also results in less uncertainty in the prediction of the next optimal batch runs as compared with the case that only output sensitivity is considered, which can be inferred from the smaller confidence intervals shown in Figure 3.3.

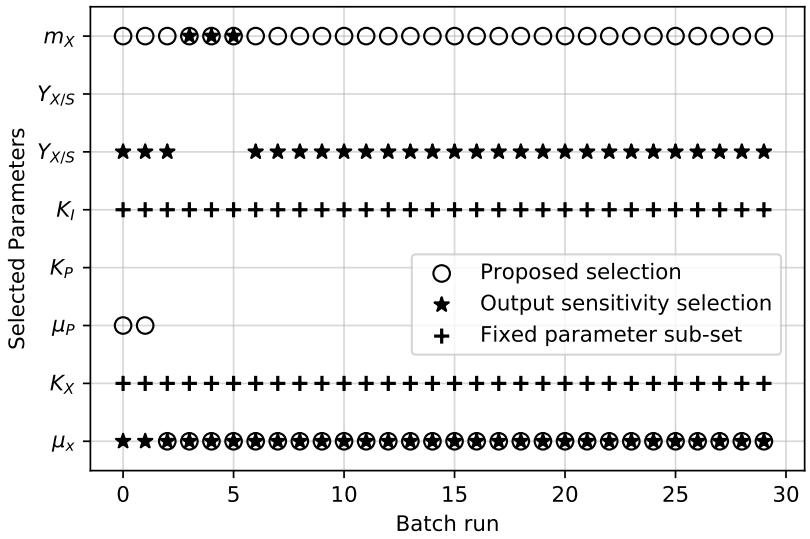

Figure 3.4.: Example of selected parameters in each iteration.

Finally, Table 3.3 compares the three approaches, i.e. fixed parameters' set, parameters selected based on output sensitivity and parameters selected based on the proposed output and gradient sensitivities. The comparison includes the integral absolute error (IAE), which is the error between the predicted and the actual process optimum, averaged over all iterations and simulations. The average gradient error is the relative error between the corrected and the measured cost function gradients, averaged over all iterations and simulations. It is evident that, for both criteria, the

proposed selection outperforms the parameter selection that is based only on output sensitivities and the fixed parameter sub-set used in previously reported studies. In conclusion, by updating model parameters based on sensitivities of both outputs and gradients, it is possible to improve speed and reduce uncertainty in the convergence of a parameter adaption scheme for the model-based optimization under model-plant mismatch.

Method	**IAE**	**Gradient Error**
Proposed output and gradient based sensitivity selection	3.6313	0.8273
Output based sensitivity selection	4.9796	1.6983
Fixed sub-set from previous works	9.5030	4.6350

Table 3.3.: Comparison of methods in terms of IAE and average gradient error.

3.5.3. Robust Optimization Results

In addition to structural model-plant mismatch, we now consider uncertainties in initial biomass and initial substrate concentrations. A robust optimization is carried out using the robust optimization methodology based on the PCE approach from section 3.4. The properties of the stochastic distributions of the uncertain initial conditions in biomass and substrate are given in Table 3.1. Performing a robust optimization as shown in (3.56) amounts to finding the operating point for which the amount of penicillin is maximized in the worst-case scenario as defined by the expectation minus the variance multiplied by w. In figure 3.5, the results are illustrated for different weights w on the variance of the cost. As in the previous subsection, the robust optimization results are averaged over 10 different noise realizations. We used the proposed parameter selection based on sensitivities of output and gradients as shown in the previous section for updating the model and correction of gradients.

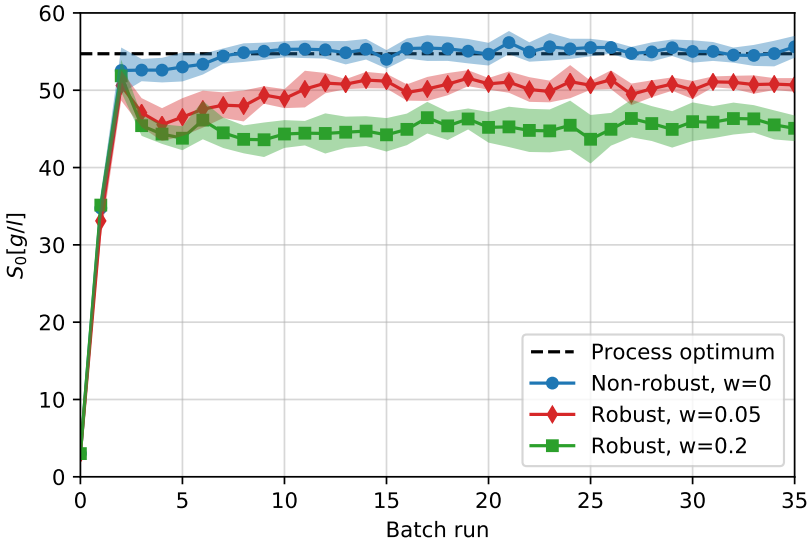

Figure 3.5.: Results of the robust optimization for different weights on the variance of the cost.

By increasing the weight on the variation due to input uncertainty, the final amount of substrate will be reduced. To explain this result, it is important to notice that the production of penicillin is governed by a substrate inhibition effect. Thus, the production of penicillin is significantly reduced if there is an excess of substrate in the reactor according to the expression in the denominator of the kinetic term in (3.50). Thus by increasing the robustness of the solution and lowering the amount of the initial amount of substrate, we avoid a possible excess of substrate that would result in inhibition in the production of penicillin. On the other hand, too small levels of initial substrate would result in smaller productivity according to the substrate dependency in the numerator of the kinetic expression in (3.50).

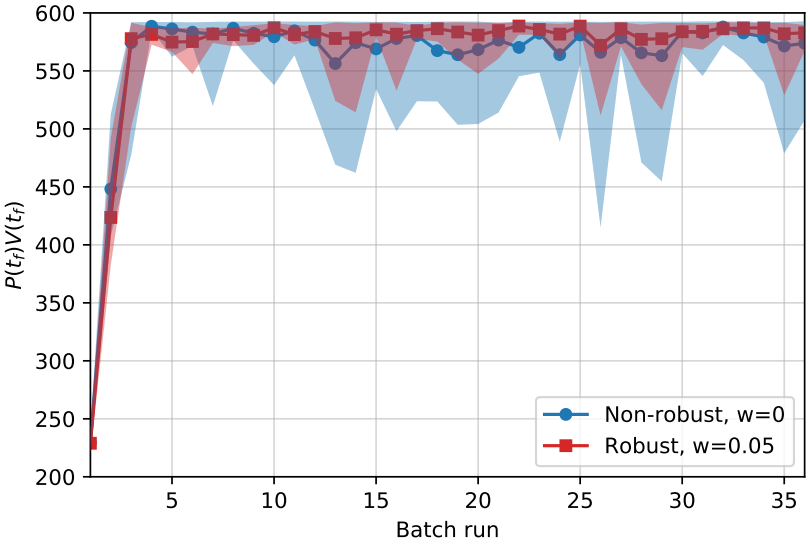

Figure 3.6.: Average amount and deviations (shaded area) of penicillin over 10 realizations of the input uncertainty at non-robust ($w = 0$) and robust ($w = 0.05$) operating points from figure 3.5.

This is confirmed, as shown in figure 3.6, by running the simulator for samples of the input distributions at the corresponding operating points obtained from the non-robust ($w = 0$) and robust ($w = 0.05$) case shown in figure 3.5. In order to assess the overall performance of the robust solution, $10,000$ MC simulations were performed at the final respective non-robust ($w = 0$) and robust ($w = 0.05$) solutions. This results in an average final amount of penicillin of $574.3\,\mathrm{g}$ for the non-robust case and $583.4\,\mathrm{g}$ for the robust solution. This is also illustrated in figure 3.7, where the non-robust output distribution exhibits a much broader tail leading to a lower mean value when compared to the robust solution. In conclusion, although the robust operating point leads to less penicillin during nominal conditions, it is much less affected by disturbances in inlet concentration, which results in superior performance for extreme cases and in a higher average over a large number of runs.

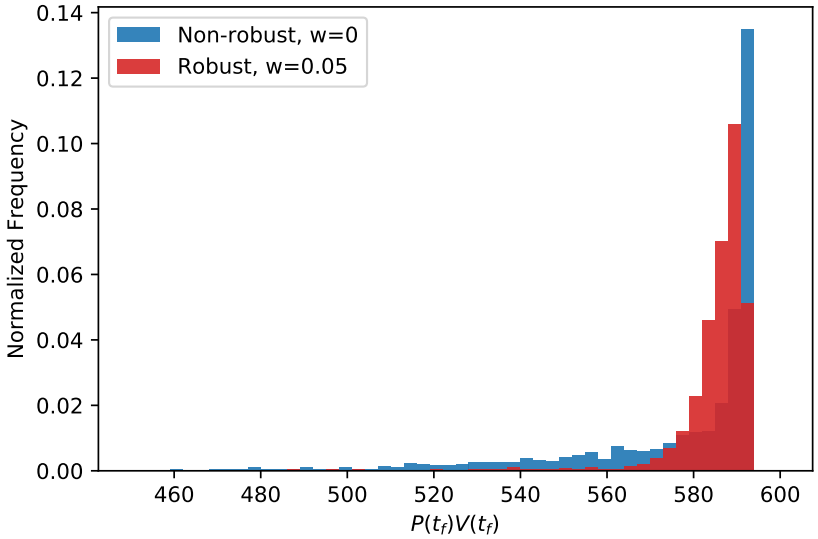

Figure 3.7.: Distributions of the amount of penicillin obtained from MC simulations for the non-robust $(w = 0)$ and robust $(w = 0.05)$ case.

3.6. Conclusions

This chapter presents a new methodology for selecting a subset of model parameters to be updated in each iteration of a batch-to-batch optimization in the presence of model-plant mismatch. The parameters are ranked according to their overall effect on the model output and the gradients of the cost function and constraints. An orthogonalization procedure is applied to account for correlation between the parameters' effect. It was shown that the selection of appropriate parameters leads to an improvement in the correction of the predicted gradients. This results in a significant speed up in convergence to the process optimum. It is demonstrated that it is very important to consider sensitivities of both outputs and gradients of cost and constraints to maximize improvements in convergence. Furthermore, robustness with respect to uncertainties in initial batch conditions was implemented by using a robust formulation of the objective. The propagation of uncertainty in

initial conditions onto the output was achieved by using a PCE based approach. A robust solution was obtained, which results in an improved performance, not only in the worst-case scenario, but also on average over a large number of batch runs. In conclusion, by using the proposed parameter selection approach in combination with the robust formulation of the objective, it is possible to obtain operating points, which are less susceptible to input uncertainty.

4. Simultaneous Identification and Optimization Using Output Uncertainty Bounds

Overview

The method of simultaneous identification and optimization aims at satisfying the conditions of optimality while providing accurate predictions of the process outputs. The model parameters are updated in a run-to-run procedure as to account for changes in operating points and to correct for errors in the predicted gradients of the cost-function and constraints. To make this parameter updating step more robust, we propose an objective function for parameter identification consisting of the ratio of the sum of squared errors to the parametric gradient sensitivity function. This results in an identified set of parameters which provide larger sensitivities for the subsequent gradient correction step thus leading to faster convergence to the optimum. Uncertainty bounds on the model outputs are utilized to enforce an adequate model fitting and are also used at each run to decide whether it is required to update the model parameters. This is especially valuable when identifying dynamic metabolic models with many parameters. The resulting improvements are

Adapted from Hille, R., and Budman, H. (2017). Simultaneous identification and optimization of biochemical processes under model-plant mismatch using output uncertainty bounds. *Computers & Chemical Engineering*, [submitted].

illustrated using two simulated cell culture processes.

4.1. Introduction

Mathematical models play a key role in the design and operation of biochemical processes. Compared to the largely continuous process operations in the chemical industries, biotechnological processes are typically still operated in batch or fed-batch modes (Croughan et al., 2015). The operation of batch processes is usually challenging due to nonlinear process behavior, an extensive cost and duration of performing experiments and the lack of a steady-state (Bonvin, 1998). Mathematical models consequently offer an opportunity to gain a better process understanding and to enhance product yields. In terms of their nature, one can generally distinguish between black-box (data-driven) (Box & Draper, 1987) and fundamental (knowledge-driven) models. The latter type, generally based on first principles balances, require an intricate understanding of the process and therefore involve an extensive development. Nonetheless, when compared to black-box models, they generally offer a larger domain of validity (Walter & Kieffer, 2007; Bonvin et al., 2016) and thus are particularly suited for searching for optima that may lie outside the data used for model calibration (Yip & Marlin, 2004).

When dealing with complex nonlinear models, the identification of the model parameters might be impaired due to lack of experimental data or occurrence of unmeasured disturbances. Moreover, due to simplifications and assumptions in the modelling step, discrepancies between the model and the process under study may arise. This so-called model-plant mismatch has the implication that the parameters' values that provide an adequate fit to the measured process outputs, do not necessarily result in an accurate prediction of the gradients of the cost-function and constraints, which are necessary for optimization (Srinivasan & Bonvin, 2002). This lack of synergy between identification and optimization objectives therefore presents a challenge when carrying out a model-based optimization. Hence, under such conditions, a standard two-step procedure (Chen & Joseph, 1987),

involving repeated identification and optimization steps, may result in a sub-optimal operating policy.

For a model to be useful in the context of optimization, it should have the capability to provide, at least in the neighbourhood of the optimum, an accurate prediction of the process optimality conditions (Forbes et al., 1994; Biegler et al., 1985). In the presence of model-plant mismatch, this implies that the model must be adjusted so as to achieve a correct prediction of the Karush-Kuhn-Tucker (KKT) conditions at the process optimum. Considering these challenges, several methods, such as *Modifier Adaptation* (MA) (Marchetti et al., 2009; Gao et al., 2016) have been developed to compensate for errors in the predicted optimality conditions. In MA, additional modifiers are added to the cost-function and constraints to update the predicted gradients based on newly available data. Instead of modifying the cost and constraint functions, the algorithm for the simultaneous identification and optimization (Mandur & Budman, 2015b) aims at reconciling these two objectives through an adaptation of the model parameters. The rationale for pursuing accurate model predictions is that the model is often needed for goals other than optimization such as testing what-if scenarios, computing set-point trajectories for closed loop operation and for designing soft sensors.

The algorithm for simultaneous identification and optimization used in this study involves 3 steps: I - model identification based on fitting of model outputs, II - parameter adaptation to correct for errors in predicted gradients of cost-function and constraints and III - optimization using the model with updated parameters. The overall objective of this chapter is to find a set of parameter values which provide an accurate prediction of the cost-function around the current operating point in the presence of model-plant mismatch. However, for complex nonlinear models, a repeated re-estimation of model parameters from uncertain measurements in each iteration can result in increased run-to-run variability in the parameter values. This in turn leads to increased uncertainty in the calculated optimal conditions for the next batch run. To mitigate this limitation, it is therefore essential to implement a parameter estimation procedure that is robust to model uncertainty.

Regarding a parameter identification from output measurements, set-based constraints have been applied to biochemical systems as a mechanism for model discrimination and invalidation (Walter & Kieffer, 2007; Rumschinski et al., 2010). For example, constraints on the model outputs, derived from experimental data, have been used for identifying feasible model parametrizations (Paulen et al., 2013) that can satisfy given output restrictions. Recently, set-based bounds have also been used for identifying dynamic metabolic flux models (Villegas et al., 2017). In this case, the constraints on the model outputs enable the use of a modified version of the standard model output fitting objective, frequently defined as the minimization of the sum of squared errors (SSE). Instead, Villegas et al. (2017) defined an objective function for the parameter estimation problem that weights the SSE objective by a parametric sensitivity measure. This results in improved estimates of model parameters for the case that the time varying trajectories are assumed to be uniformly distributed within the permissible boundaries.

In this study, we use a modified version of the approach from Villegas et al. (2017) for identifying a model to be used for optimization under model-plant mismatch. Since the model is required to meet the optimality conditions at the optimum, the model parameters are adapted to correct for errors in the predicted cost-function and constraint gradients. To facilitate the gradient correction step, we propose to modify the parameter estimation objective commonly used for identification, i.e. the SSE between model predictions and data, by dividing the latter by a lumped measure of the parametric sensitivities of the cost-function and constraint gradients. Combined with the set-based constraints, which enforce bounds on the model outputs, the approach provides higher robustness to measurement noise and unmeasured disturbances thus resulting in an improved performance in terms of convergence of a run-to-run optimization procedure. Furthermore, the uncertainty bounds on the process outputs are used to define a criterion that allows updating the parameters to re-fit the model outputs only if constraints are violated. By doing so, unnecessary re-estimation steps of the model parameters can be avoided. This is particularly useful for the identification of more complex biochemical models with a large number of parameters. To deal with large scale metabolic

models in batch-to-batch optimization, a sequential parameter estimation strategy is implemented that splits the estimation into problems with smaller number of parameters. In each instance of the estimation procedure, the use of uncertainty bounds on the model outputs guarantee a feasible prediction of metabolites' trajectories, ultimately resulting in an adequate prediction of outputs in each iteration of the run-to-run optimization. Moreover, using the uncertainty bounds, re-estimation is limited only to the metabolites which require updating. This is especially of importance, considering that an iterative optimization under model-plant mismatch leads to frequent updating of model parameters to compensate for the lack of fit between model predictions and data.

In chapter 3, it was shown that an appropriate parameter selection can speed-up convergence as highly sensitive parameters will improve the amount of gradient correction. In contrast to that, this chapter presents a methodology to obtain more suitable parameter values from the identification step itself. This approach thus represents and addition to the overall run-to-run optimization framework and is independent of the subset that has been selected a priori.

The chapter is organized as follows. Section 2 presents a brief review of the algorithm for the simultaneous identification and optimization. The proposed approach for parameter updating using set based constraints is presented in section 3 followed by the application of the proposed approach in section 4 using two cell culture case studies. Section 4 also outlines the sequential parameter estimation procedure when dealing with dynamic metabolic models containing many parameters. Conclusions are provided in section 5.

4.2. Simultaneous Identification and Optimization Methodology

The method for simultaneous identification and optimization (Mandur & Budman, 2015a) consists, in each iteration (batch run), of the following steps: I – parameter identification, II - gradient correction and III - model-based optimization to determine the input for the next batch run.

4.2.1. Parameter Identification

We consider the problem of identifying parameter values of an available model from experimental data. For this purpose, measurements of the process outputs are obtained by performing one or several batch experiments at a given operating point characterized by the input $\boldsymbol{u}_k \in \mathbb{R}^{n_u}$. A typical model-fitting objective is the minimization of the SSE between the obtained measurements and the model predictions (Bard, 1974). Such an SSE objective can be defined as follows:

$$\phi_{SSE}(\boldsymbol{\theta}) = \sum_{i=1}^{n_t} \left\| \boldsymbol{y}_{p,k}(t_i) - \boldsymbol{y}_k(\boldsymbol{\theta}, t_i) \right\|^2 \tag{4.1}$$

where $\boldsymbol{y}_{p,k} \in \mathbb{R}^{n_t \times n_y}$ are the acquired process output measurements and $\boldsymbol{y}_k \in \mathbb{R}^{n_t \times n_y}$ the outputs of the model at the specified sampling times t_i with $i \in \{1, \ldots, n_t\}$. Using the objective in (4.1), it is possible to fit the model outputs to the experimental data and thus determine an estimate of the model parameters by solving the following problem:

$$\begin{aligned} \boldsymbol{\theta}_k \quad = \quad & \arg\min_{\boldsymbol{\theta}} \phi_{SSE}(\boldsymbol{\theta}) \\ \text{s.t.} \quad & \dot{\boldsymbol{x}}_k = \boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{\theta}) \\ & \boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k) - \boldsymbol{c}_{k-1} \\ & \boldsymbol{\theta} \in [\boldsymbol{\theta}_{lb}, \boldsymbol{\theta}_{ub}] \end{aligned} \tag{4.2}$$

where $\boldsymbol{x}_k \in \mathbb{R}^{n_t \times n_x}$ are the model states and $\boldsymbol{\theta}_k \in \mathbb{R}^{n_\theta}$ a set of parameter values which provide an adequate fit according to the objective in (4.1). The bounds $\boldsymbol{\theta}_{lb}$ and $\boldsymbol{\theta}_{ub}$ present the permissible range for the model parameters. The correction term $\boldsymbol{c}_{k-1} \in \mathbb{R}^{n_t \times n_y}$ is further explained in the gradient correction section below and is required to maintain the fitting accuracy when performing a gradient correction.

As mentioned in the introduction, in the presence of model-plant mismatch, the parameters ob-

tained from fitting the model outputs may not provide a correct prediction of the gradients of the cost-function and constraints (Srinivasan & Bonvin, 2002). This error therefore can result in a mismatch between predicted and measured gradients:

$$\nabla \boldsymbol{\phi}(\boldsymbol{y}_k(\boldsymbol{\theta}), \boldsymbol{u}_k) \neq \nabla \boldsymbol{\phi}_p(\boldsymbol{u}_k) \tag{4.3}$$

and/or

$$\nabla \boldsymbol{g}_j(\boldsymbol{y}_k(\boldsymbol{\theta}), \boldsymbol{u}_k) \neq \nabla \boldsymbol{g}_{p,j}(\boldsymbol{u}_k) \tag{4.4}$$

where $\phi \in \mathbb{R}$ denotes the predicted cost-function and $\nabla \boldsymbol{\phi} \in \mathbb{R}^{n_u}$ its gradient vector at input $\boldsymbol{u}_k$. Given a set of predicted process constraints $\boldsymbol{g} \in \mathbb{R}^{n_g}$, their respective gradient is given by $\nabla \boldsymbol{g}_j \in \mathbb{R}^{n_u}$ with $j \in \{1, \ldots, n_g\}$. The subscript p denotes measured quantities.

4.2.2. Gradient Correction

To converge to the optimum in the presence of a structural model error, it is necessary to adjust the values $\boldsymbol{\theta}_k$, obtained in problem (4.2), such that the error between the respective gradients described by (4.3) and (4.4) can be minimized. This goal can be achieved by performing the following parameter adjustment:

$$\begin{aligned}
\Delta\boldsymbol{\theta}_k \quad = \quad & \arg\min_{\Delta\boldsymbol{\theta}} \Bigg(\boldsymbol{w}_\phi^T \left| \nabla \boldsymbol{\phi}_p(\boldsymbol{u}_k) - \nabla \boldsymbol{\phi}(\boldsymbol{y}_k(\boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}), \boldsymbol{u}_k) \right| \\
& + \sum_{j=1}^{n_g} \boldsymbol{w}_{g,j}^T \left| \nabla \boldsymbol{g}_{p,j}(\boldsymbol{u}_k) - \nabla \boldsymbol{g}_j(\boldsymbol{y}_k(\boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}), \boldsymbol{u}_k) \right| \Bigg) \\
\text{s.t.} \quad & \dot{\boldsymbol{x}}_k = \boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}) \\
& \boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k) - \boldsymbol{c}_k \\
& \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta} \in [\boldsymbol{\theta}_{lb}, \boldsymbol{\theta}_{ub}] \\
& \left\| \boldsymbol{\epsilon}_k^T \right\|_\infty \leq \epsilon_{max}
\end{aligned} \tag{4.5}$$

where the weights $\boldsymbol{w}_\phi \in \mathbb{R}^{n_u}$ and $\boldsymbol{w}_{g,j} \in \mathbb{R}^{n_u}$ are used to scale the respective cost-function and constraint gradients. As the parameter values are changed by the amount $\Delta\boldsymbol{\theta}_k$, it is essential to introduce the correction term $\boldsymbol{c}_k$ into the model output in order to maintain the fitting accuracy achieved in the previous identification step (4.2). The correction term is updated in each iteration and is calculated using a first order Taylor approximation:

$$\boldsymbol{c}_k(t_i) = \boldsymbol{c}_{k-1}(t_i) + \boldsymbol{D}y_k(\boldsymbol{\theta}_k, t_i)\Delta\boldsymbol{\theta}_k \tag{4.6}$$

where $\boldsymbol{D}y_k(\boldsymbol{\theta}_k, t_i) \in \mathbb{R}^{n_y \times n_\theta}$ is the Jacobian of the model at sampling time t_i. Moreover, to avoid overfitting of noisy gradient measurements, an upper bound on the relative truncation error ϵ_{max} is used to limit the amount of gradient correction. The truncation error is the relative error introduced into the model when using the approximation in (4.6) and is calculated as follows:

$$\boldsymbol{\epsilon}_k(t_i) = [\boldsymbol{y}_k(\boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}_k, t_i) - \boldsymbol{D}y_k(\boldsymbol{\theta}_k, t_i) - \boldsymbol{y}_k(\boldsymbol{\theta}_k, t_i)] \cdot [\mathrm{diag}\,(\boldsymbol{y}_k(\boldsymbol{\theta}_k, t_i))]^{-1} \tag{4.7}$$

The result of the gradient correction step is an updated set of parameter values given as follows:

$$\boldsymbol{\theta}'_k = \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}_k \tag{4.8}$$

Which, for the current operating point $\boldsymbol{u}_k$, provide a simultaneous fitting of model outputs, due to $\boldsymbol{c}_k$, as well as of the gradients of the cost-function and constraints. The next step is to perform an optimization based on the updated model, i.e. the model with the parameter values calculated from the gradient correction step, in order to determine the optimal input for the next batch run.

4.2.3. Model-based Optimization

The overall goal of the algorithm is to reach the process optimum. Using the locally corrected parameter values from (4.8), we can determine the optimal input for the next batch run during each

iteration of the run-to-run procedure as follows:

$$\begin{aligned} \boldsymbol{u}_{k+1} \quad &= \quad \arg\min_{\boldsymbol{u}} \phi(\boldsymbol{y}(\boldsymbol{\theta}'_k), \boldsymbol{u}) \\ &\text{s.t.} \quad \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\theta}'_k) \\ &\qquad \boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{c}_k \\ &\qquad \boldsymbol{g}(\boldsymbol{y}(\boldsymbol{\theta}'_k), \boldsymbol{u}) \leq \boldsymbol{0} \\ &\qquad \boldsymbol{u}^L \leq \boldsymbol{u} \leq \boldsymbol{u}^U \end{aligned} \tag{4.9}$$

where the bounds $\boldsymbol{u}^L$ and $\boldsymbol{u}^U$ present the permissible search range for the optimal input. The following assumptions are necessary for convergence of the outlined procedure to the process optimum when considering the presence of model-plant mismatch.

Assumption 3.1: For the convergence of a model-based optimization scheme to the process optimum $\boldsymbol{u}^*$, a necessary condition is that, for a set of model parameters $\boldsymbol{\theta}^*$, the model adequacy conditions (Forbes et al., 1994) can be satisfied:

$$\begin{aligned} &\exists \boldsymbol{\theta}^* \in [\boldsymbol{\theta}_{lb}, \boldsymbol{\theta}_{ub}], \boldsymbol{u}^* : \\ &\nabla \boldsymbol{\phi}(\boldsymbol{y}(\boldsymbol{\theta}^*), \boldsymbol{u}^*) = \nabla \boldsymbol{\phi}_p(\boldsymbol{u}^*) \\ &\nabla \boldsymbol{g}(\boldsymbol{y}(\boldsymbol{\theta}^*), \boldsymbol{u}^*) = \nabla \boldsymbol{g}_p(\boldsymbol{u}^*) \\ &\nabla^2 \boldsymbol{\phi}(\boldsymbol{y}(\boldsymbol{\theta}^*), \boldsymbol{u}^*) > 0 \end{aligned} \tag{4.10}$$

Furthermore, in addition to the satisfaction of the optimality conditions at the process optimum in (4.10), to converge to a neighbourhood of the optimum it is also necessary to achieve a sufficient gradient correction in each iteration. This step requires the following assumption.

Assumption 3.2: For each operating point $\boldsymbol{u}_k$ considered during the run-to-run optimization, a set of corresponding model parameters $\boldsymbol{\theta}'_k$ and an upper bound on the relative truncation ϵ_{max} provides

sufficient gradient correction:

$$\exists \boldsymbol{\theta}_k^{'} \in [\boldsymbol{\theta}_{lb}, \boldsymbol{\theta}_{ub}], \epsilon_{max} :$$
$$\left| \left[\nabla \boldsymbol{\phi}_p(\boldsymbol{u}_k) - \nabla \boldsymbol{\phi}(\boldsymbol{y}_k(\boldsymbol{\theta}_k'), \boldsymbol{u}_k) \right] + \sum_{j=1}^{n_g} \boldsymbol{w}_j^T \left[\nabla \boldsymbol{g}_{p,j}(\boldsymbol{u}_k) - \nabla \boldsymbol{g}_j(\boldsymbol{y}_k(\boldsymbol{\theta}_k'), \boldsymbol{u}_k) \right] \right| < \varepsilon \qquad (4.11)$$

With the normalizing weight given by $\boldsymbol{w}_j = [\mathrm{diag}\,(\boldsymbol{w}_\phi)]^{-1} \boldsymbol{w}_{g,j}$. The choice of ε will determine the region of convergence around the true optimum.

4.2.4. Model Parameter Selection

When dealing with the identification of nonlinear models, not all model-parameters need or can be identified. This is mainly due to correlation of the parameters' effect on the model output, lack of excitation and sensitivity to measurement noise. Hence, it is often necessary to select a subset of parameters which can be identified and that can provide a significant effect on the model output (McLean & McAuley, 2012). In case of run-to-run optimization under model-plant mismatch, it is also important to consider sensitive parameters which provide a large effect on the gradients of cost-function and constraints as to satisfy the conditions in (4.10) and (4.11). To conform to these demands on the parameters, an appropriate subset of model parameters can be selected based on the procedure previously outlined in Hille et al. (2017).

4.3. Proposed Methodology

In the run-to-run optimization framework, outlined in section 4.2, the model is updated at each iteration by re-estimating the parameters in order to re-fit model outputs to process measurements. This is generally necessary as the model parameters have to compensate for the structural model error. However, measurement noise and disturbances entering the process lead to a considerable

degree of uncertainty in the parameter values. The result is an increased batch-to-batch variability in the parameter values which affects the performance of the subsequent gradient correction step. This in turn can lead to a slower or oscillatory convergence to the process optimum. For that reason, we propose the use of uncertainty bounds on the process outputs, which enable the formulation of a model-fitting criterion that involves the parametric gradient sensitivities. We are particularly interested in using a measure of the fitting based on a combination of the SSE and parametric gradient sensitivities, where the latter are specifically relevant for optimization. Furthermore, the availability of uncertainty bounds on the outputs also allows for the definition of a model-update criterion. This way, model outputs only have to be updated when it is strictly necessary. This is especially useful for the repeated parameter estimation of large metabolic models of biochemical processes as illustrated in the case study.

4.3.1. Output Uncertainty Bounds

Experimental data from biochemical processes frequently exhibit a lack of repeatability due to sensor error, unmeasured disturbances and a high sensitivity to environmental conditions. Set-based constraints have been introduced to provide bounds on the process outputs and subsequently determine a possible model parametrization that can satisfy the uncertainty bounds (Rumschinski et al., 2012; Paulen et al., 2013). In addition, such bounds on the process outputs provide the possibility of modifying a standard model-fitting objective (4.1), i.e. the SSE between model predictions and data, so as to account for a parametric sensitivity objective (Villegas et al., 2017).

In this work, we assume that a set of process output measurements $\boldsymbol{y}_{p,k}(t_i)$ can be obtained at an operating point given by the decision variables' vector $\boldsymbol{u}_k$. In addition, due to measurement noise and input uncertainties, we assume the error of the output measurements to be bounded as follows:

$$\boldsymbol{\mathcal{E}}_k = \left[\boldsymbol{e}_k^L, \boldsymbol{e}_k^U\right] \tag{4.12}$$

For the operating point $\boldsymbol{u}_k$ and all sampling times along the batch, a permissible range of process outputs can be defined using uncertainty bounds as follows (Paulen et al., 2013):

$$\boldsymbol{\mathcal{Y}}_k = \left\{ \boldsymbol{\mathcal{Y}}_k^i = \boldsymbol{y}_{p,k}(t_i) + \left[\boldsymbol{e}_k^L, \boldsymbol{e}_k^U \right] | i \in \{1, \ldots, n_t\} \right\} \tag{4.13}$$

where the error vectors $\boldsymbol{e}_k \in \boldsymbol{\mathcal{E}}_k$ describe worst-case uncertainty bounds on the process outputs and t_i the sampling times along the batch. Accordingly, the process output at each time interval $i \in \{1, \ldots, n_t\}$ can be bounded by and upper and lower bound (Streif et al., 2016):

$$\underline{\boldsymbol{y}}_k^i \leq \boldsymbol{y}_k^i \leq \overline{\boldsymbol{y}}_k^i \tag{4.14}$$

In a similar way, we typically assume a permissible range for the model parameter values:

$$\boldsymbol{\Theta}_0 = [\boldsymbol{\theta}_{lb}, \boldsymbol{\theta}_{ub}] \tag{4.15}$$

Assumption 3.3: Note that, although we consider the case of structural mismatch between the model used for optimization and a simulator describing the actual process in this work, we assume that the discrepancy and the error bounds on the outputs are such that for every operating point $\boldsymbol{u}_k$ we can state:

$$\begin{aligned} &\exists \boldsymbol{\theta}_k \in \boldsymbol{\Theta}_0, \boldsymbol{e}_k \in \boldsymbol{\mathcal{E}}_k : \\ &\dot{\boldsymbol{x}}_k = \boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{\theta}_k) \\ &\boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k) - \boldsymbol{c}_{k-1} \\ &\boldsymbol{y}_k \in \boldsymbol{\mathcal{Y}}_k \end{aligned} \tag{4.16}$$

Thus, we do not consider the problem of model invalidation (Prajna, 2006; Rumschinski et al., 2012), but rather we take into account a bounded error for which there exist a feasible model

parametrization where the model outputs can satisfy the available worst-case bounds on the process outputs.

4.3.2. Modified Parameter Estimation Problem

Compared to a parameter estimation that is minimizing the SSE according to (4.1), the general goal of a set-based (Rumschinski et al., 2010) or guaranteed parameter estimation (Paulen et al., 2013) can be expressed as follows (Mukkula & Paulen, 2017):

$$\begin{aligned} \boldsymbol{\Theta}_k \quad &= \quad \arg \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0}{\text{find}} \ \text{all}\, \boldsymbol{\theta} \\ &\text{s.t.} \quad \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}_k, \boldsymbol{\theta}) \\ &\qquad\ \ \boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}) \\ &\qquad\ \ \boldsymbol{y} \in \boldsymbol{\mathcal{Y}}_k \end{aligned} \tag{4.17}$$

Here, the goal is to determine the complete set of model parametrizations which are able to satisfy the given measured outputs and error bounds. However, in the context of simultaneous identification and optimization, the primary objective is to reconcile the objectives of identification with those of optimization. We are thus mainly interested in a set of parameter values, which can simultaneously satisfy the uncertainty bounds and at the same time, allow for an improved gradient correction. It should be noticed that, for the purpose of enhancing the gradient correction step in (4.5), we require parameters which provide large parametric gradient sensitivities since for large parametric gradient sensitivities, given by $\partial \nabla \boldsymbol{\phi} / \partial \boldsymbol{\theta}$ and $\partial \nabla \boldsymbol{g} / \partial \boldsymbol{\theta}$, it is possible to achieve a significant gradient correction for smaller changes of $\Delta \boldsymbol{\theta}$ in the model parameters. Achieving gradient matching with smaller changes of the parameter values with respect to the values obtained in the estimation step results in smoother convergence of the run-to-run procedure. One possibility for enhancing the gradient correction is to select a subset of highly sensitive model parameters, which

has been addressed in Hille et al. (2017). The second option, pursued in the present work, is to find values for a particular subset of parameters which lead to higher gradient sensitivities when fitting model outputs, while satisfying the set-based uncertainty bounds.

Towards the formulation of the proposed parameter estimation problem, we first define a scaled local sensitivity of a cost-function gradient $\nabla\phi_l$ with respect to a model parameter θ_i as follows:

$$S^{\theta_i}_{\nabla\phi_l}(\boldsymbol{u},\boldsymbol{\theta}) = \frac{\partial\left(\nabla\phi_l(\boldsymbol{u},\boldsymbol{\theta})\right)}{\partial\theta_i}\left|\frac{\theta_i}{\nabla\phi_l(\boldsymbol{u},\boldsymbol{\theta})}\right| \tag{4.18}$$

With $i \in \{1,\ldots,n_\theta\}$ and $l \in \{1,\ldots,n_u\}$. The scaling is performed with respect to the magnitude in both parameters and gradient. For all gradients and parameters, the sensitivity in (4.18) can be calculated and arranged into a matrix as follows:

$$\boldsymbol{S}_{\nabla\phi} = \begin{bmatrix} S^{\theta_1}_{\nabla\phi_1} & \cdots & S^{\theta_i}_{\nabla\phi_1} & \cdots & S^{\theta_{n_\theta}}_{\nabla\phi_1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ S^{\theta_1}_{\nabla\phi_l} & \ddots & \ddots & \ddots & S^{\theta_{n_\theta}}_{\nabla\phi_l} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ S^{\theta_1}_{\nabla\phi_{n_u}} & \cdots & S^{\theta_i}_{\nabla\phi_{n_u}} & \cdots & S^{\theta_{n_\theta}}_{\nabla\phi_{n_u}} \end{bmatrix} \tag{4.19}$$

In a similar way, the parametric sensitivity of a constraint gradient ∇g_{jl} with respect to model parameter θ_i is given as:

$$S^{\theta_i}_{\nabla g_{jl}}(\boldsymbol{u},\boldsymbol{\theta}) = \frac{\partial\left(\nabla g_{jl}(\boldsymbol{u},\boldsymbol{\theta})\right)}{\partial\theta_i}\left|\frac{\theta_i}{\nabla g_{jl}(\boldsymbol{u},\boldsymbol{\theta})}\right| \tag{4.20}$$

With $j \in \{1,\ldots,n_g\}$. In the same way as (4.19), the sensitivity matrix for constraint j is given by

$\boldsymbol{S}_{\nabla g_j}$. For all constraints, the sensitivity matrix is then described by:

$$\boldsymbol{S}_{\nabla g} = \begin{bmatrix} \boldsymbol{S}_{\nabla g_1} \\ \vdots \\ \boldsymbol{S}_{\nabla g_j} \\ \vdots \\ \boldsymbol{S}_{\nabla g_{n_g}} \end{bmatrix} \tag{4.21}$$

Following the definitions of the sensitivity matrices in (4.19) and (4.21), we define an overall parametric gradient sensitivity measure by summing up over of all elements of the sensitivity matrices:

$$\phi_{S^\nabla}(\boldsymbol{u}, \boldsymbol{\theta}) = \sum_{i=1}^{n_u} \sum_{j=1}^{n_\theta} |s_{\nabla\phi,ij}(\boldsymbol{u}, \boldsymbol{\theta})| + \sum_{i=1}^{n_u \times n_g} \sum_{j=1}^{n_\theta} |s_{\nabla g,ij}(\boldsymbol{u}, \boldsymbol{\theta})| \tag{4.22}$$

Finally, using the scalar measure describing the sensitivity function (4.22), the proposed parameter estimation problem can be stated as follows:

$$\begin{aligned} \boldsymbol{\theta}_k \quad = \quad & \arg\min_{\boldsymbol{\theta}} \left(\frac{\phi_{SSE}(\boldsymbol{u}, \boldsymbol{\theta})}{\phi_{S^\nabla}(\boldsymbol{u}, \boldsymbol{\theta})} \right) \\ \text{s.t.} \quad & \dot{\boldsymbol{x}}_k = \boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{\theta}) \\ & \boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k) - \boldsymbol{c}_{k-1} \\ & \boldsymbol{\theta} \in \boldsymbol{\Theta}_0 \\ & \boldsymbol{y}_k \in \boldsymbol{\mathcal{Y}}_k \end{aligned} \tag{4.23}$$

where the standard SSE objective is now divided by the parametric gradient sensitivity function. In this manner, we seek parameter values which result in an adequate model-fitting by minimizing the SSE while maximizing the sensitivities of the cost-function and constraint gradients. By using this modified objective function, smaller parameter corrections $\Delta\boldsymbol{\theta}$ will be required in the subsequent gradient correction step (4.5), allowing for an improved reconciliation of the two objectives of

identification and gradient correction. Since the optimization cost was modified from a norm of the prediction errors as in (4.1) to a norm of the errors divided by the sensitivity function as in (4.23), the uncertainty bounds (set-based bounds) help to ensure that the predicted outputs remain relatively close to the measured process outputs.

4.3.3. Model-Update Criterion

In addition to the proposed identification problem (4.23), the use of the uncertainty bounds on the process outputs also allow for the definition of a model-update criterion. Since the method pursues simultaneous identification and optimization there may be conflictive demands from the parameters to satisfy each one of these two objectives. Thus, the rationale behind the use of a model update criterion is that the model-fitting should only be performed when it is absolutely necessary. This would save computational time and avoid changes in parameter values due to measurement noise. As the gradient correction step is still maintained, the model update criterion does not affect the update of the predicted gradients of the cost-function and constraints. Accordingly, we define the model-update criterion as follows:

$$\begin{aligned} & \boldsymbol{y}'_k \in \boldsymbol{\mathcal{Y}}_k \\ \text{s.t.} \quad & \dot{\boldsymbol{x}}_k = \boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{\theta}'_{k-1}) \\ & \boldsymbol{y}'_k = \boldsymbol{h}(\boldsymbol{x}_k) - \boldsymbol{c}_{k-1} \\ & \boldsymbol{\theta}'_{k-1} \in \boldsymbol{\Theta}_0 \end{aligned} \tag{4.24}$$

If condition (4.24) holds, we consider the identification objective as satisfied and thus simply update the parameter values as follows:

$$\boldsymbol{\theta}_k = \boldsymbol{\theta}'_{k-1} \tag{4.25}$$

The model-update criterion combined with the above identification approach ensures that the identification objective, as given by the set-based constraints, is satisfied and allows for an improved prediction of the next optimal input due to a superior gradient correction.

4.3.4. Summary of the Proposed Model Update Methodology

The proposed extension to the simultaneous identification and optimization method can be summarized as per the steps described in Algorithm 4.1. The required assumptions for convergence to the process optimum, given in (4.10) and (4.11), remain unchanged for the proposed method.

Algorithm 4.1 Summary of the proposed approach

1: Initialize algorithm with $k = 1$, $\boldsymbol{\theta}'_0 = \boldsymbol{\theta}_0$ and $\boldsymbol{c}_0 = \mathbf{0}$
2: Perform several experiments at input $\boldsymbol{u}_k$ to acquire process uncertainty bounds $\boldsymbol{\mathcal{Y}}_k$
3: **if** $\boldsymbol{y}'_k \in \boldsymbol{\mathcal{Y}}_k$ **then**
4: Model-update criterion (4.25): $\boldsymbol{\theta}_k = \boldsymbol{\theta}'_{k-1}$
5: **else**
6: Multi-objective identification: Solve (4.23) to find $\boldsymbol{\theta}_k$ such that $\boldsymbol{y}_k \in \boldsymbol{\mathcal{Y}}_k$
7: **end if**
8: Gradient correction: Solve (4.5) to obtain $\boldsymbol{\theta}'_k = \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}_k$
9: Model-based optimization: Solve (4.9) to find $\boldsymbol{u}_{k+1}$
10: Implement new optimal input and go back to step 2: $k+ = 1$

4.4. Results and Discussion

This section presents the results and a discussion of the proposed identification approach including the model update criterion. Two cell-culture case studies are considered for illustration. The first case study deals with a penicillin process model and the second case study presents a relatively more complex dynamic metabolic model of a Chinese Hamster Ovary (CHO) cell line. To obtain feasible model parametrizations and avoid needless re-estimation steps for the latter case study, a sequential parameter estimation procedure is implemented that makes use of the already introduced output uncertainty bounds.

4.4.1. Penicillin Process

A model of a penicillin fed-batch process has been proposed in (Birol et al., 2002). It is used as a case study to perform a run-to-run optimization. The process can be described by the following set of ODEs:

$$\frac{dX}{dt} = \left(\frac{\mu_X S X}{K_X X + S}\right) - \frac{X}{V}\frac{dV}{dt} \tag{4.26}$$

$$\frac{dP}{dt} = \left(\frac{\mu_P S X}{K_P + S + \frac{S^2}{K_I}}\right) - K_H P - \frac{P}{V}\frac{dV}{dt} \tag{4.27}$$

$$\begin{aligned}\frac{dS}{dt} = &-\left(\frac{1}{Y_{X/S}}\frac{\mu_X S X}{K_X X + S}\right) - \left(\frac{1}{Y_{P/S}}\frac{\mu_P S X}{K_P + S + \frac{S^2}{K_I}}\right) \\ &-m_X X + \frac{F s_f}{V} - \frac{S}{V}\frac{dV}{dt}\end{aligned} \tag{4.28}$$

$$\frac{dV}{dt} = F - V 6.226 \cdot 10^{-4} \tag{4.29}$$

where the biomass is given by X, the penicillin concentration by P, the substrate concentration by S and volume by V. The model parameters are as follows: μ_X is the specific growth rate of biomass, μ_P is the specific rate of penicillin production, K_X and K_P are saturation constants, K_I is a substrate inhibition constant, K_H is a constant representing the rate of consumption of penicillin by hydrolysis, $Y_{X/S}$ and $Y_{P/S}$ are the yields per unit mass of substrate for the biomass and penicillin respectively, m_X is the consumption rate of substrate for maintaining the biomass, F is the constant feed rate and s_f represents the concentration of substrate in the feed.

The equations (4.26) – (4.29) describe the process simulator and are used to generate *in silico* data for use as process output and gradient measurements by the proposed run-to-run optimization procedure. Gaussian noise with a standard deviation corresponding to $10\,\%$ of magnitude of the process outputs is added. In addition, we assume stochastic disturbances in the initial batch

concentrations of biomass and substrate, both following normal distributions. Information on the standard deviations of disturbances and on the other initial batch conditions can be found in table 4.1.

Biomass conc. (X_0)	$0.1\,\mathrm{g/l}$
Biomass standard dev. (σ_{X_0})	$0.0006\,\mathrm{g/l}$
Substrate conc. (S_0)	$1\,\mathrm{g/l}$
Substrate standard dev. (σ_{S_0})	$0.4\,\mathrm{g/l}$
Product conc. (P_0)	$0\,\mathrm{g/l}$
Volume (V_0)	$100\,\mathrm{l}$
Input Feed (F)	$0.04\,\mathrm{l/h}$

Table 4.1.: Initial batch conditions for the penicillin process.

4.4.1.1. Model-Plant Mismatch

Model-plant mismatch is deliberately introduced by eliminating the hydrolysis term in the penicillin equation in (4.27). Assuming that the user is not aware of the existence of the penicillin hydrolysis phenomena, the penicillin equation assumed in the model is given by:

$$\frac{dP}{dt} = \left(\frac{\mu_P S X}{K_P + S + \frac{S^2}{K_I}} \right) - \frac{P}{V}\frac{dV}{dt} \tag{4.30}$$

Accordingly, the model to be calibrated and utilized for optimization of the penicillin process is given by equations (4.26) and (4.28) - (4.30). The goal of this run-to-run optimization is to run repeated batch experiments as to maximize the amount of penicillin at the final batch time t_f. To accomplish this task, the available manipulated variables are the initial substrate concentration S_0 and constant feed rate F. Thus, the optimization problem to be solved in each iteration of the

run-to-run optimization is stated as:

$$\begin{aligned} \min_{S_0,F} \quad & -P(\boldsymbol{x}, S_0, F, \boldsymbol{\theta}, t_f) \\ \text{s.t.} \quad & (4.26) \text{ and } (4.28) - (4.30) \\ & V(\boldsymbol{x}, S_0, F, \boldsymbol{\theta}, t_f) \leq V_{max} \end{aligned} \tag{4.31}$$

where a constraint on the volume of the reactor is given by $V_{max} = 120\,\mathrm{l}$. We assume that four batch experiments are performed at each operating point to obtain output measurements and the process uncertainty bounds. Furthermore, a finite difference approach is used to obtain the gradients of the cost at each operating point. In order to reduce the effect of noise and input disturbances on the gradient estimate, step sizes of $\Delta S_0 = 5\,\mathrm{g/l}$ and $\Delta F = 0.5\,\mathrm{l/h}$ are used for the respective initial substrate concentration and feed rate. The initial model parameters used in the first iteration are given in table 4.2.

μ_X	K_X	μ_P	K_P	K_I	$Y_{X/s}$	$Y_{P/S}$	m_X
0.092	0.15	0.008	0.0002	0.1	0.45	0.9	0.014

Table 4.2.: Initial values of the penicillin model parameters.

4.4.1.2. Output Uncertainty Bounds

As discussed in section 4.3.1, the worst-case uncertainty bounds are used to enforce boundedness of the model outputs in each iteration. In this work, we calculate the worst-case error term (4.12) for each operating point as follows. First, we note that from the $n_b = 4$ batch experiments per operating point $\boldsymbol{u}_k$ we can calculate the average output values at each sampling time:

$$\hat{\boldsymbol{y}}_{p,k} = \frac{1}{n_b} \sum_{i=1}^{n_b} \boldsymbol{y}_{p,k}^{i} \tag{4.32}$$

Then, the worst-case errors for the outputs and input $\boldsymbol{u}_k$ are calculated as:

$$\boldsymbol{e}_k^L = \min_{i\in\{1,\dots,n_b\}} \left(\boldsymbol{y}_{p,k}^i - \hat{\boldsymbol{y}}_{p,k}\right) \tag{4.33}$$

$$\boldsymbol{e}_k^U = \max_{i\in\{1,\dots,n_b\}} \left(\boldsymbol{y}_{p,k}^i - \hat{\boldsymbol{y}}_{p,k}\right) \tag{4.34}$$

Consequently, the worst-case deviations from the mean are used as the bounded errors on the output measurements. An example of the obtained uncertainty bounds for the penicillin process and feasible trajectories are given in figure 4.1. The main reason for using these bounded errors is to provide a more even and broader range for the predicted outputs that will permit to satisfy the model feasibility condition in (4.16) despite the existing model mismatch.

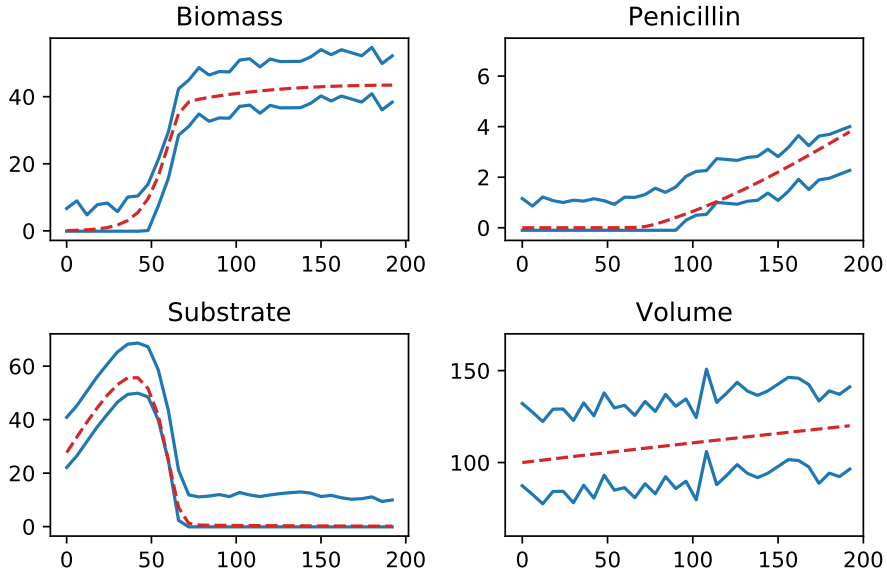

Figure 4.1.: Illustration of uncertainty bounds (-) and feasible model-fitting (--) for the penicillin process. Biomass, penicillin and substrate in [g/l]. Volume in [l] and time in [h].

4.4.1.3. Run-to-Run Optimization Results

In the following, we compare the proposed method to the previous version of the simultaneous identification and optimization method for two different parameter subsets. Subset I, consisting of K_I and K_X has been implemented in previous applications (Mandur & Budman, 2015b), while subset II, consisting of m_X, μ_P and μ_X, is based on an appropriate parameter selection (Hille et al., 2017).

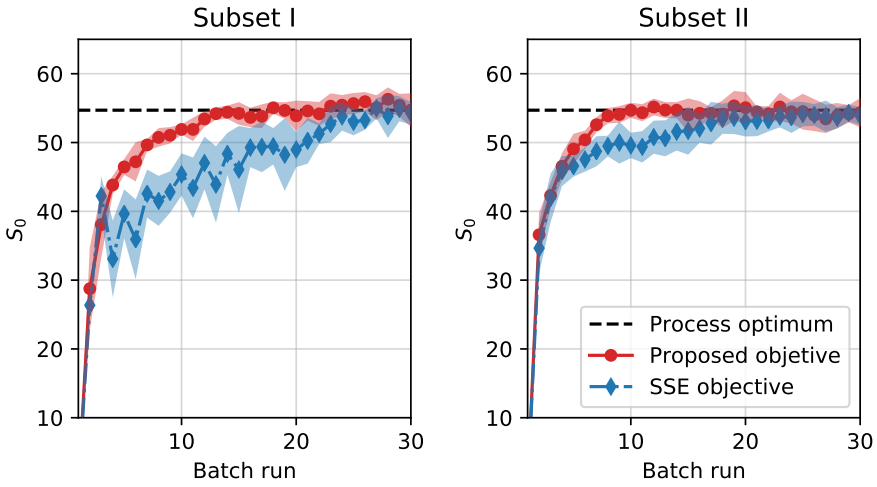

Figure 4.2.: Comparison of convergence of S_0 for parameter subset I and II.

Due to the constraint on the volume of the reactor, the optimal constant feed rate of $F^* = 0.1728\,\mathrm{l/h}$ can be obtained within one iteration. However, the model-plant mismatch has a significant effect on the prediction of the optimal substrate concentration of $S_0^* = 54.72\,\mathrm{g/l}$ and thus requires numerous batch experiments. The upper bound on the relative truncation error in (4.5) is selected as $\epsilon_{max} = 0.01$ and it is thus lower than the magnitude of measurement noise of $10\,\%$. This is done to maintain a model-fitting accuracy that is sufficient to satisfy constraints and other relevant prediction objectives. Finally, in order to account for different noise and disturbance realizations,

the results are averaged over 10 simulations.

The performance in convergence to the process optimum for parameter subset I is illustrated on the left graph in figure 4.2. The shaded areas represent the $95\,\%$ confidence regions obtained from the 10 simulations. It can be seen that the proposed objective leads to a faster and smoother convergence to the process optimum. This result corroborates that, by using an objective function for identification that consists of the SSE divided by the gradient sensitivity to parameter changes, we have obtained a better gradient correction. This is especially evident from the reduction in the confidence intervals. Overall, the proposed method leads to a reduction in the IAE (Integral Absolute Error) of $33\,\%$ (where the IAE is calculated based on the errors with respect to the true optimum S_0^*) and a decrease in the gradient error of $37\,\%$ for subset I.

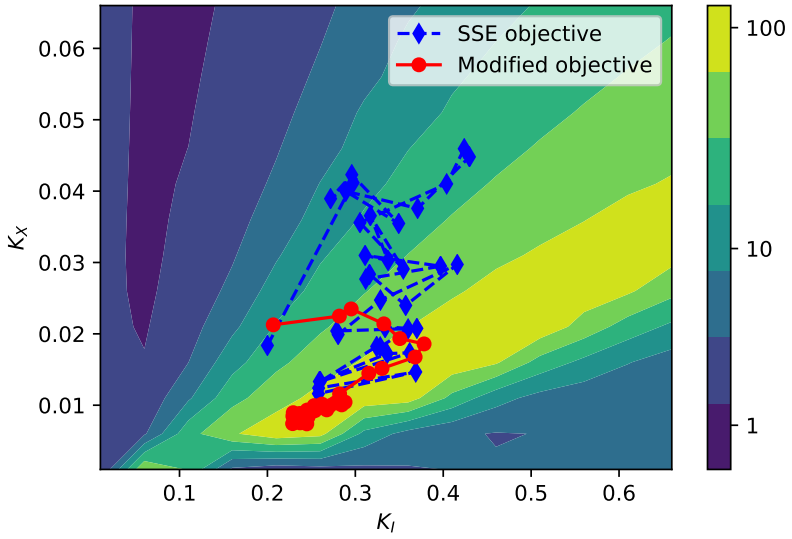

Figure 4.3.: Comparison of K_I and K_X in terms of the gradient sensitivity objective function ϕ_{S^∇}.

The main improvement stems of the fact that the inclusion of a gradient sensitivity objective forces the parameters into a region of high sensitivity. This is illustrated in figure 4.3, where the gradient sensitivity objective function ϕ_{S^∇} and the trajectories of the parameter values of K_I and K_X are

illustrated for both the SSE and the proposed modified identification objective. It is clear, that by using the proposed objective, the parameter values converge to a region of high gradient sensitivity, while the estimated parameter values from an SSE objective mostly lie in regions of lower gradient sensitivity.

The results for parameter subset II are presented on the right graph of figure 4.2. Note that this subset already results in an improved performance as compared to subset I, since the selected model parameters provide larger parametric gradient sensitivities. However, even for this case, the proposed approach leads to an additional speed-up in convergence. Furthermore, as it is also the case for parameter subset I, the new estimation objective results in less uncertainty of the next optimal batch run as illustrated by the narrower confidence regions. Overall, the presented method reduces the IAE by $24\,\%$ and improves the gradient correction by $36\,\%$ for subset II.

4.4.2. CHO Cell Cultivation Process

An important class of proteins with therapeutic potential are humanized monoclonal antibodies (MAb). In recent years, increasing attention has been paid to improve the production of MAb by cultivation in bioreactors. To this end, dynamic metabolic models (Provost et al., 2006; Zamorano et al., 2013; Yahia et al., 2015) have been developed to obtain a better understanding of these processes. For that reason, this second case study is of a cell cultivation process using a dynamic metabolic model previously developed for a CHO cell line (Aghamohseni et al., 2014). The model describes the dynamics of MAb and other extra cellular metabolites during a batch experiment. Dynamic metabolic models can be developed using the metabolic flux analysis (MFA) methodology (Naderi et al., 2011). To reduce the complexity of the entire reaction network, MFA is utilized to determine metabolites and fluxes which are important for explaining the distribution of carbon and nitrogen in the cultivation process. In this way, insignificant fluxes can be eliminated leading to a reduction of the network to a smaller set of macro-reactions, where each of those can be described

by Monod-type kinetics. Consequently, this kind of model reduction also leads to a certain degree of structural mismatch as only the important reactions are included in the model. The model development is further complicated by the fact that not all metabolites can be measured, further limiting prediction capabilities.

The process simulator is taken from (Aghamohseni et al., 2014) and is defined by the following set of mathematical equations:

$$\frac{dfgr}{dt} = -K_{11}\frac{fgr}{1+\left(\frac{[Glc]}{K_{12}}\right)} \tag{4.35}$$

$$\frac{dX_v}{dt} = \mu \cdot fgr \cdot X_v \left(\frac{[Glc]\,[Gln]}{(K_{21}+[Glc])\,(K_{26}+[Gln])}\frac{1}{1+\left(\frac{[Amm]}{K_{23}}\right)}\frac{1}{1+\left(\frac{[Glc]}{K_{22}}\right)}\right) - \frac{dX_d}{dt} \tag{4.36}$$

$$\frac{dX_d}{dt} = k_d\,(1-fgr)\,X_v\left(\frac{1}{1+\left(\frac{K_{24}}{[Amm]}\right)^n}+\frac{K_{25}}{[Glc]}\right) \tag{4.37}$$

$$\frac{d\,[Glc]}{dt} = -\left(\frac{K_{31}\,[Glc]\,[Gln]}{(K_{32}+[Glc])\,(K_{36}+[Gln])}\frac{K_{33}\,[Glc]}{(K_{34}+[Glc])}\right)X_v - K_{35}X_v \tag{4.38}$$

$$\frac{d\,[Gln]}{dt} = -\left(\frac{K_{41}\,[Glc]\,[Gln]}{(K_{42}+[Glc])\,(K_{43}+[Gln])}\right)X_v \tag{4.39}$$

$$\frac{d\,[Lac]}{dt} = -\left(\frac{K_{51}\,[Glc]}{(K_{52}+[Glc])}\frac{d\,[Glc]}{dt}\right)X_v \tag{4.40}$$

$$\frac{d\,[Asn]}{dt} = -\left(\frac{K_{61}\,[Asn]}{(K_{62}+[Asn])}\right)X_v \tag{4.41}$$

$$\frac{d\,[Asp]}{dt} = \left(\frac{K_{61}\,[Asn]}{(K_{62}+[Asn])}+\frac{K_{63}\,[Glc]\,[Gln]}{(K_{64}+[Glc])\,(K_{65}+[Gln])}-\frac{K_{71}\,[Asp]}{(K_{72}+[Asp])}\right)X_v \tag{4.42}$$

$$\frac{d\left[Ala\right]}{dt} = \left(\frac{K_{63}\left[Glc\right]\left[Gln\right]}{\left(K_{64}+\left[Glc\right]\right)\left(K_{65}+\left[Gln\right]\right)} - \frac{K_{81}\left[Ala\right]}{\left(K_{82}+\left[Ala\right]\right)}\right) X_v \tag{4.43}$$

$$\frac{d\left[Amm\right]}{dt} = -K_{91}\frac{d\left[Gln\right]}{dt} + K_{92}\left(\frac{K_{61}\left[Asn\right]}{\left(K_{62}+\left[Asn\right]\right)} + \frac{K_{71}\left[Asp\right]}{\left(K_{72}+\left[Asp\right]\right)} + \frac{K_{81}\left[Ala\right]}{\left(K_{82}+\left[Ala\right]\right)}\right) X_v \tag{4.44}$$

$$\frac{d\left[Mab\right]}{dt} = \left(K_{101}+K_{102}\left[Gln\right]\right) X_v \tag{4.45}$$

where fgr represents the fraction of growing cells, X_v the viable cell density (VCD) and X_d the dead cell density (DCD). Furthermore, the concentrations of metabolites are defined as follows: $[Glc]$ - glucose, $[Gln]$ - glutamine, $[Lac]$ - lactate, $[Asn]$ - asparagine, $[Asp]$ - aspartate, $[Ala]$ - alanine, $[Amm]$ - ammonia and $[Mab]$ – MAb. The kinetic parameters are given by K_{11} to K_{102}. As in the case of the penicillin process, the simulator (4.35) – (4.45) is used to generate in silico measurements of the process outputs and gradients. For uncertainty in measurements, additive Gaussian noise of $10\,\%$ of the average output values is assumed as well as stochastic disturbances in initial batch conditions. Gradients are estimated using finite differences by running an additional batch experiment and then calculating the difference in the cost over the perturbation step size. The initial media composition for the CHO study are given in table 4.3.

Fraction of growing cells (fgr)	1
Viable Cell Density (X_v)	$0.0024\,10^6 \mathrm{cells/ml}$
Dead Cell Density (X_d)	$0.0001\,10^6 \mathrm{cells/ml}$
Glucose ($[Glc]$)	$60\,\mathrm{mmol/l}$
Glutamine ($[Gln]$)	$3.48\,\mathrm{mmol/l}$
Lactate ($[Lac]$)	$0.22\,\mathrm{mmol/l}$
Ammonia ($[Amm]$)	$0.71\,\mathrm{mmol/l}$
Aspartate ($[Asp]$)	$1.22\,\mathrm{mmol/l}$
Alanine ($[Ala]$)	$0.66\,\mathrm{mmol/l}$
Asparagine ($[Asn]$)	$0.67\,\mathrm{mmol/l}$
Monoclonal Antibodies ($[Mab]$)	$0.1\,\mu\mathrm{g/ml}$

Table 4.3.: Initial batch conditions for the CHO process.

The parameters of the process simulator have been identified from experimental data obtained by running several batch experiments in shaker flasks. It is important to point out that, in comparison to the model given in (Aghamohseni et al., 2014), we have also included an inhibition effect in growth with respect to large concentrations of glucose in the current process model. This behavior was observed during the experiments for larger glucose concentrations, but not included in the original model due to limited available experimental data. The parameter used for generating the output and gradient measurements are given in table 4.4.

K_{11}	0.1996	K_{42}	10.1826
K_{12}	1.7519	K_{43}	4.9627
μ	4.1363	K_{51}	0.5338
K_{21}	11.9183	K_{52}	$1 \cdot 10^{-7}$
K_{22}	60	K_{61}	1.9053
K_{23}	2.0097	K_{62}	27.0327
K_{24}	4.8981	K_{63}	0.3899
k_d	2.1627	K_{64}	$1 \cdot 10^{-5}$
n	0.1046	K_{65}	0.0306
K_{25}	0.0281	K_{71}	3.0264
K_{26}	0.1449	K_{72}	7.8973
K_{31}	25.4691	K_{81}	5.5223
K_{32}	27.3228	K_{82}	34.7467
K_{33}	0.8686	K_{91}	0.4451
K_{34}	29.7875	K_{92}	0.2498
K_{35}	0.1964	K_{101}	1
K_{36}	0.1456	K_{102}	0.7
K_{41}	18.1200		

Table 4.4.: Parameter values used in the simulator (4.35) – (4.45).

4.4.2.1. Model-Plant Mismatch

In addition to the process simulator that is used to generate in silico data, we also define a model of the process that is used for calibration and optimization purposes. To introduce structural mismatch between the model and process, we assume that glutamine ($[Gln]$) measurements are not available and that its effects on the cell metabolism are unknown. This is motivated by the general observation that the concentrations of many media components are frequently not measured throughout the cultivation and that their effects on the cell metabolism are often not well understood. Therefore, the model equations are similar to the simulator, with the difference that all the parts containing a

product involving glutamine concentration are made a function of only the glucose concentration and that all dependencies on glutamine alone are eliminated. This results in a structural error for which the remaining model parameters must compensate for to achieve a precise model fitting. Furthermore, this impairs the prediction capability in terms of the location of the process optimum. Thus, the model of the process used for identification and optimization is described by the following ODEs:

$$\frac{dfgr}{dt} = -K_{11}\frac{fgr}{1+\left(\frac{[Glc]}{K_{12}}\right)} \tag{4.46}$$

$$\frac{dX_v}{dt} = \mu \cdot fgr \cdot X_v \left(\frac{[Glc]}{(K_{21}+[Glc])} \frac{1}{1+\left(\frac{[Amm]}{K_{23}}\right)} \frac{1}{1+\left(\frac{[Glc]}{K_{22}}\right)} \right) - \frac{dX_d}{dt} \tag{4.47}$$

$$\frac{dX_d}{dt} = k_d\left(1-fgr\right)X_v \left(\frac{1}{1+\left(\frac{K_{24}}{[Amm]}\right)^n} + \frac{K_{25}}{[Glc]} \right) \tag{4.48}$$

$$\frac{d\left[Glc\right]}{dt} = -\left(\frac{K_{33}\left[Glc\right]}{(K_{34}+[Glc])} \right) X_v - K_{35}X_v \tag{4.49}$$

$$\frac{d\left[Lac\right]}{dt} = -\left(\frac{K_{51}\left[Glc\right]}{(K_{52}+[Glc])} \frac{d\left[Glc\right]}{dt} \right) X_v \tag{4.50}$$

$$\frac{d\left[Asn\right]}{dt} = -\left(\frac{K_{61}\left[Asn\right]}{(K_{62}+[Asn])} \right) X_v \tag{4.51}$$

$$\frac{d\left[Asp\right]}{dt} = \left(\frac{K_{61}\left[Asn\right]}{(K_{62}+[Asn])} + \frac{K_{63}\left[Glc\right]}{(K_{64}+[Glc])} - \frac{K_{71}\left[Asp\right]}{(K_{72}+[Asp])} \right) X_v \tag{4.52}$$

$$\frac{d\left[Ala\right]}{dt} = \left(\frac{K_{63}\left[Glc\right]}{(K_{64}+[Glc])} - \frac{K_{81}\left[Ala\right]}{(K_{82}+[Ala])} \right) X_v \tag{4.53}$$

$$\frac{d\left[Amm\right]}{dt} = K_{92}\left(\frac{K_{61}\left[Asn\right]}{(K_{62}+\left[Asn\right])} + \frac{K_{71}\left[Asp\right]}{(K_{72}+\left[Asp\right])} + \frac{K_{81}\left[Ala\right]}{(K_{82}+\left[Ala\right])}\right) X_v \tag{4.54}$$

$$\frac{d\left[Mab\right]}{dt} = K_{101} X_v \tag{4.55}$$

The goal of the run-to-run optimization is to maximize the amount of MAb at the end of the batch time of $t_f = 9\,\mathrm{days}$. The decision variable in this case is the initial concentration of the main substrate glucose $[Glc]_0$. Thus, the optimization problem to be solved in each iteration can be stated as:

$$\begin{aligned} \min_{[Glc]_0} \quad & -\left[Mab\right](x, [Glc]_0, \theta, t_f) \\ \text{s.t.} \quad & (4.46) - (4.55) \end{aligned} \tag{4.56}$$

For the CHO case study, we assume that five batch experiments are performed at each operating point to obtain output measurements and the process uncertainty bounds. To obtain an estimate of the cost gradient, a step size of $\Delta[Glc]_0 = 4\,\mathrm{mmol/l}$ is used for the initial glucose concentration. Furthermore, the initial model parameters are given in table 4.5. It should be noticed that this example is considerably more complex, as compared to the first case study, in terms of the number of equations and the corresponding larger number of parameters that are candidates for model calibration and updating. Due to the complexity and number of parameters for this model, the parameter estimation step is divided into smaller problems as outlined below.

K_{11}	0.0140	K_{51}	0.2788
K_{12}	42.4918	K_{52}	1.8564
μ	3.2719	K_{61}	9.6146
K_{21}	53.6983	K_{62}	23.7654
K_{22}	10	K_{63}	0.5120
K_{23}	0.0405	K_{64}	8.5160
K_{24}	47.5529	K_{71}	3.2121
k_d	2.2	K_{72}	9.6238
n	0.2103	K_{81}	5.9924
K_{25}	12.9023	K_{82}	34.7341
K_{33}	36.1942	K_{92}	0.8071
K_{34}	55.1386	K_{101}	1.0023
K_{35}	0.0375		

Table 4.5.: Initial model parameters for the CHO model used in the run-to-run optimization.

The subset of parameters that is used in the gradient correction step is determined from the parameter selection outlined in Hille et al. (2017). For this purpose, we only selected parameters which provide large parametric sensitivities with respect to the gradients of the cost-function. Consequently, parameters K_{11}, μ, K_{23}, K_{25}, K_{22} and K_{101} are used in the gradient correction step for this case study. A threshold of $1\,\%$ of the magnitude of the most gradient sensitive parameter has been used to determine the number of parameters used in the gradient correction step.

4.4.2.2. Output Uncertainty Bounds

The uncertainty bounds for the CHO process have been derived in a similar fashion as described in section 4.4.1.2 for the penicillin process. Figure 4.4 illustrates the obtained uncertainty bounds for the CHO process together with one feasible trajectory resulting from model fitting, i.e. a trajectory that satisfies the set based bounds. The results show that the model can provide a feasible fitting

despite the structural mismatch. However, due to the missing glutamine metabolism in the model, the prediction capabilities of the fitted model are only valid for a small range of inputs.

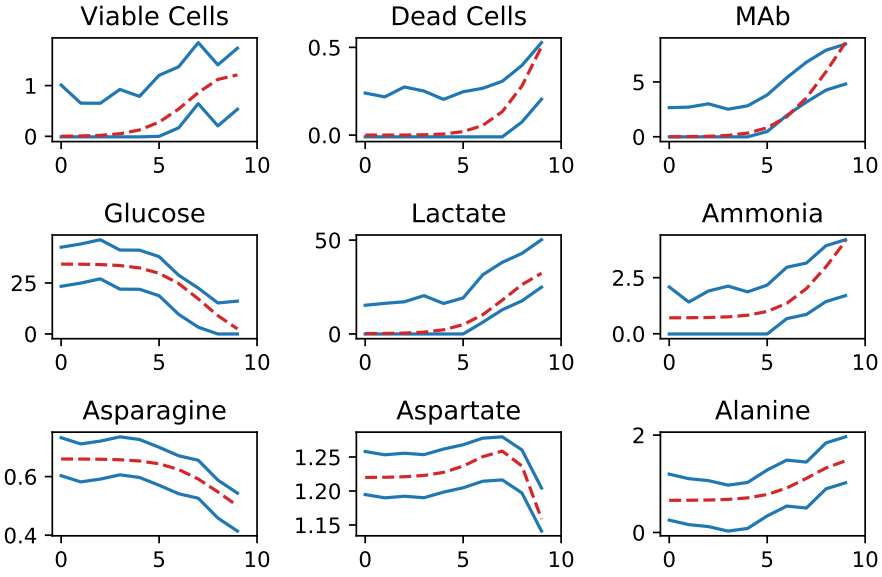

Figure 4.4.: Illustration of uncertainty bounds (-) and feasible model-fitting (--) for the CHO process. Units: Viable and dead cell densities in $[10^6\ \mathrm{cells/ml}]$; glucose, lactate, ammonia, asparagine, aspartate and alanine in $[\mathrm{mmol/l}]$; MAb in $[\mu\mathrm{g/ml}]$ and time in $[\mathrm{d}]$.

4.4.2.3. Model Parameter Estimation Using Uncertainty Bounds

When considering a biochemical model with many equations and parameters, it is not trivial to reliably identify all model parameters simultaneously. For that reason, we are making use of the introduced uncertainty bounds to facilitate the parameter estimation for the dynamic metabolic model. This is especially important considering that the parameters must be re-estimated each time the model is updated in the run-to-run optimization. The main motivation behind the presented procedure is that the dynamics of most metabolites depend on the behavior in VCD, as can be seen from the presence of X_v term in each equation ((4.46) - (4.55)). Therefore, in order to obtain initial parameter values for the extracellular metabolite equations ((4.49) – (4.55)), a piecewise

interpolation of the experimental VCD profile can be used for a preliminary fitting. Given input $\boldsymbol{u}_k$, let the measured profile be given by $\boldsymbol{X}_{p,v}$, a piecewise interpolation $\hat{\boldsymbol{X}}_{p,v}$ can then be obtained such that:

$$\underline{\boldsymbol{X}}_v \leq \hat{\boldsymbol{X}}_{p,v} \leq \overline{\boldsymbol{X}}_v \tag{4.57}$$

where the definition of the uncertainty bounds guarantees that the interpolated trajectory satisfies the output constraints $\underline{\boldsymbol{X}}_v$ and $\overline{\boldsymbol{X}}_v$. Using the approximation of the VCD trajectory, it is possible to obtain initial parameter values such that the output constraints can be satisfied for all metabolites. This task is divided up into smaller problems, where the dynamics of the main substrates, for which uptake/production it is not coupled to the other metabolites are estimated first. Finally, the dynamics of the remaining species are estimated. This results in adequate initial model prediction and estimates of parameter values which are then used to identify the dynamics of the viable and dead cell densities. This separate estimation procedure of model parameters for metabolites and for viable and dead cells is repeated iteratively until convergence. However, parameters of metabolites are only re-estimated if the corresponding output uncertainty bounds are violated. The procedure is stopped once all set-based constraints have been satisfied.

Mathematically, the identification procedure is implemented as follows. We first note that the dynamic model, given by (4.46) - (4.55), is derived from a set of macro-reactions which can be stated as follows (Naderi et al., 2011):

$$\frac{d\boldsymbol{\xi}(t)}{dt} = \boldsymbol{K}\boldsymbol{r}(t) \tag{4.58}$$

where $\boldsymbol{\xi} \in \mathbb{R}^{n_\xi}$ is the vector of extracellular metabolites, $\boldsymbol{K} \in \mathbb{R}^{n_\xi \times n_r}$ the matrix composed of the stoichiometric coefficients of the macro-reactions and $\boldsymbol{r} \in \mathbb{R}^{n_r}$ a vector of reaction rates that are expressed by Monod-type kinetics. The assumed macro-reactions for the model used for optimization are given in table 4.6. For more information on the reactions used in the original model used as the simulator, see Aghamohseni et al. (2014).

#	Reaction
1	$Glc \rightarrow 2Lac$
2	$Glc \rightarrow 6CO_2$
3	$Asn \rightarrow Asp + NH_3$
4	$Ala \rightarrow NH_3 + 3CO_2$
5	$Glc \rightarrow Asp + Ala + CO_2$
6	$Asp \rightarrow NH_3 + 4CO_2$

Table 4.6.: Macro-reactions assumed for the model used in the run-to-run optimization.

From the macro-reactions outlined in table 4.6, we can obtain the matrix of stoichiometric coefficients as follows:

$$\boldsymbol{K} = \begin{array}{c} \\ Glc \\ Lac \\ Asn \\ Asp \\ Ala \\ Amm \\ CO_2 \end{array} \begin{bmatrix} -1 & -1 & 0 & 0 & -1 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 6 & 0 & 3 & 1 & 4 \end{bmatrix} \tag{4.59}$$

where each column describes the corresponding reaction in table 4.6. Using the matrix in (4.59), we define a vector of a subset of substrates $\boldsymbol{\varphi}$ whose evolution depends only on biomass but not on the concentration of other substrates, i.e. $\boldsymbol{\varphi} = f(\boldsymbol{X}_v)$. This corresponds to species associated with rows that do not contain any positive values. It is important to point out here, that the lack of dependency on other substrates is related to the assumption that the uptake kinetics are always assumed to be a function of the reactant but not of the product of each reaction. Formally, the related indices can be defined as:

$$\boldsymbol{\mathcal{M}} = \{i \in \boldsymbol{\mathcal{R}} | \nexists k_{ij} > 0, \forall j \in \boldsymbol{\mathcal{C}}\} \tag{4.60}$$

With the row and column indices given by $\boldsymbol{\mathcal{R}} = \{1, \ldots, n_\xi\}$ and $\boldsymbol{\mathcal{C}} = \{1, \ldots, n_r\}$. Subsequently, the corresponding subset of metabolites is given by:

$$\boldsymbol{\varphi} = \{\boldsymbol{\xi}_i | i \in \boldsymbol{\mathcal{M}}\} \tag{4.61}$$

For which the corresponding uncertainty bounds, obtained from measurements at a given input $\boldsymbol{u}_k$, are denoted by $\boldsymbol{\Phi}_k$. For this case study, the relevant metabolites which are not affected by the evolution of the other metabolites are glucose (Glc) and asparagine (Asn). Next, we are interested in metabolites whose dynamics only depend on the species in $\boldsymbol{\varphi}$ and biomass, i.e. $\boldsymbol{\psi} = f(\boldsymbol{X}_v, \boldsymbol{\varphi})$. In other words, metabolites that are produced as a direct result of consumption of metabolites in set $\boldsymbol{\mathcal{M}}$. These indices are defined by:

$$\boldsymbol{\mathcal{N}} = \{i \in \boldsymbol{\mathcal{R}} | (\exists k_{ij} > 0 \wedge \nexists k_{lj} < 0, \forall l \in \boldsymbol{\mathcal{R}} \setminus \boldsymbol{\mathcal{M}}), j \in \boldsymbol{\mathcal{C}}\} \tag{4.62}$$

From which we obtain:

$$\boldsymbol{\psi} = \{\boldsymbol{\xi}_i | i \in \boldsymbol{\mathcal{N}}\} \tag{4.63}$$

where the corresponding uncertainty bounds at input $\boldsymbol{u}_k$ are denoted by $\boldsymbol{\Psi}_k$. From (4.59), the relevant metabolites included in (4.63) are lactate (Lac), aspartate (Asp) and alanine (Ala). Finally, the remaining metabolites can be determined from:

$$\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{R}} \setminus (\boldsymbol{\mathcal{M}} \cup \boldsymbol{\mathcal{N}}) \tag{4.64}$$

With the corresponding vector of metabolite concentrations given by:

$$\boldsymbol{\omega} = \{\boldsymbol{\xi}_i | i \in \boldsymbol{\mathcal{P}}\} \tag{4.65}$$

For which we get $\boldsymbol{\omega} = f(\boldsymbol{X}_v, \boldsymbol{\varphi}, \boldsymbol{\psi}, \boldsymbol{\omega})$ and the associated uncertainty bounds $\boldsymbol{\Omega}_k$. In the current

case study, the only remaining metabolites are ammonia (Amm) and MAb (Mab).

To summarize the described procedure, the sequential estimation algorithm is outlined in Algorithm 4.2. Note that, for the dynamic metabolic model, the presented estimation algorithm replaces step 2-7 in the run-to-run optimization framework shown in Algorithm 4.1.

Algorithm 4.2 Sequential parameter estimation in the presence of model-plant mismatch.

1: Obtain uncertainty bounds by performing experiments at input $\boldsymbol{u}_k$
2: Approximate $\boldsymbol{X}_{p,v}$ by a piecewise interpolation $\hat{\boldsymbol{X}}_{p,v}$
3: **if** $\boldsymbol{\varphi} \notin \boldsymbol{\Phi}_k$ **then**
4: Solve (4.23) such that $\boldsymbol{\varphi}_k(\boldsymbol{X}_v, \boldsymbol{\theta}_k) \in \boldsymbol{\Phi}_k$
5: **end if**
6: **if** $\boldsymbol{\psi} \notin \boldsymbol{\Psi}_k$ **then**
7: Solve (4.23) such that $\boldsymbol{\psi}_k(\boldsymbol{X}_v, \boldsymbol{\varphi}, \boldsymbol{\theta}_k) \in \boldsymbol{\Psi}_k$
8: **end if**
9: **if** $\boldsymbol{\omega} \notin \boldsymbol{\Omega}_k$ **then**
10: Solve (4.23) such that $\boldsymbol{\omega}_k(\boldsymbol{X}_v, \boldsymbol{\varphi}, \boldsymbol{\psi}, \boldsymbol{\omega}, \boldsymbol{\theta}_k) \in \boldsymbol{\Omega}_k$
11: **end if**
12: Use predicted dynamics of VCD: $\boldsymbol{X}_v$
13: **if** $\boldsymbol{X}_v \notin [\underline{\boldsymbol{X}}_v, \overline{\boldsymbol{X}}_v]$ **then**
14: Solve (4.23) such that $\underline{\boldsymbol{X}}_v \leq \boldsymbol{X}_v(\boldsymbol{\varphi}, \boldsymbol{\psi}, \boldsymbol{\omega}, \boldsymbol{\theta}_k) \leq \overline{\boldsymbol{X}}_v$
15: **end if**
16: Go back to step 3 and repeat estimation until feasible trajectories have been found.

A key benefit of the proposed method is that it is possible to obtain feasible model parametrizations with an initial estimation using an experimental VCD profile. By dividing the identification into smaller problems, only a subset of parameters is estimated in each step. Furthermore, an update of parameters is only performed if uncertainty bounds are not satisfied. This reduces the computational effort and facilitates finding a feasible model prediction during each iteration of the run-to-run optimization procedure.

4.4.2.4. Run-to-Run Optimization Results

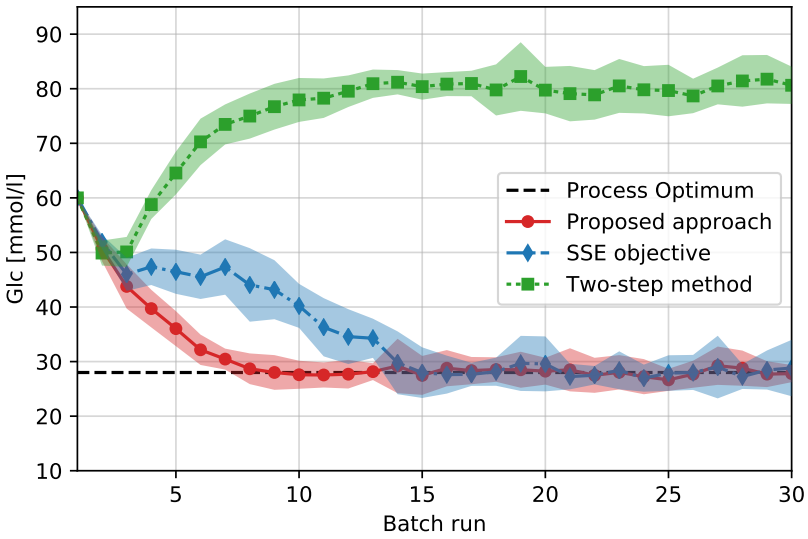

Figure 4.5.: Comparison of the run-to-run optimization results in terms of the initial glucose concentration.

The batch-to-batch optimization results are presented in figure 4.5. We compare the proposed approach using the algorithm to the original simultaneous identification/optimization algorithm used by Mandur & Budman (2015b), where the latter is based on the minimization of the SSE for identification. In addition, results are shown when performing a run-to-run optimization when just using the two-step approach (repeated identification and optimization) (Chen & Joseph, 1987). Again, the results are averaged over 10 simulations and the shaded areas illustrate the $95\,\%$ confidence intervals. The left graph in figure 4.5 shows the convergence results in terms of the initial glucose concentration $[Glc]_0$, where the optimal concentration is given by $[Glc]_0^* = 28.17\,\mathrm{mM}$. It is clear that the proposed approach outperforms the original SSE based identification/optimization procedure both, in terms of speed and confidence in the prediction of the next input. Overall, the results show an improvement of $52\,\%$ and $38\,\%$ in the IAE and confidence intervals, respectively.

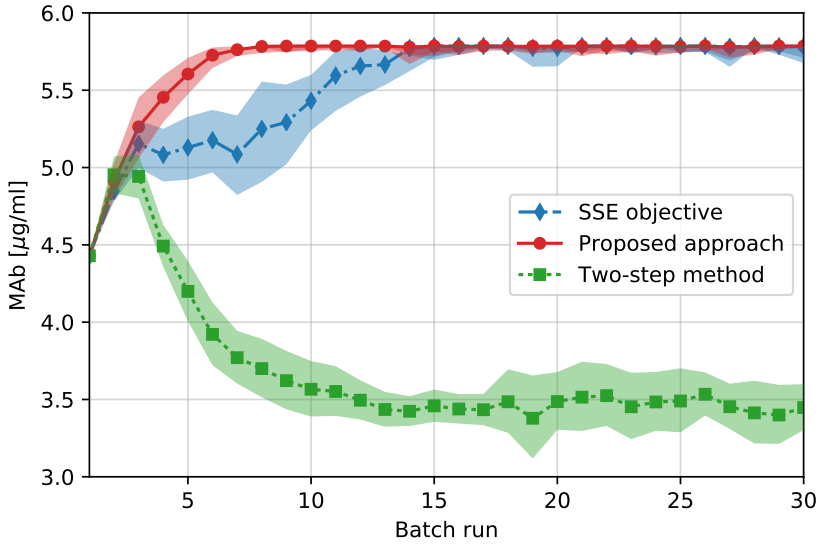

Figure 4.6.: Comparison of the run-to-run optimization results in terms of the final amount of MAb.

Furthermore, figure 4.6 illustrates the performance of the three approaches in terms of the final amount of MAb. When using the two-step method, i.e. not correcting for errors in the predicted gradients, the obtained operating policy is sub-optimal and far from the process optimum leading to a loss in product yield of $37\,\%$.

4.5. Conclusions

In this chapter we presented a parameter identification method tailored for the simultaneous identification and optimization of batch processes in the presence of model-plant mismatch. The method introduced a modified objective function for identification that is different from the standard SSE minimization used in the original approach, consisting of the SSE divided by the gradient sensitivity to parameter changes. Since the SSE objective is modified, uncertainty bounds on time trajectories are imposed to ensure an acceptable level of fitting between model predictions and

measurements. At the same time, the use of these uncertainty bounds on the process outputs allow for the formulation of a model-update criterion that determines whether model update is needed or not. This is especially useful when dealing with the identification of biochemical models with a larger number of parameters. Overall, the presented approach leads to an improved reconciliation of the objectives of identification and optimization, resulting in a more robust (less oscillatory) and much faster convergence to the process optimum as compared to the original method with less uncertainty in the prediction of the optimal input for the next batch run. Future research focuses on the incorporation of a design of experiments approach to obtain a better cost prediction over a wider range of inputs and gradient measurements that are more informative for obtaining higher parameter precision.

5. Design of Experiments for Optimization Under Model-Plant Mismatch

Overview

Model-plant mismatch commonly arises from simplifications and assumptions during the development of first-principles models. Hence, when employing such models in iterative optimization schemes, structural mismatch may lead to inaccurate prediction of the necessary conditions of optimality. This results in convergence to a predicted optimum which does not coincide with the actual process optimum. The method of simultaneous identification and optimization, used in the current work, aims to correct for errors in the predicted gradients of the cost and constraints by adapting the model parameters. In a former implementation of this approach, the gradients have been corrected only locally at the current operating point. To achieve a better prediction of the cost function over a wider range of input conditions, we propose to consider cost measurements from previous batch experiments combined with an optimal experimental design of future experiments. Using this approach, it is possible to achieve a better prediction, especially around the optimum, and to make the gradient correction step less susceptible to uncertainty in local gradient measurements. The improvements are illustrated using three simulated run-to-run optimization case studies.

Adapted from Hille, R., and Budman, H. (2018). Experimental design in simultaneous identification and optimization of batch processes under model-plant mismatch. *IFAC Symposium on Advanced Control of Chemical Processes 2018*, [submitted].

5.1. Introduction

Mathematical models play an essential role in the optimal design and operation of chemical processes and are typically classified as either black-box or first-principles models (Bonvin et al., 2016). While first-principles models require process knowledge and a rigorous development, they offer the benefit of superior extrapolation abilities compared to black-box models thus offering a better chance of predicting an optimum that lies outside the range of data used for model calibration (Yip & Marlin, 2004).

However, due to simplifications and assumptions during model development, there regularly exists structural mismatch between the model and the process. As a result, the model parameters' values that minimize the errors between measured and predicted process outputs (identification) may not be equal to the values that result in a correct prediction of the gradients of the cost function and constraints (optimization). In this case, an optimization method that is based on successive identification and optimization steps may fail to converge to the process optimum (Srinivasan & Bonvin, 2002).

When the main use of the model is optimization, methods such as Modifier Adaptation (Marchetti et al., 2009; Gao et al., 2016) have been proposed to deal with structural mismatch. On the other hand, for some cases, a model is sought both for optimization as well as for predicting the process behavior around the optimum. For such cases the method for simultaneous identification and optimization (Mandur & Budman, 2015b) has been proposed that aims at finding a set of parameter values which simultaneously predicts the model outputs as well as fits the gradients of cost-function and constraints as to correctly predict the necessary conditions of optimality (NCOs). However, in our previous studies, only the most recent gradient measurements have been used for the gradient correction, thus not making use of information already acquired through past experimental effort. In addition, the choice of the location of next gradient measurement, as realized in the Modifier Adaptation algorithm (Costello et al., 2016), has so far not been addressed in the simultaneous

identification and optimization framework used in the current work.

The Design of Experiments (DoE) methodology, first derived for data-driven models (Box & Draper, 1987), is also an established method for reducing parameter uncertainty in the estimation of nonlinear first-principles models (Franceschini & Macchietto, 2008). The focus of these methods is the minimization of an estimation related criterion associated with the parameter covariance matrix, which is typically approximated using the inverse of the Fisher Information Matrix (FIM). In the context of batch process optimization, such an experimental design approach has been applied for the selection of sampling times that are important for an optimization objective (Martinez et al., 2009). However, as mentioned above, in the presence of model-plant mismatch, a precise fitting of model outputs does not necessary result in an accurate prediction of cost and constraint gradients. This is also relevant if an economic design objective is considered when designing experiments for model output fitting (Houska et al., 2015). Hence, the goal of the simultaneous identification and optimization framework used in this work, is both to fit model outputs to measured ones and matching the measured gradients of the cost-function and the constraints. To improve the prediction of the process cost-function, model parameters' values can be sought that can reduce the parametric uncertainty when fitting the cost function and the constraint gradients. Towards that goal, we make use of a covariance matrix derived from the parametric sensitivities of the gradients of the cost-function and constraints in combination with a suitable experimental design criterion. The overall objective is to determine at each iteration new experiments that provide valuable gradient information in addition to the past experiments.

In summary, this chapter represents an addition to the methodologies outlined in chapter 3 and 4. As a first step, we use information about costs and constraints already gathered from past experiments when correcting the predicted gradients. In a second step, future optimal experiments, necessary for obtaining better gradient estimates, are determined based on a design of experiments approach. It is shown that the presented approach leads to the following improvements:

i. The effect of gradient uncertainty is significantly reduced when cost measurements from previous batch runs are considered.

ii. The improved parameter precision leads to a better prediction of the cost function near the process optimum.

For illustration purposes, run-to-run optimization studies are performed using simulated case studies of a simple synthetic batch process, a penicillin process and a CHO cell cultivation process.

5.2. Simultaneous Identification and Optimization Methodology

The method for simultaneous identification and optimization (Mandur & Budman, 2015b) has been recently extended to a parameter identification using set-based constraints (Hille & Budman, 2017). The main steps are briefly reviewed below.

5.2.1. Identification Using Set-Based Bounds

Suppose we perform several experiments (batch runs) at a given operation point $\boldsymbol{u}_k \in \mathbb{R}^{n_u}$. The collection of measurements for all sampling times t_i can then be defined as (Rumschinski et al., 2010):

$$\boldsymbol{\mathcal{Y}}_k = \left\{ \boldsymbol{\mathcal{Y}}_k^i \in \mathbb{R}^{n_y} | i \in \{1, \ldots, n_t\} \right\} \tag{5.1}$$

where the set-based bounds provide an upper and lower bound for the permissible range of model outputs at each sampling time such that (Streif et al., 2016):

$$\underline{\boldsymbol{y}}_k^i \leq \boldsymbol{y}_k^i \leq \overline{\boldsymbol{y}}_k^i \tag{5.2}$$

With the model outputs given by $\boldsymbol{y}_k \in \mathbb{R}^{n_y}$. Set-based bounds have been found to be particularly well suited for describing experimental data in biological systems (Rumschinski et al., 2010).

When estimating model parameters, a typical model fitting objective is given by the sum of squared errors (SSE) between process outputs and model predictions:

$$\phi_{SSE}(\boldsymbol{\theta}) = \sum_{i=1}^{n_t} \left\| \boldsymbol{y}_{p,k}(t_i) - \boldsymbol{y}_k(\boldsymbol{\theta}, t_i) \right\|^2 \tag{5.3}$$

where $\boldsymbol{y}_{p,k} \in \mathbb{R}^{n_t \times n_y}$ are the plant measurements and $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$ are the set of model parameters. In contrast to a standard identification problem where only model fitting is required, the goal in simultaneous identification and model-based optimization is to find parameter values which yield both good model fitting and a correct prediction of the gradients of the cost function and constraints. To obtain parameter values from the identification step which enhance the performance of the subsequent gradient correction step, Hille & Budman (2017) proposed the following parametric sensitivity objective:

$$\phi_{S^\nabla}(\boldsymbol{u}, \boldsymbol{\theta}) = \sum_{i=1}^{n_u} \sum_{j=1}^{n_\theta} \left| s_{ij}^{\nabla\phi}(\boldsymbol{u}, \boldsymbol{\theta}) \right| + \sum_{i=1}^{n_u \times n_g} \sum_{j=1}^{n_\theta} \left| s_{ij}^{\nabla g}(\boldsymbol{u}, \boldsymbol{\theta}) \right| \tag{5.4}$$

where $s_{ij}^{\nabla\phi}$ and $s_{ij}^{\nabla g}$ are elements of the scaled cost-function and constraint gradient sensitivity matrices:

$$s_{ij}^{\nabla\phi} = \frac{\partial \left(\nabla\phi_i \right)}{\partial \theta_j} \left| \frac{\theta_j}{\nabla\phi_i} \right| \tag{5.5}$$

Equation (5.4) defines a scalar measure of the parametric cost function and constraint sensitivities where large gradient sensitivities are desired since smaller parameter deviations from the ones obtained in the identification step are required for matching of the gradients. Avoiding large deviations between the parameter values required for fitting the outputs and the parameter values required for matching gradients result in smaller oscillations and smoother convergence of the run to run optimization procedure. An objective combining the model-fitting goal (5.3) and the maximization of

the sensitivity measure in (5.4) is subsequently defined as:

$$\begin{aligned} \boldsymbol{\theta}_k \quad &= \quad \arg\min_{\boldsymbol{\theta}} \left(\frac{\phi_{SSE}(\boldsymbol{u}, \boldsymbol{\theta})}{\phi_{S^\nabla}(\boldsymbol{u}, \boldsymbol{\theta})} \right) \\ &\text{s.t.} \quad \dot{\boldsymbol{x}}_k = \boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{\theta}) \\ &\qquad \boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k) - \boldsymbol{c}_{k-1} \\ &\qquad \boldsymbol{\theta} \in \boldsymbol{\Theta}_0 \\ &\qquad \boldsymbol{y}_k \in \boldsymbol{\mathcal{Y}}_k \end{aligned} \tag{5.6}$$

where $\boldsymbol{x}_k \in \mathbb{R}^{n_t \times n_y}$ are the states and $\boldsymbol{y}_k \in \mathbb{R}^{n_t \times n_y}$ the outputs of the model. The correction term $\boldsymbol{c}_{k-1} \in \mathbb{R}^{n_t \times n_y}$ is defined in the gradient correction step described below. The set $\boldsymbol{\Theta}_k$ presents a permissible space for the parameter values. According to (5.6), the goal of the set-based parameter estimation is to fit the model predictions to model outputs while penalizing parameter values which lead to a reduction in the gradient sensitivities. It should be noticed that, since the optimization cost was modified from a norm of the prediction errors as in (5.3) to a norm of the errors divided by the sensitivity function as in (5.6), the set-based bounds are necessary to enforce that the predicted outputs remain reasonably close to the process outputs.

Furthermore, using the output uncertainty bounds (5.1), it is possible to define a model-update criterion so that the multi-objective identification step (5.6) is only performed when it is strictly necessary to update the outputs, i.e. when the uncertainty bounds are violated. For more information, see Hille & Budman (2017).

5.2.2. Gradient Correction

As mentioned above, a correct prediction of the process optimum requires that the predicted gradients at each iteration coincide with that of the process. To satisfy this condition, a gradient

correction step is performed as follows:

$$\begin{aligned}
\Delta\boldsymbol{\theta}_k \quad = \quad & \arg\min_{\Delta\boldsymbol{\theta}} \Bigg(\boldsymbol{w}_\phi^T \left| \nabla\boldsymbol{\phi}_p(\boldsymbol{u}_k) - \nabla\boldsymbol{\phi}(\boldsymbol{y}_k(\boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}), \boldsymbol{u}_k) \right| \\
& + \sum_{j=1}^{n_g} \boldsymbol{w}_{g,j}^T \left| \nabla\boldsymbol{g}_{p,j}(\boldsymbol{u}_k) - \nabla\boldsymbol{g}_j(\boldsymbol{y}_k(\boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}), \boldsymbol{u}_k) \right| \Bigg) \\
\text{s.t.} \quad & \dot{\boldsymbol{x}}_k = \boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}) \\
& \boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k) - \boldsymbol{c}_k \\
& \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta} \in \boldsymbol{\Theta}_0 \\
& \left\| \boldsymbol{\epsilon}_k^T \right\|_\infty \leq \epsilon_{max}
\end{aligned} \tag{5.7}$$

where $\nabla\boldsymbol{\phi} \in \mathbb{R}^{n_u}$ and $\nabla\boldsymbol{g}_j \in \mathbb{R}^{n_u}$ with $j = 1, \ldots, n_g$ are the cost and constraint gradients. The measured gradients are denoted by the subscript p. The errors in gradients are normalized using the respective weights $\boldsymbol{w}_\phi \in \mathbb{R}^{n_u}$ and $\boldsymbol{w}_{g,j} \in \mathbb{R}^{n_u}$. A correction factor $\boldsymbol{c}_k$ is introduced into the model outputs so as to preserve the fitting accuracy that has been achieved in the identification step (5.6). The correction term is derived from a first order Taylor expansion:

$$\boldsymbol{c}_k(t_i) = \boldsymbol{c}_{k-1}(t_i) + \boldsymbol{D}y_k(\boldsymbol{\theta}_k, t_i)\Delta\boldsymbol{\theta}_k \tag{5.8}$$

where $\boldsymbol{D}y_k(\boldsymbol{\theta}_k, t_i) \in \mathbb{R}^{n_y \times n_\theta}$ is the Jacobian of the model at sampling time t_i. The upper bound ϵ_{max} on the relative truncation error is a user selected parameter that determines the maximum error allowed for the model fitting and indirectly determines the allowable amount of gradient correction. The relative truncation error is defined as the error introduced by the linear correction term as follows:

$$\boldsymbol{\epsilon}_k(t_i) = [\boldsymbol{y}_k(\boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}_k, t_i) - \boldsymbol{D}y_k(\boldsymbol{\theta}_k, t_i) - \boldsymbol{y}_k(\boldsymbol{\theta}_k, t_i)] \cdot [\operatorname{diag}(\boldsymbol{y}_k(\boldsymbol{\theta}_k, t_i))]^{-1} \tag{5.9}$$

Thus, after the additional gradient correction step, the adapted parameter values at batch run k are given by:

$$\boldsymbol{\theta}'_k = \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}_k$$

5.2.3. Model-based Optimization

Following the identification (5.6) and the gradient correction steps (5.7), a model based optimization is performed as follows:

$$\begin{aligned} \boldsymbol{u}_{k+1} \quad = \quad & \arg\min_{\boldsymbol{u}} \phi(\boldsymbol{y}(\boldsymbol{\theta}'_k), \boldsymbol{u}) \\ \text{s.t.} \quad & \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\theta}'_k) \\ & \boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{c}_k \\ & \boldsymbol{g}(\boldsymbol{y}(\boldsymbol{\theta}'_k), \boldsymbol{u}) \leq \boldsymbol{0} \\ & \boldsymbol{u}^L \leq \boldsymbol{u} \leq \boldsymbol{u}^U \end{aligned} \tag{5.10}$$

where $\boldsymbol{u}_{k+1}$ presents the optimal input for the next batch run. The lower and upper bounds $\boldsymbol{u}^L$ and $\boldsymbol{u}^U$ limit the search space for the optimal input.

5.3. Experimental Design Methodology

Although the simultaneous identification and optimization methodology already provides some robustness to gradient uncertainty due to the use of a bound on the relative truncation error in (5.7), one drawback is that the gradients are only corrected at the current operating point. Information from past operating points is thus not taken into consideration when using the most recent gradient measurements. However, correcting only at the current operating point may lead to more uncertainty in the prediction of the next optimal batch run due to overfitting of the local gradient.

Furthermore, a local correction may lead to an adequate local prediction, but does not guarantee an accurate prediction of the cost function at other operating points around the process optimum. To introduce additional robustness to uncertainty in gradient measurements and to increase parameter precision, we therefore propose to match the predicted gradients not only locally but also consider cost measurements from previous batch runs. Moreover, the inputs for the next experimental batch runs, necessary for gradient measurements, will be determined based on an optimal experimental design approach. The goal is an improved prediction capability of the model for a wider range of inputs and lower sensitivity to uncertainty in local gradient measurements, especially in the neighbourhood of the process optimum.

5.3.1. Local Gradient Correction

In the parameter adaptation methodology outlined in section 5.2, gradient measurements are required to satisfy the necessary conditions of optimality as per the gradient correction step described in (5.7). Regarding the cost function, a gradient is defined as the derivative with respect to the decision variables:

$$\nabla\phi_m(\boldsymbol{u}_k) = \frac{\partial\phi}{\partial u_m}(\boldsymbol{u}_k) \tag{5.11}$$

where $m = 1, \ldots, n_u$ denotes the respective decision variable. The gradient at operating point $\boldsymbol{u}_k$ can be estimated by performing a step change Δu_m in the direction of each decision variable. At input $\boldsymbol{u}_k$ and using finite differences, the derivative can be approximated as a normalized difference coefficient between two operating points:

$$\alpha_{k,m} = \frac{\phi(\boldsymbol{u}_k + \Delta u_m \boldsymbol{e}_m) - \phi(\boldsymbol{u}_k)}{\|\Delta u_m \boldsymbol{e}_m\|} \tag{5.12}$$

where $\boldsymbol{e}_m \in \mathbb{R}^{n_u}$ is the identity vector in the direction of the respective decision variable. Using this approach, the standard gradient correction problem from (5.7) for the cost-function gradients

can thus also be expressed as follows:

$$\begin{aligned}
\Delta\boldsymbol{\theta}_k \quad &= \quad \arg\min_{\Delta\boldsymbol{\theta}} \sum_{m=1}^{n_u} w_m^{\alpha} \left\| \alpha_{k,m}^{p}(\boldsymbol{u}_k) - \alpha_{k,m}(\boldsymbol{u}_k, \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}_k) \right\|^2 \\
&\text{s.t.} \quad \dot{\boldsymbol{x}}_k = \boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}) \\
&\qquad \boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k) - \boldsymbol{c}_k \\
&\qquad \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta} \in \boldsymbol{\Theta}_0 \\
&\qquad \left\| \epsilon^T \right\|_{\infty} \leq \epsilon_{max}
\end{aligned} \tag{5.13}$$

where w_m^{α} is a normalizing weight and the superscript p denotes the approximated cost-function derivative (5.12) estimated from plant measurements.

5.3.2. Consideration of Information from Prior Experiments

Besides the gradient measurements that can be acquired by perturbing the plant at operating point $\boldsymbol{u}_k$ (5.12), additional cost-function measurements are already available from past experiments. Let us define a vector whose elements are the differences between the measured cost at the current operating point $\boldsymbol{u}_k$ and past ones as follows:

$$\Delta\boldsymbol{\Phi}_k = \begin{bmatrix} \phi_k - \phi_{k-1} & \phi_k - \phi_{k-2} & \cdots & \phi_k - \phi_{k-n_b-1} \end{bmatrix}^T \tag{5.14}$$

where n_b is the number of past operating points at which experiments have been performed and whose measurements are available. Similarly, we define a matrix containing the differences between the current and past decision variables as:

$$\Delta\boldsymbol{\mathcal{U}}_k = \begin{bmatrix} \boldsymbol{u}_k - \boldsymbol{u}_{k-1} & \boldsymbol{u}_k - \boldsymbol{u}_{k-2} & \cdots & \boldsymbol{u}_k - \boldsymbol{u}_{k-n_b-1} \end{bmatrix}^T \tag{5.15}$$

Following the finite difference approach (5.12), for any two operating points k and $k-l$ with $l \in \{1, \ldots, n_b - 1\}$, we can define the following normalized cost-difference coefficient:

$$\beta_{k,l} = \frac{\Delta \boldsymbol{\Phi}_{k,l}}{\|\Delta \boldsymbol{\mathcal{U}}_{k,l}\|} = \frac{\phi_k - \phi_{k-l}}{\|\boldsymbol{u}_k - \boldsymbol{u}_{k-l}\|} \tag{5.16}$$

Due to the uncertainty in the measured cost, we are ultimately interested in considering only past operating points that are sufficiently far away. In other words, we want to reduce the effect of gradient uncertainty when considering past cost-function evaluations. For that reason, from the available coefficients of previous experiments (5.16), we select only the ones for which the increase in predicted cost is beyond the magnitude of the measurement noise. Accordingly, we solely consider the points belonging to the following set:

$$\boldsymbol{\mathcal{L}}_{\boldsymbol{\varepsilon}} = \left\{ l \in \{1, \ldots, n_\beta\} \,\middle|\, 1 - \frac{\phi(\boldsymbol{u}_k)}{\phi(\boldsymbol{u}_{k-l})} \geq \varepsilon_\phi \right\} \tag{5.17}$$

where $n_\beta \leq n_b - 1$ describes the maximum number of past points to be considered and the bound ε_ϕ determines the minimum deviation in the inputs from the current operating point. This limit can be estimated from cost-function measurements as follows:

$$\varepsilon_\phi = \frac{\sigma_\phi}{\mu_\phi} \tag{5.18}$$

where σ_ϕ is the estimated standard deviation of the measurement noise and μ_ϕ the average cost magnitude. Hence, the minimum range for operating points considered from past experiments is controlled by the increase in the predicted cost corresponding to the magnitude in the expected standard deviation of the measurement noise.

It is important to point out that, instead of a fixed number of past operating points n_β to be considered in (5.17), it is also possible to determine the maximum permissible range by checking for which operating point $\boldsymbol{u}_{k-l}$ the corresponding set-based constraints $\boldsymbol{\mathcal{Y}}_{k-l}$ can no longer be satisfied

with the current set of parameter values (i.e. $\boldsymbol{y}(\boldsymbol{u}_{k-l}, \boldsymbol{\theta}_k) \notin \boldsymbol{\mathcal{Y}}_{k-l}$). In that sense, the minimum deviation is still given by the inequality in (5.17), however, the maximum deviation is determined based on the range for which the model can provide an adequate output prediction.

5.3.3. Design of New Experiments

In addition to using past cost measurements as outlined above, we propose to use optimal DoE to acquire future cost information. For the latter, the goal is to identify future operating points for gradient experiments that are more informative in terms of cost information instead of the fixed perturbations done at current k as done in our earlier studies as per (5.12). As before, these experiments are run in addition to the experiments conducted at the current optimal input determined by the model-based optimization (5.10). Thus, the goal of the proposed experimental design is to replace the experiments involving fixed perturbations in the direction of each decision variable with experiments involving perturbations that are more informative as per an experimental design criterion.

To quantify the level of information, we first define a parametric sensitivity matrix $\boldsymbol{S}_\beta$ of the coefficients of past experiments (5.16), whose elements are defined as follows:

$$\boldsymbol{S}_\beta = \left[\frac{\partial \beta_{k,l}}{\partial \theta_j}\right] \quad \begin{matrix} \forall l \in \boldsymbol{\mathcal{L}}_{\boldsymbol{\varepsilon}} \\ \forall j \in \{1, \ldots, n_\theta\} \end{matrix} \tag{5.19}$$

Using the sensitivity matrix, let us define a D-optimal design criterion (Franceschini & Macchietto, 2008) which seeks to minimize the following measure of the parameter covariance matrix:

$$\psi = \det\left(\boldsymbol{V}_\theta\right) = \det\left(\left[\boldsymbol{S}_\beta^T \boldsymbol{\Sigma}_\beta^{-1} \boldsymbol{S}_\beta\right]^{-1}\right) \tag{5.20}$$

where the measurement error matrix $\boldsymbol{\Sigma}_\beta$ of the gradient measurements can be obtained from the

cost-function measurement noise. The variance of the gradient measurement is therefore given as follows:

$$\sigma^2_{\beta_{k,l}} = \frac{2\sigma^2_\phi}{\|\boldsymbol{u}_k - \boldsymbol{u}_{k-l}\|^2} \tag{5.21}$$

where we assume that the measurement noise in the cost, necessary for the gradient estimation, remains unchanged and is uncorrelated.

Thus, the aim of the experimental design is primarily to find the plant perturbation vectors which provide information that will complement the information already gained from past experiments. To this end, we note that the gradient estimator in (5.12) can also be formulated for directions other than the directions associated with the individual decision variables as follows:

$$\gamma_{k,q}(\boldsymbol{v}_q) = \frac{\phi(\boldsymbol{u}_k + \boldsymbol{v}_q) - \phi(\boldsymbol{u}_k)}{\|\boldsymbol{v}_q\|} \tag{5.22}$$

where $q = 1, \ldots, n_{DOE}$ presents the number of plant perturbations at each operating point implemented to acquire the gradient information. The vector $\boldsymbol{u}_k + \boldsymbol{v}_q$ presents a perturbation of the plant in the neighbourhood of $\boldsymbol{u}_k$. The experimental design goal is to select the appropriate perturbation vectors $\boldsymbol{v}_q$ so as to increase parameter precision when performing the fitting of measured gradients. Similar to (5.21), the variance of (5.22) can be estimated by:

$$\sigma^2_{\gamma_{k,q}} = \frac{2\sigma^2_\phi}{\|\boldsymbol{v}_q\|^2} \tag{5.23}$$

Notice that the gradient variance approaches infinity for $\boldsymbol{v}_q \to \boldsymbol{0}$. For that reason, it is desired to introduce a minimum distance when designing new experiments in order to reduce the estimated variance of the gradient measurement (5.23). At the same time, there is often a cost associated with performing new experiments. In other words, experiments for gradient information should not be performed too far away from the optimum as it could result in a significant deterioration (increase) in cost. Therefore, to implement a minimum step-size due to gradient uncertainty while

limiting the distance from the current optimum to avoid an increase in cost, we propose to enforce the following equality when designing new experiments:

$$1 - \frac{\phi(\boldsymbol{u}_k)}{\phi(\boldsymbol{u}_k + \boldsymbol{v}_q)} = \varepsilon_\phi \tag{5.24}$$

where the ε_ϕ is defined in (5.18). Notice that this bound is similar to the inequality used for past operating points in (5.17), but in this case the equality is used to obtain a lower and upper bound for the step-size.

Finally, using the D-optimality design criterion in (5.20), we can formulate a sequential procedure to select the most informative input perturbations as follows:

1. Initialize by setting $q = 0$, $\boldsymbol{S}_\phi^q = \boldsymbol{S}_\beta$ and $\boldsymbol{\Sigma}_\phi^q = \boldsymbol{\Sigma}_\beta$.

2. Based on the considered prior experiments, find the perturbation which provides the most additional information as per the D-optimality criterion:

$$\begin{aligned}
\boldsymbol{v}_q \quad = \quad & \arg\min_{\boldsymbol{v}} \det\left(\left[\boldsymbol{S}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{S}\right]^{-1}\right) \\
\text{s.t.} \quad & \boldsymbol{S} = \begin{bmatrix} \boldsymbol{S}_\phi^q \\ \boldsymbol{s}_\gamma \end{bmatrix}, \; \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_\phi^q & \boldsymbol{0} \\ \boldsymbol{0} & \sigma_\gamma^2 \end{bmatrix} \\
& \boldsymbol{s}_\gamma = \begin{bmatrix} \frac{\partial \gamma}{\partial \theta_1} & \cdots & \frac{\partial \gamma}{\partial \theta_{n_\theta}} \end{bmatrix} \\
& \gamma(\boldsymbol{v}) = \frac{\phi(\boldsymbol{u}_k + \boldsymbol{v}) - \phi(\boldsymbol{u}_k)}{\|\boldsymbol{v}\|} \\
& 1 - \frac{\phi(\boldsymbol{u}_k)}{\phi(\boldsymbol{u}_k + \boldsymbol{v})} = \varepsilon_\phi
\end{aligned} \tag{5.25}$$

3. Update the sensitivity matrix $\boldsymbol{S}_\phi^{q+1} = \boldsymbol{S}$, $\boldsymbol{\Sigma}_\phi^{q+1} = \boldsymbol{\Sigma}$ and set $q = q + 1$.

4. Go back to step 2 until the number of gradient measurements n_{DOE} is reached.

Thus, using the outlined procedure, we are essentially seeking input perturbations around $\boldsymbol{u}_k$ that

increase the available information content given by (5.20).

5.3.4. Extended Gradient Correction

After carrying out the set of cost-function measurements from the optimal plant perturbations provided by the procedure in (5.25), the standard gradient correction method from (5.13) can be extended to consider the normalized differences to cost-function measurements of past operating points as well as gradient measurements resulting from the experimental design as follows:

$$\begin{aligned}
\Delta\boldsymbol{\theta}_k \quad = \quad & \arg\min_{\Delta\boldsymbol{\theta}} \sum_{q=1}^{n_{DOE}} w_q^{\gamma} \left\| \gamma_{k,q}^{p}(\boldsymbol{u}_k) - \gamma_{k,q}(\boldsymbol{u}_k, \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}_k) \right\|^2 \\
& + \sum_{l \in \boldsymbol{\mathcal{L}_\varepsilon}} w_l^{\beta} \left\| \beta_{k,l}^{p}(\boldsymbol{u}_k) - \beta_{k,l}(\boldsymbol{u}_k, \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}_k) \right\|^2 \\
\text{s.t.} \quad & \dot{\boldsymbol{x}}_k = \boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}) \\
& \boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k) - \boldsymbol{c}_k \\
& \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta} \in \boldsymbol{\Theta}_0 \\
& \left\| \boldsymbol{\epsilon}_k^T \right\|_{\infty} \leq \epsilon_{max}
\end{aligned} \tag{5.26}$$

In this way, we simultaneously consider available information from past batch runs and valuable information from specifically designed experiments. This results in more robustness to gradient uncertainty and an improved parameter precision as further illustrated by the case studies at the end of this chapter.

5.4. Results and Discussion

The proposed design of experiments methodology is illustrated in the following using three different case studies. While first study consists of a simple synthetic process, the remaining studies

describe two cell culture processes.

5.4.1. Synthetic Batch Process

This simple case study provides an illustration of the algorithm for the simultaneous identification and optimization in combination with the proposed design of experiments approach. The synthetic batch process under investigation is defined by the following ODE:

$$\frac{dy_p}{dt} = \beta_1 u - \beta_2 y_p + \beta_3 t \tag{5.27}$$

where $y_p \in \mathbb{R}$ describes the measured process output, $u \in \mathbb{R}$ the manipulated input and β_1, β_2 and β_3 the process parameters. We assume that measurements are taken along the duration of the batch at particular sampling times given by $t_i \in \{t_1, \ldots, t_f\}$, where t_f represents the final batch time. In addition to the process output (5.27), a cost function to be minimized is given by:

$$\phi(u, y_p(t_f)) = \gamma u^2 - y_p(u, t_f) \tag{5.28}$$

With the additional cost-function parameter γ. In the following we assume that ϕ can measured for each input u. The values of the process parameters are given in table 5.1.

β_1	β_2	β_3	γ
1	2	0.3	0.05

Table 5.1.: Process parameter values for the synthetic batch process.

In the following, we assume that the actual process representation (5.27) is unknown and that only a model of the process is available. However, due to a lack of knowledge about the exact behavior, there exist a discrepancy between model and process. The known process model is defined as

follows:

$$\frac{dy}{dt} = \hat{\beta}_1 u - \hat{\beta}_2 y \tag{5.29}$$

Notice that model-plant mismatch is deliberately introduced by eliminating the time-dependent term $\beta_3 t$ in (5.27). Furthermore, the available cost-function used in the optimization is given by:

$$\phi(y(u, \boldsymbol{\theta}, t_f), u, \boldsymbol{\theta}) = \hat{\gamma} u^2 - y(u, \boldsymbol{\theta}, t_f) \tag{5.30}$$

where the model parameters $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\gamma}$ are expressed by the vector $\boldsymbol{\theta} = [\ \hat{\beta}_1 \quad \hat{\beta}_2 \quad \hat{\gamma}\ ]^T$. Following the identification step outlined in (5.6) and the subsequent gradient correction (5.7), for each input u_k, we obtain the adapted set of parameters $\boldsymbol{\theta}'_k$. Using the updated parameters, the overall goal of the run-to-run optimization is to iteratively minimize the cost-function (5.30) so as to find the optimal input u^*, which minimizes the actual cost-function (5.28). The problem to be solved in each iteration k can therefore be stated as follows:

$$\begin{aligned} u_{k+1} \quad &= \quad \arg\min_u \phi\left(y(u, \boldsymbol{\theta}'_k, t_f), u, \boldsymbol{\theta}'_k\right) \\ &\text{s.t.} \quad (5.29) \text{ and } (5.30) \\ &\qquad u^L \leq u \leq u^U \end{aligned} \tag{5.31}$$

The model parameters and initial conditions for this case study are given in table 5.2. Here u_0 presents the input for the initial batch run.

$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\gamma}$	y_0	u_0
4	2	0.05	8	20

Table 5.2.: Initial conditions and model parameters.

5.4.1.1. Run-to-Run Optimization Results

In the following, we compare the results of using the extended gradient correction (5.26) to case of the standard gradient correction (5.13). The upper bound of the truncation error was selected to be $\epsilon_{max} = 0.01$, while the standard deviation of the measurement noise σ_ϕ corresponds to $10\,\%$ of the magnitude in the cost. Although the parameters $\hat{\beta}_1$ and $\hat{\beta}_2$ already provide enough degrees of freedom for an adequate gradient correction to achieve convergence to the process optimum, we opted to also include the cost-function parameter $\hat{\gamma}$ when performing the gradient correction. Including parameter $\hat{\gamma}$ results in a higher sensitivity to gradient measurement uncertainty with respect to which the extended gradient correction should provide more robustness. Furthermore, the number of past operating points considered for the experimental design and the extended gradient correction was limited to five, i.e. $n_\beta = 5$.

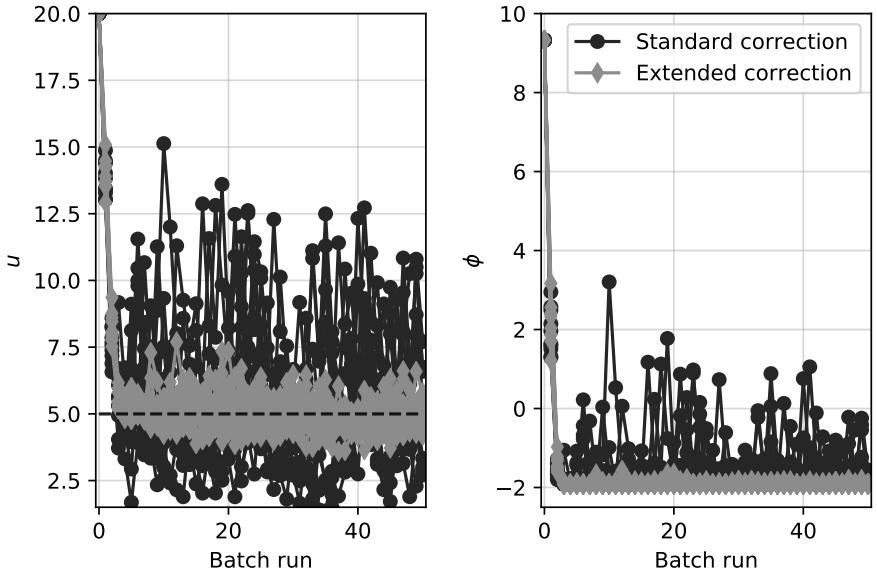

Figure 5.1.: Comparison of convergence results using the standard gradient correction and extended correction with DoE. The left graph shows the optimal input prediction over 10 noise realizations, while the right graph illustrates the corresponding cost.

The results are illustrated in figure 5.1, where the left graph is showing the convergence for the two gradient correction methods for all 10 different noise realizations. As shown in the figure, by using the proposed method, it is possible to significantly reduce the uncertainty in the prediction of the next optimal input around the optimum. This is also evident from the corresponding cost evaluations, shown on the right graph in figure 5.1, where the proposed method results in a significant reduction in deviation from the optimal cost.

Method	**IAE**	σ_u
Standard Gradient Correction (5.13)	1.38	6.41
Extended Gradient Correction (5.26)	0.41	0.38

Table 5.3.: Integral absolute error (IAE) and variance of predicted inputs u.

To quantify the performance, table 5.3 provides the integral absolute error (IAE) and the variance in the predicted optimal input σ_u for the gradient correction methods. This confirms the very significant improvement in performance since the proposed method leads to a reduction of $70\,\%$ and $94\,\%$ in IAE and σ_u respectively.

5.4.2. Penicillin Process Case Study

The case study under investigation is of a fed-batch penicillin process. The following set of equations define the process simulator (Birol et al., 2002):

$$\frac{dX}{dt} = \left(\frac{\mu_X S X}{K_X X + S}\right) - \frac{X}{V}\frac{dV}{dt} \tag{5.32}$$

$$\frac{dP}{dt} = \left(\frac{\mu_P S X}{K_P + S + \frac{S^2}{K_I}}\right) - K_H P - \frac{P}{V}\frac{dV}{dt} \tag{5.33}$$

$$\begin{aligned}\frac{dS}{dt} = &-\left(\frac{1}{Y_{X/S}}\frac{\mu_X S X}{K_X X + S}\right) - \left(\frac{1}{Y_{P/S}}\frac{\mu_P S X}{K_P + S + \frac{S^2}{K_I}}\right) \\ &-m_X X + \frac{F s_f}{V} - \frac{S}{V}\frac{dV}{dt}\end{aligned} \tag{5.34}$$

$$\frac{dV}{dt} = F - V 6.226 \cdot 10^{-4} \tag{5.35}$$

where X is the biomass, P is the concentration of penicillin, S is the concentration of substrate and V the volume in the reactor. The constants are defined as follows: μ_X is the specific growth rate of biomass, μ_P is the specific rate of penicillin production, K_X and K_P are saturation constants, K_I is a substrate inhibition constant, K_H is a constant representing the rate of consumption of penicillin by hydrolysis, $Y_{X/S}$ and $Y_{P/S}$ are the yields per unit mass of substrate for the biomass and penicillin respectively, m_X is the consumption rate of substrate for maintaining the biomass, F is the constant feed rate and s_f represents the concentration of substrate in the feed.

The simulator (5.32) – (5.35) is used to produce in silico experimental data for model fitting and gradient correction. For that purpose, $10\,\%$ measurement noise as well as stochastic disturbances in initial biomass and substrate concentrations are realized. Based on the in silico measurements, a model is calibrated and utilized for the purpose of run-to-run optimization. Model-plant mismatch

is intentionally introduced by assuming a lack of knowledge about the hydrolysis term in the penicillin equation. Therefore, the model used in the optimization scheme is defined by (5.32), (5.34), (5.35) and:

$$\frac{dP}{dt} = \left(\frac{\mu_P S X}{K_P + S + \frac{S^2}{K_I}} \right) - \frac{P}{V}\frac{dV}{dt} \tag{5.36}$$

The goal of this run-to-run study is the maximization of penicillin at the end of the batch time. The available decision variables are the initial substrate concentration S_0 and constant feed rate F. Accordingly, we can formulate the objective as follows:

$$\begin{aligned} \min_{S_0,F} \quad & -P\left(\boldsymbol{x}, \boldsymbol{\theta}, S_0, F, t_f\right) \\ \text{s.t.} \quad & (5.32) \text{ and } (5.34) - (5.36) \\ & V(\boldsymbol{x}, \boldsymbol{\theta}, S_0, F, t_f) \leq V_{max} \end{aligned} \tag{5.37}$$

where a constraint on the volume of the reactor is given by $V_{max} = 120\,\mathrm{l}$. The initial values used for the first operating point are given in table 5.4, where S_0 and F are the decision variables to be determined by the model-based optimization.

Biomass conc. (X_0)	$0.1\,\mathrm{g/l}$
Substrate conc. (S_0)	$0.1\,\mathrm{g/l}$
Product conc. (P_0)	$0\,\mathrm{g/l}$
Volume (V_0)	$100\,\mathrm{l}$
Input Feed (F)	$0.04\,\mathrm{l/h}$

Table 5.4.: Initial batch conditions.

The initial values of the model parameters are given in table 5.5. From these eight available model parameters, only a subset is selected for performing the model update (fitting of predicted outputs to measurements) and gradient correction (fitting predicted gradients to measured ones) to avoid overfitting and sensitivity to parameter correlation. The optimal choice of a suitable subset of

parameters for updating has been addressed in Hille et al. (2017). Accordingly, the parameters μ_X, μ_P and m_X are used for model update and gradient correction.

μ_X	K_X	μ_P	K_P	K_I	$Y_{X/s}$	$Y_{P/S}$	m_X
0.092	0.15	0.008	0.0002	0.1	0.45	0.9	0.014

Table 5.5.: Initial model parameter values.

For the standard gradient correction, the gradients of the cost function are estimated by perturbing the plant in the directions of each of the decision variables as shown in (5.12). In this case, fixed step sizes of $\Delta S_0 = 2\,\mathrm{g/l}$ and $\Delta F = 0.5\,\mathrm{l/h}$ are used for the initial substrate concentration and the constant flow rate respectively. As the optimal fixed flow rate is obtained within one iteration (due to the constraint on the volume), we used the proposed approach for improving the predictions of the optimal initial substrate concentration. Accordingly, in this case, the optimal step size (or plant perturbation) ΔS_0 was determined based on the available past experimental data and DOE methodology (5.25).

5.4.2.1. Results

To compare the performance of the proposed approach, we conducted 10 run-to-run simulations for the following two cases:

1. Only the local gradient is used for correction (5.13).
2. The extended gradient correction is used involving past and current gradients using DOE (5.26).

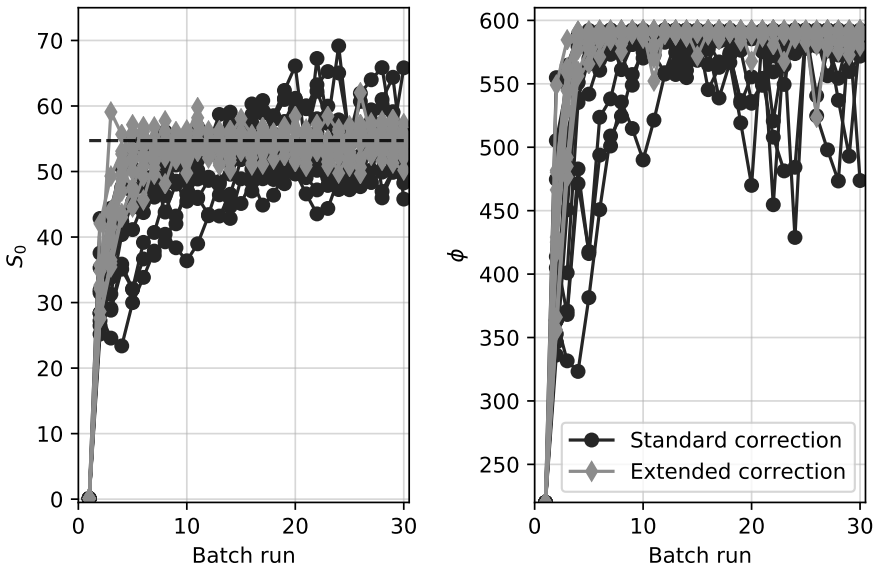

Figure 5.2.: Performance comparison of the standard correction and the proposed experimental design methodology.

The run-to-run optimization results are shown in figure 5.2, where the left-hand graph shows the comparison in terms of the manipulated variable S_0. The bound on the relative truncation error in (5.26) was selected to be $\epsilon_{max} = 0.03$ and thus it is smaller than the level of measurement noise of $10\,\%$. The proposed approach shows a speed-up in convergence to the process optimum corresponding to a ca. 28% improvement in the IAE. It is evident from this result that by considering cost information of previous experiments, it is possible to reduce the effect uncertainty in the local gradient measurement. Furthermore, the proposed approach leads to a 61% reduction in the variability in the predicted optimal input, characterized by the more precise confidence regions. In addition, the right-hand graph in figure 5.2 shows the cost-function evaluations corresponding to the optimal inputs from the left graph. The proposed approach thus leads to far superior prediction capabilities, especially around the optimum.

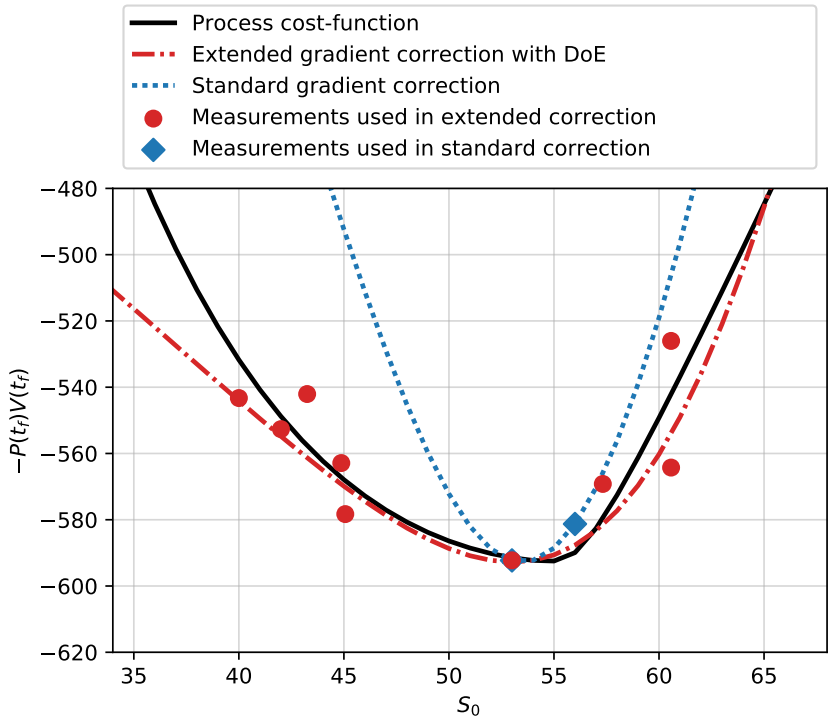

Figure 5.3.: Example of the predicted cost-functions for S_0 in the neighbourhood of the process optimum.

This fact is further supported by the cost-function prediction as shown in Figure 5.3. The use of past cost function measurements and DoE results in a decrease in the SSE of the cost-function fitting around the optimum of more than 90% when compared to the cost predicted without past gradients and DoE. Thus, when the identification of a model around the optimum is an important goal for the user beyond the optimization goal, the proposed gradient correction provides a significantly more accurate model.

5.4.3. CHO Cell Cultivation Process

The second case study describes a cell cultivation process of a CHO cell line (Aghamohseni et al., 2014). A dynamic metabolic model describes the dynamics of MAb and other extra cellular metabolites during a batch experiment. The process simulator is defined by the following set

of ODEs:

$$\frac{dfgr}{dt} = -K_{11}\frac{fgr}{1+\left(\frac{[Glc]}{K_{12}}\right)} \tag{5.38}$$

$$\frac{dX_v}{dt} = \mu \cdot fgr \cdot X_v \left(\frac{[Glc]\,[Gln]}{(K_{21}+[Glc])\,(K_{26}+[Gln])} \frac{1}{1+\left(\frac{[Amm]}{K_{23}}\right)} \frac{1}{1+\left(\frac{[Glc]}{K_{22}}\right)} \right) - \frac{dX_d}{dt} \tag{5.39}$$

$$\frac{dX_d}{dt} = k_d\,(1-fgr)\,X_v \left(\frac{1}{1+\left(\frac{K_{24}}{[Amm]}\right)^n} + \frac{K_{25}}{[Glc]} \right) \tag{5.40}$$

$$\frac{d\,[Glc]}{dt} = -\left(\frac{K_{31}\,[Glc]\,[Gln]}{(K_{32}+[Glc])\,(K_{36}+[Gln])} \frac{K_{33}\,[Glc]}{(K_{34}+[Glc])} \right) X_v - K_{35}X_v \tag{5.41}$$

$$\frac{d\,[Gln]}{dt} = -\left(\frac{K_{41}\,[Glc]\,[Gln]}{(K_{42}+[Glc])\,(K_{43}+[Gln])} \right) X_v \tag{5.42}$$

$$\frac{d\,[Lac]}{dt} = -\left(\frac{K_{51}\,[Glc]}{(K_{52}+[Glc])} \frac{d\,[Glc]}{dt} \right) X_v \tag{5.43}$$

$$\frac{d\,[Asn]}{dt} = -\left(\frac{K_{61}\,[Asn]}{(K_{62}+[Asn])} \right) X_v \tag{5.44}$$

$$\frac{d\,[Asp]}{dt} = \left(\frac{K_{61}\,[Asn]}{(K_{62}+[Asn])} + \frac{K_{63}\,[Glc]\,[Gln]}{(K_{64}+[Glc])\,(K_{65}+[Gln])} - \frac{K_{71}\,[Asp]}{(K_{72}+[Asp])} \right) X_v \tag{5.45}$$

$$\frac{d\left[Ala\right]}{dt} = \left(\frac{K_{63}\left[Glc\right]\left[Gln\right]}{\left(K_{64}+\left[Glc\right]\right)\left(K_{65}+\left[Gln\right]\right)} - \frac{K_{81}\left[Ala\right]}{\left(K_{82}+\left[Ala\right]\right)}\right) X_v$$

$$\frac{d\left[Amm\right]}{dt} = -K_{91}\frac{d\left[Gln\right]}{dt} + K_{92}\left(\frac{K_{61}\left[Asn\right]}{\left(K_{62}+\left[Asn\right]\right)} + \frac{K_{71}\left[Asp\right]}{\left(K_{72}+\left[Asp\right]\right)}\right. \tag{5.46}$$

$$\left. + \frac{K_{81}\left[Ala\right]}{\left(K_{82}+\left[Ala\right]\right)}\right) X_v \tag{5.47}$$

$$\frac{d\left[Mab\right]}{dt} = \left(K_{101}+K_{102}\left[Gln\right]\right) X_v \tag{5.48}$$

where fgr represents the fraction of growing cells, X_v the viable cell density (VCD) and X_d the dead cell density (DCD). Furthermore, the concentrations of metabolites are defined as follows: $[Glc]$ - glucose, $[Gln]$ - glutamine,$[Lac]$ - lactate, $[Asn]$ - asparagine, $[Asp]$ - aspartate, $[Ala]$ - alanine, $[Amm]$ - ammonia and $[Mab]$ – MAb. The kinetic parameters are given by K_{11} to K_{102}. As in the case of the penicillin process, the simulator (5.38) – (5.48) is used to generate in silico measurements of the process outputs and gradients. For uncertainty in measurements, additive Gaussian noise of $10\,\%$ of the average output values is assumed as well as stochastic disturbances in initial batch conditions. Gradients are estimated using finite differences by running an additional batch experiment and then calculating the difference in the cost over the perturbation step size. The initial media composition for the CHO study are given in table 5.6.

Fraction of growing cells (fgr)	1
Viable Cell Density (X_v)	$0.0024\,10^6 \mathrm{cells/ml}$
Dead Cell Density (X_d)	$0.0001\,10^6 \mathrm{cells/ml}$
Glucose ($[Glc]$)	$60\,\mathrm{mmol/l}$
Glutamine ($[Gln]$)	$3.48\,\mathrm{mmol/l}$
Lactate ($[Lac]$)	$0.22\,\mathrm{mmol/l}$
Ammonia ($[Amm]$)	$0.71\,\mathrm{mmol/l}$
Aspartate ($[Asp]$)	$1.22\,\mathrm{mmol/l}$
Alanine ($[Ala]$)	$0.66\,\mathrm{mmol/l}$
Asparagine ($[Asn]$)	$0.67\,\mathrm{mmol/l}$
Monoclonal Antibodies ($[Mab]$)	$0.1\,\mu\mathrm{g/ml}$

Table 5.6.: Initial batch conditions

5.4.3.1. Model-Plant Mismatch

In addition to the process simulator that is used to generate in silico data, we also define a model of the process that is used for calibration and optimization purposes. To introduce structural mismatch between the model and process, we assume that glutamine ($[Gln]$) measurements are not available and that its effects on the cell metabolism are unknown. This is motivated by the fact that the concentrations of many media components are frequently not measured throughout the cultivation and that their effects on the cell metabolism are often not well understood. Therefore, the model equations are similar to the simulator, with the difference that all the parts containing a product involving glutamine concentration are made a function of only the glucose concentration and that all dependencies on glutamine alone are eliminated. This results in a structural error for which the remaining model parameters must compensate for his mismatch as to achieve a precise model fitting. Furthermore, this impairs the prediction capability in terms of the location of the process optimum. Thus, the model of the process used for identification and optimization is described by

the following ODEs:

$$\frac{dfgr}{dt} = -K_{11}\frac{fgr}{1+\left(\frac{[Glc]}{K_{12}}\right)} \tag{5.49}$$

$$\frac{dX_v}{dt} = \mu \cdot fgr \cdot X_v \left(\frac{[Glc]}{(K_{21}+[Glc])}\frac{1}{1+\left(\frac{[Amm]}{K_{23}}\right)}\frac{1}{1+\left(\frac{[Glc]}{K_{22}}\right)}\right) - \frac{dX_d}{dt} \tag{5.50}$$

$$\frac{dX_d}{dt} = k_d\left(1-fgr\right)X_v\left(\frac{1}{1+\left(\frac{K_{24}}{[Amm]}\right)^n}+\frac{K_{25}}{[Glc]}\right) \tag{5.51}$$

$$\frac{d\left[Glc\right]}{dt} = -\left(\frac{K_{33}\left[Glc\right]}{(K_{34}+[Glc])}\right)X_v - K_{35}X_v \tag{5.52}$$

$$\frac{d\left[Lac\right]}{dt} = -\left(\frac{K_{51}\left[Glc\right]}{(K_{52}+[Glc])}\frac{d\left[Glc\right]}{dt}\right)X_v \tag{5.53}$$

$$\frac{d\left[Asn\right]}{dt} = -\left(\frac{K_{61}\left[Asn\right]}{(K_{62}+[Asn])}\right)X_v \tag{5.54}$$

$$\frac{d\left[Asp\right]}{dt} = \left(\frac{K_{61}\left[Asn\right]}{(K_{62}+[Asn])}+\frac{K_{63}\left[Glc\right]}{(K_{64}+[Glc])}-\frac{K_{71}\left[Asp\right]}{(K_{72}+[Asp])}\right)X_v \tag{5.55}$$

$$\frac{d\left[Ala\right]}{dt} = \left(\frac{K_{63}\left[Glc\right]}{(K_{64}+[Glc])}-\frac{K_{81}\left[Ala\right]}{(K_{82}+[Ala])}\right)X_v \tag{5.56}$$

$$\frac{d\left[Amm\right]}{dt} = K_{92}\left(\frac{K_{61}\left[Asn\right]}{(K_{62}+[Asn])}+\frac{K_{71}\left[Asp\right]}{(K_{72}+[Asp])}+\frac{K_{81}\left[Ala\right]}{(K_{82}+[Ala])}\right)X_v \tag{5.57}$$

$$\frac{d\left[Mab\right]}{dt} = K_{101}X_v \tag{5.58}$$

The goal of the run-to-run optimization is to maximize the amount of MAb at the end of the batch time of t_f over the batch time length. The decision variables in this case are the initial concentration of the main substrate glucose $[Glc]_0$ and final batch time t_f. Thus, we are seeking the initial glucose

concentration and final batch time for which the process is most cost effective. The optimization problem to be solved in each iteration is stated as:

$$\begin{aligned} \boldsymbol{u}_{k+1} \quad = \quad & \min_{[Glc]_0, t_f} -\frac{[Mab]}{t_f}\,(x, [Glc]_0, \boldsymbol{\theta}, t_f) \\ \text{s.t.} \quad & (5.49) - (5.58) \\ & 1 \leq t_f \leq 20 \end{aligned}$$

For the CHO case study, we assume that five batch experiments are performed at each operating point to generate output measurements and the process uncertainty bounds. To obtain an estimate of the cost gradient, a step size of $\Delta[Glc]_0 = 2\,\mathrm{mmol/l}$ is used for the initial glucose concentration as well as a step size of $\Delta t_f = 1\,\mathrm{d}$. Furthermore, the initial model parameters are given in table 5.7.

K_{11}	0.0140	K_{51}	0.2788
K_{12}	42.4918	K_{52}	1.8564
μ	3.2719	K_{61}	9.6146
K_{21}	53.6983	K_{62}	23.7654
K_{22}	10	K_{63}	0.5120
K_{23}	0.0405	K_{64}	8.5160
K_{24}	47.5529	K_{71}	3.2121
k_d	2.2	K_{72}	9.6238
n	0.2103	K_{81}	5.9924
K_{25}	12.9023	K_{82}	34.7341
K_{33}	36.1942	K_{92}	0.8071
K_{34}	55.1386	K_{101}	1.0023
K_{35}	0.0375		

Table 5.7.: Initial model parameters for the CHO model used in the run-to-run optimization.

The subset of parameters that is used in the gradient correction step is determined from the pa-

rameter selection outlined in Hille et al. (2017). For this purpose, we performed a selection of parameters which provide large parametric sensitivities with respect to the gradients of the cost-function. Consequently, parameters K_{11}, μ, K_{21}, K_{23}, K_{25}, K_{22} and K_{101} are used in the gradient correction step for this case study.

5.4.3.2. Results

Figure 5.4 illustrates the convergence for seven different noise realizations using the standard gradient correction. The gradient uncertainty strongly affects the performance leading to a significant variability between different realizations. This is especially evident around the optimum.

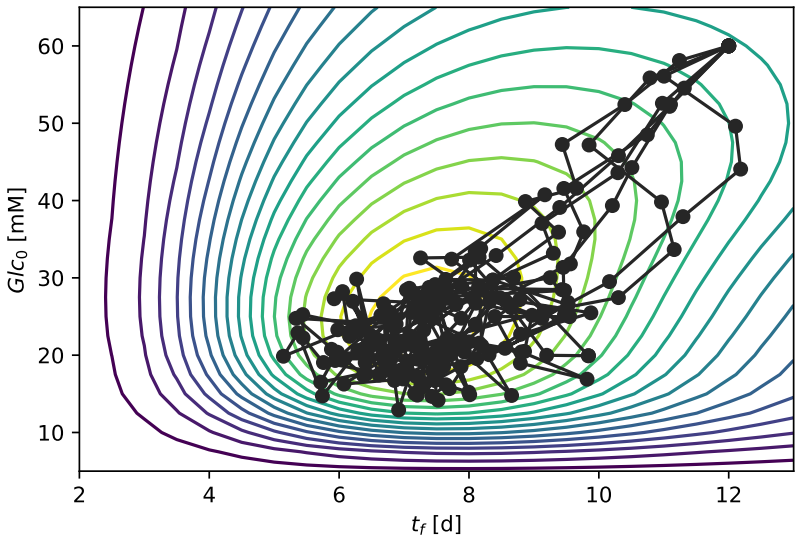

Figure 5.4.: Run-to-run optimization results with standard gradient correction.

On the other hand, figure 5.5 shows the results when using the extended gradient correction procedure. Again, five different noise realizations result in different trajectories towards the optimum. In this case however, the performance between different run-to-run optimizations is more similar.

More importantly, the proposed gradient correction method leads to a significantly better prediction in the neighbourhood of the optimum.

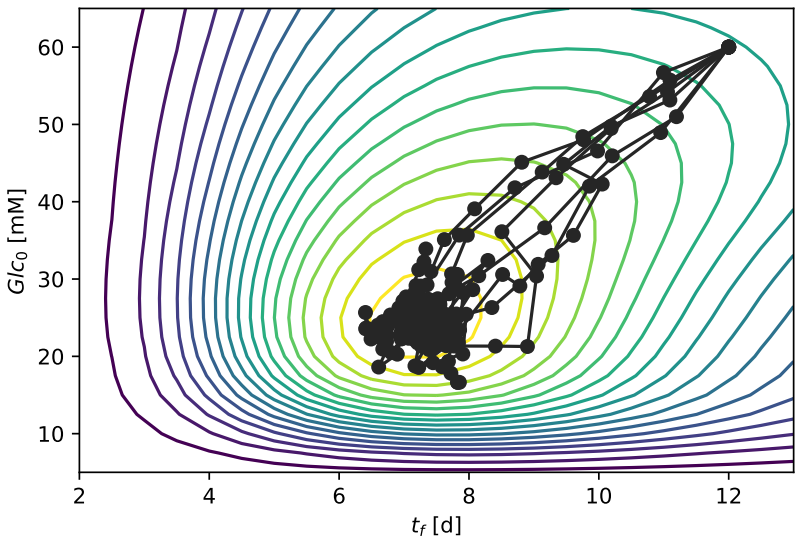

Figure 5.5.: Run-to-run optimization results using the extended gradient correction procedure.

To quantify this improvement, table 5.8 compares the two approaches in terms of the average IAE and the standard deviation of the manipulated variables over the seven simulations. Although - on average - the IAE is fairly similar between the two methods, the proposed method outperforms the standard gradient correction with regard to the standard deviation of the cost about the optimum thus further illustrating the robustness of the proposed method.

Method	**IAE**	σ_u
Standard Gradient Correction (5.13)	2.81	3.02
Extended Gradient Correction (5.26)	2.53	1.72

Table 5.8.: IAE and standard deviation of the averaged manipulated variables for the CHO case study.

5.5. Conclusions

When dealing with model-based run-to-run optimization under model-plant mismatch, it is vital to correct for errors in the predicted gradients of the cost-function and constraints. To achieve this, it is required to obtain estimates of the gradients by perturbing the plant around the current operating point. An experimental design methodology was presented, where the goal is to reduce the effect of gradient uncertainty on the convergence of the run-to-run procedure and to improve parameter precision to achieve a better prediction of the cost-function around the optimum. More specifically, based on available cost information from previous experiments, new experiments for gradient measurements are designed that are particularly valuable to complement the already available information. Based on the results from three case studies, we conclude that the proposed approach leads to smoother convergence of the run-to-run procedure and to a superior prediction of the cost-function, especially in the vicinity of the process optimum.

6. Run-to-Run Optimization Applied to Mammalian Cells in Perfusion Cultures

Overview

This chapter is based on a collaboration with an industrial partner pursued during my PhD that involved the development of a comprehensive model of a CHO cell culture system operated in perfusion mode. The comprehensive cell culture model developed here is considerably larger than the examples considered earlier in this thesis in terms of number of equations and parameters and therefore it served to test the ability of our run-to-run methods to deal with larger complexity. The industrial collaborator provided extensive data that was used to calibrate the model.

Due to a constant supply of fresh media and reactor outflow, perfusion processes can be used to avoid nutrient limitations and the build up of toxic byproducts. At the same time, a high cell density can be maintained through the use of a cell retention device which retains the cells within the system and thus prevents the loss of biomass. Hence, it is possible to maintain a steady-state in perfusion cultures, which can be beneficial for the production of recombinant proteins. The first part of this chapter presents the development of a dynamic metabolic model for the description of the dynamic behavior of CHO cells and extra cellular metabolites in a perfusion culture. Towards that goal, the established MFA (Metabolic Flux Analysis) methodology is adjusted for the application to perfusion processes. In the second part, the developed model is utilized as a case

study for a run-to-run optimization in the presence of model-plant mismatch. The methodology for simultaneous identification and optimization is used to deal with model-plant mismatch and to drive the process to the desired operating point. The optimization problem, specifically tailored for perfusion processes, is selected to take into account a cell specific perfusion rate.

6.1. Introduction

Throughout the recent decades, mammalian cells have been used as the primary platform for the production of recombinant therapeutic proteins that require human-like post-translational modifications. In particular, Chinese hamster ovary (CHO) cells are an established cell line for the industrial manufacturing of monoclonal antibodies (MAbs) (Fischer et al., 2015). Due to growing medical needs, pharmaceutical companies are faced with increasing pressure for a cost-efficient commercial production of MAbs (Farid, 2007; Butler & Meneses-Acosta, 2012). Increasing attention has been therefore paid to improve the production of antibodies through genetic engineering (Fischer et al., 2015), leading to higher recombinant protein yields (Majors et al., 2009). In addition to cell engineering, the optimization of the bioprocess operation and media design have also been widely used to improve the performance of CHO processes (Wurm, 2004).

The challenging design and operation of bioprocesses combined with their complex nature and extensive cost of experiments has motivated the development of macroscopic knowledge-driven models. The main purpose of these mechanistic models is to not only gain a better understanding of the processes under study, but also to enhance their performance (Yahia et al., 2015; Zamorano et al., 2013). Regarding the cultivation of CHO cells, metabolic flux analysis (MFA) has been successfully applied as a technique to derive metabolic models to describe the dynamics of the major metabolites (Gao et al., 2007). The MFA methodology is based on the metabolic reactions that are occurring within the cell. By assuming a steady state of intracellular metabolites an optimal flux profile can be determined which reflects the measured uptake and production rates of consumed and

secreted metabolites (Provost et al., 2006). In this way, MFA is used to determine significant intra cellular fluxes which are subsequently utilized to derive a set of macro reactions linking consumed nutrients to secreted by-products that should be included in the metabolic model.

In the past, dynamic metabolic models have been primarily used to describe batch and fed-batch operations (Dorka et al., 2009; Nolan & Lee, 2011). In contrast to their previous applications, in this work we investigate the implementation of dynamic metabolic models to describe and compare results obtained from bioreactor perfusion processes. The main difference of perfusion systems compared to typical fed-batch operations is that fresh media is continuously provided by a feeding rate, while the outflow of the reactor is regulated at a particular rate referred to as the harvest rate. Furthermore, the cells are separated from the rest of the outflow by means of a cell separation device and recirculated into the reactor. Thus, due to the constant feed of fresh media and the retaining of cells, it is possible to achieve high cell densities along with longer operation times (Chuppa et al., 1997; Clincke et al., 2013).

The first goal of the work in this chapter is therefore to investigate to which extent dynamic metabolic models are capable of describing mammalian cell cultivation in perfusion systems. To this end, a MFA has been performed using experimental data from perfusion experiments. Subsequently, the derived dynamic metabolic model is calibrated by performing a parameter estimation. In the second part of this work, the developed model is utilized for a run-to-run optimization case study of a perfusion system. In this case, it is assumed that there exist a discrepancy between the model and the process under study. In order to address the model error, it is necessary to implement a method that is robust to structural uncertainties. To this end, the method of simultaneous identification and optimization (Mandur & Budman, 2015b; Hille et al., 2017) presents a framework where the model parameters are adjusted to correct for errors in the predicted gradients of the cost-function and constraints. This is necessary for the model to accurately predict the optimality conditions of the process. The objective of the run-to-run optimization involves criteria such as the

cell specific perfusion rate and thus it is particularly targeted to perfusion processes. To facilitate the run-to-run optimization, the methodologies developed in chapters 3, 4 and 5 are implemented and compared to a previous version of the algorithm.

6.2. Model Development

6.2.1. Metabolic Flux Analysis

In recent years, MFA has been applied as a tool for systematic dynamic metabolic model development (Naderi et al., 2011) for cell culture processes. The idea behind the approach is to estimate the intracellular flux distribution by means of stoichiometric balances of intra- and extracellular reactions and measurements of extracellular metabolites. One of the fundamental assumptions of this analysis is that there is no accumulation of intracellular metabolites during the cultivation process (Provost et al., 2006). Based on the stoichiometric matrix of known intracellular reactions and following the calculation of the fluxes corresponding to each reaction, it is possible to reduce the complexity of the complete cell network to determine metabolites and fluxes which are important for explaining the distribution of carbon and nitrogen in the cultivation process. In such a way, insignificant fluxes can be eliminated leading to reduction of the network dimensionality to a set of macro-reactions, referred as "macro" since each represent more than one reaction. Then, each one of the macro-reactions can be described by kinetic expressions that are generally assumed to be of Monod-type.

As a first step of the MFA, a dynamic mass balance equation is obtained for each of the metabolites, which results in the following set of linear equations:

$$\frac{d\boldsymbol{\psi}(t)}{dt} = \boldsymbol{r}_m X_v(t) \tag{6.1}$$

where $\boldsymbol{\psi}(t) \in \mathbb{R}^{n_\psi}$ describes the time dependent vector of intra- and extracellular metabolite concentrations and $X_v(t) \in \mathbb{R}$ the total viable cell density. $\boldsymbol{r}_m \in \mathbb{R}^{n_\psi}$ is a vector of specific uptake/production rates, usually obtained during a particular phase of cell cultivation (i.e. typically exponential or post-exponential phases). Following the quasi steady-state assumption (no accumulation of intracellular metabolites), $\boldsymbol{r}_m$ is deemed to be constant during each phase and thus can be determined from the relation in (6.1). However, if operating modes other than batch are considered, it is necessary to adjust equation (6.1) for any feeding of media and outflow of the reactor. In this way, it is possible to predict the evolution of the metabolites' concentration with time as a function of changes in biomass and the metabolites' concentrations at previous times. In the case of a perfusion operation, we account for the effect of the time varying perfusion rate $P(t)$ and harvest rate $H(t)$ on each of the metabolite concentrations by modifying the left-hand side of equation (6.1) as follows:

$$\frac{\frac{d\boldsymbol{\psi}(t)}{dt} - \frac{P(t)}{V(t)}\boldsymbol{\psi}_{in} + \frac{H(t)}{V(t)}\boldsymbol{\psi}(t)}{X_v(t)} = \boldsymbol{r}_m \tag{6.2}$$

To obtain a smooth rate of change, the left-hand side of equation (6.2) can be discretized with respect to different regimes of perfusion/harvest rates and summed up to provide a cumulative production/consumption profile (Niu et al., 2013):

$$\boldsymbol{\Psi} = \sum_{k=1}^{n_t} \frac{\boldsymbol{\psi}(t_{k+1}) - \boldsymbol{\psi}(t_k) + \left[-\frac{P(t_k)}{V(t_k)}\boldsymbol{\psi}_{in} + \frac{H(t_k)}{V(t_k)}\frac{\boldsymbol{\psi}(t_{k+1})+\boldsymbol{\psi}(t_k)}{2}\right]\Delta t}{\frac{X_v(t_{k+1})+X_v(t_k)}{2}} \tag{6.3}$$

where the perfusion and harvest rates are constant for the time period of $\Delta t = t_{k+1} - t_k$. Typically, the perfusion rates are changed once per day until a desired steady-state has been achieved. A linear regression function $f_{lr}(\cdot)$ is then used to approximate the cumulative profile (6.3). From the corresponding slope we are able to obtain the rate of change of this cumulative profile as follows (Niu et al., 2013):

$$\frac{df_{lr}(\boldsymbol{\Psi}_i)}{dt} = r_{m,i} \tag{6.4}$$

With $i = 1, \ldots, n_\psi$. Accordingly, for each measured metabolite, the RHS of equation (6.4) can be evaluated to obtain the specific uptake and production rates to construct the vector $\boldsymbol{r}_m$. Given the vector $\boldsymbol{r}_m$ and the stoichiometric matrix of intracellular reactions related to external $\boldsymbol{R}$ and internal rates $\boldsymbol{S}$, it is possible to calculate the intracellular fluxes $\boldsymbol{j}$ by posing the following problem:

$$\begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{S} \end{bmatrix} \boldsymbol{j} = \begin{bmatrix} \boldsymbol{r}_m \\ \mathbf{0} \end{bmatrix} \tag{6.5}$$

As the system in (6.5) is generally undetermined, it is necessary to perform a least-squares minimization as follows:

$$\begin{aligned} \hat{\boldsymbol{j}} \quad &= \quad \arg\min_{\boldsymbol{j}} \left(\boldsymbol{r} - \boldsymbol{r}_m\right)^T \boldsymbol{W} \left(\boldsymbol{r} - \boldsymbol{r}_m\right) \\ &\text{s.t.} \quad \boldsymbol{r} = \boldsymbol{R}\boldsymbol{j} \\ &\qquad\quad \mathbf{0} = \boldsymbol{S}\boldsymbol{j} \end{aligned} \tag{6.6}$$

where $\boldsymbol{W}$ is a weighting matrix composed of the inverses of the measurement variances. The matrix relating the intracellular metabolites $\boldsymbol{S}$ is used to enforce the steady state condition of the intracellular metabolites. With the kernel of $\boldsymbol{S}$ defined as $\boldsymbol{K} = null(\boldsymbol{S})$, the solution to problem (6.6) has been derived in (Leighty & Antoniewicz, 2011):

$$\hat{\boldsymbol{j}} = \boldsymbol{K} \left(\boldsymbol{K}^T \boldsymbol{R}^T \boldsymbol{W} \boldsymbol{R} \boldsymbol{K}\right)^{-1} \boldsymbol{K}^T \boldsymbol{R}^T \boldsymbol{W} \boldsymbol{r}_m \tag{6.7}$$

By calculating the fluxes using (6.7), the goal is to eliminate insignificant fluxes to simplify the metabolic network. For example, fluxes that contribute less than 1% to the total fluxes are considered to be insignificant and can subsequently be removed from the reaction network (Naderi et al., 2011). Based on the network reduction, it is possible to formulate the macro reactions which involve the remaining significant intracellular reactions (macro-reactions are shown in Section 3).

Finally, a dynamic metabolic model is developed by assuming Monod-type kinetics for each of the extracellular metabolites.

6.2.2. Macro-Reactions

Once the cumulative profiles have been obtained by using (6.3), the specific uptake and production rates can be determined from (6.4). Following that, we calculated the profile of intracellular fluxes using (6.7). To this end, we used a matrix of intracellular reactions containing 27 metabolites and 35 reactions (Niu et al., 2013). For more information on the metabolic network, consult the supplementary material in the appendix A.1. It is important to point out that the matrix of stoichiometric coefficients includes a reaction for the glutamine synthesis as the cell line has been introduced with a glutamine synthetase (GS) gene.

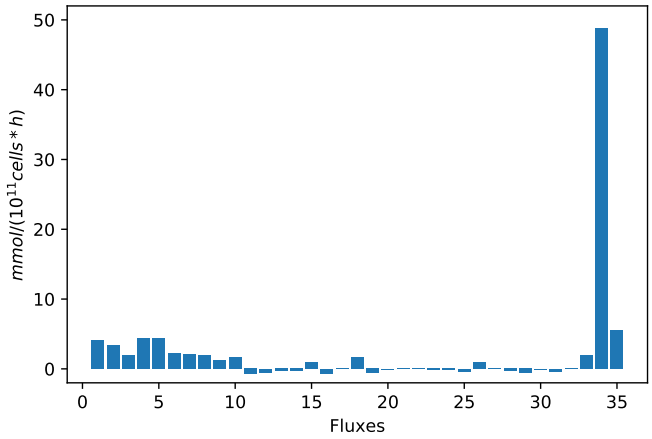

Figure 6.1.: Optimal flux distribution that explains the observed uptake/production rates.

By using eq. (6.7), it is possible to obtain the optimal flux distribution that can describe the observed uptake and production rates as shown in figure 6.1. Based on the most significant fluxes, we are able to construct the macro-reactions by relating substrates to products by combining reactions and including in these combinations all of the significant fluxes at least once (Naderi et al., 2011).

The result is the following macro-reactions shown in table 6.1, which are subsequently used to formulate the dynamic metabolic model.

#	Macro Reaction	Included Fluxes
1	$Glc + 2ADP \rightarrow 2Lac + 2ATP$	1,2,3
2	$Asn \rightarrow Asp + NH_3$	18
3	$Ser + NH_3 + CO_2 + NADH \rightarrow 2Gly + NAD^+$	19
4	$Ser + NADH \rightarrow Ala + NAD^+$	12, 15, 26
5	$Asp + 0.5Glc + 3NAD^+ + ADP \rightarrow Glu + 2CO_2 + 3NADH + ATP$	1, 2, 4, 5, 6, 16
6	$Glu + NH_3 \rightarrow Gln$	11
7	$Glc + 12NAD + 2ADP \rightarrow 6CO_2 + 12NADH + 2ATP$	1, 2, 4, 5, 6, 7, 8, 9, 10, 33
8	$Biomass$	34
9	IgG	35

Table 6.1.: Macro-reactions based on significant fluxes

The macro reactions in table 6.1 provide an understanding of the dominant intracellular reactions of this cell line. For example, it is evident that lactate is getting produced as a result of the pathway from glucose to the TCA cycle. Other significant reactions are the synthesis of glutamine from glutamate and ammonia as well as the production of aspartate and ammonia from asparagine. For most of the minor amino acids, which are not part of the macro reactions as per the MFA, we assumed a simple consumption term expressed by a Monod kinetic. We constructed dynamic balance equations as shown in the next subsection.

6.2.3. Model Equations

Based on the macro-reaction shown in table 6.1, a set of ODEs can be derived to describe the underlying dynamics. The macro-reactions are used to construct the equations of the main metabolites

whereas for the remaining minor amino acids we assumed that they are getting consumed without affecting the evolution of other metabolites. Furthermore, the bleeding rate for viable and dead cell bleeding is denoted by B, while the perfusion and harvest rate is denoted by P and H respectively.

$$\begin{aligned}
\frac{dX_v}{dt} &= \mu X_v \left(\frac{[Glc]}{(K_{21}+[Glc])} \frac{1}{1+\left(\frac{[Amm]}{K_{23}}\right)} \frac{1}{1+\left(\frac{[Lac]}{K_{25}}\right)} \right) \\
&\quad -k_d X_v^2 \left(\frac{1}{1+\left(\frac{K_{24}}{[Amm]}\right)^n} + \frac{K_{26}}{[Glc]} \right) - \frac{B}{V} X_v \qquad (6.8)
\end{aligned}$$

$$\frac{dX_d}{dt} = k_d X_v^2 \left(\frac{1}{1+\left(\frac{K_{24}}{[Amm]}\right)^n} + \frac{K_{26}}{[Glc]} \right) - K_{lys} X_d - \frac{B}{V} X_d \qquad (6.9)$$

$$\frac{d[Glc]}{dt} = -K_{41} X_v - \left(\frac{K_{42}[Glc]}{K_{43}+[Glc]} \right) X_v + \frac{P}{V}[Glc]_{in} - \frac{H}{V}[Glc] \qquad (6.10)$$

$$\frac{d[Lac]}{dt} = K_{51} X_v + \left(\frac{K_{52}[Glc]}{K_{43}+[Glc]} - \frac{K_{53}[Lac]}{K_{54}+[Lac]} \right) X_v - \frac{H}{V}[Lac] \qquad (6.11)$$

$$\begin{aligned}
\frac{d[Amm]}{dt} &= \left(\frac{K_{64}[Asn]}{K_{72}+[Asn]} - \frac{K_{62}[Amm][Glu]}{(K_{61}+[Amm])(K_{63}+[Glu])} \right. \\
&\quad \left. - \frac{K_{65}[Amm][Ser]}{(K_{133}+[Amm])(K_{134}+[Ser])} \right) X_v - \frac{H}{V}[Amm] \qquad (6.12)
\end{aligned}$$

$$\frac{d[Asn]}{dt} = -\left(\frac{K_{71}[Asn]}{K_{72}+[Asn]} \right) X_v + \frac{P}{V}[Asn]_{in} - \frac{H}{V}[Asn] \qquad (6.13)$$

$$\frac{d[Asp]}{dt} = \left(\frac{K_{81}[Asn]}{K_{72}+[Asn]} - \frac{K_{82}[Asn]}{K_{83}+[Asn]} \right) X_v + \frac{P}{V}[Asp]_{in} - \frac{H}{V}[Asp] \qquad (6.14)$$

$$\begin{aligned}
\frac{d[Ala]}{dt} &= K_{91} \left(\frac{K_{42}[Glc]}{K_{43}+[Glc]} + \frac{K_{53}[Lac]}{K_{54}+[Lac]} - \frac{K_{92}[Ala]}{K_{93}+[Ala]} \right) X_v \\
&\quad + \frac{P}{V}[Ala]_{in} - \frac{H}{V}[Ala] \qquad (6.15)
\end{aligned}$$

$$\frac{d[Glu]}{dt} = \left(\frac{K_{101}[Glc]}{K_{43}+[Glc]} - \frac{K_{104}[Lac]}{K_{54}+[Lac]} + \frac{K_{102}[Asn]}{K_{83}+[Asn]} \right.$$

$$\left. - \frac{K_{103}[Amm][Glu]}{(K_{61}+[Amm])(K_{63}+[Glu])} \right) X_v + \frac{P}{V}[Glu]_{in} - \frac{H}{V}[Glu] \quad (6.16)$$

$$\frac{d[Pro]}{dt} = -\left(\frac{K_{110}[Pro]}{K_{111}+[Pro]} \right) X_v + \frac{P}{V}[Pro]_{in} - \frac{H}{V}[Pro] \quad (6.17)$$

$$\frac{d[Ser]}{dt} = -\left(\frac{K_{120}[Ser]}{K_{121}+[Ser]} + \frac{K_{122}[Amm][Ser]}{(K_{133}+[Amm])(K_{134}+[Ser])} \right) X_v$$

$$+\frac{P}{V}[Ser]_{in} - \frac{H}{V}[Ser] \quad (6.18)$$

$$\frac{d[Gly]}{dt} = \left(\frac{K_{130}[Amm][Ser]}{(K_{133}+[Amm])(K_{134}+[Ser])} - \frac{K_{131}[Gly]}{K_{132}+[Gly]} \right) X_v$$

$$+\frac{P}{V}[Gly]_{in} - \frac{H}{V}[Gly] \quad (6.19)$$

$$\frac{d[Gln]}{dt} = \left(\frac{K_{140}[Amm][Glu]}{(K_{61}+[Amm])(K_{63}+[Glu])} - \frac{K_{143}[Gln]}{K_{144}+[Gln]} \right) - \frac{H}{V}[Gln] \quad (6.20)$$

$$\frac{d[Thr]}{dt} = -\left(\frac{K_{150}[Thr]}{K_{151}+[Thr]} \right) X_v + \frac{P}{V}[Thr]_{in} - \frac{H}{V}[Thr] \quad (6.21)$$

$$\frac{d[His]}{dt} = -\left(\frac{K_{160}[His]}{K_{161}+[His]} \right) X_v + \frac{P}{V}[His]_{in} - \frac{H}{V}[His] \quad (6.22)$$

$$\frac{d[Arg]}{dt} = -\left(\frac{K_{170}[Arg]}{K_{171}+[Arg]} \right) X_v + \frac{P}{V}[Arg]_{in} - \frac{H}{V}[Arg] \quad (6.23)$$

$$\frac{d[Cys]}{dt} = -\left(\frac{K_{180}[Cys]}{K_{181}+[Cys]} \right) X_v + \frac{P}{V}[Cys]_{in} - \frac{H}{V}[Cys] \quad (6.24)$$

$$\frac{d[Lys]}{dt} = -\left(\frac{K_{190}[Lys]}{K_{191}+[Lys]} \right) X_v + \frac{P}{V}[Lys]_{in} - \frac{H}{V}[Lys] \quad (6.25)$$

$$\frac{d[Tyr]}{dt} = -\left(\frac{K_{200}[Tyr]}{K_{201}+[Tyr]}\right)X_v + \frac{P}{V}[Tyr]_{in} - \frac{H}{V}[Tyr] \quad (6.26)$$

$$\frac{d[Met]}{dt} = -\left(\frac{K_{210}[Met]}{K_{211}+[Met]}\right)X_v + \frac{P}{V}[Met]_{in} - \frac{H}{V}[Met] \quad (6.27)$$

$$\frac{d[Val]}{dt} = -\left(\frac{K_{220}[Val]}{K_{221}+[Val]}\right)X_v + \frac{P}{V}[Val]_{in} - \frac{H}{V}[Val] \quad (6.28)$$

$$\frac{d[Ile]}{dt} = -\left(\frac{K_{230}[Ile]}{K_{231}+[Ile]}\right)X_v + \frac{P}{V}[Ile]_{in} - \frac{H}{V}[Ile] \quad (6.29)$$

$$\frac{d[Leu]}{dt} = -\left(\frac{K_{240}[Leu]}{K_{241}+[Leu]}\right)X_v + \frac{P}{V}[Leu]_{in} - \frac{H}{V}[Leu] \quad (6.30)$$

$$\frac{d[Phe]}{dt} = -\left(\frac{K_{250}[Phe]}{K_{251}+[Phe]}\right)X_v + \frac{P}{V}[Phe]_{in} - \frac{H}{V}[Phe] \quad (6.31)$$

$$\frac{d[Trp]}{dt} = -\left(\frac{K_{260}[Trp]}{K_{261}+[Trp]}\right)X_v + \frac{P}{V}[Trp]_{in} - \frac{H}{V}[Trp] \quad (6.32)$$

$$\frac{d[Mab]}{dt} = -\left(K_{270}+K_{271}[Glc]\right)X_v - \frac{H}{V}[Mab] \quad (6.33)$$

where X_v represents the viable cell density (VCD) and X_d the dead cell density (DCD). Furthermore, the concentrations of metabolites are defined as follows: $[Glc]$ - glucose,$[Lac]$ - lactate, $[Amm]$ - ammonia, $[Asn]$ - asparagine, $[Asp]$ - aspartate, $[Ala]$ - alanine, $[Glu]$ - glutamate, $[Pro]$ - proline, $[Ser]$ - serine, $[Gly]$ - glycine, $[Gln]$ - glutamine, $[Thr]$ - threonine, $[His]$ - histidine, $[Arg]$ - arginine, $[Cys]$ - cysteine, $[Lys]$ - lysine, $[Tyr]$ - tyrosine, $[Met]$ - methionine, $[Val]$ - valine, $[Ile]$ - Isoleucine, $[Leu]$ - leucine, $[Phe]$ - phenylalanine, $[Trp]$ - tryptophan and $[Mab]$ – MAb. The kinetic parameters, to be estimated from the experimental data, are given by K_{21} to K_{271}.

6.2.4. Parameter Estimation

Performing a parameter estimation involving such a large number of ODEs and parameters as provided by the model described in subsection 6.2.3 is a computational extensive task. This problem is even more challenging due to the coupling between equations and sparse measurements in case of most of the amino acids. Trying to estimate all parameters at the same time is thus an intractable task for such a large system due to the nonlinear nature of the optimization problem. For that reason, we divided the parameter estimation problem into successive optimization problems involving a reduced number of parameters. This is similar to the approach presented in subsection 4.4.2.3. The main motivation behind this sequential model calibration procedure is that the dynamics of most metabolites mainly depend on the change in biomass. This allows us to divide the problem into the following steps:

1. Approximation of the measured viable cell density profile by a simple piecewise linear interpolation to capture the dynamic behaviour of biomass.
2. Successive estimation of all the minor amino acids which do not contribute to the evolution of other metabolites in the sense that their fluxes are insignificant as per the MFA. These smaller parameter estimation steps can be executed independently from each other.
3. Estimation of major metabolites which are involved in the macro-reactions, where the dynamics of coupled metabolites are estimated simultaneously.
4. Finally, separate estimation of the parameters related to the dead cell and viable cell density equations.

6.2.5. Results

In the following, the results of the model-fitting are presented. One set of experimental data is used for calibration while a different set is used for validation. The estimated parameter values are given

in the appendix in table A.2.

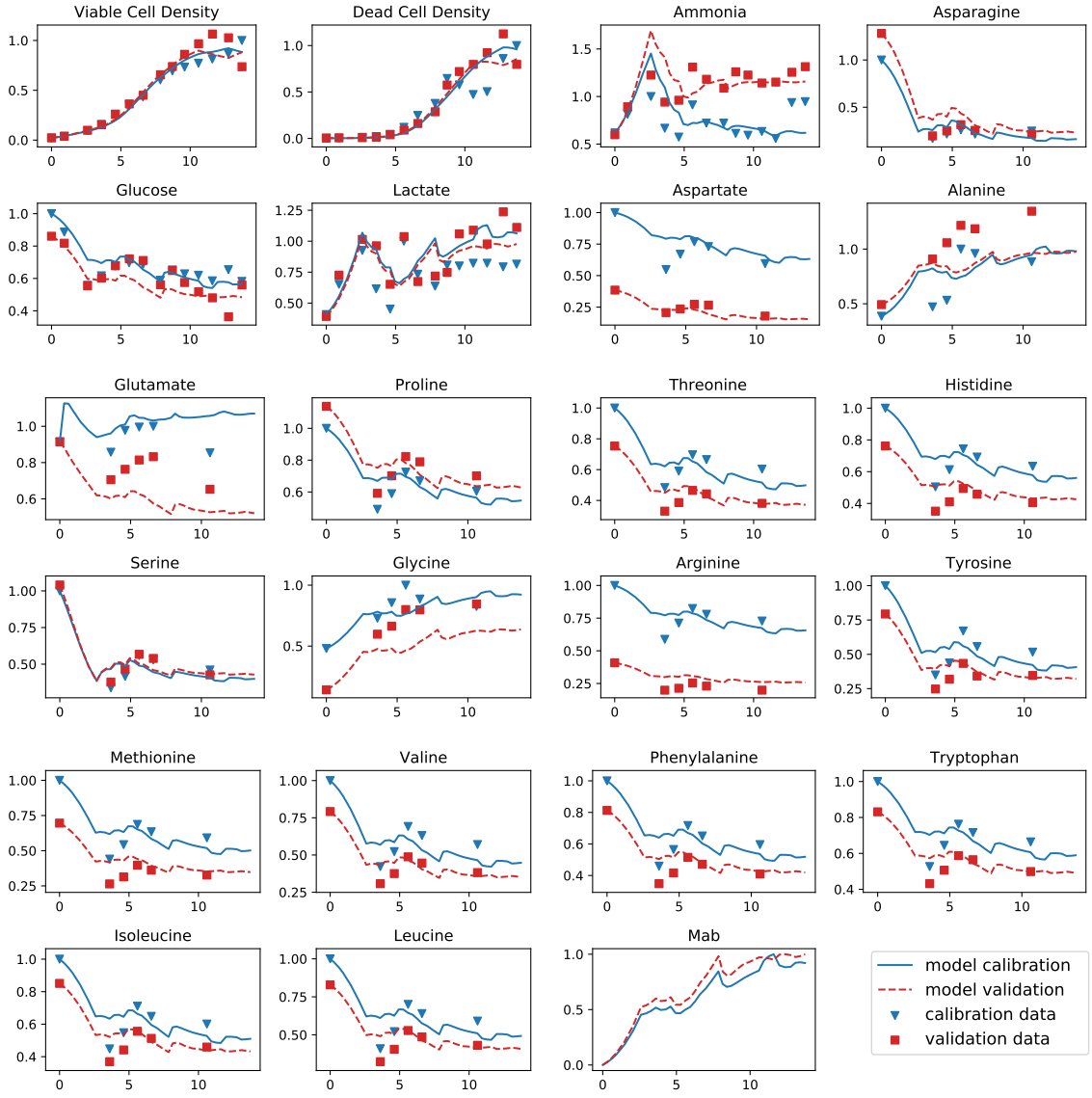

Figure 6.2.: Calibration and validation results of the developed dynamic metabolic model for perfusion systems with normalized units.

The developed model provides an adequate fit (qualitatively) of the metabolites for both the calibration and validation data. For example, one major difference between the two media which is

captured by the model is the different rate of ammonia production. Some metabolites such as alanine and glycine seem to be more difficult to predict accurately. Note that the parameters of the Mab equation (6.33) haven been calibrated from a different data set (not shown).

6.3. Run-to-Run Optimization

The aim of this section is to utilize the developed model for the purpose of a run-to-run optimization under model-plant mismatch case study. Specifically, we implement the methodologies proposed in the chapters 3, 4 and 5, which are based on the algorithm developed by Mandur & Budman (2015b). The model equations in (6.8) - (6.33) thus define the model for the perfusion process simulation and are utilized to generate experimental data. We assume that the measurements are corrupted by Gaussian noise with standard deviation of $10\,\%$ of the respective average output magnitude. Before providing more detail of the optimization objective, we first discuss the assumed model-plant mismatch.

6.3.1. Model-Plant Mismatch

One challenge in the modelling of cell culture processes is the description of the viable and dead cell densities. This is especially true for mammalian cells who exhibit complex biological phenomena such as apoptosis (programmed cell death) (Meshram et al., 2012). Due to the complex behaviour of biomass formation, it is common to construct semi-empirical relations for the VCD and DCD as shown in (6.8) and (6.9). For that reason, we assume that there exist a structural error in the biomass behavior. Instead of the DCD being a function of X_v^2, we just assume a linear dependence of the death term with respect to viable cell concentration. Therefore, the respective

model equations to be used for optimization purposes are given by:

$$\frac{dX_v}{dt} = \mu X_v \left(\frac{[Glc]}{(K_{21}+[Glc])} \frac{1}{1+\left(\frac{[Amm]}{K_{23}}\right)} \frac{1}{1+\left(\frac{[Lac]}{K_{25}}\right)} \frac{1}{1+\frac{[Glc]}{K_{22}}} \right) - k_d X_v \left(\frac{1}{1+\left(\frac{K_{24}}{[Amm]}\right)^n} + \frac{K_{26}}{[Glc]} \right) - \frac{B}{V} X_v \tag{6.34}$$

$$\frac{dX_d}{dt} = k_d X_v \left(\frac{1}{1+\left(\frac{K_{24}}{[Amm]}\right)^n} + \frac{K_{26}}{[Glc]} \right) - K_{lys} X_d - \frac{B}{V} X_d \tag{6.35}$$

Notice, that we have also introduced an inhibition term in (6.34) for large glucose concentrations. This is necessary to obtain a distinct optimum for the optimization objective described in the next section. For the optimization case study, this inhibition term has also been introduced into the simulator equation (6.8).

6.3.2. Optimization Objective

The main advantage of a perfusion system compared to batch or fed-batch operations is the fact that it is possible to avoid nutrient limitations due to a continuous supply of fresh media and discharge of toxic byproducts. Furthermore, maintaining a steady-state of the VCD is achievable through the bleeding of cells. As a result, perfusion cultures can be maintained at optimal conditions for long periods of time which leads to increased recombinant protein yields. However, it is first of all necessary to reach the desired cell density level before a steady-state is to be maintained. In order to do that, the perfusion rate is gradually increased in a step-wise fashion. For example, the normalized perfusion rate profile corresponding to the validation data shown in figure 6.2 is illustrated on the left-hand graph in figure 6.3.

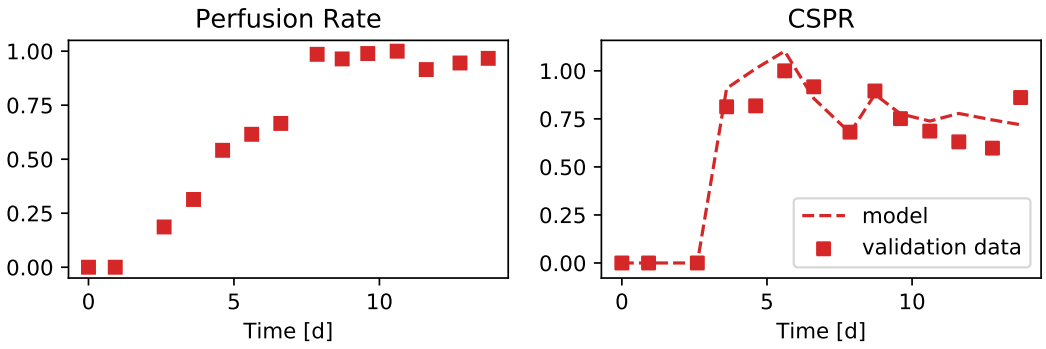

Figure 6.3.: Normalized perfusion rate and cell specific perfusion rate (CSPR) corresponding to the validation data shown in figure 6.2.

The reason for implementing this gradual increase is to maintain a certain cell specific perfusion rate (CSPR), which is defined as (Ozturk, 1996):

$$CSPR = \frac{D}{X_v} \tag{6.36}$$

where D is the dilution rate. For illustration, the measured and (from the model) predicted normalized CSPR are shown on the right graph in figure 6.3. Hence, a goal in the operation of perfusion processes is to maintain a desired CSPR to avoid nutrient limitations and the build-up of toxic byproducts (Clincke et al., 2013).

For this case study, we define as the objective the maximization of the amount of harvested product per day during the steady-state period. The decision variables for optimization are the media glucose concentration $[Glc]_{in}$ and the perfusion rate at steady-state P_{ss}. To achieve this, we implement a step-wise increasing perfusion profile, similar to the left graph in figure 6.3, where we assume that the perfusion rate is changed on a daily basis until the steady-state is reached within t_{ss} days. Moreover, to limit the amount of perfused media, a constraint on the maximum CSPR at steady-state is provided. The optimization problem solved during each perfusion run can therefore

be formulated as follows:

$$\begin{aligned} \boldsymbol{u}_{k+1} \quad = \quad & \min_{[Glc]_0, P_{ss}} -P_{ss}\,[Mab]\,(x, [Glc]_0, P_{ss}, \boldsymbol{\theta}, t_{ss}) \\ \text{s.t.} \quad & (6.10) - (6.35) \\ & CSPR_{ss}(x, [Glc]_0, P_{ss}, t_{ss}) \leq 0.08\,\mathrm{nL/cell/day} \end{aligned} \tag{6.37}$$

Where $\boldsymbol{u}_{k+1}$ presents the input for the next perfusion run. The steady-state CSPR is defined as $CSPR = P_{ss}/X_{v,ss}$. Furthermore, for simplicity, the bleed rate is defined as a function of the perfusion rate as $B = 0.1P$.

6.3.3. Results

For updating model outputs, it is possible to apply the sequential procedure outlined in (Hille & Budman, 2017). However, before an optimization can be performed, it is first necessary to determine the subset of parameters to be adjusted for gradient correction. For that reason, the parameter selection procedure proposed in (Hille et al., 2017) is utilized to determine, at each iteration, the particular subset of parameters associated with large gradient sensitivities. When applying the procedure to the model $(6.10) - (6.35)$, a set of highly sensitive parameters in the first iteration is given by μ, K_{21}, K_{43}, K_{25} and K_{270}. On the other hand, upon converging to the process optimum, the optimal subset for gradient correction is given by μ, K_{21}, K_{22}, k_d and K_{53}. Considering the selection procedure is based on local sensitivities, this illustrates, that it is particularly important to adjust the parameters associated with glucose inhibition, i.e. K_{21} and K_{22}, to obtain an accurate prediction around the optimum.

It is important to point out that, due to the large number of equations and parameters in $(6.10) - (6.35)$, there is a significant increase in computational time in each of the steps involved in the simultaneous identification and optimization framework. Compared to the penicillin process, in-

volving 8 model parameters, presented in chapter 3, just the computational time of the parameter selection increases by a factor of 30.

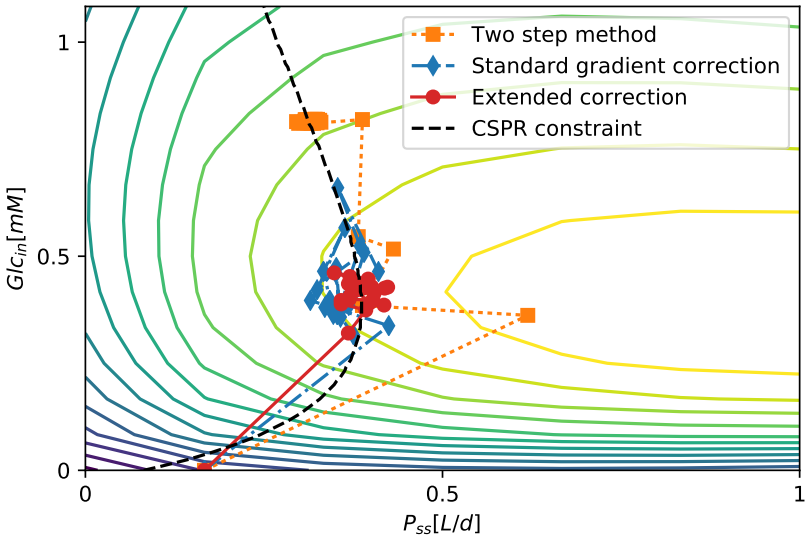

Figure 6.4.: Run-to-run optimization results (normalized) of the perfusion process. Contour lines represent the objective function described in (6.37).

Finally, the optimization results are shown in figure 6.4, where three optimization schemes are compared. The two-step method (Chen & Joseph, 1987), consisting of a repeated identification step followed by a model-based optimization, converges to a sub-optimal operating point. This shows that a minor model-plant mismatch such as the wrongly assumed dependence of cell death with respect to cell density can have significant impact in iterative optimization schemes.

When using the approach for simultaneous identification and optimization on the other hand, convergence to the optimal operating point is achieved. The standard gradient correction method was applied where only the most recent gradient measurements are used in the gradient correction step. The extended gradient correction is based on the procedure outlined in chapter 5 and includes a design of experiments approach to determine the optimal input perturbation for more informative

cost gradient measurements. From 6.4 it can be seen that the effect of uncertainty in gradient measurements can be mitigated by applying the proposed procedure for gradient correction with design of experiments. This illustrates the effectiveness of the parameter selection, estimation and experimental design techniques presented in this work.

6.4. Conclusions

This chapter presented two contributions. In the first part, a dynamic metabolic model was developed to describe the dynamic behavior of biomass and extra cellular metabolites during a perfusion cultivation of CHO cells. A metabolic flux analysis was adjusted to account for the effects of perfusion and harvest rates. The obtained model showed a good agreement with the experimental data.

In the second part of this chapter, the derived model was utilized for a case study of a run-to-run optimization in the presence of model-plant mismatch. To illustrate the effectiveness of the run-to-run procedure in the presence of model error, a lack of knowledge about the accurate effect of apoptosis was assumed. Furthermore, an objective was defined involving the maximization of antibody production while maintaining a specified CSPR. Overall, the results illustrate the importance of optimization methods that provide robustness to model-plant mismatch. To this end, it is shown in this chapter that the methods developed in this work are particularly valuable when complex models with numerous parameters are used in iterative optimization schemes.

Acknowledgements

We would like to thank Delia Lyons, Jeremiah Riesberg and Dustin Davis from MilliporeSigma, SAFC/Sigma-Aldrich Corp. for their collaboration and for providing the experimental data.

7. Conclusions and Outlook

7.1. Overview

Biochemical processes present a platform for the production of specific pharmaceuticals such as humanized monoclonal antibodies. Due to rising demand for medical applications, there exist an urgency to increase product yields. Model-based optimization presents an opportunity to improve the process performance in a methodical approach. This is especially true for biotechnological processes that are mostly operated in batch or fed-batch modes, which makes the run-to-run optimization approach a particularly attractive method. Following the run-to-run approach, it is possible to utilize the information from past experiments to update an existing model to predict the optimal input for the next batch-run.

However, several challenges have to be addressed to achieve a successful convergence to the process optimum. These challenges are mostly related to the existence of model-plant mismatch and the occurrence of unmeasured disturbances that impair the prediction capabilities of the involved models. For that reason, it is vital to implement optimization methods that provide the required robustness to such uncertainties. To this end, the method of simultaneous identification and optimization presents a framework where the model parameters are adjusted to correct for errors in the predicted cost-function and constraint gradients. This enforces model adequacy and drives the process to the desired operating point. The present work builds up on this framework and extends it in the following ways.

Chapter 3 presents a novel methodology for selecting a subset of model parameters to be updated in each iteration of a batch-to-batch optimization in the presence of model-plant mismatch. The parameters are ranked according to their overall effect on the model output and the gradients of the cost-function and constraints. Furthermore, robustness with respect to uncertainties in initial batch conditions is implemented by using a robust formulation of the objective. The propagation of uncertainty in initial conditions onto the output was achieved by using an approach based on Polynomial Chaos Expansions.

Chapter 4 introduces set-based uncertainty bounds into the run-to-run optimization framework. Using the bounds on the model outputs, a parameter estimation problem is proposed that minimizes the ratio of the error between model predictions and measurements to a lumped measure of the parametric sensitivities of cost-function and constraint gradients subject to constraints on time varying trajectories (set based constraints) of output variables. In this way, trajectories within the set-based constraints are found that provide high gradient sensitivities that are necessary for correcting the gradients with small deviation of the parameter values with respect to the values obtained in the identification step. As a result, we are able to mitigate the conflicting objectives of identification and gradient correction which results in a smoother and faster convergence to the process optimum. Furthermore, using the output uncertainty bounds, a model-update criterion is implemented to only update the specific model outputs where it is strictly necessary. This is especially of use for larger biochemical models where the identification step can be divided into smaller problems with a reduced number of parameters.

Experimental design is vital for acquiring additional information with the aim of reducing parametric uncertainty. To this end, most design criteria evolve around the design of new experiments deemed more informative in terms of model identification. However, in the presence of model-plant mismatch there exist a lack of synergy between identification and optimization. Chapter 5 therefore outlines a design of experiments approach that focuses on the design of new batch exper-

iments that are most informative with respect to the gradients of cost and constraints. In that sense, parameter precision is being improved by using available information of past batch-runs combined with new experiments. This use of an experimental design approach leads to additional robustness against gradient uncertainty and superior model prediction capabilities, especially around the process optimum.

Finally, chapter 6 presents the model development of a perfusion operation of a mammalian cell culture and the application of the run-to-run optimization procedures developed in this thesis. This part of the work was developed in collaboration with an industrial partner that provided perfusion data for model development and calibration. In contrast to regular batch processes, perfusion operations provide a constant supply of fresh media and a constant outflow through a perfusion and harvest rate respectively. At the same time, the discharge of cells is prevented through a filter leading to high cell densities during cultivation. Perfusion processes thus present an attractive system for the industrial production of bio manufactured products. The illustrated model development and optimization case study show that model-based optimization presents an opportunity to enhance the performance of perfusion processes, while providing robustness to model uncertainty.

7.2. Concluding Remarks

In summary, it can be concluded that an appropriate parameter selection is essential in iterative optimization schemes where the model is updated and corrected by adjusting the model parameters. As shown in chapter 3, the main benefit in selecting a set of highly sensitive parameters is a major speed-up in convergence leading to a reduction in the integral absolute error (IAE), i.e. the difference between the predicted and optimal input averaged over all iterations, of more than $50\,\%$. This is mainly due to the fact that more sensitive parameters result in a significantly improved gradient correction. In other words, for the same upper bound on the relative truncation error, the proposed selection leads to parameters that will provide larger effect on the gradient with small deviation in

parameter values from the ones calculated in the earlier identification step. In this way we reduce the conflict between the identification and the gradient matching step thus resulting in smoother convergence.

Another reason for selecting only a subset of parameters is the effect of overfitting and computational time. In that sense, overfitting of noisy measurements can be avoided by only considering the parameter that provide the most effect on model outputs. At the same time, as computational time increases monotonically with the number of parameters that are updated at each iteration, there is a motivation to limit the number of considered parameters . This is especially important when employing models of larger complexity as illustrated by the dynamic metabolic models in chapter 4, 5 and 6. In terms of model identification, the parameter selection problem has been mainly addressed by a sequential procedure that divides the identification in smaller problems with a reduced number of parameters. Nonetheless, the parameter selection technique from chapter 3 can still be applied to determine the subset of parameters to be utilized in the gradient correction step. This can be of substantial help to the user, as it is often unclear, which set of parameters are ideally suited for performing the gradient matching.

As illustrated in chapter 4, it is also possible to further improve convergence to the optimum when using suitable or even sub-optimal parameter sets. This can be achieved by using a modified parameter estimation objective, i.e. the ratio of the sum of squared errors to a measure of parametric sensitivity, which leads to parameter values that are better suited for the subsequent gradient correction step. The strength of the modified parameter estimation is that it not only leads to a speed-up of convergence in terms of the IAE, but also results in a reduction of the variance of the predicted input for the next batch-run. This is made possible by utilizing worst-case bounds on the process outputs (set-based constraints) to avoid unnecessary re-estimation steps and to also find time varying trajectories that provide larger parametric sensitivities. However, a trade-off exists here since additional experiments have to be performed in order to obtain the set-based constraints.

While the methods developed in chapter 3 and 4 help in enforcing model adequacy, a persisting problem in run-to-run optimization is given by the uncertainty in gradient measurements. To this end, the design of experiments approach outlined in chapter 5 introduces additional robustness by incorporating past and future cost measurements into the gradient correction step. The main benefit here is not necessarily related to improvement in the IAE, as shown by the CHO case study in chapter 5 for example, but rather in the reduction of the variance in the prediction of the next optimal input for which reductions of over $90\,\%$ were observed. This effect is most prominent around the optimum, where the variance in gradient measurements is especially large. One drawback of the design of experiment approach is the increase in computational time of the gradient correction step resulting from additional function evaluations as more gradient measurements are considered. In presence of time constraints, this increase in computational time can be especially prohibitive for large scale models presented in chapter 6.

Finally, the impact of stochastic uncertainty in initial batch conditions has been investigated in chapter 3. By performing a robustness analysis, it was shown that the effect of input uncertainties can be mitigated leading, on average, to higher product yields.

7.3. Future Work

Through this work we have identified a number of challenges that should be addressed in the future:

1. **Model adequacy:** One of main aspects discussed in this work is the enforcement of model adequacy in model-based run-to-run optimization schemes. In order to achieve that, it was shown that it is critical to utilize the particular model parameters associated with large gradient sensitivities. However, for this procedure to be successful, it is necessary that there exist a suitable set of model parameters that are available for adaptation. To facilitate this step in the future, it will be useful to address this issue already during the model development. This

implies that additional parameters may have to be included as to provide enough sensitivity for optimization objectives. To this end, it will be necessary to investigate what kind of model structure is best suited for this purpose. This issue is also related to the feasibility of set-based constraints which is currently considered a key challenge in the use of uncertainty bounds for parameter estimation.

2. **Development of growth media for cell culture processes:** One example where the run-to-run approach is directly applicable is the model-based design of new media for cell cultivation processes. It is generally the case that cultivation media contain a large number of different components. Many of those are not included in the kind of fundamental models discussed throughout this work, as their mechanistic effects are usually not well understood. However, if it is desired to perform a model-based media design, it is essential to introduce additional model components that have been ignored but may have large impact on growth and productivity. To address this problem without significantly increasing the dimensions of the model one option is to use Principal Component Analysis (PCA) to reduce the input dimension into a few principal components (PCs) and relate those to the observed uptake and production rates exhibited by the cells for different media. Then, the mechanistic model could be augmented with respect to the scores of the main principal components to result in a hybrid model that combines the major amino acids and principal scores, where the latter describe a large combination of minor elements. This way, a measure of the effect of all media components can be incorporated, which enables a model-based optimization by searching for an optimum within the space spanned by the PCs and the major amino acids.

3. **Real Time Optimization (RTO):** A major challenge of presented method to RTO applications lies in the determination of the optimal trade-off between exploration and exploitation. In RTO, the plant has to be perturbed to introduce excitation necessary for a parameter estimation. At the same time, when dealing with model-plant mismatch, plant perturbation are

also necessary to obtain gradient measurements. Hence, there exist two objectives for which the optimal plant perturbation directions have to be determined. In this context, perfusion processes present an innovative opportunity for RTO applications. Typically, after an initial period, perfusion processes are able to reach a steady-state where the cell density is maintained around a specific level. During the steady-state there exist an opportunity to increase the production of the desired proteins by changing process conditions such as the temperature. For this purpose, a model could be calibrated around the steady-state and used for RTO, provided the presence of sufficient excitation. Furthermore, an available model can also be used for process control purposes. In that regard, it is often desired to reach a certain cell specific perfusion rate, which could be maintained by a controller.

4. **Economic Model Predictive Control (EMPC):** MPC presents an established methodology in advanced control of chemical processes. An extension of the method, referred to as EMPC, that also takes into account an economic cost of the control action, has recently gained much popularity. Similar to run-to-run optimization, EMPC presents an approach where a plant is driven to an optimal operating point over the course of a finite horizon. However, in presence of a structural error in the model used by the controller such as the model errors considered in the this thesis, the process may converge to a sub-optimal operating point. Hence, there exist an opportunity to introduce robustness to model-plant mismatch in EMPC by applying the methods developed in this work to correct the involved model. In this case, since the optimization will be performed online and will be typically applied to continuous processes, a number of challenges will have to be addressed such as computational time and persistence of excitation that will enable parameter update and speed of parameter adaptation during transient operation.

Bibliography

Agarwal, M. (1997). Feasibility of on-line reoptimization in batch processes. *Chemical Engineering Communications*, *158*, 19–29.

Aghamohseni, H., Ohadi, K., Spearman, M., Krahn, N., Moo-Young, M., Scharer, J. M., Butler, M., & Budman, H. M. (2014). Effects of nutrient levels and average culture ph on the glycosylation pattern of camelid-humanized monoclonal antibody.

Ariyur, K. B., & Krstic, M. (2003). *Real-time optimization by extremum-seeking control*. Hoboken, N.J.: Wiley-Interscience.

Atkinson, A. C., & Donev, A. N. (1992). *Optimum experimental designs*. Oxford: Clarendon Press.

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

Biegler, L. T., Grossmann, I. E., & Westerberg, A. W. (1985). A note on approximation techniques used for process optimization. *Computers & Chemical Engineering*, *9*, 201–206.

Birol, G., Undey, C., & Cinar, A. (2002). A modular simulation package for fed-batch fermentation: penicillin production. *Computers & Chemical Engineering*, *26*, 1553–1565.

Bonvin, D. (1998). Optimal operation of batch reactors - a personal view. *Journal of Process Control*, *8*, 355–368.

Bonvin, D., & Francois, G. (2017). Control and optimization of batch chemical processes. Coulson and Richardson's Chemical Engineering (pp. 441–503). Butterworth-Heinemann Ltd.

Bonvin, D., Georgakis, C., Pantelides, C. C., Barolo, M., Grover, M. A., Rodrigues, D., Schneider, R., & Dochain, D. (2016). Linking models and experiments. *Industrial & Engineering Chemistry Research*, *55*, 6891–6903.

Box, G., & Draper, N. (1987). *Empirical model-building and response surfaces*. Wiley.

Box, G. E. P., & Draper, N. R. (1969). *Evolutionary Operation: A Statistical Method for Process Improvement*. New York: John Wiley.

Brdys, M. A., Chen, S., & Roberts, P. D. (1986). An extension to the modified 2-step algorithm for steady-state system optimization and parameter-estimation. *International Journal of Systems Science*, *17*, 1229–1243.

Brdys, M. A., & Tatjewski, P. (2005). *Iterative Algorithms for Multilayer Optimizing Control*.

Brun, R., Kuhni, M., Siegrist, H., Gujer, W., & Reichert, P. (2002). Practical identifiability of asm2d parameters - systematic selection and tuning of parameter subsets. *Water research*, *36*, 4113–4127.

Butler, M., & Meneses-Acosta, A. (2012). Recent advances in technology supporting biopharmaceutical production from mammalian cells. *Applied Microbiology and Biotechnology*, *96*, 885–894.

Camacho, J., Pico, J., & Ferrer, A. (2007). Self-tuning run to run optimization of fed-batch processes using unfold-pls. *AIChE Journal*, *53*, 1789–1804.

Chachuat, B., Srinivasan, B., & Bonvin, D. (2009). Adaptation strategies for real-time optimization. *Computers & Chemical Engineering*, *33*, 1557–1567.

Chen, C. Y., & Joseph, B. (1987). On-line optimization using a two-phase approach: An application study. *Industrial & Engineering Chemistry Research*, *26*, 1924–1930.

Chen, J., & Liu, K.-C. (2002). On-line batch process monitoring using dynamic pca and dynamic pls models. *Chemical Engineering Science*, *57*, 63–75.

Chu, Y., & Hahn, J. (2007). Parameter set selection for estimation of nonlinear dynamic systems. *AICHE Journal*, *53*, 2858–2870.

Chu, Y., & Hahn, J. (2012). Generalization of a parameter set selection procedure based on orthogonal projections and the d-optimality criterion. *AIChE Journal*, *58*, 2085–2096.

Chuppa, S., Tsai, Y.-S., Yoon, S., Shackleford, S., Rozales, C., Bhat, R., Tsay, G., Matanguihan, C., Konstantinov, K., & Naveh, D. (1997). Fermentor temperature as a tool for control of high-density perfusion cultures of mammalian cells. *Biotechnology and bioengineering*, *55*, 328–338.

Clincke, M.-F., Mölleryd, C., Zhang, Y., Lindskog, E., Walsh, K., & Chotteau, V. (2013). Very high density of cho cells in perfusion by atf or tff in wave bioreactor^{TM}. part i. effect of the cell density on the process. *Biotechnology Progress*, *29*, 754–767.

Costello, S., Francois, G., & Bonvin, D. (2016). A directional modifier-adaptation algorithm for real-time optimization. *Journal of Process Control*, *39*, 64–76.

Cougnon, P., Dochain, D., Guay, M., & Perrier, M. (2011). On-line optimization of fedbatch bioreactors by adaptive extremum seeking control. *Journal of Process Control*, *21*, 1526–1532.

Croughan, M. S., Konstantinov, K. B., & Cooney, C. (2015). The future of industrial bioprocessing: Batch or continuous? *Biotechnology and bioengineering*, *112*, 648–651.

Degenring, D., Froemel, C., Dikta, G., & Takors, R. (2004). Sensitivity analysis for the reduction of complex metabolism models. *Journal of Process Control*, *14*, 729–745.

Dong, McAvoy, T. J., & Zafiriou, E. (1996). Batch-to-batch optimization using neural network models. *Industrial & Engineering Chemistry Research*, *35*, 2269–2276.

Dorka, P., Fischer, C., Budman, H., & Scharer, J. M. (2009). Metabolic flux-based modeling of

mab production during batch and fed-batch operations. *Bioprocess and Biosystems Engineering*, *32*, 183–196.

Doyle, F. J., Harrison, C. A., & Crowley, T. J. (2003). Hybrid model-based approach to batch-to-batch control of particle size distribution in emulsion polymerization. *Computers & Chemical Engineering*, *27*, 1153 – 1163.

Du, Y., Budman, H., & Duever, T. A. (2017). Comparison of stochastic fault detection and classification algorithms for nonlinear chemical processes. *Computers & Chemical Engineering*, *106*, 57 – 70.

Du, Y., Duever, T. A., & Budman, H. (2016). Generalized polynomial chaos-based fault detection and classification for nonlinear dynamic processes. *Industrial & Engineering Chemistry Research*, *55*, 2069–2082.

Duarte, B., & Saraiva, P. (2003). Hybrid models combining mechanistic models with adaptive regression splines and local stepwise regression. *Industrial & Engineering Chemistry Research*, *42*, 99–107.

Duran-Villalobos, C. A., & Lennox, B. (2013). Iterative learning modelling and control of batch fermentation processes. *IFAC Proceedings Volumes*, *46*, 511 – 516. 10th IFAC International Symposium on Dynamics and Control of Process Systems.

Eldred, M. (2009). Recent advances in non-intrusive polynomial chaos and stochastic collocation methods for uncertainty analysis and design. In *50th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*.

Farid, S. S. (2007). Process economics of industrial monoclonal antibody manufacture. *Journal of Chromatography B*, *848*, 8–18.

Fischer, S., Handrick, R., & Otte, K. (2015). The art of cho cell engineering: A comprehensive retrospect and future perspectives. *Biotechnology Advances*, *33*, 1878–1896.

Forbes, J. F., & Marlin, T. E. (1996). Design cost: A systematic approach to technology selection for model-based real-time optimization systems. *Computers & Chemical Engineering*, *20*, 717–734.

Forbes, J. F., Marlin, T. E., & MacGregor, J. F. (1994). Model adequacy requirements for optimizing plant-operations. *Computers & Chemical Engineering*, *18*, 497–510.

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

Francois, G., Srinivasan, B., & Bonvin, D. (2005). Use of measurements for enforcing the necessary conditions of optimality in the presence of constraints and uncertainty. *Journal of Process Control*, *15*, 701–712.

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

Gao, J., Gorenflo, V. M., Scharer, J. M., & Budman, H. M. (2007). Dynamic metabolic modeling for a mab bioprocess. *Biotechnology progress*, *23*, 168–181.

Gao, W., Wenzel, S., & Engell, S. (2016). A reliable modifier-adaptation strategy for real-time optimization. *Computers & Chemical Engineering*, *91*, 318–328.

Gao, W. H., & Engell, S. (2005). Iterative set-point optimization of batch chromatography. *Computers & Chemical Engineering*, *29*, 1401–1409.

Garcia, C. E., & Morari, M. (1981). Optimal operation of integrated processing systems .1. open-loop online optimizing control. *AIChE Journal*, *27*, 960–968.

Georgakis, C. (2013). Design of dynamic experiments: A data-driven methodology for the optimization of time-varying processes. *Industrial & Engineering Chemistry Research*, *52*, 12369–12382.

Hille, R., & Budman, H. M. (2017). Run-to-run optimization of batch process using set-based constraints. *In Proceedings IFAC World Congress*, (pp. 4678–4683).

Hille, R., Mandur, J., & Budman, H. M. (2017). Robust batch-to-batch optimization in the presence of model-plant mismatch and input uncertainty. *AIChE Journal*, *63*, 2660–2670.

Houska, B., Telen, D., Logist, F., Diehl, M., & Impe, J. F. M. V. (2015). An economic objective for the optimal experiment design of nonlinear dynamic processes. *Automatica*, *51*, 98–103.

Kim, K. K. K., & Braatz, R. D. (2012). Probabilistic analysis and control of uncertain dynamic systems: Generalized polynomial chaos expansion approaches. *2012 American Control Conference (Acc)*, (pp. 44–49).

Leighty, R. W., & Antoniewicz, M. R. (2011). Dynamic metabolic flux analysis (dmfa): A framework for determining fluxes at metabolic non-steady state. *Metabolic engineering*, *13*, 745–755.

Ljung, L. (1999). *System Identification: Theory for the User*. (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Lund, B. F., & Foss, B. A. (2008). Parameter ranking by orthogonalization - applied to nonlinear mechanistic models. *Automatica*, *44*, 278–281.

Ma, D. L., Chung, S. H., & Braatz, R. D. (1999). Worst-case performance analysis of optimal batch control trajectories. In *1999 European Control Conference (ECC)* (pp. 3256–3261).

Majors, B. S., Betenbaugh, M. J., Pederson, N. E., & Chiang, G. G. (2009). Mcl-1 overexpression leads to higher viabilities and increased production of humanized monoclonal antibody in chinese hamster ovary cells. *Biotechnology progress*, *25*, 1161–1168.

Mandur, J., & Budman, H. (2014). Robust optimization of chemical processes using bayesian description of parametric uncertainty. *Journal of Process Control*, *24*, 422–430.

Mandur, J. S., & Budman, H. M. (2015a). Robust algorithms for simultaneous model identification

and optimization in the presence of model-plant mismatch. *Industrial & Engineering Chemistry Research*, *54*, 9382–9393.

Mandur, J. S., & Budman, H. M. (2015b). Simultaneous model identification and optimization in presence of model-plant mismatch. *Chemical Engineering Science*, *129*, 106–115.

Marchetti, A., Chachuat, B., & Bonvin, D. (2009). Modifier-adaptation methodology for real-time optimization. *Industrial & Engineering Chemistry Research*, *48*, 6022–6033.

Marchetti, A., Chachuat, B., & Bonvin, D. (2010). A dual modifier-adaptation approach for real-time optimization. *Journal of Process Control*, *20*, 1027–1037.

Marlin, T. E., & Hrymak, A. N. (1997). Real-time operations optimization of continuous processes. *In American Institute of Chemical Engineering Symposium Series*, *93*.

Martinez, E. C., Cristaldi, M. D., & Grau, R. J. (2009). Design of dynamic experiments in modeling for optimization of batch processes. *Industrial & Engineering Chemistry Research*, *48*, 3453–3465.

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, *21*, 239–245.

McLean, K. A. P., & McAuley, K. B. (2012). Mathematical modelling of chemical processes: obtaining the best model predictions and parameter estimates using identifiability and estimability procedures. *Canadian Journal of Chemical Engineering*, *90*, 351–366.

Meshram, M., Naderi, S., McConkey, B., Budman, H., Scharer, J., & Ingalls, B. (2012). Population-based modeling of the progression of apoptosis in mammalian cell culture. *Biotechnology and Bioengineering*, *109*, 1193–1204.

Montgomery, D. C. (2012). *Design and Analysis of Experiments*. (8th ed.). John Wiley & Sons Inc.

Morari, M., Arkun, Y., & Stephanopoulos, G. (1980). Studies in the synthesis of control-structures for chemical processes .1. formulation of the problem - process decomposition and the classification of the control tasks - analysis of the optimizing control-structures. *AIChE Journal*, *26*, 220–232.

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, *33*, 161–174.

Mukkula, A. R. G., & Paulen, R. (2017). Model-based design of optimal experiments for nonlinear systems in the context of guaranteed parameter estimation. *Computers & Chemical Engineering*, *99*, 198–213.

Naderi, S., Meshram, M., Wei, C., McConkey, B., Ingalls, B., Budman, H., & Scharer, J. (2011). Development of a mathematical model for evaluating the dynamics of normal and apoptotic chinese hamster ovary cells. *Biotechnology progress*, *27*, 1197–1205.

Nagy, Z. K., & Braatz, R. D. (2007). Distributional uncertainty analysis using power series and polynomial chaos expansions. *Journal of Process Control*, *17*, 229–240.

Najm, H. N. (2009). Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics. *Annual Review of Fluid Mechanics*, *41*, 35–52.

Navia, D., Briceno, L., Gutierrez, G., & de Prada, C. (2015). Modifier-adaptation methodology for real-time optimization reformulated as a nested optimization problem. *Industrial & Engineering Chemistry Research*, *54*, 12054–12071.

Niu, H., Amribt, Z., Fickers, P., Tan, W., & Bogaerts, P. (2013). Metabolic pathway analysis and reduction for mammalian cell cultures - towards macroscopic modeling. *Chemical Engineering Science*, *102*, 461–473.

Nolan, R. P., & Lee, K. (2011). Dynamic model of cho cell metabolism. *Metabolic Engineering*, *13*, 108 – 124.

Nomikos, P., & MacGregor, J. F. (1995). Multi-way partial least squares in monitoring batch processes. *Chemometrics and Intelligent Laboratory Systems*, *30*, 97–108.

O'Hagan, A. (2013). Polynomial chaos: A tutorial and critique from a statistician's perspective, .

Oliveira, R. (2004). Combining first principles modelling and artificial neural networks: a general framework. *Computers & Chemical Engineering*, *28*, 755 – 766.

Ozturk, S. S. (1996). Engineering challenges in high density cell culture systems. *Cytotechnology*, *22*, 3–16.

Paulen, R., Villanueva, M., Fikar, M., & Chachuat, B. (2013). Guaranteed parameter estimation in nonlinear dynamic systems using improved bounding techniques. *Proceedings of European Control Conference (ECC)*, (pp. 4514–4519).

Paulson, J. A., & Mesbah, A. (2017). An efficient method for stochastic optimal control with joint chance constraints for nonlinear systems. *International Journal of Robust and Nonlinear Control*, (pp. 1–21).

Prajna, S. (2006). Barrier certificates for nonlinear model validation. *Automatica*, *42*, 117–126.

Provost, A., Bastin, G., Agathos, S. N., & Schneider, Y. J. (2006). Metabolic design of macroscopic bioreaction models: application to chinese hamster ovary cells. *Bioprocess and Biosystems Engineering*, *29*, 349–366.

Roberts, P. D. (1979). An algorithm for steady-state system optimization and parameter estimation. *International Journal of Systems Science*, *10*, 719–734.

Roberts, P. D., & Williams, T. W. C. (1981). On an algorithm for combined system optimization and parameter-estimation. *Automatica*, *17*, 199–209.

Rodrigues, A. E., & Minceva, M. (2005). Modelling and simulation in chemical engineering: Tools for process innovation. *Computers & Chemical Engineering*, *29*, 1167–1183.

Rumschinski, P., Borchers, S., Bosio, S., Weismantel, R., & Findeisen, R. (2010). Set-base dynamical parameter estimation and model invalidation for biochemical reaction networks. *Bmc Systems Biology*, *50*, 1151–1159.

Rumschinski, P., Streif, S., & Findeisen, R. (2012). Combining qualitative information and semi-quantitative data for guaranteed invalidation of biochemical network models. *International Journal of Robust and Nonlinear Control*, *22*, 1157–1173.

Ruppen, D., Benthack, C., & Bonvin, D. (1995). Optimization of batch reactor operation under parametric uncertainty - computational aspects. *Journal of Process Control*, *5*, 235–240.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2008). *Global sensitivity analysis : the primer*. Chichester, England ; Hoboken, NJ: John Wiley.

Schittkowski, K. (2007). Experimental design tools for ordinary and algebraic differential equations. *Industrial & Engineering Chemistry Research*, *46*, 9137–9147.

Schittkowski, K. (2008). Parameter identification and model verification in systems of partial differential equations applied to transdermal drug delivery. *Mathematics and Computers in Simulation*, *79*.

Skogestad, S. (2000). Self-optimizing control: the missing link between steady-state optimization and control. *Computers & Chemical Engineering*, *24*, 569–575.

Smoljak, S. (1963). Quadrature and interpolation formulas for tensor products of certain classes of functions. *Soviet Math. Dokl*, *4*, 240–243.

Spanos, P., & Ghanem, R. (1991). *Stochastic Finite Elements: A Spectral Approach*. Berlin: Springer.

Srinivasan, B., & Bonvin, D. (2002). Interplay between identification and optimization in run-

to-run optimization schemes. *In Proceedings of the 2002 American Control Conference*, (pp. 2174–2179).

von Stosch, M., Oliveira, R., Peres, J., & de Azevedo, S. F. (2014). Hybrid semi-parametric modeling in process systems engineering: Past, present and future. *Computers & Chemical Engineering*, *60*, 86–101.

Streif, S., Kim, K. K. K., Rumschinski, P., Kishida, M., Shen, D., Findeisen, R., & Braatz, R. (2016). Robustness analysis, prediction, and estimation for uncertain biochemical networks: An overview. *Journal of Process Control*, *42*, 14–34.

Tatjewski, P. (2002). Iterative optimizing set-point control - the basic principle redesigned. *In Proceedings of the 15th Triennial IFAC World Congress*, .

Teixeira, A. P., Clemente, J. J., Cunha, A. E., Carrondo, M. J. T., & Oliveira, R. (2006). Bioprocess iterative batch-to-batch optimization based on hybrid parametric/nonparametric models. *Biotechnology Progress*, *22*, 247–258.

Verma, A. K. (2014). *Process Modelling and Simulation in Chemical, Biochemical and Environmental Engineering*. CRC Press.

Villegas, R. M., Budman, H., & Elkamel, A. (2017). Identification of dynamic metabolic flux balance models based on parametric sensitivity analysis. *Industrial & Engineering Chemistry Research*, *56*, 1911–1919.

Walter, E., & Kieffer, M. (2007). Guaranteed nonlinear parameter estimation in knowledge-based models. *Journal of Computational and Applied Mathematics*, *199*, 277–285.

Walter, E., & Pronzato, L. (1990). Qualitative and quantitative experiment design for phenomenological models - a survey. *Automatica*, *26*, 195–213.

Weijers, S. R., & Vanrolleghem, P. A. (1997). A procedure for selecting best identifiable parameters

in calibrating activated sludge model no.1 to full-scale plant data. *Water Science and Technology*, *36*, 69–79.

Wiener, N. (1938). The homogeneous chaos. *American Journal of Mathematics*, *60*, 897–936.

Wurm, F. (2004). Production of recombinant protein therapeutics in cultivated mammalian cells. *Nature biotechnology*, *22*, 1393–1398.

Xiu, D. (2010). *Numerical Methods for Stochastic Computations: A Spectral Method Approach*. Princeton: Princeton University Press.

Xiu, D. B., & Karniadakis, G. E. (2002a). Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos. *Computer Methods in Applied Mechanics and Engineering*, *191*, 4927–4948.

Xiu, D. B., & Karniadakis, G. E. (2002b). The wiener-askey polynomial chaos for stochastic differential equations. *Siam Journal on Scientific Computing*, *24*, 619–644.

Yahia, B. B., Malphettes, L., & Heinzle, E. (2015). Macroscopic modeling of mammalian cell growth and metabolism. *Applied Microbiology and Biotechnology*, *99*, 7009–7024.

Yao, K. Z., Shaw, B. M., Kou, B., McAuley, K. B., & Bacon, D. W. (2003). Modeling ethylene/butene copolymerization with multi-site catalysts: Parameter estimability and experimental design. *Polymer Reaction Engineering*, *11*, 563–588.

Yip, W. S., & Marlin, T. E. (2004). The effect of model fidelity on real-time optimization performance. *Computers & Chemical Engineering*, *28*, 267–280.

Zamorano, F., Wouwer, A. V., Jungers, R. M., & Bastin, G. (2013). Dynamic metabolic models of cho cell cultures through minimal sets of elementary flux modes. *Journal of Biotechnology*, *164*, 409–422.

Zhang, T., Guay, M., & Dochain, D. (2003). Adaptive extremum seeking control of continuous stirred-tank bioreactors. *AIChE Journal*, *49*, 113–123.

A. Supplementary Material for Chapter 6

A.1. Metabolic Network

The following metabolic network and corresponding reactions, used for the analysis and model development in chapter 6, have been taken from (Niu et al., 2013).

Table A.1.: Metabolic network based on 35 metabolic reactions.

Flux #	**Pathway**	**Reaction**
1	$f_{Glc-G6P}$	$Glc_{ex} + ATP \rightarrow G6P + ADP$
2	$f_{G6P-Pyr}$	$G6P + 2NAD^{+} + 3ADP \rightarrow 2Pyr + 2NADH + 3ATP$
3	$f_{Pyr-Lac}$	$Pyr + NADH \rightarrow Lac_{ex} + NAD^{+}$
4	$f_{Pyr-Pyr_m}$	$Pyr \rightarrow Pyr_m$
5	$f_{Pyr_m-AcC_m}$	$Pyr_m + NAD_m^{+} \rightarrow AcC_m + CO_2 + NADH_m$
6	$f_{Oaa_m-\alpha KG_m}$	$AcC_m + OAA_m + NAD_m^{+} \rightarrow \alpha KG_m + CO_2 + NADH_m$
7	$f_{\alpha KG_m-Suc_m}$	$\alpha KGm + NAD_m^{+} \rightarrow Suc_m + CO_2 + NADH_m$
8	$f_{Suc_m-Fum_m}$	$Suc_m + FAD_m + GDP \rightarrow Fum_m + FADH_{2m} + GTP$
9	$f_{Fum_m-Mal_m}$	$Fum_m \rightarrow Mal_m$

Table A.1.: Metabolic network based on 35 metabolic reactions.

Flux #	Pathway	Reaction
10	$f_{Mal_m-Oaa_m}$	$Mal_m + NAD^+_m \rightarrow OAA_m + NADH_m$
11	$f_{Gln-Glu}$	$Glu_{ex} + NH_{3ex} + ATP \rightarrow Gln_{ex} + ADP$
12	$f_{Glu-\alpha KG_m}$	$Glu_{ex} + NAD^+_m \rightarrow \alpha KG_m + NH_{3ex} + NADH_m$
13	f_{Mal_m-Mal}	$Mal_m \rightarrow Mal$
14	$f_{Mal-Pyr}$	$Mal + NAD^+ \rightarrow Pyr + CO_2 + NADH$
15	$f_{Pyr-Ala}$	$Pyr + Glu_{ex} \rightarrow Ala_{ex} + \alpha KG_m$
16	$f_{Oaam-Asp}$	$OAA_m + Glu_{ex} \rightarrow Asp + \alpha KG_m$
17	$f_{Arg-Glu}$	$Arg + 2NAD^+_m \rightarrow Glu_{ex} + 3NH_{3ex} + CO_2 + 2NADH_m$
18	$f_{Asn-Asp}$	$Asn \rightarrow Asp + NH_{3ex}$
19	$f_{Gly-Ser}$	$2Gly + NAD^+_m \rightarrow Ser + NH_{3ex} + CO_2 + NADH_m$
20	$f_{His-Glu}$	$His + NAD^+_m \rightarrow Glu_{ex} + 2NH_{3ex} + CO_2 + NADH_m$
21	$f_{Ile-Suc_m}$	$Ile + 2NAD^+_m \rightarrow AcC_m + Suc_m + NH_{3ex} + 2NADH_m$
22	$f_{Leu-AcC_m}$	$Leu + 3NAD^+_m \rightarrow 3AcC_m + 3NADH_m$
23	$f_{Lys-AcC_m}$	$Lys + 6NAD^+_m \rightarrow 2CO_2 + 2AcC_m + 6NADH_m$
24	$f_{Met-Suc_m}$	$Met + 4NAD^+_m \rightarrow CO2 + NH_{3ex} + Suc_m + 4NADH_m$
25	$f_{Phe-Tyr}$	$Phe + NAD^+_m \rightarrow Tyr + NADH_m$
26	$f_{Ser-Pyr}$	$Ser \rightarrow Pyr + NH_{3ex}$
27	$f_{Thr-Suc_m}$	$Thr + NAD^+_m \rightarrow Suc_m + NH_{3ex} + NADH_m$

Table A.1.: Metabolic network based on 35 metabolic reactions.

Flux #	Pathway	Reaction
28	$f_{Trp-AcC_m}$	$Trp + 19NAD_m^+ \rightarrow 3AcC_m + 5CO_2 + 19NADH_m$
29	$f_{Tyr-Fum_m}$	$Tyr + 5NAD_m^+ \rightarrow Fum_m + 2AcC_m + CO_2 + 5NADH_m$
30	$f_{Val-Suc_m}$	$Val + 5NAD_m^+ \rightarrow CO2 + Suc_m + 5NADH_m + NH^{3ex}$
31	$f_{NADH-NADH_m}$	$NADH + NAD_m^+ \rightarrow NADH_m + NAD^+$
32	f_{NADH_m-ATP}	$NADH_m + 0.5O_2 + (P/O)ADP \rightarrow NAD_m^+ + (P/O)ATP$
33	$f_{FADH_{2m}-NADH_m}$	$FADH_{2m} + GTP + NAD_m^+ \rightarrow NADH_m + GDP + FAD_m$
34	f_{bio}	$\gamma_{G6P,bio}G6P + \gamma_{Gln,bio}Gln_{ex} + \gamma_{Glu,bio}Glu_{ex} +$ $\gamma_{Ala,bio}Ala_{ex} + \Sigma\gamma_{AAi,bio}C_{wi}H_{xi}N_{yi}O_{zi} + (\kappa_{bio} -$ $6\gamma_{G6P,bio}\kappa_{G6P} - 5\gamma_{Gln,bio}\kappa_{Gln} - 5\gamma_{Glu,bio}\kappa_{Glu} -$ $3\gamma_{Ala,bio}\kappa_{Ala} - \Sigma\gamma_{AAi}wi\kappa_{AAi})NADH^+ y'_{atpbio}ATP \rightarrow$ $Biomass(CH_{\alpha,bio}N_{\beta,bio}O_{\gamma,bio}) + (\kappa_{bio} - 6\gamma_{G6P,bio}\kappa_{G6P} -$ $5\gamma_{Gln,bio}\kappa_{Gln} - 5\gamma_{Glu,bio}\kappa_{Glu} - 3\gamma_{Ala,bio}\kappa_{Ala} -$ $\Sigma\gamma_{AAi}wi\kappa_{AAi})NAD^+ + y_{atpbio}ADP$
35	f_{mab}	$\gamma_{Gln,mab}Gln_{ex} + \gamma_{Glu,mab}Glu_{ex} +$ $\gamma_{Ala,mab}Ala_{ex} + \Sigma\gamma_{AAi,mab}C_{wi}H_{xi}N_{yi}O_{zi} +$ $y_{atpmab}ATP \rightarrow MAb(CH_{\alpha,mab}N_{\beta,mab}O_{\gamma,mab}) + y_{atpmab}ADP$

A.2. Model Parameters

Parameter	Value	Parameter	Value	Parameter	Value
μ	2.52	K_{83}	1.28	K_{161}	11.76
K_{21}	2.53	K_{91}	1.17	K_{170}	0.69
K_{23}	100.04	K_{92}	1.36	K_{171}	14.76
K_{24}	66.75	K_{93}	13.68	K_{180}	7.06
K_{25}	3.22	K_{101}	18.49	K_{181}	20.14
K_{26}	0.10	K_{102}	23.04	K_{190}	1.31
k_d	0.01	K_{103}	64.97	K_{191}	16.66
K_{lys}	0.71	K_{104}	1.00	K_{200}	1.42
K_{41}	0.02	K_{110}	0.92	K_{201}	12.64
K_{42}	5.98	K_{111}	13.55	K_{210}	1.09
K_{43}	91.27	K_{120}	17.34	K_{211}	14.46
K_{51}	0.58	K_{121}	28.71	K_{220}	1.58
K_{52}	4.64	K_{122}	3.21	K_{221}	15.78
K_{53}	1.24	K_{130}	0.16	K_{230}	1.27
K_{54}	87.56	K_{131}	0.06	K_{231}	16.26
K_{61}	0.33	K_{132}	9.57	K_{240}	3.92
K_{62}	0.27	K_{133}	3.91	K_{241}	49.17
K_{63}	2.52	K_{134}	0.10	K_{250}	0.81
K_{64}	0.61	K_{140}	0.01	K_{251}	10.92
K_{65}	0.76	K_{143}	0.29	K_{260}	1.91
K_{71}	0.99	K_{144}	98.95	K_{261}	36.85
K_{72}	1.78	K_{150}	1.29	K_{270}	26.38
K_{81}	0.03	K_{151}	15.79	K_{271}	0.10
K_{82}	0.33	K_{160}	0.72		

Table A.2.: Estimated parameter values from the results shown in figure 6.2.

B. Availablility of the Implementation

Most of the documented methodologies and case studies were implemented in MATLAB (2015b) and are available under https://github.com/RHille/R2R_Batch_Optimization.