

# Experimental Design in Simultaneous Identification and Optimization of Batch Processes under Model-Plant Mismatch

Rubin Hille, Hector M. Budman

Department of Chemical Engineering, University of Waterloo, Waterloo, ON N2L 3G1  
Canada (Tel: 519-888-4567, x36980; e-mail: hbudman@uwaterloo.ca)

---

**Abstract:** 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* 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 a simulated run-to-run optimization study of a cell-culture process.

**Keywords:** Batch Processes, Run-to-Run Optimization, Model-plant Mismatch, Model Correction, Design of Experiments

---

## 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 offer the potential for process optimization (Yip and Marlin, 2004).

However, due to simplifications and assumptions during model development, there regularly occurs 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 and Bonvin, 2002).

When the main use of the model is optimization, methods such as *Modifier Adaptation* (Marchetti et al., 2009) 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 behaviour around the optimum. For such cases the method for *simultaneous identification and optimization* (Mandur and Budman, 2015) 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, up to this point, 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 (Marchetti et al., 2010; Costello et al., 2016), has so far not been addressed in the simultaneous identification and optimization framework.

The Design of Experiments (DoE) methodology, first derived for data-driven models (Box and Draper, 1987), is also an established method for reducing parameter uncertainty in the estimation of nonlinear mechanistic models (Franceschini and 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). However, as mentioned above, in the presence of model-plant mismatch, a precise fitting of model outputs does not necessarily 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). Consequently, the goal of the simultaneous identification and optimization framework used in this work, is not only to fit model outputs to measured ones, but also to match 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 work presents an approach to address the limitations of the previous implemented parameter adaptation methodology by incorporating a design of experiments approach. 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 also considered.
- ii. The improved parameter precision leads to a better prediction of the cost function near the process optimum.

For illustration purposes, a run-to-run optimization study is performed using a simulated case study of a fed-batch penicillin process.

## 2. SIMULTANEOUS IDENTIFICATION AND OPTIMIZATION METHODOLOGY

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

### 2.1 Identification Using Set-Based Bounds

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

$$\mathbf{y}_k = \{\mathbf{y}_k^i \in \mathbb{R}^{n_y} | i \in \{1, \dots, n_t\}\} \quad (1)$$

where the set-based bounds provide an upper and lower bound for the permissible range of model outputs at each sampling time such that:

$$\underline{\mathbf{y}}_i \leq \mathbf{y}_i \leq \bar{\mathbf{y}}_i \quad (2)$$

with the model outputs given by  $\mathbf{y} \in \mathbb{R}^{n_y}$ . 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:

$$\phi_{SSE}(\boldsymbol{\theta}) = \sum_{i=1}^{n_t} \|\mathbf{y}_p(\mathbf{u}_k, t_i) - \mathbf{y}(\mathbf{u}_k, \boldsymbol{\theta}, t_i)\|^2 \quad (3)$$

where  $\mathbf{y}_p \in \mathbb{R}^{n_y}$  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 and Budman (2017) proposed the following parametric sensitivity objective:

$$\phi_{S^\nabla}(\boldsymbol{\theta}) = \sum_{i=1}^{n_u} \sum_{j=1}^{n_\theta} |s_{ij}^{\nabla\phi}(\boldsymbol{\theta})| + \sum_{i=1}^{n_u \times n_g} \sum_{j=1}^{n_\theta} |s_{ij}^{\nabla g}(\boldsymbol{\theta})| \quad (4)$$

where  $s_{ij}^{\nabla\phi}$  and  $s_{ij}^{\nabla g}$  are elements of the scaled cost-function and constraint gradient sensitivity matrices, i.e.  $s_{ij}^{\nabla\phi} = \frac{\partial(\nabla\phi_i)}{\partial\theta_j} \left| \frac{\theta_j}{\nabla\phi_i} \right|$ .

Equation (4) defines a scalar measure of the parametric cost function and constraint sensitivities where large gradient sensitivities are desired to obtain parameter values which lead to a more accurate matching of the gradients. An objective combining the model-fitting goal (3) and the maximization of the sensitivity measure in (4) is subsequently defined as:

$$\begin{aligned} \boldsymbol{\theta}_k &= \arg \min_{\boldsymbol{\theta}} \left( \frac{\phi_{SSE}(\mathbf{u}_k, \boldsymbol{\theta})}{\phi_{S^\nabla}(\mathbf{u}_k, \boldsymbol{\theta})} \right) \\ \text{s. t.} \quad &\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}_k, \boldsymbol{\theta}) \\ &\mathbf{y} = \mathbf{h}(\mathbf{x}) - \mathbf{c}_{k-1} \\ &\boldsymbol{\theta} \in \boldsymbol{\Theta}_0 \\ &\mathbf{y} \in \mathcal{Y}_k \end{aligned} \quad (5)$$

The correction term  $\mathbf{c}_{k-1}$  is defined in the gradient correction step described below. According to (5), 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. Note that since the optimization cost was modified from a norm of the prediction errors as in (3) to a norm of the errors divided by the sensitivity function as in (5), the set-based bounds are necessary to enforce that the predicted outputs remain reasonably close to the process outputs.

### 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:

$$\Delta\boldsymbol{\theta}_k = \arg \min_{\Delta\boldsymbol{\theta}} (\mathbf{w}_\phi^T |\nabla\phi_p(\mathbf{u}_k) - \nabla\phi(\mathbf{u}_k, \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta})|)$$

$$+ \mathbf{w}_g^T |\nabla \mathbf{g}_p(\mathbf{u}_k) - \nabla \mathbf{g}(\mathbf{u}_k, \boldsymbol{\theta}_k + \Delta \boldsymbol{\theta})|$$

$$\begin{aligned} \text{s. t. } \quad & \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}_k, \boldsymbol{\theta}_k + \Delta \boldsymbol{\theta}) \\ & \mathbf{y} = \mathbf{h}(\mathbf{x}) - \mathbf{c}_k \\ & \|\boldsymbol{\epsilon}^T\|_\infty \leq \epsilon_{max} \end{aligned} \quad (6)$$

where  $\nabla \boldsymbol{\phi}$  and  $\nabla \mathbf{g}$  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  $\mathbf{w}_\phi$  and  $\mathbf{w}_g$ . A correction factor  $\mathbf{c}_k$  is introduced into the model outputs as to preserve the fitting accuracy that has been achieved in the identification step (5). The correction term is derived from a first order Taylor expansion:

$$\mathbf{c}_k(t_i) = \mathbf{c}_{k-1}(t_i) + \mathbf{D}\mathbf{y}(\boldsymbol{\theta}_k, t_i)\Delta\boldsymbol{\theta}_k \quad (7)$$

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

$$\begin{aligned} \boldsymbol{\epsilon}(t_i) &= [\mathbf{y}(\mathbf{u}_k, \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}_k, t_i) - \mathbf{D}\mathbf{y}(\boldsymbol{\theta}_k, t_i)\Delta\boldsymbol{\theta}_k \\ &\quad - \mathbf{y}(\mathbf{u}_k, \boldsymbol{\theta}_k, t_i)] \cdot [\text{diag}(\mathbf{y}(\mathbf{u}_k, \boldsymbol{\theta}_k, t_i))]^{-1} \end{aligned} \quad (8)$$

### 2.3 Model-based Optimization

Following the identification (5) and the gradient correction steps (6), a model-based optimization is performed by using the updated parameter values  $\boldsymbol{\theta}'_k = \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}_k$ :

$$\begin{aligned} \mathbf{u}_{k+1} &= \arg \min_{\mathbf{u}} \phi(\mathbf{u}, \mathbf{y}(\mathbf{u}, \boldsymbol{\theta}'_k)) \\ \text{s. t. } \quad & \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta}'_k) \\ & \mathbf{y} = \mathbf{h}(\mathbf{x}) - \mathbf{c}_k \\ & \mathbf{g}(\mathbf{y}(\mathbf{u}, \boldsymbol{\theta}'_k), \mathbf{u}) \leq \mathbf{0} \\ & \mathbf{u}^L \leq \mathbf{u} \leq \mathbf{u}^U \end{aligned} \quad (9)$$

## 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 in (6), one drawback is that the gradients are only corrected at the current operating point. In other words, information from past operating points is 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 noisy gradient measurement. 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. Therefore, to introduce additional robustness to uncertainty in gradient measurements and to acquire parameter values that lead to a more accurate representation of the process cost-function, we 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 will be determined based on an optimal experimental design approach. Thus, 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.

### 3.1 Local Gradient Correction

In the simultaneous identification and optimization approach, gradient measurements are required to satisfy the necessary conditions of optimality as per the gradient correction step described in (6). Regarding the cost function, a gradient can be defined as the derivative with respect to the decision variables:

$$\nabla_{u_m} \phi(\mathbf{u}_k) = \frac{\partial \phi(\mathbf{u}_k)}{\partial u_m} \quad (10)$$

where  $m \in \{1, \dots, n_u\}$  denotes the respective decision variable. The gradient at operating point  $\mathbf{u}_k$  can be estimated by performing a step change  $\Delta u_m$  in the direction of each decision variable. Using finite differences, the derivative can be approximated by:

$$\alpha_{k,m}(\mathbf{u}_k) = \frac{\phi(\mathbf{u}_k + \Delta u_m \mathbf{e}_m) - \phi(\mathbf{u}_k)}{\|\Delta u_m \mathbf{e}_m\|} \quad (11)$$

where  $\mathbf{e}_m$  is the identity vector in the direction of the respective decision variable. Accordingly, the standard gradient correction of cost-function gradients from (6) can be reformulated as follows:

$$\begin{aligned} \Delta\boldsymbol{\theta}_k &= \arg \min_{\Delta\boldsymbol{\theta}} \sum_{m=1}^{n_u} w_m^\alpha \|\alpha_{k,m}^p(\mathbf{u}_k) - \alpha_{k,m}(\mathbf{u}_k, \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta})\|^2 \\ \text{s. t. } \quad & \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}_k, \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}) \\ & \mathbf{y} = \mathbf{h}(\mathbf{x}) - \mathbf{c}_k \\ & \|\boldsymbol{\epsilon}^T\|_\infty \leq \epsilon_{max} \end{aligned} \quad (12)$$

where  $w_m^\alpha$  is a normalizing weight and the superscript  $p$  denotes the approximated cost function derivative (11) estimated from plant measurements.

### 3.2 Consideration of Information from Prior Experiments

Besides the gradient measurements that can be acquired by perturbing the plant at operating point  $\mathbf{u}_k$  (11), 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  $\mathbf{u}_k$  and past ones as follows:

$$\Delta\boldsymbol{\Phi}_k = [\phi_k - \phi_{k-1} \quad \phi_k - \phi_{k-2} \quad \dots \quad \phi_k - \phi_{k-n_b-1}]^T \quad (13)$$

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

$$\Delta \mathbf{u}_k = [\mathbf{u}_k - \mathbf{u}_{k-1} \quad \mathbf{u}_k - \mathbf{u}_{k-2} \quad \dots \quad \mathbf{u}_k - \mathbf{u}_{k-n_b-1}]^T \quad (14)$$

Following (11), for any two operating points  $k$  and  $k - l$  we can define the following finite difference:

$$\beta_{k,l}(\mathbf{u}_k) = \frac{\Delta \Phi_{k,l}}{\|\Delta \mathbf{u}_{k,l}\|} = \frac{\phi_k - \phi_{k-l}}{\|\mathbf{u}_k - \mathbf{u}_{k-l}\|} \quad (15)$$

with  $l = 1, \dots, n_b - 1$ . Due to the uncertainty in the measured cost, we are ultimately interested in past operating points that are sufficiently far away so as to reduce the effect of gradient uncertainty when considering past cost-function evaluations. For that reason, from the differences with respect to previous experiments (15), 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:

$$\mathcal{L}_\varepsilon = \left\{ l \in \{1, \dots, n_\beta\} \mid 1 - \frac{\phi(\mathbf{u}_k)}{\phi(\mathbf{u}_{k-l})} \geq \varepsilon_\phi \right\} \quad (16)$$

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

$$\varepsilon_\phi = \frac{\sigma_\phi}{\mu_\phi} \quad (17)$$

### 3.3 Design of New Experiments

In addition to using past cost measurements as above, we propose to use optimal DoE to acquire future cost information. The goal is to identify future operating points for experiments that are more informative in terms of cost information instead of the fixed perturbations done at current  $k$  as per (11). As before, these experiments are run in addition to the experiments conducted at the current optimal input determined by the model-based optimization (9). Thus, the goal of the proposed experimental design is to replace the fixed perturbations in the direction of each decision variable with perturbations that are more informative as per an experimental design criterion. We first define a parametric sensitivity matrix  $\mathbf{S}_\beta$  of the coefficients of past experiments, whose elements are defined as follows:

$$\mathbf{S}_\beta = \left[ \frac{\partial(\beta_{k,l})}{\partial \theta_j} \right] \quad \forall l \in \mathcal{L}_\varepsilon \quad \forall j \in \{1, \dots, n_\theta\} \quad (18)$$

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

$$\psi = \det(\mathbf{V}_\theta) = \det([\mathbf{S}_\beta^T \mathbf{\Sigma}_\beta^{-1} \mathbf{S}_\beta]^{-1}) \quad (19)$$

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

measurement noise, i.e.  $\sigma_\beta^2 = \frac{2\sigma_\phi^2}{\|\mathbf{u}_k - \mathbf{u}_{k-l}\|^2}$ . Where the measurement noise in the cost, necessary for the gradient estimation, remains unchanged and is uncorrelated. Therefore, the aim of the experimental design is primarily to find the plant perturbation vectors which provide information as to complement the information already gained from past experiments. To this end, we note that the gradient estimator in (11) can also be formulated for directions other than the directions associated with the individual decision variables as follows:

$$\gamma_{k,q}(\mathbf{u}_k, \mathbf{v}_q) = \frac{\phi(\mathbf{u}_k + \mathbf{v}_q) - \phi(\mathbf{u}_k)}{\|\mathbf{v}_q\|} \quad (20)$$

where  $q \in \{1, \dots, n_{DOE}\}$  presents the number of plant perturbations at each operating point implemented to acquire the gradient information. The vector  $\mathbf{u}_k + \mathbf{v}_q$  presents a perturbation of the plant in the neighbourhood of  $\mathbf{u}_k$ . The experimental design goal is to select the appropriate perturbation vectors  $\mathbf{v}_q$  so as to increase parameter precision when performing the fitting of measured gradients.

The procedure for designing new experiments can be summarized as follows:

1. Initialize by setting  $q = 0$ ,  $\mathbf{S}_\phi^q = \mathbf{S}_\beta$  and  $\mathbf{\Sigma}_\phi^q = \mathbf{\Sigma}_\beta$ .
2. Based on the considered prior experiments, find the perturbation which provides the most additional information as per the D-optimality criterion:

$$\begin{aligned} \mathbf{v}_q &= \arg \min_{\mathbf{v}} \det([\mathbf{S}^T \mathbf{\Sigma}^{-1} \mathbf{S}]^{-1}) \\ \text{s. t. } \mathbf{S} &= \begin{bmatrix} \mathbf{S}_\phi^q \\ \mathbf{s}_\gamma \end{bmatrix}, \mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_\phi^q & \mathbf{0} \\ \mathbf{0} & \sigma_\gamma^2 \end{bmatrix} \\ \mathbf{s}_\gamma &= \begin{bmatrix} \frac{\partial \gamma}{\partial \theta_1} & \dots & \frac{\partial \gamma}{\partial \theta_{n_\theta}} \end{bmatrix} \\ \gamma(\mathbf{u}_k, \mathbf{v}) &= \frac{\phi(\mathbf{u}_k + \mathbf{v}) - \phi(\mathbf{u}_k)}{\|\mathbf{v}\|} \\ 1 - \frac{\phi(\mathbf{u}_k)}{\phi(\mathbf{u}_k + \mathbf{v})} &= \varepsilon_\phi \end{aligned} \quad (21)$$

3. Update the sensitivity matrix  $\mathbf{S}_\phi^{q+1} = \mathbf{S}$ ,  $\mathbf{\Sigma}_\phi^{q+1} = \mathbf{\Sigma}$  and set  $q = q + 1$ .
4. Go back to step 2 until the number of gradient measurements  $n_{DOE}$  is reached.

Notice that we incorporated an equality constraint in (21), as it is desired to introduce a minimum distance when designing new experiments to reduce the estimated variance of the gradient measurement. 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.

### 3.4 Extended Gradient Correction

After carrying out the set of cost function measurements from the optimal plant perturbations provided by the procedure in (21), the standard gradient correction method from (12) 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:

$$\Delta \boldsymbol{\theta}_k = \arg \min_{\Delta \boldsymbol{\theta}} \sum_{q=1}^{n_{DOE}} w_q^\gamma \|\gamma_{k,q}^p(\mathbf{u}_k) - \gamma_{k,q}(\mathbf{u}_k, \boldsymbol{\theta}_k + \Delta \boldsymbol{\theta})\|^2 + \sum_{l \in \mathcal{L}_\varepsilon} w_l^\beta \|\beta_{k,l}^p(\mathbf{u}_k) - \beta_{k,l}(\mathbf{u}_k, \boldsymbol{\theta}_k + \Delta \boldsymbol{\theta})\|^2$$

$$\text{s. t. } \begin{aligned} \dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x}, \mathbf{u}_k, \boldsymbol{\theta}_k + \Delta \boldsymbol{\theta}) \\ \mathbf{y} &= \mathbf{h}(\mathbf{x}) - \mathbf{c}_k \\ \|\boldsymbol{\epsilon}^T\|_\infty &\leq \epsilon_{max} \end{aligned} \quad (22)$$

## 5. RESULTS AND DISCUSSION

### 5.1 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):

$$\frac{dX}{dt} = \left( \frac{\mu_X SX}{K_X X + S} \right) - \frac{X}{V} \frac{dV}{dt} \quad (23)$$

$$\frac{dP}{dt} = \left( \frac{\mu_P SX}{K_P + S + \frac{S^2}{K_I}} \right) - K_H P - \frac{P}{V} \frac{dV}{dt} \quad (24)$$

$$\frac{dS}{dt} = -\frac{1}{Y_{X/S}} \left( \frac{\mu_X SX}{K_X X + S} \right) - \frac{1}{Y_{P/S}} \left( \frac{\mu_P SX}{K_P + S + \frac{S^2}{K_I}} \right) - m_X X + \frac{F S_f}{V} - \frac{S}{V} \frac{dV}{dt} \quad (25)$$

$$\frac{dV}{dt} = F - 6.226 \cdot 10^{-4} V \quad (26)$$

where  $X$ ,  $P$  and  $S$  describe the respective concentrations of biomass, penicillin and substrate. The volume in the reactor is given by  $V$ . The simulator (23) – (26) 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 part of the process under study. We propose to ignore the hydrolysis term in the penicillin equation. Therefore, the model used in the optimization scheme is defined by (23), (25), (26) and:

$$\frac{dP}{dt} = \left( \frac{\mu_P SX}{K_P + S + \frac{S^2}{K_I}} \right) - \frac{P}{V} \frac{dV}{dt} \quad (27)$$

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

$$\min_{S_0, F} -P(\mathbf{x}, \boldsymbol{\theta}, S_0, F, t_f)$$

$$\text{s. t. } (23) \text{ and } (25) - (27)$$

$$V(\mathbf{x}, \boldsymbol{\theta}, S_0, F, t_f) \leq V_{max} \quad (28)$$

where a constraint on the volume of the reactor is given by  $V_{max} = 120L$ . The initial values used for the first operating point are given in table 1, where  $S_0$  and  $F$  are determined by the model-based optimization.

Biomass conc. ( $X_0$ )	0.1 (g/L)
Substrate conc. ( $S_0$ )	0.1 (g/L)
Product conc. ( $P_0$ )	0 (g/L)
Initial culture volume ( $V_0$ )	100 (L)
Input Feed ( $F$ )	0.04 (L/h)

Table 1: Initial operating conditions for the simulator

The initial values of the model parameters are given in table 2. 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  $\mu_X$ ,  $\mu_P$  and  $m_X$  are used for model update and gradient correction.

$\mu_X$	$K_X$	$\mu_P$	$K_P$	$K_I$	$Y_{X/S}$	$Y_{P/S}$	$m_X$
0.092	0.37	0.008	0.0002	0.1	0.45	0.9	0.014

Table 2: Initial 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 (11). In this case, a fixed step size of  $\Delta S_0 = 2$  g/L and  $\Delta F = 0.5$  L/h is used for the initial substrate concentration and the constant flow rate respectively. Note that, as the optimal fixed flowrate of  $F^* = 0.17$  L/h is obtained within one iteration (due to the constraint), we used the proposed approach for the improving the predictions capabilities in terms of the initial substrate concentration. Accordingly, in this case, the optimal step size (or plant perturbation)  $\Delta S_0$  was determined based on the available past experimental data and DoE methodology (21).

### 5.2 Results

To compare the performance of the proposed approach, we conducted 10 run-to-run simulations for the following two cases: I - only the local gradient is used for correction. II - extended gradient correction is used involving past and current gradients using DoE. The upper bound on the relative truncation error in (22) was selected to be  $\epsilon_{max} = 0.03$  while the maximum number of considered past experiments in (16) is  $n_\beta = 20$ .

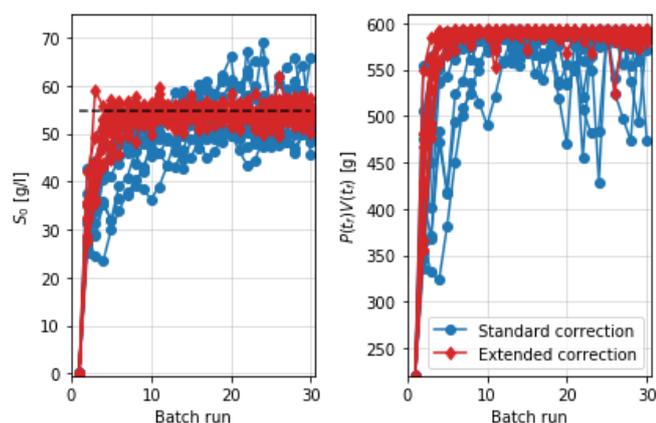


Figure 1: Performance comparison when using information from prior experiments and experimental design.

The run-to-run optimization results are shown in figure 1, where the left-hand graph shows the comparison in terms of the manipulated variable  $S_0$ . The proposed approach shows a speed-up in convergence to the process optimum, which leads to a ca. 28% improvement in the IAE. Since 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, leading to higher average yield as shown on the right graph in figure 1. This is especially evident around the optimum (“plateau”), as shown in Figure 2, where the use of already available 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.

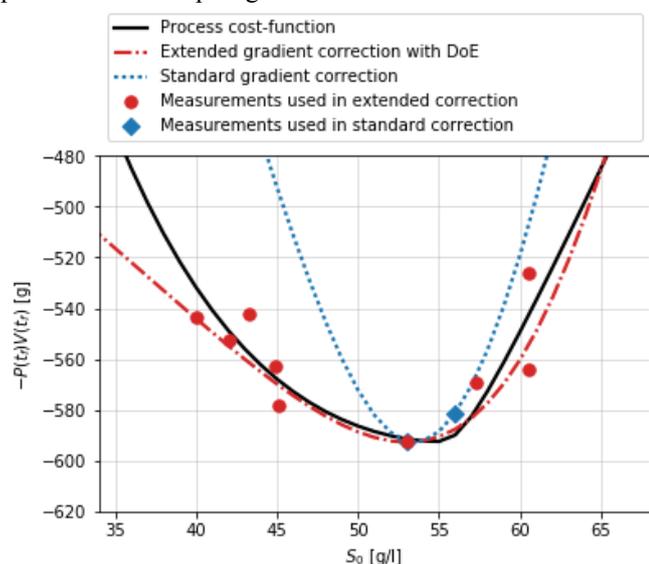


Figure 2: Example of the predicted cost-functions for  $S_0$  in the neighbourhood of the process optimum.

## CONCLUSIONS

An experimental design methodology was presented, where the goal is to achieve a better prediction of the cost-function around the optimum. Based on available cost information from previous experiments, a DoE is implemented to determine the

next optimal experiments. Subsequently, an extended gradient correction is implemented to estimate parameter values that lead to a more accurate representation of the process cost-function compared to a correction that only uses local gradient measurements. The proposed approach shows promising results and could also be useful for the optimization of continuous processes.

## ACKNOWLEDGEMENTS

NSERC (National Science and Engineering Research Council of Canada) is acknowledged for the financial support.

## REFERENCES

- Biol, G., Undey, C., and Cinar, A. (2002). A modular simulation package for fed-batch fermentation: penicillin production. *Comput Chem Eng*, 26 (11), 1553-1565.
- Bonvin, D., Georgakis, C., Pantelides, C. C., Barolo M., Grover, M.A., Rodrigues, D., Schneider, R., and Dochain, D. (2016). Linking models and experiments. *Ind Eng Chem Res*, 55, 6891-6903.
- Box, G., and Draper, N.R. (1987). *Empirical Model-Building and Response Surfaces*. John Wiley & Sons, New York.
- Costello, S., Francois, G., and Bonvin, D. (2016). A directional modifier-adaptation algorithm for real-time optimization. *J Process Control*, 39, 64-76.
- Franceschini, G., and Macchietto, S. (2008). Model-based design of experiments for parameter precision: State of the art. *Chem Eng Sci*, 63, 4846-4872.
- Hille, R., and Budman, H.M. (2017). Run-to-run optimization of batch process using set-based constraints. *IFAC World Congress*, [in press].
- Hille, R., Mandur, J. Budman, H.M. (2017). Robust batch-to-batch optimization in the presence of model-plant mismatch and input uncertainty. *AIChE J*, 63, 2660-2670.
- Houska, B., Telen, D., Logist, F., Diehl, M., and Van Impe, J.F.M (2015). An economic objective for the optimal design of nonlinear dynamic processes. *Automatica*, 51, 98-103.
- Mandur, J., and Budman, H. (2015). Simultaneous model identification and optimization in presence of model-plant mismatch. *Chem Eng Sci*, 129, 106-115.
- Marchetti, A., Chachuat, B., and Bonvin, D. (2009). Modifier-adaptation methodology for real-time optimization. *Ind Eng Chem Res*, 48, 6022-6033.
- Marchetti, A., Chachuat, B., and Bonvin, D. (2010). A dual modifier-adaptation approach for real-time optimization. *J Process Control*, 20, 1027-1037.
- Rumschinski, P., Borchers, S., Bosio, S., Weismantel, R., and Findeisen, R. (2010). Set-base dynamical parameter estimation and model invalidation for biochemical reaction networks. *Bmc Systems Biology*, 50, 1151-1159.
- Srinivasan, B., and Bonvin, D. (2002). Interplay between identification and optimization in run-to-run optimization schemes. *In Proceedings ACC*, 2174-2179.
- Yip, W. S., and Marlin, T. E. (2004). The effect of model fidelity on real-time optimization performance. *Comput Chem Eng*, 28, 267-280.