Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Selection of Parameter Subsets and Design of Experiments for Estimation of Nonlinear Dynamic Systems

Yunfei Chu and Juergen Hahn

Texas A&M University, College Station, TX 77843 USA
(Tel: 979-845-3568; fax: 979-845-6446; e-mail: hahn@tamu.edu)

Abstract: Models describing complex processes often contain a large number of parameter and may need to describe nonlinear behavior of the system. It is usually not possible in practice to identify all parameters due to the number and quality of measurement data as well as interactions among the parameters. A common approach is to select a set of parameters for estimation while other parameters are fixed at their nominal values. Such a parameter selection procedure is often based on sensitivity analysis; however, the determined sensitivity value depends on an assumed distribution of values of the parameters, initial states and known trajectories of the input signals. This work addresses some of the mentioned issues and presents a procedure which combines parameter selection for estimation with experimental design. Additionally, the effect that uncertainty in the parameter values has on the parameter set selection is also taken into account. An optimization problem is formulated whose solution is the optimal set of parameters to be estimated and the experimental conditions required for determining this set of parameters.

1. INTRODUCTION

Mathematical modeling plays an important role in study of complex dynamic systems and parameter estimation forms an essential component of deriving mathematical models. However, estimation of large-scale models can still pose a major challenge, even though a large number of estimation methods have been developed (Ljung, 1999; Nelles, 2001). One of the reasons for this is that a model often consists of tens or even hundreds of parameters while at the same time only limited experimental data may be available. Therefore not all the parameters are identifiable in practice and parameter estimation may result in an ill-conditioned problem. Since it is difficult to estimate all the parameters accurately, a widely used strategy is to select a subset of identifiable parameters for estimation while other parameters are fixed at a nominal value. This method requires that a decision needs to be made about which parameters should be selected for parameter estimation. It is the purpose of this work to develop a new approach for determining sets of parameters that should be estimated simultaneously with the experimental conditions used for generating data sets for parameter estimation.

Many methods for parameter selection are based on sensitivity analysis. Some examples are the collinearity index (Brun *et al*., 2001), a column pivoting method (Velez-Reyes and Verghese, 1995), an extension of the relative gain array (Sandink *et al*., 2001), a Gram-Schmidt orthogonalization method (Yao *et al*., 2003) and a recursive approach based upon principal component analysis (Li *et al*., 2004). A more systematic approach for parameter selection is based on experimental optimality criteria. The inverse of the Fisher information matrix (FIM) provides a lower bound for the asymptotic covariance matrix of parameter estimators (Walter and Pronzato, 1990) and it can serve as a measure for the quality of a parameter set. A subset of identifiable parameters can be selected based upon optimizing certain criteria of the Fisher information matrix such as the *D*-optimality criterion or the modified E-optimality criterion (Weijers and Vanrolleghem, 1997; Brun *et al*., 2002).

However, most parameter set selection procedures suffer from one major drawback: some knowledge about the parameter values needs to be available to evaluate the sensitivity vectors or the Fisher information matrix. This is a problem insofar as the exact values of the parameters are not known and hence one is interested in estimating the values. A common assumption is that the true parameter value is near the assumed nominal value, or that it follows a certain distribution around the nominal value, and that the results returned by sensitivity analysis will also hold for the true value. However, there is no guarantee in practice that the true parameter value will be in the vicinity of the nominal value. In our previous work (Chu and Hahn, 2007) the effects of uncertainty of the value of parameters, initial states, and input signals on the parameter selection has investigated. It was demonstrated in examples that significant changes of the parameter set to be determined can occur if parameters differ by 10% from their nominal values.

This work investigates a different aspect of this problem while also dealing with uncertainty in the parameter values. Since it has been shown that parameter sensitivity results are affected by the operating conditions of a process, which in turn are determined by manipulating the inputs, it can be beneficial to determine the optimal operating conditions for creating a data set simultaneously with the parameters to be determined. It is the aim of this paper to develop a method for selecting parameter sets for estimation of nonlinear dynamic systems simultaneously with the operating

conditions for creating a data set and while uncertainty in the values of the process parameters are taken into account. This type of problem can be formulated as an optimization problem, where the trajectories of the inputs can be manipulated in order to maximize a criterion that determines how much information about the system can be captured by estimating process parameters.

2. BACKGROUNDS

2.1 Experimental Optimality Criteria

The measured output is a function of the parameters and is affected by measurement noise

$$\tilde{\mathbf{y}} = \mathbf{y}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon}, \tag{1}$$

where $\tilde{\mathbf{y}} = [\tilde{y}(t_1), \cdots, \tilde{y}(t_n)]^{\mathrm{T}}$ is the observation of the output, $\mathbf{y}(\boldsymbol{\theta}) = [y(t_1, \boldsymbol{\theta}), \cdots, y(t_n, \boldsymbol{\theta})]^{\mathrm{T}}$ is the true value of the output and $\boldsymbol{\varepsilon} = [\varepsilon(t_1), \cdots, \varepsilon(t_n)]^{\mathrm{T}}$ is the measurement noise. In practice the measurement noise is often assumed to be normally distributed with zero mean and a covariance matrix $\boldsymbol{\Sigma}$. For this case, the measurements are also normally distributed and their density function is given by

$$p(\tilde{\mathbf{y}} \mid \boldsymbol{\theta}) = \left[(2\pi)^n \det \boldsymbol{\Sigma} \right]^{-1/2} \cdot \exp\left[-1/2 \left(\tilde{\mathbf{y}} - \mathbf{y}(\boldsymbol{\theta}) \right)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \left(\tilde{\mathbf{y}} - \mathbf{y}(\boldsymbol{\theta}) \right) \right]. \tag{2}$$

The partial derivative of the logarithm of the density function with respect to the parameter $\boldsymbol{\theta}$ can be written as

$$\frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \ln p(\tilde{\mathbf{y}} \mid \boldsymbol{\theta}) = \left(\tilde{\mathbf{y}} - \mathbf{y}(\boldsymbol{\theta}) \right)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^{\mathrm{T}}}, \tag{3}$$

which can be used for computation of the Fisher information matrix:

$$\begin{aligned} \mathbf{FIM} &= \mathrm{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}} \ln p(\tilde{\mathbf{y}} \mid \boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \ln p(\tilde{\mathbf{y}} \mid \boldsymbol{\theta}) \right] \\ &= \left(\frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \right)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \mathrm{E}\left[\left(\tilde{\mathbf{y}} - \mathbf{y}(\boldsymbol{\theta}) \right) \left(\tilde{\mathbf{y}} - \mathbf{y}(\boldsymbol{\theta}) \right)^{\mathrm{T}} \right] \boldsymbol{\Sigma}^{-1} \frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \\ &= \left(\frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \right)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^{\mathrm{T}}}. \end{aligned} \tag{4}$$

In many cases the noise is uncorrelated and the covariance matrix is a diagonal matrix with the variance along the diagonal. The Fisher information matrix can then be written as product of the sensitivity matrix with its transpose

$$\mathbf{FIM} = \left(\frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \right)^{\mathrm{T}} \frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^{\mathrm{T}}}, \tag{5}$$

where the variances of the covariance matrix are viewed as scaling factors that can be omitted as they does not affect the experimental design. The inverse of the Fisher information matrix is the Cramer-Rao lower bound (Walter and Pronzato, 1990).

A set of real functions of the Fisher information matrix need to be defined to compare the information content of the matrix for experimental design. These functions are called the experimental optimality criteria and are named alphabetically (Kiefer, 1959; Pukelsheim, 1993; Silvey, 1980). The most popular criterion is the *D*-optimality criterion which maximizes the logarithm of the determinant of the Fisher information matrix:

$$\varphi_D{}^* = \max \varphi_D(\mathbf{FIM}) = \max \log \det(\mathbf{FIM}). \tag{6}$$

This criterion minimizes the volume of the confidence ellipsoid with an arbitrary fixed confidence level for a least square estimator. This criterion is used in this work, however, the presented techniques can be easily generalized to other criteria.

2.2 Simultaneous Perturbation Stochastic Approximation (SPSA)

The parameter selection and optimal experimental design procedure results in an objective function where the expectation of the criterion function needs to be maximized

$$\max_{\mathbf{w}} \mathrm{E}_{\mathbf{v}} \varphi(\mathbf{w}, \mathbf{v}), \tag{7}$$

where $\mathbf{w}$ is a vector of the decision variables and $\mathbf{v}$ is a vector of random variables according to some distribution. The expectation is operated on $\mathbf{v}$ and returns a function of $\mathbf{w}$.

To evaluate the objective function directly a multi-dimension integral is required to calculate. A much more efficient method is the stochastic approximation as described by Robbins and Monro (1951). The method avoids evaluating the expectation and solves such kind of optimization problem using the update of

$$\mathbf{w}_{k+1} = \mathbf{w}_k + a_k \hat{\mathbf{g}}_k, \tag{8}$$

where $\hat{\mathbf{g}}_k$ is the gradient of the criterion function at some value of $\mathbf{v}$:

$$\hat{\mathbf{g}}_k(\mathbf{v}) = \frac{\partial}{\partial \mathbf{w}} \varphi(\mathbf{w}_k, \mathbf{v}) . \quad (9)$$

The gradient of the criterion function is calculated in a next step. One possible procedures for this is to approximate the gradient using the method of simultaneous perturbation (Spall, 1992) which only uses 2 evaluations of the criterion function to calculate the gradient

$$\hat{\mathbf{g}}_k = \begin{bmatrix} \dfrac{\varphi(\mathbf{w}_k + c_k\boldsymbol{\Delta}_k, \mathbf{v}_k) - \varphi(\mathbf{w}_k - c_k\boldsymbol{\Delta}_k, \mathbf{v}_k)}{2c_k\Delta_{k1}} \\ \vdots \\ \dfrac{\varphi(\mathbf{w}_k + c_k\boldsymbol{\Delta}_k, \mathbf{v}_k) - \varphi(\mathbf{w}_k - c_k\boldsymbol{\Delta}_k, \mathbf{v}_k)}{2c_k\Delta_{kp}} \end{bmatrix}, \quad (10)$$

where $\boldsymbol{\Delta}_k = [\Delta_{k1}, \cdots, \Delta_{kp}]^{\mathrm{T}}$ refers to the perturbation. A sampling point $\mathbf{v}_k$ is generated according to the distribution for each iteration to evaluate $\hat{\mathbf{g}}_k$.

The parameters used for SPSA can be selected as

$$a_k = a \Big/ (k + 1 + A)^{\alpha} , \quad (11)$$

and

$$c_k = c \Big/ (k + 1)^{\gamma} , \quad (12)$$

where the values α=0.602 and γ=0.101 are often recommended in practice (Spall, 1998). The variables a, c are initial values of the coefficients and A is chosen to reduce the value of a_k from one iteration to the next. Those values of these variables can be determined by trial and error. Each component of the perturbation $\boldsymbol{\Delta}_k$ can use a Bernoulli ±1 distribution with probability of 1/2 for each ±1 outcome.

3. SIMULTANEOUS SELECTION OF PARAMETER SUBSETS AND EXPERIMENTAL DESIGN

The following four types of variables exist which affect state estimation: time-varying adjustable variables whose values can change over time and can be manipulated, time-invariant adjustable variables which are also adjustable variables but which cannot be changed after time zero, parameters which are estimated and the unknown factors whose values are not known nor need to be known. The sensitivity values and the FIM are evaluated along the state trajectories so the FIM is a dependent upon the choice of these four types of variables

$$\mathbf{FIM} = \mathbf{FIM}(\mathbf{u}(t), \mathbf{v}, \boldsymbol{\theta}, \boldsymbol{\varphi}) , \quad (13)$$

where $\mathbf{u}(t)$ is a vector of time-varying adjustable variables, $\mathbf{v}$ is a vector of time-invariant adjustable variables, $\boldsymbol{\theta}$ is a vector of the parameters, and $\boldsymbol{\varphi}$ is a vector of the unknown factors.

Though the values of the parameters and the unknown factors are not known when the FIM is computed, some information about their uncertainty such as the range of the values or their distribution is often available. Therefore, these two types of variables can be described as random variables according to some distribution based on the knowledge of the uncertainty.

Some criterion of the FIM can be used to denote how good a subset of parameters is for parameter estimation. A selected parameter subset should have a large mean criterion value over the uncertainty range of the unknown variables so that the chosen parameter set has a high average value of the criterion.

The experimental conditions for generating a data set for parameter estimation are determined by the trajectories of the manipulated variables. Since the manipulated variables can be adjusted, they should be chosen such that a trajectory has a beneficial effect on parameter estimation. The optimal trajectory of the manipulated variables can be determined by optimizing the experimental criterion. It needs to be emphasized here that the selection of parameter sets is dependent on the trajectories of the controlled variables while on the other hand the optimal trajectory of the manipulated variables is dependent on the parameters selected for estimation. Therefore, these two procedures are performed simultaneously.

A new formulation of the parameter set selection that takes into account the effect that these four types of variables have on the FIM is given by (14). The binary variable $\mathbf{z}$ denotes if a parameter i is selected for estimation (z_i=1) or is set to its constant value (z_i=0). The first two constraints represent the system equations while the third and the fourth constraint are the sensitivity equations. The system equations and the sensitivity equations are solved simultaneously to compute the sensitivity value. The sensitivity matrix is formed by combining the sensitivity values at different time points. Some columns of the sensitivity matrix are selected according to the decision variables $\mathbf{z}$ to compute the Fisher information matrix. The parameters $\boldsymbol{\theta}$ and other unknown variables $\boldsymbol{\varphi}$ are assumed to be random variables with distribution functions based on knowledge of their uncertainties. Solving this optimization problem determines the subset of parameters which can be estimated most accurately in the sense of averaging over the unknown variables. The optimal trajectories of the manipulated variables for generating a data set for parameter estimation are also computed.

The optimization problem (14) is a mixed-integer nonlinear programming (MINLP) problem which also includes the solution of a set of differential equations. This type of problem is difficult to solve by standard methods. However, according to the special features of the parameter selection, a two-stage meta-heuristic method is developed in this work.

$$
\begin{aligned}
&\mathbf{z}^*, \mathbf{u}(t)^*, \mathbf{v}^* = \arg\max_{\mathbf{z},\mathbf{u}(t),\mathbf{v}} \ \mathrm{E}_{\boldsymbol{\theta},\boldsymbol{\varphi}}\left[\phi\left(\mathbf{F}(\mathbf{z},\mathbf{u}(t),\mathbf{v},\boldsymbol{\theta},\boldsymbol{\varphi})\right)\right] \\
&s.t. \quad \frac{d\mathbf{x}}{dt} = \mathbf{f}\left(\mathbf{x},\mathbf{u}(t),\mathbf{v},\boldsymbol{\theta},\boldsymbol{\varphi}\right), \quad \mathbf{x}(0) = \mathbf{x}_0\left(\mathbf{v},\boldsymbol{\varphi}\right) \\
&\quad y = h\left(\mathbf{x},\mathbf{u}(t),\mathbf{v},\boldsymbol{\theta},\boldsymbol{\varphi}\right) \\
&\quad \frac{d}{dt}\frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}^{\mathrm{T}}}\frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} + \frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}^{\mathrm{T}}}, \quad \frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}^{\mathrm{T}}}(0) = \mathbf{0} \\
&\quad \frac{\partial y}{\partial \boldsymbol{\theta}^{\mathrm{T}}} = \frac{\partial h}{\partial \mathbf{x}^{\mathrm{T}}}\frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} + \frac{\partial h}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \\
&\quad \mathbf{S} = \left[\frac{\partial y(t_1)}{\partial \boldsymbol{\theta}} \ \cdots \ \frac{\partial y(t_n)}{\partial \boldsymbol{\theta}}\right]^{\mathrm{T}} \\
&\quad \mathbf{F} = \left(\mathbf{SL}\right)^{\mathrm{T}}\left(\mathbf{SL}\right) \\
&\quad \mathbf{L} = \left[\mathbf{e}_{i_1} \ \ \mathbf{e}_{i_2} \ \cdots \ \mathbf{e}_{i_m}\right], \text{ with } i_j \text{ that } z_{i_j} = 1 \\
&\quad \mathbf{e}_i \text{ is the } i\text{-th column of the identity matrix} \\
&\quad z_i \in \{0,1\} \text{ and } L_z \le \sum_i z_i \le U_z \\
&\quad \mathbf{L}_{\mathbf{u}} \le \mathbf{u}(t) \le \mathbf{U}_{\mathbf{u}} \\
&\quad \mathbf{L}_{\mathbf{v}} \le \mathbf{v} \le \mathbf{U}_{\mathbf{v}} \\
&\quad \boldsymbol{\theta} \text{ is a random variable with p.d.f. } p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \\
&\quad \boldsymbol{\varphi} \text{ is a random variable with p.d.f. } p_{\boldsymbol{\varphi}}(\boldsymbol{\varphi}).
\end{aligned}
\qquad (14)
$$

The nominal criterion value, which is calculated while all the variables are fixed at their nominal values, can be used to identify a collection of the promising candidate subset which are likely to become the optimal subset of the problem (14), even though the nominal criterion value by itself is not likely to be useful for determining the optimal subset due to uncertainty in the parameter values and changes in the inputs. In practice it is unlikely that a subset which has a very low nominal criterion value will have a very high mean criterion value even when the adjustable variables are optimized. Selection of a collection of the promising subsets is a prescreening method which eliminates unimportant subsets from consideration and decreases the number of the subsets to be searched. A genetic algorithm can be used to determine such a collection of parameter sets (Chu and Hahn, 2007). The more subsets are included, the more likely it is that the optimal parameter set will be included in this collection of sets.

Determining a collection of parameter sets for estimation of their optimal input trajectories, fixes the values of the discrete variable **z** for each individual parameter set. As a result, the optimization problem described by (14) simplifies to the one given by (7). SPSA can then be applied to solve this optimization problem to determine the optimal trajectories of the manipulated variables. The mean criterion value for uncertainty in the parameter values can then be computed for each set of parameters for the optimal input trajectory. The parameter set which has the largest mean criterion among all the ones selected by the first step of this procedure should be chosen as the one used for parameter estimation. A flowchart of the described procedure is shown in Fig. 1.

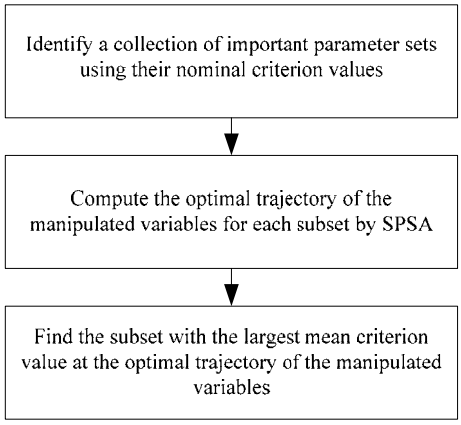


Fig. 1. Flowchart of the presented method

Another advantage of this procedure, when compared to methods which are only performed at the nominal values of the parameters, is that a collection of parameter sets rather than just one optimal set is determined. This is important in practice because the criterion value of the optimal set may be just slightly higher than that of a suboptimal one. In this case it is likely that the two parameter sets will be equally good for all practical purposes. Additionally, it is possible that some practical considerations play a role in the decision making process as well. If two parameter sets have similar values for their statistical criteria then a parameter set can be selected among the group of good parameter sets using other criteria.

4. CASE STUDY

A model of an exothermic continuously-stirred tank reactor (Muske and Georgakis, 2003) is used to illustrate the technique for parameter set selection. The differential equations are

$$
\begin{aligned}
\dot{c}_A &= \frac{F}{V}(c_A^f - c_A) - ke^{-E/RT}c_A \\
\dot{T} &= \frac{F}{V}(T^f - T) + \frac{\Delta H}{C_P}ke^{-E/RT}c_A - \frac{hA}{C_P V}(T - T_c), \\
\dot{T}_c &= \frac{F_c}{V_c}(T_c^f - T_c) + \frac{hA}{C_{P_c}V_c}
\end{aligned}
\qquad (15)
$$

The three states of the system are the concentration of one of the components, the temperature of the reactor, and the temperature of the coolant jacket. The reactor temperature is measured and is the only output of the system. The variables and their nominal values are listed in Table 1.

In the analysis all the variables are normalized by their nominal values to eliminate effects caused by different units. The uncertain parameters and the unknown factor are assumed to be uniformly distributed in the range from 20% to 180% of their nominal values. The range of the manipulated variables is assumed to be from 50% to 150% of their

nominal values. The time window for generating data for parameter estimation is set to 8 hr and the time-varying inputs are assumed to be step functions where a step can occur every hour. The algorithm is implemented in Matlab.

Table 1. Nominal value of the variables in the CSTR model

Parameter	Sym-bol	Nominal value	Type
Feed temperature	T^f	20 ºC	Parameter
Feed composition	c^f_A	2500 mol/m^3	
Fluid heat capacity	C_P	1600 kJ/ m^3•ºC	
Heat of reaction	ΔH	160 kJ/mol	
Activation energy	E/R	255 K	
Pre-exponential factor	k	2.5 h^{-1}	
Coolant inlet temperature	T^f_c	10 ºC	
Coolant heat capacity	C_{Pc}	1200 kJ/ m^3•ºC	
Heat transfer coefficient	h	1000 W/m^2•ºC	
Initial value of composition	c_{A0}	1000 mol/m^3	Unknown factor
Feed flow rate	F	0.1 m^3/h	Time-varying input
Coolant flow rate	F_c	0.15 m^3/h	
Initial value of reactor temperature	T_0	20 ºC	Time-invariant input
Initial value of coolant temperature	T_{c0}	20 ºC	
Reactor volume	V	0.2 m^3	Design parameter
Cooling volume	V_c	0.055 m^3	
Heat transfer area	A	4.5 m^2	

Table 2 presents the singular values of the sensitivity matrix at the nominal point. The smallest singular value is near zero indicating that the sensitivity matrix is nearly rank deficient and not all the parameters are estimable in practice. The sum of the first five singular values is more than 99% of the sum of all the singular values. Accordingly, the number of parameters to be estimated can be chosen to be equal to five with little loss of information that cannot be represented by this set of parameters.

Table 2. Singular values of the sensitivity matrix at the nominal point

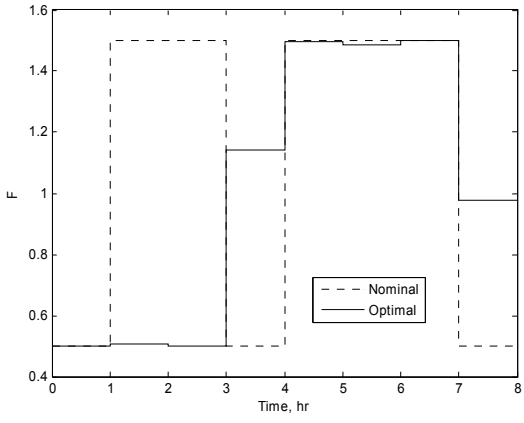
No.	1	2	3	4	5	6	7	8	9
s.v.	18.1	2.88	0.99	0.67	0.56	0.16	0.05	0.02	0.00

The 10 parameter sets with the largest nominal criterion value are selected by an exhaustive search and are shown in Table 3. While an exhaustive search can be performed for the system shown here with a reasonably small number of parameters, an approach using heuristic optimization procedures, e.g., genetic algorithms, can be used for systems with a larger number of parameters (Chu and Hahn, 2007). Once a parameter set is selected then the optimal input trajectory can be found by SPSA. The time-dependent profiles of the two manipulated variables, i.e., the feed flow rate F and the coolant flow rate F_c, are shown in Fig. 2 for the parameter set $\{c^f_A, \Delta H, T^f_c, C_{Pc}, h\}$ which is the parameter set with the largest mean criterion value at the optimal setting of the manipulated variables.

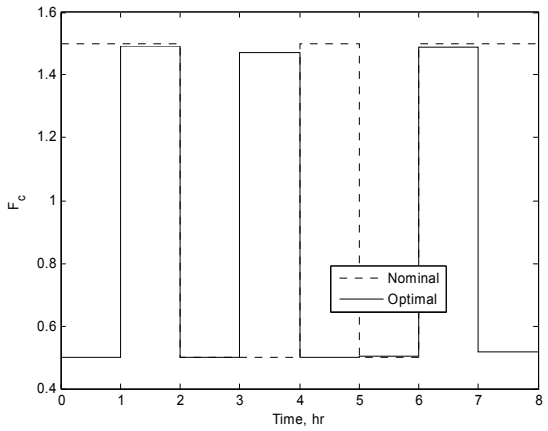
Table 3. The selected parameter subsets

No.	Subsets	Optimal mean D-criterion	Mean D-criterion	Nominal D-criterion
1	$c^f_A, \Delta H, T^f_c, C_{Pc}, h$	5.22	3.74	3.43
2	$c^f_A, C_P, \Delta H, T^f_c, C_{Pc}$	4.93	3.83	3.43
3	$c^f_A, C_P, T^f_c, C_{Pc}, h$	4.80	3.78	3.43
4	$c^f_A, k, T^f_c, C_{Pc}, h$	4.59	4.39	3.86
5	c^f_A, C_P, k, T^f_c, h	4.58	4.16	3.74
6	$c^f_A, E/R, T^f_c, C_{Pc}, h$	4.49	3.65	3.44
7	$c^f_A, C_P, \Delta H, T^f_c$, h	4.44	3.71	3.43
8	$c^f_A, C_P, k, T^f_c, C_{Pc}$	4.36	4.27	3.78
9	$c^f_A, C_P, E/R, T^f_c, C_{Pc}$	4.11	3.54	3.41
10	$c^f_A, C_P, E/R, T^f_c, h$	4.03	3.41	3.36

The 10 selected parameter sets in Table 3 are ordered by the mean criterion value for the optimal input trajectory. The mean criterion is calculated from 1000 sampling points for varying values of the uncertain parameters. The mean criterion value at the nominal setting of the manipulated variables and the criterion value evaluated with the parameters fixed at their nominal values are also shown in the table. It can be seen that there are clear differences between the three methods and which set of parameters should be estimated. The parameter set $\{c^f_A, k, T^f_c, C_{Pc}, h\}$ should be estimated if an input trajectory is fixed and it is assumed that the parameters are at or near their nominal values. A local approach to parameter set selection would return this result. While this result coincidentally turns out to also be the best parameter set if uncertainty in the parameter values is taken into account, it is nevertheless not the optimal parameter set if the input trajectory is allowed to change. There are three sets of parameters which are better for parameter estimation for a more appropriate choice of the input trajectory.



(a)



(b)

Fig. 2. Trajectories of (a) optimal feed flow rate F and (b) optimal coolant flow rate F_c

It can also be concluded that the criterion value can significantly increase if an optimal input trajectory is used. In fact, each of the 10 parameter sets shown in Table 3 has a larger mean value of the D-optimality criterion for its optimal input trajectory than the best parameter set has for a the fixed input trajectory and with parameters at the nominal point. It should be noted that this is the case despite the fact that the original input trajectory excited by the system with values ranging from the smallest to the largest value for both inputs.

The subset of $\{c^f_A, \Delta H, T^f_c, C_{Pc}, h\}$ has the largest mean criterion value for its optimal input trajectory, however, when the input trajectories are assumed to be fixed and the parameters are at their nominal point then this parameter set is only ranked as the sixth best one and would likely be overlooked. Therefore, it is important to integrate optimal experimental design and investigation of the effect that uncertainty in the model parameters has on parameter sensitivity analysis into the parameter set selection procedure. Additionally, it can be useful to investigate a collection of parameter sets instead of focus on one optimal set as many factors can influence the choice of a parameter set (What is the uncertainty range of the parameters? To what degree can an input trajectory be changed?) as it is important to make a comparison between results computed for different situations.

5. CONCLUSION

A prerequisite for successful parameter estimation is appropriate selection of parameters which can be accurately estimated from data. Though many methods for parameter selection have been developed and applied to a wide range of systems, most of them are based upon optimizing a certain criterion value, where the computation of the value is in many cases based upon local information. However, it can be shown the criterion value varies with the assumed nominal values of the parameters, of the system initial conditions, and the values of the input signals for nonlinear systems. Accordingly, there is a need to develop techniques which simultaneously compute the set of parameters to be estimated and the input trajectories for generating a data set. This paper addresses this point by formulating an optimization problem in addition to presenting a solution technique for this problem. A collection of suboptimal sets of parameters is identified and the optimal input trajectories for each parameter set are computed using SPSA. The parameter set to be estimated is selected by the largest mean criterion value for an optimal input trajectory. The technique was illustrated by applying it to the model of a CSTR.

REFERENCES

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

Brun, R., M. Kuhni, H. Siegrist, W. Gujer and P. Reichert (2002). Practical identifiability of ASM2d parameters - systematic selection and tuning of parameter subsets. *Water Res.*, **38**, pp. 4113-4127.

Chu, Y. and J. Hahn (2007). Parameter set selection for estimation for nonlinear dynamic systems. *AIChE J.*, **53**, pp. 2858-2870.

Kiefer, J. (1959). Optimum experimental designs. *J. R. Stat. Soc. Ser. B-Stat. Methodol.*, **21**, pp. 272-319.

Li, R.J., M.A. Henson and M.J. Kurtz (2004). Selection of model parameters for off-line parameter estimation. *IEEE Trans. Control Syst. Technol*., **12**, pp. 402-412.

Ljung, L. (1999). *System Identification: Theory for the User. 2nd ed*. Prentice Hall PTR, Upper Saddle River, NJ.

Muske, K.R. and C. Georgakis (2003). Optimal measurement system design for chemical processes. *AIChE J*., **49**, pp. 1488-1494.

Nelles, O. (2001). *Nonlinear System Identification: from Classical Approaches to Neural Networks and Fuzzy Models*. Springer, Berlin.

Pukelsheim, F. (1993). *Optimal design of experiments*. Wiley, New York.

Robbins, H. and S. Monro (1951). A stochastic approximation method. *Ann. Math. Stat*., **22**, pp. 400-407.

Sandink, C.A., K.B. McAuley and P.J. McLellan (2001). Selection of parameters for updating in on-line models *Ind. Eng. Chem. Res*., **40**, pp. 3936-3950.

Silvey, S.D. (1980). *Optimal design: an introduction to the theory for parameter estimation*. Chapman and Hall, London.

Spall, J.C. (1992). Multivariate stochastic-approximation using a simultaneous perturbation gradient approximation. *IEEE Trans. Autom. Control*, **37**, pp. 332-341.

Spall, J.C. (1998). Implementation of the simultaneous perturbation algorithm for stochastic optimization. *IEEE Trans. Aerosp. Electron. Syst*., **34**, pp. 817-823.

Walter, E. and L. Pronzato (1990). Qualitative and quantitative experiment design for phenomenological models - A survey. *Automatica*, **26**, pp. 195-213.

Weijers, S.R. and P.A. Vanrolleghem (1997). A procedure for selecting best identifiable parameters in calibrating activated sludge model no.1 to full-scale plant data. *Water Sci. and Tech*., **36**, pp. 69-79.

Velez-Reyes, M. and G.C. Verghese (1995). Subset selection in identification, and application to speed and parameter estimation for induction machines. In *Proc. 4th IEEE Conf. on Control Applications*, Albany, pp. 991 – 997.

Yao, K.Z., B.M. Shaw, B. Kou, K.B. McAuley and D.W. Bacon (2003). Modeling ethylene/butene copolymerization with multi-site catalysts: Parameter estimability and experimental design. *Poly. Reac. Eng*., **11**, pp. 563-588.