

Auto-updating of Sampling Time Redesign for System Identification under Parameter Uncertainty

Ke Wang and Hong Yue

Department of Electronic & Electrical Engineering
University of Strathclyde, Glasgow, UK
Email: k.wang@strath.ac.uk; hong.yue@strath.ac.uk

Abstract— In this paper, the mid-term experimental redesign of sampling strategy has been studied to improve design robustness under model uncertainties. With this method, the whole design time horizon is divided into several zones. Parameter estimation is made with the data in each zone and the updated model is used for experimental design in the next zone until the experiment is complete. A novel auto-updating strategy is developed to determine the length of each zone to assure the identifiability for parameter estimation. No pre-settings from users are required as in previous redesign algorithms. Simulation studies on an enzyme reaction system have been conducted. The results demonstrate that, compared to the conventional offline design and the standard online redesign methods, the proposed mid-term redesign with auto-updating of zones produces data that lead to more accurate parameter identification under large uncertainties.

Keywords – parameter estimation; mid-term experimental redesign; model uncertainty; parameter identifiability; time sampling strategy

I. INTRODUCTION

Building reliable mathematical models for dynamic systems is not only beneficial to scientific research and development, for example, it can provide the information of underlying systems that are not apparent in laboratory experimental studies due to the limitation of equipment or other resources, but also useful for industries as it will help instruct the manufacturing process to obtain the improved products, or even create new products [1]. In most cases, the mathematical model is formed by a set of differential equations according to the underlying physical laws with a bunch of unknown parameters. Thus, one main focus of the model building problem is to find the credible parameter values according to the observation data [2], which is also referred as parameter identification.

Since the observation data used for modelling are often sparse and polluted by measurement noises, optimal experimental design (OED) has been proposed to acquire the most informative experimental data for model discrimination and parameter estimation. This technology is especially useful when limited time and resources are available for experiments. A typical OED is model-based and the design result is affected by model uncertainties. Robust experimental design (RED) methods are developed for situations where large model uncertainties need to be considered. Iterative experimental design methods can be used to overcome the limitations due to lack of model

knowledge. According to the availabilities of equipment and/or time, sequential iterative design (less demanding on equipment), parallel iterative design (less demanding on time), or sequential-parallel mixed design can be employed [3]. One benefit of those iterative designs is that parameter estimations can be updated in each iteration. Thus, more accurate parameters will be adopted for OED in the next iteration. This improvement is achieved with a price of costing more time and resources. The worst case design and the pseudo-Bayesian design are alternative methods for RED under model uncertainties. Similar to the iterative designs, these two methods are also “offline” in nature which use information covering the complete experiment process.

More recently, an “online” optimal experimental redesign has been proposed, in which the information content is updated over the running process of the experiment following the pre-set time zones [4]. Therefore, the information content contained in the data is used progressively over the experiment process. The main idea of this mid-term experimental redesign is that the whole experimental process is divided into several design stages, which will be called sub-zones or sub-experiments in the following. Model parameters will be estimated using all historical data collected from the previous sub-experiments, and the updated model is then used as a priori for the next sub-experiment design. This idea was originally proposed in [5-7] for linear systems, and later on extended to nonlinear systems [8]. For nonlinear systems, issues such as ill conditioning need to be addressed. In [9], the disturbance estimation was embedded in the model-based experimental design. In [10], the online redesign strategy was improved by adding a procedure of subset selection to select the identifiable parameters so that the well-conditioned optimisation problem can be guaranteed. The “online” experimental redesign was also extended to parallel experiments [11, 12], in which the sliding-window optimal experimental redesign was proposed.

The above mid-term experimental redesign methods show good potential for experimental design under model uncertainties. However, the choice of sub-experiment zones, including the number of zones and the length of each zone, needs to be determined with the preliminary knowledge of the system. A lack of such knowledge may lead to experimental data with insufficient information for parameter estimation for each zone. To tackle this problem, in this work, an auto-updating strategy is proposed based

on the identifiability criterion for parameter estimation under model uncertainties. This novel strategy is tested on an enzyme reaction system and compared with both offline design methods and online redesign with user-set updating strategy.

The remaining of the paper is organised as follows. In Section II, the methodology for model-based experimental design including OED, RED, online experimental redesign, and the novel auto-updating redesign, is presented. In Section III, the mid-term experimental redesign is tested on an enzyme reaction system. Its robustness to parameter uncertainty is compared with the conventional OED and the redesign with pre-setting of sub-experiments. Besides, the auto-updating strategy is added to the mid-term experimental redesign procedure, and the performance is compared with the user pre-setting strategy. Conclusions and future work are given in Section IV.

II. METHODOLOGY

A. Offline Robust Exeprimental Design

In model-based OED, the model is often described by a series of nonlinear ordinary differential equations (ODEs):

$$\dot{\mathbf{X}}(t) = \mathbf{f}(\mathbf{X}(t), \boldsymbol{\theta}), \mathbf{X}(t_0) = \mathbf{X}_0 \quad (1)$$

$$\mathbf{Y}(t) = \mathbf{h}(\mathbf{X}(t), \boldsymbol{\theta}) + \boldsymbol{\xi}(t) \quad (2)$$

where $\mathbf{X} = [x_1, x_2, \dots, x_n]^T$ is a $n \times 1$ vector of state variables, $\mathbf{f}(\cdot)$ is a set of state transition functions of the system dynamics which are assumed to be continuous and first-order differentiable, $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_p]^T$ is a $p \times 1$ vector of parameters, $\mathbf{Y} = [y_1, y_2, \dots, y_m]^T$ is a $m \times 1$ vector of measurable outputs, $\mathbf{h}(\cdot)$ is normally a selection function and $\boldsymbol{\xi}$ is a vector of measurement errors which is assumed to be independent and identically distributed, zero mean Gaussian noise.

According to the Cramer-Rao lower bound inequality, the Fisher information matrix (FIM) can be used as an evaluation of the information content. For linear models, FIM is approximately the inverse of the parameter estimation error covariance matrix \mathbf{Q} with the assumption of measurement error being uncorrelated and Gaussian distributed and the assumption of unbiased parameter estimation [2]. FIM can be calculated using the sensitivity matrix:

$$\mathbf{F} = \sum_{i=1}^N \mathbf{S}^T(t_i) \cdot \mathbf{Q}^{-1} \cdot \mathbf{S}(t_i) \quad (3)$$

where \mathbf{S} is the local sensitivity matrix with elements $s_{ij} = \frac{\partial x_i}{\partial \theta_j}$. Denoting the design factors which characterise the experiment into a vector, $\boldsymbol{\varphi}$, the OED can be formed as:

$$\boldsymbol{\varphi}^* = \arg \max_{\boldsymbol{\varphi} \in \Phi} \psi(\mathbf{F}) \quad (4)$$

where $\psi(\cdot)$ is a scaler function that can evaluate the ‘‘size’’ of the information matrix, Φ is the admissible space of the design factors. Standard criteria including D-, A-, E-optimality are devised for features such as the determinant, the trace and the minimum eigenvalue of FIM respectively.

When parameter uncertainty exists, RED methods are needed. Typical RED methods include the worst-case design and the pseudo-Bayesian design [13]. The worst-

case method, which can also be called the max-min method, adopts the idea of optimizing the worst case of the design problem to reduce the influence caused by parameter uncertainties. The optimisation problem can be formed as:

$$\boldsymbol{\varphi}^* = \arg \max_{\boldsymbol{\varphi} \in \Phi} \min_{\boldsymbol{\theta} \in \Theta} \psi(\mathbf{F}) \quad (5)$$

where Θ is the range for $\boldsymbol{\theta}$. The pseudo-Bayesian design is often called the expected value method. It obtains the design vector by averaging the classical design criteria over prior distributions of uncertain parameters, e.g. the probability density function $p(\boldsymbol{\theta})$, to account for the effect of parameter uncertainties.

$$\boldsymbol{\varphi}^* = \arg \max_{\boldsymbol{\varphi} \in \Phi} \int \psi(\mathbf{F}) \cdot p(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (6)$$

B. Online Experimental Redesign

A general procedure of online model-based experimental redesign is shown as a flowchart in Fig. 1.

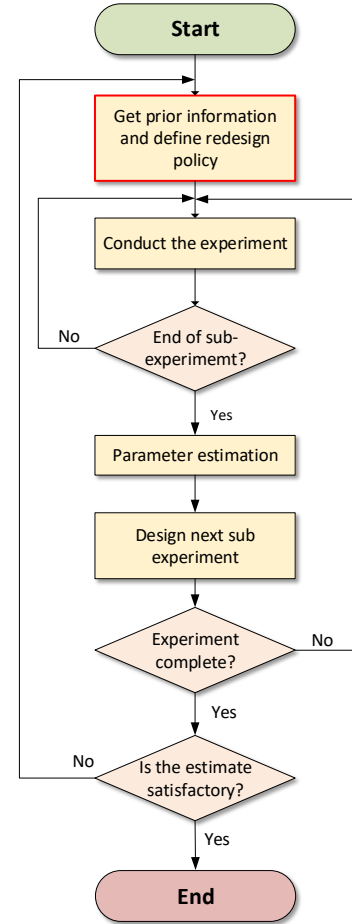


Figure 1. General procedure of online experimental redesign

The redesign policy, as shown in the red square, needs to be given by users before the experiment begins. Besides, identifiability problem will easily arise if any sub-experiments do not cover adequate dynamics of the system for parameters to be estimated.

C. Auto-updating Strategy for Redesign

In this section, the length of each sub-experiment will be selected according to the identifiability of parameters. Correspondingly, the number of sub-experiments will be

determined once the whole length and the sub-experiments are both known.

According to the subset selection method, which is proposed by [10], the identifiable parameters will be found by checking the condition number $\kappa(\mathbf{S})$ and the collinearity index $\gamma(\mathbf{S})$ when a fixed amount of data is available. In our case, the parameters that need to be estimated are already known. It is the data length of every sub-experiment that should be decided to ensure those parameters are identifiable.

In this case, in order to decide the length of the k th sub-experiment, the local sensitivity will be calculated using parameters estimated at the end of the $(k-1)$ th sub-experiment. Then, the singular value decomposition (SVD) will be conducted on local sensitivity matrix \mathbf{S} as a rank-revealing factorisation:

$$\mathbf{S} = \mathbf{U}_D \cdot \boldsymbol{\Sigma}_D \cdot \mathbf{V}_D^T \quad (7)$$

where \mathbf{U}_D and \mathbf{V}_D are real or complex unitary matrix; \mathbf{V}_D^T is the conjugate transpose of \mathbf{V}_D ; $\boldsymbol{\Sigma}_D$ is a rectangular diagonal matrix with non-negative elements $\sigma_1 \geq \sigma_2 \dots \geq \sigma_p \geq 0$ on the diagonal. Those diagonal elements are singular values of matrix \mathbf{S} . After the decomposition, the parameter θ_i that should be selected into the identifiable subset must satisfy two requirements:

$$\kappa_i = \frac{\sigma_1}{\sigma_i} \leq \kappa^{\max} \quad (8)$$

$$\gamma_i = \frac{1}{\sigma_i} \leq \gamma^{\max} \quad (9)$$

The threshold of κ_i which can be written as κ^{\max} is defined as 1000; the threshold of γ_i which can be written as γ^{\max} should be 10-15 empirically according to [10].

All condition numbers and collinearity indexes will be checked for parameters that need to be estimated. If any of the thresholds are not satisfied, more data should be included in this sub-experiment, otherwise this sub-experiment can be ended. This procedure will continue until the end of the whole experiment. Therefore, the sub-experiments are decided automatically.

Since identifiability is the only criterion in the sub-experiment selection procedure, information content will be taken into consideration when choosing the number of samplings in every sub-experiment. Suppose n_1 samplings are selected in the first sub-experiment. Adopting E-optimality, the number of samplings of the k th sub-experiment should be:

$$n_k = \frac{\lambda_{\max}(F_k)}{\lambda_{\max}(F_{k-1})} \cdot n_{k-1}, \quad k = 2, 3, \dots, N_T \quad (10)$$

where N_T is the number of sub-experiments.

With this auto-updating strategy, only the sampling number of the first sub-experiment needs to be defined by the user in the auto-updating strategy. The length of the following sub-experiments and the number of sub-experiments will be decided according to the information of the system automatically. Therefore, less intervene of the user is involved.

III. CASE STUDY

A case study is conducted using mid-term model-based experimental redesign for an enzyme reaction system. The detailed reaction process, model description, nominal values of parameters and the initial condition of states can be found in [14]. After the sensitivity analysis, three parameters, k_2 , k_{-3} and k_5W are determined as the three most important parameters to be estimated. The nominal values of these parameters are $k_2 = 100$, $k_{-3} = 200$ and $k_5W = 5000$. The objective of RED is the sampling strategy in time. Five concentrations, S, N, P, Q, R are measured using the same sampling time profile.

Firstly, we will check the online experimental redesign's ability of dealing with the parameter uncertainty. Then we will test if our auto-updating strategy improved the performance of online experimental redesign.

A. Experimental Redesign under Parameter Uncertainty

In this section, the mid-term redesign method (with pre-set updating policy) will be compared with the offline experimental design and also the non-design method to test its ability and reliability for parameter estimation under parameter uncertainty. In order to give a more convincing conclusion, those strategies will be applied to circumstances with small parameter uncertainty and large uncertainty.

The settings for those three design strategies are summarized as follows.

- Online model-based experimental redesign (it will be called online design in figures for convenience). In this design, the total length of 6000 seconds is divided into three sub-experiments: sub-experiment 1 (0s-2000s), sub-experiment 2 (2000s-4000s), and sub-experiment 3 (4000s-6000s). In each sub-experiment, seven sampling time points will be chosen for all the five measurable states. Therefore, there will be 21 sampling times in total. The parameter estimation is made at the end of each sub experiment and the updated model is used for the sampling design in the next sub experiment until the whole experiment length is complete.
- Offline experimental design. This strategy will select 21 sampling times for all the five measurable states that have the maximum information content at the end of the experiment.
- Non-design strategy. This strategy will measure the five states in equal space along the time horizon, i.e., (0:300:6000).

Both the online and offline design will adopt Powell's method to search for the most informative sampling times, and the D-optimal criterion is used in the design. Assume 601 points are possible sampling times, which is 0:10:6000, and the initial values of all the states are set as in [14].

(A1) Design under small parameter uncertainty

In this section, parameters are set at their nominal values. The design results of the three strategies are shown in Table I. It can be seen that offline experimental design has the largest D-value, followed by online experimental

redesign and non-design method, which indicates that online experimental design has the best performance for parameter estimation when the initial parameter guess is very close to the nominal value.

In order to analyse the relationships between the chosen sampling times and the system dynamics, the sampling patterns are illustrated together with the relative sensitivities of parameters on state Q and shown in Fig.2.

TABLE I. SAMPLING TIME DESIGN (PARAMETERS SET AT NOMINAL VALUES)

Design strategy	Sampling times	D-value
Online redesign	[610:10:630], [1940:10:1980], [2600:10:2620], [3940:10:3980], [4590:10:4610], [5940:10:5980]	1.32e-05
Offline design	[560:10:630], [2830:10:2890], [4250:10:4330]	3.24e-05
Non-design	[250:250:6000]	8.32e-08

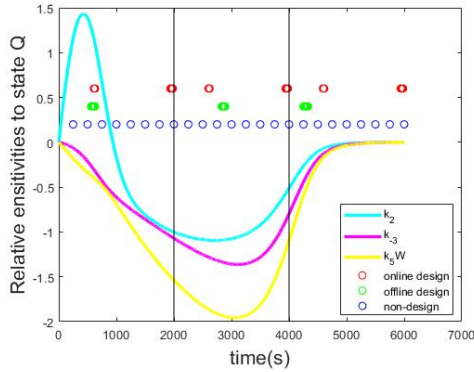


Figure 2. Sampling patterns and relative sensitivities of state Q with online redesign, offline design and non-design methods (parameters set at nominal values)

It can be seen from Fig.2 that the design points obtained by offline design can be divided into three sections, all located near the positions where the sensitivities to one or more parameters are relatively large. However, the design pattern of the online redesign policy can be divided into six sections (each sub-experiment has two), and some sampling points, such as those in the last two sections, locate in the region where all three parameters have very small local sensitivities. In order to further examine the data quality for parameter estimation, the confidential interval for parameters k_2 , k_{-3} and k_5W are calculated using the design patterns in Table I, and the results are presented in Fig.3.

It can be observed from Fig.3 that the offline design has the smallest confidence interval, which indicates that measurements collected according to those offline design sampling times will be the most informative than the other two strategies. The non-design strategy has the largest confidence interval because it sample the measurements in equal space, which may miss some informative points. The confidence interval of the online design has a medium size since it captures some informative points but may lose

others due to the update of parameters (which may change the system dynamics) and the new uncertainties caused in the procedure of updating. Therefore, the update time in online design strategy is crucial. If the update time is at or

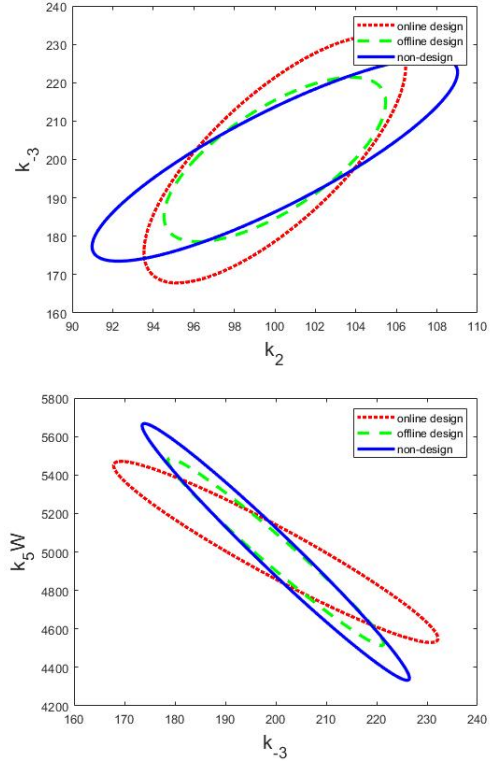


Figure 3. Confidence intervals of k_2 , k_{-3} and k_5W with online redesign, offline design and non-design methods (parameters set at nominal values)

right before the most dynamic sensitive part of the experiments, this part might be considered unimportant due to the uncertainties in the updated parameters, therefore causes loss of useful information.

From the discussion above, it can be concluded that online model-based experimental design may not be the best choice when the initial guess of the (unknown) parameters is very close to the true value. The offline design should be chosen in this circumstance.

(A2) Design under large parameter uncertainty.

In this section, the online redesign, offline design and non-design strategies are tested on situations that the initial parameter guesses stay far from the true values. Assume that the uncertainties of k_2 , k_{-3} and k_5W are 0.5-1.5 times of their nominal values. In order to ensure the robustness for the whole possible parameter region, the concept of the worst case design has been adopted which selects the parameter values with the worst performance on simulating the system as the initial parameter values. After searching through the whole possible parameter region, we chose $k_2 = 150$, $k_{-3} = 300$ and $k_5W = 7500$. The designed sampling patterns of all the three design strategies are shown in Table II. The design results show that the performance of the online redesign is the best when the initial parameters are far from the true values as indicated by the largest D-value in the design.

In order to compare the results more straightforwardly, the joint confidence regions for all three design strategies are plotted in Fig. 4.

TABLE II. SAMPLING TIME DESIGN (PARAMETERS SET AT WORST-CASE IN THE UNCERTAIN REGION)

Design strategy	Sampling times	D-value
Online redesign	[460:10:480], [1200:10:1210], [1960:10:1980], [2820:10:2840], [3940:10:3980], [4760:10:4780], [5840:10:5980]	7.82e-05
Offline design	[430:10:500], [1310:10:1370], [2080:10:2160]	2.84e-06
Non-design	[250:250:6000]	8.32e-08

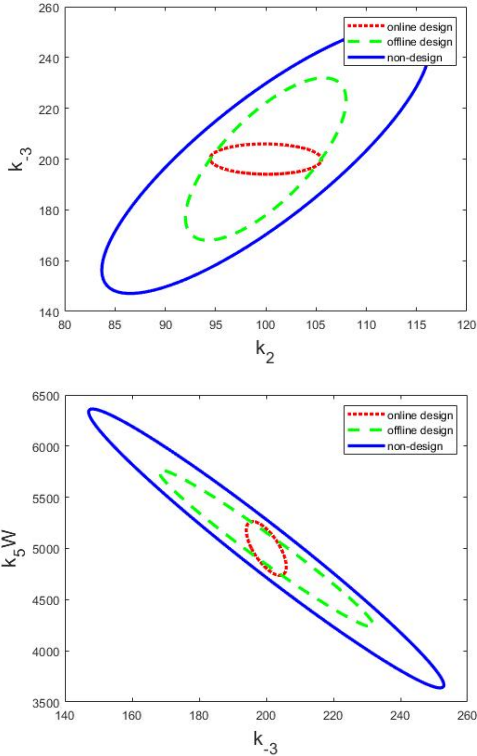


Figure 4. Confidence intervals of k_2 , k_{-3} and k_5W with online redesign, offline design and non-design methods (parameters set at worst-case)

It can be seen from Fig.4 that the online redesign has the smallest CI compared to the offline design and the non-design strategy.

To analyse this result, we first look at the sampling patterns from these strategies together with the relative sensitivities of parameters using the state variable Q (shown in Fig.5). Through the sensitivity curves we can see that the most sensitive part for k_{-3} and k_5W located in the middle (near 3000s). However, the sampling pattern of the offline design is all located in the beginning part (before 2200s) due to the wrong sensitivity calculation caused by large uncertainties in parameters. This means that the offline design has overlooked many informative data points.

On the contrary, the online redesign, even though with wasting points in the last parts, has not missed out the most informative parts because it contains a fixed number of

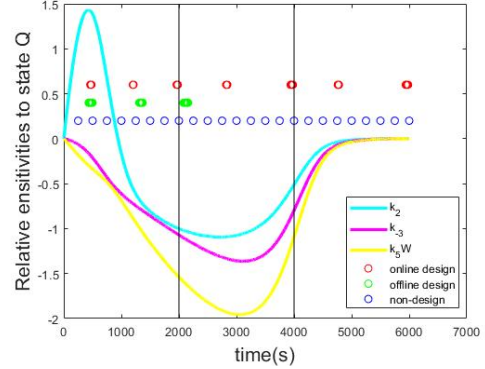


Figure 5. Sampling patterns and relative sensitivities of state Q with online redesign, offline design and non-design methods (parameters set at worst-case)

samples in all three sub experiments as in the pre-setting. Besides, compared to the offline design, parameters are modified in each sub-experiments and therefore reduce uncertainties for the design in the next sub experiment.

B. Online Redesign using Auto-updating Strategy

In this case study, eight samples will be obtained in the first sub-experiment. Consider the large parameter uncertainty case, the initial guess of the parameters are selected using the worst-case method. The values are again: $k_2 = 150$, $k_{-3} = 300$ and $k_5W = 7500$. Three design methods, (i) the offline design, (ii) the online redesign with pre-defined updating strategy, and (iii) the online redesign with auto-updating strategy, will be implemented and compared. The total sampling number of these three design methods are kept as the same.

In the pre-defined online redesign updating strategy, the whole experiment is divided into three sub-experiments: (0-2000), (2000-4000), and (4000-6000). According to the simulation results, the auto-updating strategy returns four sub-experiments (0-1960), (1970-3370), (3380-4720), and (4730-6000). The total number of samplings is 28. The results with the three design methods are listed in Table III.

TABLE III. SAMPLING TIME DESIGN USING PRE-DEFINED, AUTO-UPDATING ONLINE REDESIGN AND OFFLINE DESIGN

Design strategy	Sampling times	D-value
Online redesign (pre-defined)	[460:10:480], 1200, [1960:10:1980], [2590:10:2610], [3940:10:3980], [4580:10:4600], [5940:10:5980]	2.89e-05
Offline design	[430:10:500], [1330:10:1390], [2090:10:2160]	2.50e-06
Online redesign (auto)	[470:10:480], 1130, [1910:10:1940], [2530, 2540], [3330:10:3350], 3900, 3910, [4680:10:4700], 5230, 5240, [5960:10:5980]	4.72e-05

It can be seen from the D-value of the FIM that data collected using the online redesign with auto-updating

strategy contains the most useful information. The confidence interval of the three design methods are plotted in Fig. 6. The joint-confidence region of the online redesign with auto-updating strategy is the smallest, followed by the

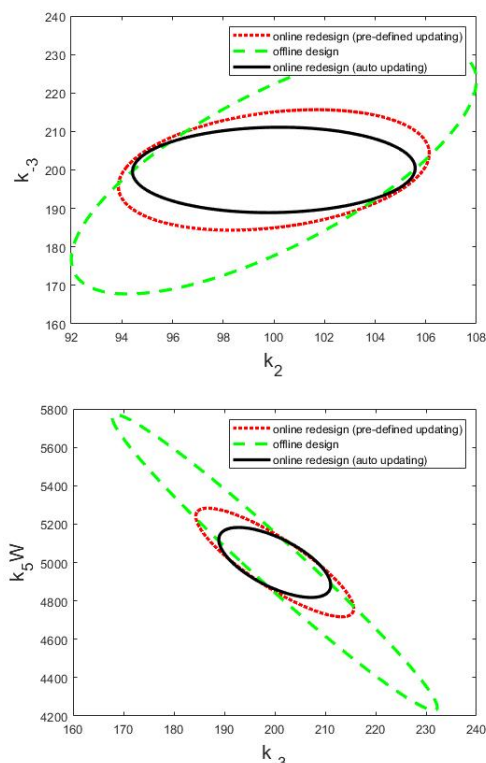


Figure 6. Confidence intervals of k_2 , k_{-3} and k_5W with online redesign with auto-updated strategy, online design with user pre-set updating strategy and offline design (parameters set at worst case)

online redesign with pre-defined updating strategy. It shows that the potential of obtaining more accurate parameter estimation is further increased when the auto-updating strategy is applied in the online redesign procedure.

IV. CONCLUSION AND FUTURE WORK

In this work, the robustness of the online experimental redesign is studied considering parameter uncertainty. A redesign method is proposed by introducing the auto-updating idea. Several design methods are tested on an enzyme reaction system for sampling time design of the measurements. Simulation studies show that the online experimental redesign has a better performance on parameter estimation than the conventional OED when the initial parameter guess is far from the true values. This benefit comes from the parameter updates in the middle of the experiment since the updated model is used for the next stages of sampling design. With a better knowledge of the model, the sampling design is improved from the learning in the previous sub experiments.

It should be noted that the online experimental redesign may not necessarily improve the data quality when the initial parameter guess is already very close to the true values. This is because the mid-term update itself could

introduce modelling errors due to the numerical procedure and added complexity.

This paper also proposes an improvement for online experimental redesign. Other than taking the user pre-defined mid-term redesign strategy, an auto-updating strategy is developed by checking the identifiability conditions of the parameters. The length of each sub-experiment must contain sufficient data to assure identifiability of parameters. This auto-updating strategy not only avoids the lack of identifiability problem in each sub-experiment, but also improves the performance of the online experimental redesign in that the data collected could lead to more accurate parameter estimation.

The method of auto-updating for online experimental redesign is mostly suitable for sampling time design due to its nature in the design along time horizon. For RED of other experimental factors such as measurement set selection under large model uncertainties, the use of historical data information need to be further investigated so as to compensate the effects of model uncertainty on the experimental design.

REFERENCES

- [1] X.-S. Yang, *Mathematical modeling with multidisciplinary applications*. Wiley Online Library, 2013.
- [2] L. Ljung, *System Identification: Theory for the User*. Pearson Education Canada, 1987.
- [3] G. Franceschini and S. Macchietto, "Model-based design of experiments for parameter precision: State of the art," *Chemical Engineering Science*, vol. 63, no. 19, pp. 4846-4872, 2008.
- [4] F. Galvanin, M. Barolo, G. Pannocchia, and F. Bezzo, "Online model-based redesign of experiments with erratic models: a disturbance estimation approach," *Computers & Chemical Engineering*, vol. 42, pp. 138-151, 2012.
- [5] R. Mehra, "Optimal input signals for parameter estimation in dynamic systems--Survey and new results," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 753-768, 1974.
- [6] L. Keviczky, "'Design of Experiments' for the identification of linear dynamic systems," *Technometrics*, vol. 17, no. 3, pp. 303-308, 1975.
- [7] L. Gerencsér and H. Hjalmarsson, "Adaptive input design in system identification," in *Proceedings of the 44th IEEE Conference on Decision and Control*, 2005: IEEE, pp. 4988-4993.
- [8] J. Stigter, D. Vries, and K. Keesman, "On adaptive optimal input design: a bioreactor case study," *AIChE Journal*, vol. 52, no. 9, pp. 3290-3296, 2006.
- [9] F. Galvanin, M. Barolo, and F. Bezzo, "Online model-based redesign of experiments for parameter estimation in dynamic systems," *Industrial & Engineering Chemistry Research*, vol. 48, no. 9, pp. 4415-4427, 2009.
- [10] T. Barz, D. C. López Cárdenas, H. Arellano - Garcia, and G. Wozny, "Experimental evaluation of an approach to online redesign of experiments for parameter determination," *AIChE Journal*, vol. 59, no. 6, pp. 1981-1995, 2013.
- [11] M. Cruz Bournazou, T. Barz, D. Nickel, and P. Neubauer, "Sliding - window optimal experimental re - design in parallel microbioreactors," *Chemie Ingenieur Technik*, vol. 86, no. 9, pp. 1379-1380, 2014.
- [12] M. Cruz Bournazou *et al.*, "Online optimal experimental re - design in robotic parallel fed - batch cultivation facilities," *Biotechnology and bioengineering*, vol. 114, no. 3, pp. 610-619, 2017.
- [13] S. Asprey and S. Macchietto, "Designing robust optimal dynamic experiments," *Journal of Process Control*, vol. 12, no. 4, pp. 545-556, 2002.
- [14] H. Yue, P. Halling, and H. Yu, "Model development and optimal experimental design of a kinetically controlled synthesis system," *IFAC Proceedings Volumes*, vol. 46, no. 31, pp. 327-332, 2013.