Optimal Experimental Design for an Enzymatic Biodiesel Production System

Hui Yu*, Hong Yue*, Peter Halling**

** Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow G1 1XW, UK*
** Department of Pure and Applied Chemistry, *University of Strathclyde, Glasgow G1 1XL, UK*

(e-mail: hui.yu.100@strath.ac.uk, hong.yue@strath.ac.uk, p.j.halling@strath.ac.uk)

Abstract: Two optimal experimental design (OED) problems for an enzymatic biodiesel production system are investigated to improve parameter estimation quality. An orthogonalized sensitivity analysis method is firstly implemented to select important parameters. Next the design of measurement set and sampling strategy is developed in the form of two convex optimization problems which are solved by the interior-point algorithm and the Powell's method, respectively. Simulation results demonstrate the function of OED in reducing parameter estimation errors. The biodiesel concentration is identified to be the most valuable state variable observation, and the parameter estimation accuracy can be improved through optimal sampling design.

Keywords: optimal experimental design (OED), enzymatic biodiesel reaction system, measurement set selection, optimal sampling strategy, parameter estimation, local sensitivity analysis (LSA).

1. INTRODUCTION

The purpose of optimal experimental design (OED) is to devise necessary dynamic experiments in such a way that model parameters can be estimated from the resulting experimental data with the best possible statistical quality. There is a growing interest in OED in recent years particularly in biological and biochemical systems where performing experiments to obtain rich data are usually time-consuming and cost expensive. Informative measurement data can be generated for parameter identification and model calibration through the model-based experimental design. Also experiment efforts can be reduced since the modelling efficiency is improved. Various OED methods have been developed for nonlinear dynamic systems and many have been successfully applied to a wide range of systems (Hagen et al., 2013; Martinez et al., 2009; Atkinson and Bogacka, 2002). Useful reviews can be found in (Franceschini and Macchietto, 2008; Chaloner and Verdinelli, 1995; Maria, 2004; Kreutz and Timmer, 2009), to name a few.

In general, experimental design for parameter estimation can be divided into two categories according to design factors. One is on design of manipulations such as initial conditions, input variables, length of perturbation time, etc., which are factors that drive/excite the dynamic processes (Balsa-Canto et al., 2007; Faller et al., 2003; Asprey and Macchietto, 2002; Banga et al., 2002). The other category is on design of measurements which is to answer the question of what, where and when to measure in order to collect the most 'useful' data. Two challenging problems in the latter category are sampling (time) scheduling and selection of measured variables. In chemical and biochemical processes, uniformed sampling in time domain is widely accepted which is convenient for operation but may not be the best solution for parameter estimation. Through design of optimal sampling points, the parameter estimation accuracy can be improved and the number of samples to be drawn is reduced (Bauer et al., 2000; Kutalik et al., 2004; Pagendam and Pollett, 2013; Asyali, 2010). On the other hand, the variables to be measured are normally determined following expert knowledge and are mainly selected from the control and monitoring point of view rather than modelling. Earlier work on optimal design of measurement set selection can be found in (He et al., 2010; Yue et al., 2008), where the design tasks were formulated as constrained nonlinear optimization problems.

In this work, we aim to tackle the design problems of non-uniform sampling scheduling and selection of measurement variables for a biodiesel production system. In the design of optimal sampling strategy, multiple measurement variables will be considered. The employed model was developed for an enzyme-catalyzed biodiesel process and tested on lab-scaled fed-batch experiments for the transesterification of rapeseed oil with methanol using Callera™ Trans L (Price et al., 2013). In this reaction scheme, there are a lot of unknown kinetic parameters need to be estimated from experimental data, where accurate estimation is highly dependent on the experimental data. A systematic experimental design is therefore crucial for assuring modelling quality.

The rest of the paper is organized as follows. Preliminaries on relevant methodologies are briefed in Section 2. The OED techniques on sampling time design and measurement set selection are presented in Section 3. In Section 4, OED for the enzymatic biodiesel production system is implemented. Finally, conclusions and discussions are made in Section 5.

2. PRELIMINARIES

A general ordinary differential equations (ODEs) model is considered for nonlinear dynamic systems:

$$\dot{\mathbf{X}} = \mathrm{f}\left(\mathbf{X}, \boldsymbol{\theta}, \mathbf{u}, t\right), \quad \mathbf{X}\left(t_0\right) = \mathbf{X}_0 \tag{1}$$

$$\mathbf{Y} = \mathrm{h}\left(\mathbf{X}, \boldsymbol{\theta}, t\right) + \boldsymbol{\xi} \tag{2}$$

where $\mathrm{f}\left(\cdot\right)$ is the set of state transition functions which is assumed to be continuous and first-order derivative. $\mathbf{X} \in \mathbb{R}^n$ denotes the vector of state variables with initial condition $\mathbf{X}_0$, and n is the number of state variables. $\boldsymbol{\theta} \in \mathbb{R}^m$ is the vector of model parameters with m being the number of parameters. $\mathbf{u} \in \mathbb{R}^r$ represents the vector of input variables, r is the number of input variables. $\mathbf{Y} \in \mathbb{R}^p$ is the measurement output vector with p being the number of measurement variables, and $\mathrm{h}\left(\cdot\right)$ is the measurement function, normally used for selecting which variables to be measured. $\boldsymbol{\xi}$ is the measurement error assumed to be independently and identically distributed (i.i.d.), zero-mean Gaussian noise. In practice, unknown parameters can be estimated by comparing the output values from the model prediction with the measurement data. The commonly used least-squares algorithm can be employed to estimate those practically identifiable parameters through minimizing the sum of the squared residuals.

Most OED techniques are developed based on measures of the Fisher information matrix (FIM) which quantifies information content of parameter estimation. A formulated scalar function of FIM contains experimental design factors, and the design process is to optimize those design factors so that parameter estimation errors are minimized. The local (parametric) sensitivity analysis (LSA) plays a major part in formulating FIM, thus is crucial for performing experimental design. LSA is also used to identify key parameters that strongly affect the system output. The local sensitivity is defined as the partial derivative of the output states with respect to system parameters. Denoting $\mathbf{X} = [x_1, x_2, \cdots, x_n]^{\mathrm{T}}$ and $\boldsymbol{\theta} = [k_1, k_2, \cdots k_m]^{\mathrm{T}}$, the absolute sensitivity matrix is $\mathbf{S} = \partial \mathbf{X} / \partial \boldsymbol{\theta} = \left(s_{ij}\right)$ with $s_{ij} = \partial x_i / \partial k_j$ for $i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, m$. This sensitivity matrix can be easily obtained by partial differentiation of (1) with respect to $\boldsymbol{\theta}$ which results in a set of sensitivity differential equations:

$$\dot{\mathbf{S}} = \mathbf{JS} + \mathbf{F}, \quad \mathbf{S}\left(t_0\right) = \mathbf{S}_0 \tag{3}$$

where $\mathbf{J} = \partial f / \partial \mathbf{X}$ and $\mathbf{F} = \partial \mathrm{f} / \partial \boldsymbol{\theta}$ are the Jacobian matrix and the parametric Jacobian matrix, respectively. For biochemical systems, kinetic parameters often have different orders of magnitude. In order to compare their influence on the system output directly, relative sensitivities are used instead, i.e., $\overline{s}_{ij} = \left(\partial x_i / \partial k_j\right) \cdot \left(\partial k_j / \partial x_i\right)$.

For high-dimensional systems, normally not all the unknown parameters are identifiable due to: (1) small influence of some parameters on the measured system output; (2) high correlations between parameter pairs. It is therefore necessary to perform identifiability analysis. This in turn will reduce the number of parameters to be estimated. Several methods have been developed for parameter identifiability analysis and parameter reduction of complex system models such as a hybrid technique that integrates conservation analysis, LSA, principal component analysis and flux analysis together (Jia and Yue, 2008); a method using collinearity index (Brun et al., 2001); a relative gain array method (Sandink et al., 2001); a method based on Hanken singular value (Sun and Hahn, 2006), etc. In this work, the orthogonalization-based technique (Yao et al., 2003) is employed for choosing parameters that are both sensitive and identifiable. This method was applied to work on a signal transduction pathway model where both the influence of parameters to the system output and the cross-correlation between parameters were examined (Yue et al., 2006).

3. OPTIMAL EXPERIMENTAL DESIGN

3.1 Basics of Optimal Experimental Design

OED is aimed at devising dynamic experiments by optimizing design factors $\boldsymbol{\zeta}$ which include initial conditions $\mathbf{X}\left(t_0\right)$, input variables $\mathbf{u}$, sampling schedule $\mathbf{t}_{sp}$, valuable measured response $\mathbf{y}\left(\mathbf{t}_{sp}\right)$, etc. so that model parameters can be estimated most precisely.

$$\boldsymbol{\zeta} = [\mathbf{u}, \mathrm{t}_{sp}, \mathbf{X}\left(t_0\right), \mathbf{y}\left(t_{sp}\right), T] \tag{4}$$

T is the experimental duration. The FIM which combines parameter influence with measurement noise is represented as a nonlinear function of local sensitivity matrix.

$$\mathbf{FIM}\left(\boldsymbol{\theta}, \boldsymbol{\zeta}\right) = \mathbf{S}\left(\boldsymbol{\theta}, \boldsymbol{\zeta}\right)^{\mathrm{T}} \mathbf{Q}^{-1} \mathbf{S}\left(\boldsymbol{\theta}, \boldsymbol{\zeta}\right) \tag{5}$$

$\mathbf{Q}$ is a weighting matrix which is usually chosen to be the measurement error covariance matrix. When the model is linear in parameters, and the measurement noise is additive i.i.d. Gaussian white noise, the inverse of FIM is approximately equal to the lower bound of the parameter estimation error covariance matrix (Cramer-Rao bound). The OED problem can therefore be cast as minimization of parameter covariance matrix measure or maximization of measures of FIM, i.e.

$$\boldsymbol{\zeta}^* = \arg\max_{\boldsymbol{\zeta} \in \boldsymbol{\Omega}} \Phi\left(\mathbf{FIM}\left(\boldsymbol{\theta}, \boldsymbol{\zeta}\right)\right) \tag{6}$$

where $\boldsymbol{\Omega}$ is the admissible space of design parameters. $\Phi\left(\cdot\right)$ is a function used to scalarize the information content. The most commonly used 'alphabet' optimization criteria are A-optimal, D-optimal, E-optimal, and modified E-optimal design. There are also other scalar functions developed for OED based on FIM which can been seen in (Ljung, 1998) and other papers. All these criteria have advantages and disadvantages and some may be superior to others for certain systems. By using these design criteria, the OED problem can be transferred into a convex optimization problem when the FIM is an appropriate function of the design parameters.

3.2 Design of Measurement Set Selection

The purpose of measurement set selection is to choose the most informative measurable state variables as observations to get the best parameter estimation. This will also reduce the experiment cost when only relevant variables are measured. The measurement set selection design can be represented as (He et al., 2010):

$$\begin{aligned} &\xi = \begin{Bmatrix} x_1 & \cdots & x_n \\ \lambda_1 & \cdots & \lambda_n \end{Bmatrix} \\ &\min \sum = \sigma^2 \left(\sum_{i=1}^{n} \lambda_i S_i^T S_i \right)^{-1} \\ &s.t. \quad \lambda_i \in \{0,1\} \\ &\quad \mathbf{1}^T \boldsymbol{\lambda} = n \end{aligned} \tag{7}$$

where λ_i is an integer weight with values of 0 or 1, relating to the *i*-th state variable. The variance of measurement noise, σ^2, is taken as constant and the same for all the noise channels therefore has no effect to the optimization design. This integer optimization problem can be transferred into a continuous optimization problem by relaxing the weighting factors to a continuous value within the range of [0, 1]. By using different scalar OED criteria, the problem of measurement set selection can be written into different optimization problems (Boyd and Vandenberghe, 2004). For instance, the E-optimal design of the covariance matrix can be cast as a semi-definite program (SDP):

$$\begin{aligned} &\min \ -t \\ &s.t. \ \sum_{i=1}^{n} \lambda_i S_i^T S_i \succ t\mathbf{I} \\ &\quad \lambda_i \succ 0, \ \forall i; \ \mathbf{1}^T \lambda = 1 \end{aligned} \tag{8}$$

This can be easily solved by optimization tools such as SeDuMi. When using the D-optimal criterion, the design problem can be written as a finite dimensional constrained linear optimization problem which can be solved by the interior-point method.

3.3 Design of Optimal Sampling Strategy

The task of optimal sampling design is to find the best sampling strategy for the measurement variables that will give most informative experimental data for parameter estimation. This design can be formulated as an integer optimization problem:

$$\begin{aligned} &\boldsymbol{\zeta} = \begin{Bmatrix} t_1 & t_2 & \cdots & t_N \\ \omega_1 & \omega_2 & \cdots & \omega_N \end{Bmatrix} \\ &\boldsymbol{\zeta}^* = \arg\min_{\boldsymbol{\omega} \to \boldsymbol{\Omega}} \left(\sigma^2 \sum_{i=1}^{N} \omega_i \mathbf{S}(t_i)^T \mathbf{S}(t_i) \right) \\ &s.t. \ \omega_i \in \{0,1\} \\ &\quad \mathbf{1}^T \boldsymbol{\omega} = N_{sp} \end{aligned} \tag{9}$$

where $\boldsymbol{\omega} = [\omega_1 \cdots \omega_N]^T$ is the weighting vector for available measurement points, the number of which is *N*. N_{sp} is the number of sampling points to be selected. In this case, $\omega_i = 1$ means the *i*-th time point is selected, while those sampling points with weight value of 0 are not selected. This integer optimization problem can be relaxed to a continuous optimization problem by applying rounding heuristics to the solution (Bauer et al., 2000). However, the optimal solution may be affected by the rounding. A more computationally efficient procedure, named Powell's quadratically convergent method, is introduced in (Kutalik et al., 2004). In the latter algorithm, an initial sampling strategy is given to start with (normally an equally-spaced sampling). In each iteration, one sampling point in the selected sequence is replaced by a sampling point from the available measurement points which gives the best result. The iteration process will continue until the optimal solution is obtained.

4. EXPERIMENTAL DESIGN OF AN ENZYMATIC BIODIESEL PRODUCTION SYSTEM

4.1 Model Description

The kinetic model used in this work was established to describe an enzymatic transesterification reaction of rapeseed oil with methanol in a biphasic oil–water system using a liquid lipase, Callera™ Trans L, based on several assumptions (Price et al., 2014). In this reaction scheme, the free enzyme (E_{bulk}) contained in the polar phase is absorbed at the water oil interface (A_f) and forms the penetrated enzyme (E), which further reacts with triglyceride (T), diglyceride (D) and monoglyceride (M) to form enzyme substrate complexes ET, ED and EM. Then these enzyme substrates can be decomposed into the acyl enzyme complex and D, M and glycerol (G), respectively. The acyl enzyme complex can then react with water (W) or methanol (CH) and produce the free fatty acid (FFA) and biodiesel (BD). Additionally, the competitive methanol inhibition is also considered in this reaction process in which CH reacts with E to form ECH. From these kinetic reactions a set of ordinary differential equations (ODEs) can be formulated following the mass-balance principle (Price et al., 2014).

4.2 Sensitivity Analysis and Parameter Identifiability

In this fed-batch process, the experiment length is set to be 25 hours and sampling takes place every 15 minutes in the first hour and then once each hour. The unit for all reactant concentrations is in mol/L. The initial condition and the feeding rates are provided in **Error! Reference source not found.** in the Appendix. Fig. **1** illustrates the concentration time profiles of the five measurable state variables where the red points represent the real experimental values and the blue lines describe the simulated concentration trajectories. It can be seen that the model predicts well the trends of the experimental data except for FFA which shows a clear deviation. This over prediction of FFA may be due to

processes that are not taken into account (such as the change of viscosity of the reaction media) or some statistical measurement errors.

In order to compare the effects of different kinetic parameters on the system output variables, relative local sensitivities and their 2-norms are first calculated. A bar chart is shown in Fig. 2 to demonstrate the overall influence of each parameter to all the 5 measurable states, from which it can be observed that k_5, k_8 and k_{-8} are the most sensitive parameters.

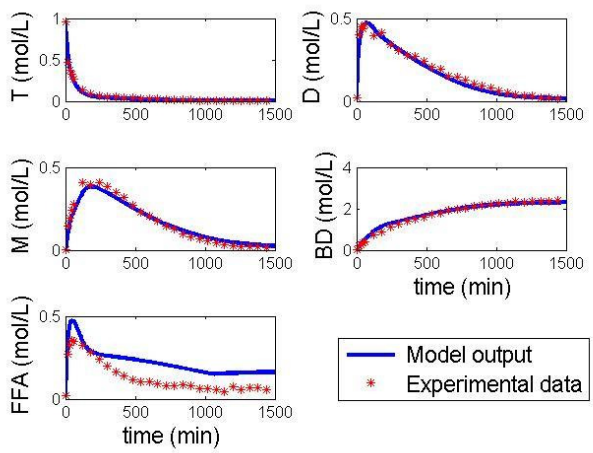

Fig. 1. Time profiles for 5 measurable state variables

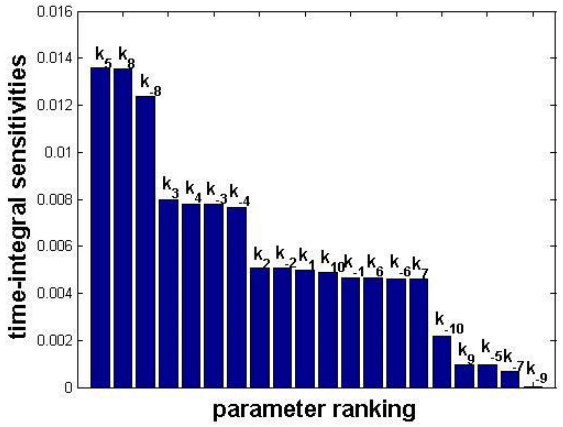

Fig. 2. Overall parameter ranking via LSA

LSA suggests that not all the parameters are influential to the measurable outputs and some of the parameters may be non-identifiable. To further examine correlations between parameters, the collinearity index was calculated to determine estimable parameters for this system (Price et al., 2014). It was found that the maximum number of parameters that can be estimated is 10, but it was still difficult to estimate all the 10 parameters using the experimental data. In this work, based on the LSA results, the orthogonalization-based method (Yao et al., 2003) is applied to examine parameter correlations and to rank parameters so as to select the set of identifiable parameters. This alternative method gives consistent results regarding the 10 estimable parameters using the collinearity index. The 3 most important parameters identified in this analysis are k_5, k_8 and k_6 (shown in Fig. 3). Compared to the LSA ranking results (Fig. 2), both k_5 and k_8 are always regarded as the most important parameters. The influence of k_{-8} is reduced in the latter group of analysis mainly because it is partially correlated with k_8.

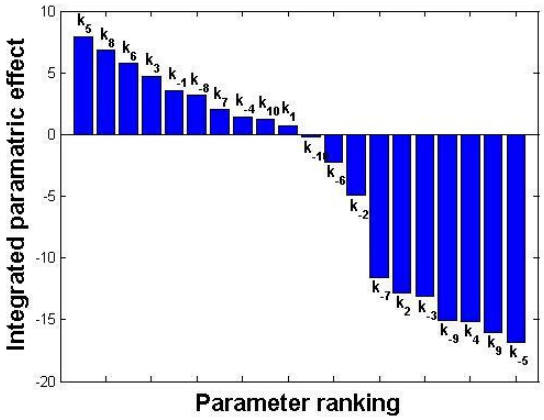

Fig. 3. Parameter ranking via orthogonalization

4.3 *Design of Measurement Set*

Taking the three most important parameters, k_5, k_8 and k_6, into the parameter estimation scheme, OED has been applied to determine the most valuable observation from the five measurable state variables. The optimal weights calculated from E- and D-optimal design are listed in Table 1.

The E-optimal design result shows that the state variable T has the largest weight (more than 0.9) and it should be selected as the most valuable observation. However, the D-optimal design reveals that the state variable BD is the most important measurement target and FFA which has a weight of 0.33 can also contribute considerable data information, while the state variable T is not important at all. To analyze these conflicting results, parameter estimation errors are compared by using different measurement sets. Fig. 4 compares the confidence intervals (CI) using the parameter pair (k_5, k_8) with different set of observations. The largest dash-dot ellipsoid corresponds to the situation when only T is used as the measurement signal. The solid curve corresponds to the results by using BD as the only measurement variable. The smallest dashed ellipsoid is from the estimation when all the 5 measurements are used. It can be observed that using BD as the observation leads to a smaller parameter estimation error compared with simply using T. Also, the results from the D-optimal design are very close to that including all the five measurable state variables. From numerical viewpoint, the E-optimal design only focuses on the improvement of the most uncertain parameter, therefore the generated measurement data from E-optimal design may lack information for other parameters contained in FIM. Therefore, the D-optimal design is regarded as the most suitable for this system.

Table 1 Weighting coefficients for measurable states from measurement set selection

	T	D	M	BD	FFA
E-	0.902	0.087	0.009	0.001	6.23e-4
D-	1.539e-7	1.705e-7	1.842e-6	0.671	0.329

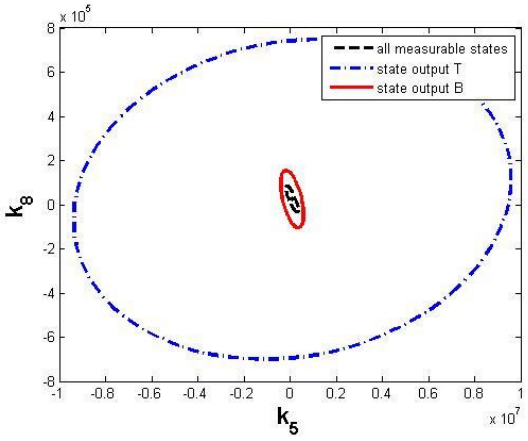

Fig. 4. CI ellipsoids for (k_5, k_8) with different observations

4.4 Design of Optimal Sampling Strategy

D-optimal design is employed to determine the optimal sampling strategy, i.e., at which time points to collect the measurement data. Without loss of generality, it is assumed that the measurement errors are time independent and are equal for each observation. The design problem can be formulated as the following optimization problem:

$$\max \det\left(\sum_{i=1}^{N}\mathbf{S}(t_i)^T\mathbf{S}(t_i)\right)$$

$$s.t. \quad t_l - t_{l-1} \geq 5 \qquad (10)$$
$$t_1 \geq 0$$
$$t_N \leq 1500$$
$$N = 28$$

The minimum sampling interval is set to be 5 minutes. The total number of (time) sampling points is set to be 28 which was used in the lab experiments (see Table 2). By using Powell's quadratically convergent algorithm, the optimal sampling points are calculated as given in Table 2.

Table 2 OED and Experience sampling strategies

	Measurement time points (unit: minutes)
D-Optimal sampling	96, 101, 106, 111, 116, 121, 126, 131, 136, 660, 665, 670, 675, 680, 685, 690, 695, 700, 705, 1460, 1465, 1470, 1475, 1480, 1485, 1490, 1495, 1500
Experimental sampling	0, 15, 30, 45, 60, 120, 180, 240, 300, 360, 420, 480, 540, 600, 660, 720, 780, 840, 900, 960, 1020, 1080, 1140, 1200, 1260, 1320, 1380, 1440

It is found that the optimal sampling strategy favours those time points where the sensitivities are relatively high for the designed parameters. The CI ellipsoids for (k_5 , k_8) are shown in Fig. 5. It is not surprising that the designed sampling points lead to smaller CIs which indicate possibly more accurate parameter estimation using the designed sampling schedule.

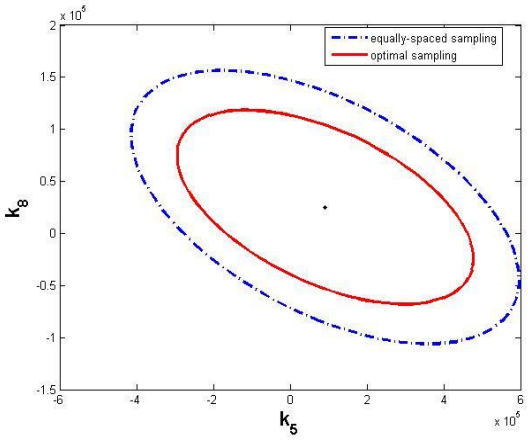

Fig. 5. CI ellipsoids for (k_5, k_8) under optimal and experience-based sampling strategies

6. CONCLUSIONS

In this work, we have developed OED methods for measurement set selection and sampling scheduling, respectively. Using a kinetic model developed for a lab-scale enzyme-catalysed biodiesel reaction process, the two OED methods are implemented following the real experimental conditions. Through the design of measurement set, it is suggested that the state variable *BD* may provide the most informative experimental data than other measurable variables for those important parameters to be estimated. The information available by observing only *BD* is close to that of using all 5 measurable output variables. Therefore, it can be considered as the major observation for modelling. It is also observed that using different criteria in OED could give different, in fact conflicting, results for the design of measurement set. Therefore, carefully choosing design criteria is important. There are no widely accepted rules for how to choose OED criteria. Trial-and-error effort is a common practice. The optimal sampling design is achieved by using Powell's method which leads to significant improvement for parameter estimation. The result shows that measurement points chosen at regions with higher parameter sensitivities can generate more informative data.

The next work is to validate and test the OED results in the experimental system. The two experimental designs were implemented separately in this work. We firstly determine the measurement set; then take the designed measurement set into OED of time sampling strategy. Although this seems to be a reasonable sequence, the OED results could be different if the sampling strategy is designed first. A better solution is to integrate multiple experimental design factors into one optimization scheme. This will be a challenging task for problem formulation and optimization. For this enzymatic biodiesel reaction system, design of the time varying input (methanol) is also an important aspect which has attracted lots of attention by experimenters and control engineers. Design of the optimal feeding strategy in order to reduce

parameter uncertainties as well as increase the production rate is undergoing research.

ACKNOWLEDGEMENT

The authors would like to thank Dr Jason Price and his colleagues from the Department of Chemical and Biochemical Engineering, Technical University of Denmark, for providing the mathematical model and also for many useful discussions.

REFERENCES

Asprey, S. P. and Macchietto, S. (2002). Designing robust optimal dynamic experiments. *Journal of Process Control,* 12(4), 545-556.

Asyali, M. H. (2010). Design of optimal sampling times for pharmacokinetic trials via spline approximation. *Turk. J. Elec. Eng. & Comp. Sci.,* 18(6), 1019-1030.

Atkinson, A. C. and Bogacka, B. (2002). Compound and other optimum designs for systems of nonlinear differential equations arising in chemical kinetics. *Chemometr. Intell. Lab.,* 61(1), 17-33.

Balsa-Canto, E., Rodriguez-Fernandez, M. and Banga, J. R. (2007). Optimal design of dynamic experiments for improved estimation of kinetic parameters of thermal degradation. *J. Food. Eng.,* 82(2), 178-188.

Banga, J. R., Versyck, K. J. and Van Impe, J. F. (2002). Computation of optimal identification experiments for nonlinear dynamic process models: a stochastic global optimization. *Ind. Eng. Chem. Res.,* 41(10), 2425-2430.

Bauer, I., Bock, H. G., Körkel, S. and Schlöder, J. P. (2000). Numerical methods for optimum experimental design in DAE systems. *J. Comput. Appl. Math.,* 120(1), 1-25.

Boyd, S. P. and Vandenberghe, L. (2004). *Convex Optimization.* Cambridge University Press.

Brun, R., Reichert, P. and Künsch, H. R. (2001). Practical identifiability analysis of large environmental simulation models. *Water Resour. Res.,* 37(4), 1015-1030.

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science,* 10(3), 273-304.

Faller, D., Klingmüller, U. and Timmer, J. (2003). Simulation methods for optimal experimental design in systems biology. *Simulation,* 79(12), 717-725.

Franceschini, G. and Macchietto, S. (2008). Model-based design of experiments for parameter precision: state of the art. *Chem. Eng. Sci.,* 63(19), 4846-4872.

Hagen, D. R., White, J. K. and Tidor, B. (2013). Convergence in parameters and predictions using computational experimental design. *Interf. focus,* 3(4).

He, F., Brown, M. and Yue, H. (2010). Maximin and Bayesian robust experimental design for measurement set selection in modelling biochemical regulatory systems. *Int. J. Robust and Nonl. Contr.,* 20(9), 1059-1078.

Jia, J. and Yue, H. (2008). Model simplification of signal transduction pathway networks via a hybrid inference strategy. *The 17th IFAC Conference.* Seoul, Korea, 10307-10312.

Kreutz, C. and Timmer, J. (2009). Systems biology: experimental design. *FEBS Journal,* 276(4), 923-942.

Kutalik, Z., Cho, K.-H. and Wolkenhauer, O. (2004). Optimal sampling time selection for parameter estimation in dynamic pathway modeling. *Biosystems,* 75(1), 43-55.

Ljung, L. (1998). *System Identification.* Springer.

Maria, G. (2004). A review of algorithms and trends in kinetic model identification for chemical and biochemical systems. *Chem. Biochem. Eng. Q.,* 18(3), 195-222.

Martinez, E. C., Cristaldi, M. D. and Grau, R. J. (2009). Design of dynamic experiments in modeling for optimization of batch processes. *Ind. Eng. Chem. Res.,* 48(7), 3453-3465.

Pagendam, D. and Pollett, P. (2013). Optimal design of experimental epidemics. *J. Stat. Plan. Infer.,* 143(3), 563-572.

Price, J., Hofmann, B., Silva, V. T., Nordblad, M., Woodley, J. M. and Huusom, J. K. (2014). Mechanistic modelling of biodiesel production using a liquid lipase formulation. *Biotechnol. Progr.*

Price, J. A., Nordblad, M., Woodley, J. and Huusom, J. K. (2013). Fed-batch feeding strategies for enzymatic biodiesel production. *19th World Congress of the International Federation of Automatic Control.*

Sandink, C., Mcauley, K. and Mclellan, P. (2001). Selection of parameters for updating in on-line models. *Ind. Eng. Chem. Res.,* 40(18), 3936-3950.

Sun, C. and Hahn, J. (2006). Parameter reduction for stable dynamical systems based on Hankel singular values and sensitivity analysis. *Chem. Eng. Sci.,* 61(16), 5393-5403.

Yao, K. Z., Shaw, B. M., Kou, B., Mcauley, K. B. and Bacon, D. (2003). Modeling Ethylene/Butene copolymerization with multi‐site catalysts: parameter estimability and experimental design. *Polym. React. Eng.,* 11(3), 563-588.

Yue, H., Brown, M., He, F., Jia, J. and Kell, D. B. (2008). Sensitivity analysis and robust experimental design of a signal transduction pathway system. *Int. J. Chem. Kinet.,* 40(11), 730-741.

Yue, H., Brown, M., Knowles, J., Wang, H., Broomhead, D. S. and Kell, D. B. (2006). Insights into the behaviour of systems biology models from dynamic sensitivity and identifiability analysis: a case study of an NF-κB signalling pathway. *Molec. Bio.,* 2(12), 640-649.

APPENDIX

Table A. 1 Initial input values and feeding rate of methanol

Initial conditions (unit: mol/L)					
T	0.9536	W	2.3854	EM	0
D	0.0195	CH	0.5850	ECH	0
M	0.0014	E	0	Ef	9.7165e-6
B	1e-4	EX	0	Vp	0.0661
FFA	0.0224	ET	0	V	1.5383
G	1e-6	ED	0		

Methanol feed rate [eq./h]	Initial dose methanol [eq.]	Water [wt.% oil]	Enzyme [wt.% oil]
0.185 first 2hrs. 0.06 thereafter	0.2	5	0.5