Parameter Set Selection for Estimation of Nonlinear Dynamic Systems

Yunfei Chu and Juergen Hahn
Dept. of Chemical Engineering, Texas A&M University, College Station, TX 77843

DOI 10.1002/aic.11295
Published online September 17, 2007 in Wiley InterScience (www.interscience.wiley.com).

A new approach is introduced for parameter set selection for nonlinear systems that takes nonlinearity of the parameter-output sensitivity, the effect that uncertainties in the nominal values of the parameters have and the effect that inputs and initial conditions have on parameter selection into account. In a first step, a collection of (sub)optimal parameter sets is determined for the nominal values of the parameters using a genetic algorithm. These parameter sets are then further analyzed for uncertainty in the parameters and changes in the initial conditions and inputs using differential analysis as well as a sampling-based approach to determine the key factors influencing sensitivity and the likelihood of a parameter set to be the optimal set under these varying conditions. The outcome of this procedure is a collection of parameter sets, which can be used for parameter estimation and additional information about how likely it is that a set is optimal for parameter estimation. Additionally, the size of the region in parameter space in which a certain set of parameters will remain optimal is determined. © 2007 American Institute of Chemical Engineers *AIChE J,* 53: 2858–2870, 2007
Keywords: parameter selection, parameter estimation, experimental design, sensitivity analysis, nonlinear dynamic systems

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, accurate estimation of parameters can be challenging as models can contain hundreds or even thousands of parameters while at the same time experimental data gathered for parameter estimation may be sparse and noisy. It is usually not possible to estimate the values of all the parameters accurately from the experimental data. It is the purpose of this work to develop a new approach for determining sets of parameters that should be estimated.

Parameter sensitivity analysis and experimental design are closely related techniques. The Fisher information matrix (FIM) serves as a measure of how much information about the parameters can be extracted from an experiment.[1–3] If the Fisher information matrix is far from being singular in some sense then parameters are practically identifiable.[4,5] A subset of parameters, which can be estimated accurately, is selected based upon optimizing certain criteria,[1,6] as it is usually not possible to estimate the values of all parameters. A combination of the *D*-optimality and the modified *E*-optimality criteria has been used to determine identifiable parameters.[7,8] If the Fisher information matrix is not close to being singular, then the norm of the sensitivity vectors is likely to be reasonably large and the angles between the sensitivity vectors are not small, either. Following these two rules, several parameter-selection techniques have been developed based on the sensitivity vectors, such as an orthogonalization method[9] and a recursive approach based upon principal component analysis.[10]

However, these parameter selection approaches are local methods, since parameter sensitivities will vary depending upon the choice of nominal values of parameters. The inherent uncertainty in the parameter values poses a challenge on parameter selection. Sequential design is the most common

Correspondence concerning this article should be addressed to J. Hahn at hahn@tamu.edu.

approach to handle the described challenge[5,11–13]: a set of initial values for the parameters is used for experimental design and to estimate parameters. The newly estimated parameter values are then used for another round of experimental design where values of the parameters are re-estimated. While such a procedure can be useful for systems where it is possible to perform a relatively large number of experiments, it can pose problems for systems such as intracellular signal pathways,[14,15] as experiments can take weeks of preparation and can be expensive. Other procedures such as Bayesian methods[16,17] and maximin methods[18,19] require intensive computation and may prohibit applicability to systems with a large number of parameters.

Another challenge that arises for dynamic systems is that sensitivities need to be calculated along state trajectories, which result in the Fisher information matrix being dependent not only on the parameter values but also on the initial states and inputs. It is the aim of this work to present a parameter set selection technique for dynamic systems described by nonlinear autonomous differential equations, which will take the effect of uncertainties of the parameter values and initial states as well as changes of the inputs into account. Analysis of possible parameter sets to determine their likelihood to be the optimal set for parameter estimation as well as the magnitude of the region in parameter space under which a set will remain optimal form important components of this work.

A collection of (sub)optimal parameter sets is investigated rather than just focusing on the "optimal" set due to the following reasons: (i) the differences in the values of the optimality criteria between the "optimal" set and a suboptimal set may be negligible and it may not be possible to distinguish between them in practice; (ii) the "optimal" set may only be the best set at the nominal point and it may be worse than a suboptimal set if the nominal values of the parameters are slightly different than it was originally thought; (iii) further analysis can concentrate on these important sets rather than considering all possible subsets of parameters; (iv) some experimental limitations may not have been taken into account when deriving the "optimal" set of parameters and determining several sets of potential candidates for parameter estimation can allow more flexibility for conducting experiments. A collection of suboptimal sets is determined by a genetic algorithm and is subsequently analyzed to determine the key factors influencing the sensitivity and to compute which parameter sets work best when uncertainty in the nominal values of the parameters is taken into account.

Preliminaries

Optimality criterion for parameter set selection

A certain class of nonlinear dynamical autonomous systems can be described by

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta}), \tag{1}$$

$$\mathbf{y} = \mathbf{g}(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta}), \tag{2}$$

where $\mathrm{x} \in R^{n_x}$ is the state vector, $\mathbf{u} \in R^{n_u}$ is the input vector, $\mathbf{y} \in R^{n_y}$ is the output vector, and $\boldsymbol{\theta} \in R^{n_\theta}$ is the parameter vector. The observations of the outputs are sampled at a series of time points and are subject to measurement noise

$$\mathbf{y}_{\mathrm{o}}(t_i) = \mathbf{y}(t_i) + \varepsilon(i), \quad i = 1 \cdots n_t.$$

In most practical applications, the noise $\{\varepsilon(i)\}$ is assumed to be white noise with a Gaussian distribution. Parameter estimation deals with the computation of the unknown values of the parameters from available input and output data. However, many systems contain a large number of parameters and it is not possible in practice to estimate the values of all of them from available data. A result of this is that in practice a subset of parameters is usually selected for estimation while other parameters are fixed at their nominal values.

A measure for evaluating the ability of a parameter set to capture changes in the output is required to compare different parameter sets. The Fisher information matrix[4,6] plays a central role in determining such a criterion as it captures the effect that parameters have on the outputs. The Fisher information matrix is calculated based on the output sensitivities

$$\mathbf{FIM} = \sum_{i=1}^{n_y} \sum_{j=1}^{n_t} \frac{1}{\sigma_{ij}^2} \frac{\partial y_i(t_j)}{\partial \boldsymbol{\theta}} \frac{\partial y_i(t_j)}{\partial \boldsymbol{\theta}^{\mathrm{T}}}. \tag{3}$$

or in the form of a sensitivity matrix

$$\mathrm{FIM} = \mathbf{S}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \mathbf{S}. \tag{4}$$

where the sensitivity matrix and the covariance matrix are

$$\mathbf{S} = \begin{bmatrix} \partial y_1(t_1)/\partial\theta_1 & \cdots & \partial y_1(t_1)/\partial\theta_{n_\theta} \\ \vdots & \ddots & \vdots \\ \partial y_1(t_{n_t})/\partial\theta_1 & \cdots & \partial y_1(t_{n_t})/\partial\theta_{n_\theta} \\ \vdots & \ddots & \vdots \\ \partial y_{n_y}(t_1)/\partial\theta_1 & \cdots & \partial y_{n_y}(t_1)/\partial\theta_{n_\theta} \\ \cdots & \ddots & \vdots \\ \partial y_{n_y}(t_{n_t})/\partial\theta_1 & \cdots & \partial y_{n_y}(t_{n_t})/\partial\theta_{n_\theta} \end{bmatrix},$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11}^2 & & & & & & \\ & \ddots & & & & & \\ & & \sigma_{1n_t}^2 & & & & \\ & & & \ddots & & & \\ & & & & \sigma_{n_y 1}^2 & & \\ & & & & & \ddots & \\ & & & & & & \sigma_{n_y n_t}^2 \end{bmatrix}$$

where σ_{ij}^2 denotes the variance of the noise in the observation of output y_i at time point t_j. The sensitivity values are often normalized by multiplying with the nominal values of parameters and by dividing through the nominal value of the outputs to ensure that different units for the parameters/outputs do not affect the sensitivity results. Furthermore, it is possible to normalize the sensitivity values by dividing with the square root of the variance assuming that the variance is known. In this case, the Fisher information matrix is the product of the transpose of the sensitivity matrix and the sensitivity matrix itself as Σ will be the identity matrix due to the normalization.

The inverse of the Fisher information matrix provides a lower bound for the asymptotic covariance matrix of parameter estimators,[4] and it can serve as a measure for the quality of a parameter set. However, a specific criterion is required to evaluate the information contained in the Fisher information matrix. Several criteria have been defined in the literature and they have been named alphabetically.[6] The most popular criterion is the *D*-optimality criterion, which minimizes the logarithm of the determinant of the inverse of the Fisher information matrix. Since the inverse of the determinant of a matrix is equal to the determinant of its inverse, the *D*-optimality criterion is equivalent to maximizing the logarithm of the determinant of the Fisher information matrix:

$$\begin{aligned}\phi_D^* &= \max \phi_D(\mathbf{FIM}) = \max\ \log\ \det(\mathrm{FIM}) \\ &= \min\ \log\ \det(\mathrm{FIM}^{-1}).\end{aligned} \tag{5}$$

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.

The Fisher information matrix of the selected parameter set will likely have to be far from being singular to maximize the optimality criterion. There are two components to achieving this: (1) the norm of the sensitivity vector of a selected parameter should be quite large and (2) the sensitivity vectors of parameters in the selected set cannot be linearly dependent. These two rules can be used to directly select the parameters from the sensitivity matrix. The orthogonalization method, which is based on the Gram–Schmidt orthogonalization, has been applied to the selection of parameters for chemical reactors[9] and also to applications in the field of systems biology.[11,12,20,21] Using this technique, the first parameter to be selected has the largest norm of the sensitivity vector. In a second step, the sensitivity vectors of other parameters are projected on to the space normal to the first sensitivity vector. From the projected sensitivity vectors, the longest one is chosen and the corresponding parameter is selected as the second parameter of the step. This same procedure is repeated to determine the remaining parameters of the set until the number of parameters to be estimated is reached or until the length of the projected sensitivity vectors decreases below a certain threshold. In fact the orthogonalization method is a sequential approach to maximize the *D*-optimality criterion as will be shown in the first subsection (*Parameter Subset Selection by GA*) under the *Presentation of a New Parameter Subset Selection Procedure* section.

First- and second-order sensitivity calculation

For systems described by Eqs. 1 and 2, the sensitivities are calculated along the output trajectories. To calculate the sensitivities the values of the parameters, inputs and initial states are required, as variations of these factors will change the sensitivity values. In this work, only autonomous systems are considered and the inputs are assumed to be constant. To simplify the expression, the initial states and the inputs are concatenated into an augmented parameter vector $\boldsymbol{\psi} \in R^{n_\theta+n_x+n_u}$

$$\boldsymbol{\psi} = [\boldsymbol{\theta}^{\mathrm{T}}, \mathbf{x}_0^{\mathrm{T}}, \mathbf{u}^{\mathrm{T}}]^{\mathrm{T}}. \tag{6}$$

The sensitivity values of the state variables $\mathbf{x}(t)$ with respect to a parameter ψ_i can be calculated by differentiating both sides of the state Eq. 1 to obtain

$$\frac{d}{dt}\frac{\partial \mathbf{x}}{\partial \psi_i} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}^{\mathrm{T}}}\frac{\partial \mathbf{x}}{\partial \psi_i} + \frac{\partial \mathbf{f}}{\partial \psi_i}, \quad i = 1 \cdots n_\theta + n_x + n_u. \tag{7}$$

If ψ_i is an initial value of a state, x_0, then the second term of the right hand side of Eq. 7 is zero. The initial conditions of the differential equations are given by

$$\left.\frac{\partial \mathbf{x}}{\partial \psi_i}\right|_{t=0} = \begin{cases} \mathbf{0}, \text{ if } \psi \in \boldsymbol{\theta} \text{ or } \psi_i \in \mathbf{u} \\ \mathbf{e}_j, \text{if } \psi_i = \mathbf{x}_0(j) \end{cases}, \tag{8}$$

where $\mathbf{e}_j \in R^{n_x}$ is a vector with entries of 1 on its *j*th element and entries of 0 on all other elements. By solving the sensitivity Eq. 7 and the state Eq. 1 simultaneously, the sensitivity values are calculated along the state/output trajectories.

Only the first-order sensitivity values are required to evaluate the Fisher information matrix for parameter selection. However, the second-order sensitivity values play an important role in the uncertainty analysis of the selection procedure and will be used in the *Presentation of a New Parameter Subset Selection Procedure* section. The second-order sensitivities can be calculated by differentiating both sides of Eq. 7

$$\begin{aligned}\frac{d}{dt}\frac{\partial^2 \mathbf{x}}{\partial \psi_i \partial \psi_j} &= \frac{\partial \mathbf{f}}{\partial \mathbf{x}^{\mathrm{T}}}\frac{\partial^2 \mathbf{x}}{\partial \psi_i \partial \psi_j} + \frac{\partial^2 \mathbf{f}}{\partial \psi_i \partial \mathbf{x}^{\mathrm{T}}}\frac{\partial \mathbf{x}}{\partial \psi_j} + \frac{\partial^2 \mathbf{f}}{\partial \mathbf{x}^{\mathrm{T}} \partial \psi_j}\frac{\partial \mathbf{x}}{\partial \psi_i} \\ &\quad + \frac{\partial^2 \mathbf{f}}{\partial \psi_i \partial \psi_j} + \left(\mathbf{I}_N \otimes \frac{\partial \mathbf{x}^{\mathrm{T}}}{\partial \psi_j}\right)\mathbf{H}\frac{\partial \mathbf{x}}{\partial \psi},\end{aligned} \tag{9}$$

$$i, j = 1 \cdots n_\theta + n_x + n_u$$

where $\mathbf{H} \in R^{n_x n_x \times n_x}$ is the Hessian matrix of $\mathbf{f}$

$$\mathbf{H} = \left[\frac{\partial^2 f_1}{\partial \mathbf{x} \partial \mathbf{x}^{\mathrm{T}}} \quad \frac{\partial^2 f_2}{\partial \mathbf{x} \partial \mathbf{x}^{\mathrm{T}}} \quad \cdots \quad \frac{\partial^2 f_{n_x}}{\partial \mathbf{x} \partial \mathbf{x}^{\mathrm{T}}}\right]^{\mathrm{T}}.$$

To calculate the second-order sensitivities, the values of the state variables and the first-order sensitivities need to be known.

Several software packages exist for efficiently solving the differential equations for the sensitivity calculations, e.g., VODE,[22] DASPK,[23] or SUNDIALS.[24] In this work, the Matlab ODE solver is used, as the sensitivity calculations can easily be integrated with the other calculations required for the presented analysis scheme.

Sampling-based method

Sampling-based methods, also called Monte Carlo simulations, are a popular approach for uncertainty and sensitivity analysis. The techniques use a set of sampling points for the uncertain parameters and compute a performance characteristic

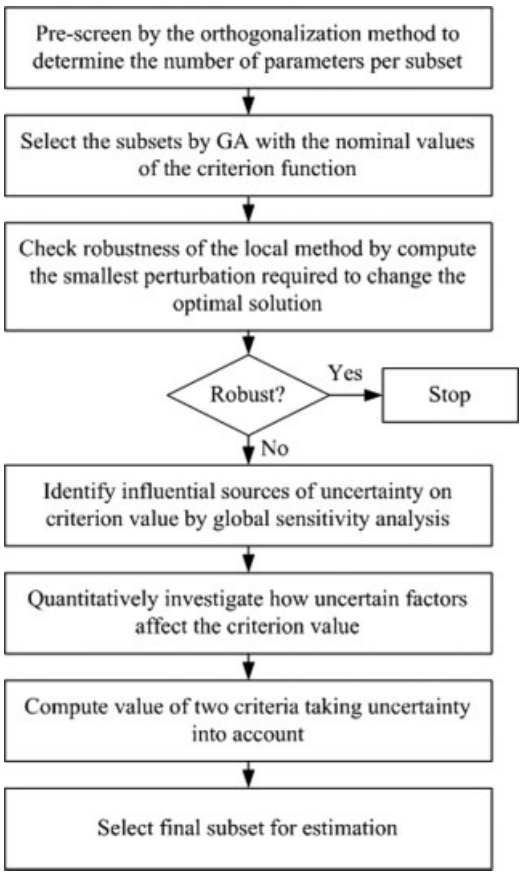


Figure 1. Flow diagram of procedure for parameter subset selection.

for each sampling point. Three sampling procedures are widely used[25]: random sampling, stratified sampling, and Latin hypercube sampling. The random sampling method generates the value of each parameter using its distribution. This method has the most easy-to-interpret statistical meaning but there is no assurance that points will be sampled from any given subregion of the sample space and the technique can be inefficient if several sampled values are close to one another. Stratified sampling divides the uncertainty space into some disjoint strata and obtains a random sample from each stratum to ensure that the important regions of the uncertainty space will be covered by sampling. Latin hypercube sampling divides the range of each parameter into several intervals of equal length. In every interval, one sample point, i.e., a value, of a parameter is selected. The sample points of one parameter are determined separately from those of other parameter and a sample vector is formed by randomly pairing the values of different parameters.

The important parameters, which have a strong influence on the output over a range of values, are identified by statistical analysis. The sampling values for each parameter are grouped into two sets, namely the "acceptable set" and the "unacceptable set." The elements in the "acceptable set" are the parameter values where the corresponding output values are below a threshold value. However, if the corresponding output value is larger than the threshold, then the parameter value is grouped into the "unacceptable set." The *KS* (Kolmogorov–Smirnov) statistic,[26] which is determined by the largest vertical gap between the cumulative functions of the values of a parameter in the two sets, is used as the sensitivity measure

$$KS = \sup|S_a(x) - S_u(x)|. \tag{10}$$

where S_a and S_u are the cumulative distribution functions respectively associated with the "acceptable set" and the "unacceptable set." The larger the value of the *KS* statistic, the more sensitive the output with respect to the factor. Details of sampling-based methods for uncertainty and sensitivity analysis can be found in the literature.[27]

Presentation of a New Parameter Subset Selection Procedure

This section presents a new procedure for parameter set selection for parameter estimation of nonlinear dynamic systems. The contribution of this technique is that it combines a method for selecting parameter sets with uncertainty analysis to determine when a parameter set that is suboptimal for the nominal values of the parameters may become optimal due to changes of the nominal values. A flow diagram of the procedure that is used in this work is shown in Figure 1.

Parameter subset selection by GA

Parameter selection procedures search for a subset of parameters, which maximizes an optimality criterion. One specific form of such an optimization problem is given by

$$\begin{aligned} &\mathbf{z}^* = \arg\max_{\mathbf{z}} \ \phi_D(\mathbf{F}(\mathbf{z})) \\ &\text{s.t. } \mathbf{F}(\mathbf{z}) = \mathbf{FIM}^{(i_1,\cdots i_{n_s})}_{(i_1,\cdots i_{n_s})} \text{ with } i_j \text{ that } z_{i_j} = 1, j = 1\cdots n_s \\ &\quad z_1 + z_2 + \cdots + z_{n_\theta} = n_s \\ &\quad z_i \in (0,1), i = 1\cdots n_\theta. \end{aligned} \tag{11}$$

The decision vector $\mathbf{z} \in \{0,1\}^{n_\theta}$ denotes whether a parameter is included in the selected parameter subset. If $z_i = 1$, then θ_i belongs to the selected subset with the size of n_s. The value of n_s can be determined through prescreening by the orthogonalization method. **FIM** is the Fisher information matrix of all parameters. $\mathbf{F}(\mathbf{z})$ is the Fisher information matrix of the parameters included in the selected subset, and it is equal to the principal submatrix of **FIM** with the indices of the nonzero decision variables (the entries of column i_j and row i_k, j, $k = 1\ldots n_s$).

This optimization problem results in a nonlinear integer programming problem. While an exhaustive search is a simple approach to find the optimal solution,[7,8] this is not a practical approach for any problem of reasonable size. Sequential methods, which add parameters to the subset one at a time, are able to significantly reduce the computational burden. It will be shown that the orthogonalization method is a sequential approach, which maximizes the *D*-criterion at each step. To elaborate on this point, the QR decomposition is used to express the orthogonalization

$$\mathbf{S} = \mathbf{QR}, \tag{12}$$

where **S** is the normalized sensitivity matrix of the selected parameters, **Q** is an orthogonal matrix, and **R** is an upper triangular matrix. The columns of **Q** form the unit orthogonal bases of the space spanned by the sensitivity vectors (the columns of **S**) and the columns of **R** are the coordinates of the sensitivity vectors on the orthogonal bases. When a new parameter is selected, its sensitivity vector is added to **S**, a new base is added to **Q**, and the coordinates of the sensitivity vector on the bases are added to **R**. The new diagonal entry of **R** denotes the projected value of the last sensitivity vector on the space normal to the sensitivity vectors of the previously selected parameter. The orthogonal method maximizes the square of the new diagonal entry of **R** at each step when a new parameter is selected. The determinant of the information matrix is related to the determinant of **R** by

$$\det(\mathbf{S}^{\mathrm{T}}\mathbf{S}) = \det(\mathbf{R}^{\mathrm{T}}\mathbf{R}) = \det(\mathbf{R})^2. \tag{13}$$

Because **R** is upper triangular, the determinant of the information matrix is equal to the product of the squared diagonal entries of **R**. Accordingly, the orthogonalization method, which maximizes the squared diagonal entry of **R** at each step, can be regarded as a sequential method that maximizes the *D*-criterion at each step. However, due to the sequential nature of the orthogonalization method, it is possible that parameter sets with even larger criterion values may be missed as they can only be found by a simultaneous approach. That being said, this procedure can still be implemented as a prescreening tool as it is straightforward to implement and does not require extensive computations.

It is important to select a set of estimable parameters for parameter estimation; however, the parameter set corresponding to the optimal criterion value at the nominal point may not always be the best choice due to the optimality criterion changing with the nominal values of the parameters. Accordingly, a procedure is required to not only to compute the optimal set of parameters but also to determine a collection of suboptimal parameter sets. This can be achieved by using a genetic algorithm (GA)[28,29] to solve the optimization problems shown in Eq. 11. One distinct property of a GA is that it involves a population of potential solutions to the problem. Multiple candidate solutions are considered simultaneously and according to the evolution law good population member has a larger chance to be preserved in the new generation than unfit members. After many generations, the population will usually contain many members with high fitness values. This property makes GA very suitable to solve the problem of subset selection. A collection of (sub)optimal solutions can be formed by choosing good candidates from each generation with a value of the optimality criterion larger than a threshold level α. This procedure will return a collection of parameter sets with near optimal value of the optimization problem shown in Eq. 11.

Determine the region in parameter space for which local results remain valid

Because of continuity of the optimality criterion, the optimal subset selected at the nominal value will still be the best set in a neighborhood around the nominal point. However, if the nominal values of the parameters can vary significantly, then the results computed by local sensitivity analysis may not be accurate over the entire range. A technique is presented in this subsection, which determines the smallest magnitude of parameter changes that is required such that the parameter set with the optimal value at the nominal point will lose its "top positions" to another set of parameters. The magnitude of the variation under which the chosen parameter set does not change is an indicator of the robustness of the results computed by the local method.

Since an analytical expression describing the relationship between the criterion function and the nominal values of to the parameters is usually not known in practice, a linear approximation of the sensitivity vectors is used:

$$\left.\frac{\partial \mathbf{y}}{\partial \theta_i}\right|_{\boldsymbol{\psi}+\Delta\boldsymbol{\psi}} = \left.\frac{\partial \mathbf{y}}{\partial \theta_i}\right|_{\boldsymbol{\psi}} + \left.\frac{\partial^2 \mathbf{y}}{\partial \boldsymbol{\psi}^{\mathrm{T}} \partial \theta_i}\right|_{\boldsymbol{\psi}} \Delta\boldsymbol{\psi}. \tag{14}$$

The sensitivity matrix contains the sensitivity vectors of a subset of parameters $\theta_{i_1}, \theta_{i_2}, ..., \theta_{i_{n_s}}$ and can be expressed by

$$\left.\left[\frac{\partial \mathbf{y}}{\partial \theta_{i_1}} \cdots \frac{\partial \mathbf{y}}{\partial \theta_{i_{n_s}}}\right]\right|_{\boldsymbol{\psi}+\Delta\boldsymbol{\psi}} = \left.\left[\frac{\partial \mathbf{y}}{\partial \theta_{i_1}} \cdots \frac{\partial \mathbf{y}}{\partial \theta_{i_{n_s}}}\right]\right|_{\boldsymbol{\psi}} + \left.\left[\frac{\partial^2 \mathbf{y}}{\partial \boldsymbol{\psi}^{\mathrm{T}} \partial \theta_{i_1}} \cdots \frac{\partial^2 \mathbf{y}}{\partial \boldsymbol{\psi}^{\mathrm{T}} \partial \theta_{i_{n_s}}}\right]\right|_{\boldsymbol{\psi}} (\mathbf{I}_{n_s} \otimes \Delta\boldsymbol{\psi}) \tag{15}$$

To simplify the notation,

$$\mathbf{S}_I = \bar{\mathbf{S}}_I + \mathbf{W}_I(\mathbf{I}_{n_s} \otimes \mathbf{d}). \tag{16}$$

will be used where

$$\mathbf{S}_I = \left.\left[\frac{\partial \mathbf{y}}{\partial \theta_{i_1}} \cdots \frac{\partial \mathbf{y}}{\partial \theta_{i_{n_s}}}\right]\right|_{\boldsymbol{\psi}+\Delta\boldsymbol{\psi}}$$

$$\bar{\mathbf{S}}_I = \left.\left[\frac{\partial \mathbf{y}}{\partial \theta_{i_1}} \cdots \frac{\partial \mathbf{y}}{\partial \theta_{i_{n_s}}}\right]\right|_{\boldsymbol{\psi}}$$

$$\mathbf{W}_I = \left.\left[\frac{\partial^2 \mathbf{y}}{\partial \boldsymbol{\psi}^{\mathrm{T}} \partial \theta_{i_1}} \cdots \frac{\partial^2 \mathbf{y}}{\partial \boldsymbol{\psi}^{\mathrm{T}} \partial \theta_{i_{n_s}}}\right]\right|_{\boldsymbol{\psi}}$$

$$\mathbf{d} = \Delta\boldsymbol{\psi}$$

$$I = \{i_1 \quad i_2 \quad \cdots \quad i_{n_s}\}$$

One should note that the linear approximation of the sensitivity vectors is used rather than the linear approximation of the optimality criterion itself as linearization of the sensitivity vector offers a more accurate approximation.

Suppose that a parameter set at the nominal point (indicated by indices *J*) has a larger criterion value than another parameter set (indicated by indices *I*). The smallest perturbation required to change the order of two parameter sets can

be calculated by the following optimization problem:

$$\begin{aligned} &\min \ \|\mathbf{d}\|_2 \\ \text{s.t. } &\phi_D(\mathbf{S}_I^{\mathrm{T}}\mathbf{S}_I) > \phi_D(\mathbf{S}_J^{\mathrm{T}}\mathbf{S}_J) \\ &\mathbf{S}_I = \bar{\mathbf{S}}_I + \mathrm{W}_I(\mathbf{I}_{n_s} \otimes \mathbf{d}) \\ &\mathbf{S}_J = \bar{\mathbf{S}}_J + \mathrm{W}_J(\mathbf{I}_{n_s} \otimes \mathbf{d}) \\ &\mathbf{d}_{\mathrm{L}} \le \mathbf{d} \le \mathbf{d}_{\mathrm{U}} \end{aligned} \tag{17}$$

The last inequality constraint provides an upper and a lower bound for variation of the parameter vector such that constraints on parameters by physics can be taken into account. For example, all the kinetic parameters in a model referring to rate constants should always be positive. It should be noted that due to the linear approximation of the sensitivity matrix it may be possible that the variation calculated may not change the order of the two subsets. In this case, the sensitivity values can be re-evaluated at the perturbed parameter value calculated by the first solution of optimization problem and the optimization problem is solved again. This is an iterative procedure that is performed until a perturbation is found that will change the order of the criterion values of the two sets.

Sampling-based method to identify sources of uncertainty that affect the value of the optimality criterion

The technique presented in this subsection uses global sensitivity analysis to determine how sensitive the optimality criterion is to sources of uncertainty. For the most part, these sources of uncertainty are due to changes in the values of the parameters; however, changes in initial conditions can also be considered.

A sampling-based method with Latin hypercube sampling is used in this work, since it is the most efficient sampling way for large systems. The optimality criterion is evaluated at each sampling point by simulating the model. The *KS* statistic of the criterion value with respect to a parameter is calculated to serve as the global sensitivity measures following the procedure described below

Step 1. Determine the uncertainty range of each parameter.

Step 2. Generate uniformly distributed samples of the parameters by Latin hypercube sampling.

Step 3. Calculate the first-order sensitivities by solving the state Eq. 1 and the sensitivity Eq. 7 simultaneously for each sample value and compute the value of the optimality criterion.

Step 4. Calculate the objective function for each sample.

$$f_\phi(k) = (\phi_D(k) - \bar{\phi}_D)^2,$$

where $\phi_D(k)$ is the criterion value calculated at the kth sample, $\bar{\phi}_D$ is the criterion value calculated at the nominal value.

Step 5. Calculate the mean value of $f_\phi(k)$ and group the sample values of each parameter into two sets. If $f_\phi(k)$ is larger than the mean value, then the kth sample value of the parameter is placed into the "unacceptable set"; otherwise, it is put into the "acceptable set."

Step 6. Compute the two cumulative distribution functions of the sample values contained in the two sets for each parameter and use Eq. 10 to calculate the *KS* statistic.

From the sampling points, the criterion functions that are not subject to the parameter uncertainty can be calculated. Because of the uncertainty, a subset of parameters can be estimated more accurate than another subset at one point but less accurate at another point. The mean criterion value of a subset indicates the overall performance of a subset. A good estimator of the mean criterion is the average criterion value on the sampling points. However, it is the case that a subset can have a large mean criterion value because it has a very large criterion value in a small range but has low criterion value over most of the parameter space. In practice, the situation where the subset has large criterion value is unlikely, and it is more likely that the subset is worse than others. One may prefer to select the subset, which has the largest probability to have the largest criterion value in the uncertain range. The probability of each subset to be the top one can be calculated from the sampling points as well. From the explanation above, the two criteria may not be completely consistent and an example is in the case study in the next section. One is often at loss to choose the criterion before selection. This is another motivation to select a collect of subsets. After calculation of the value of the two criteria of the subsets, one is easy to make a balance among different criteria.

Quantitative investigation of the effect of uncertain factors on the optimality criterion

Even though global sensitivity analysis is able to identify the important uncertain factors affecting the value of the optimality criterion, it is unable to determine quantitatively how changes in the nominal parameter values affect estimation accuracy. The gradient of the criterion function can provide such information, and it can be used directly to determine the optimal setting of adjustable variables. The mathematical procedure for this technique is provided in the following.

Assume a selected parameter subset is $\{\theta_{i_1}, \theta_{i_2}, ..., \theta_{i_{ns}}\}$ and the sensitivity matrix is

$$\mathbf{S}(\boldsymbol{\psi}) = \left[\frac{\partial \mathbf{y}}{\partial \theta_{i_1}} \quad \frac{\partial \mathbf{y}}{\partial \theta_{i_2}} \cdots \frac{\partial \mathbf{y}}{\partial \theta_{i_{ns}}}\right]\Bigg|_{\boldsymbol{\psi}=[\theta^{\mathrm{T}}, \mathbf{x}_0^T, \mathbf{u}^{\mathrm{T}}]^{\mathrm{T}}} \tag{18}$$

where $\mathbf{S}$ is evaluated at some value of the parameter $\boldsymbol{\psi}$ and the D-criterion is a function of $\boldsymbol{\psi}$

$$\phi_D(\boldsymbol{\psi}) = \log\det(\mathbf{S}(\boldsymbol{\psi})^{\mathrm{T}}\mathbf{S}(\boldsymbol{\psi})). \tag{19}$$

Differentiation of the criterion function results in

$$d\phi_D(\boldsymbol{\psi}) = 2\mathrm{trace}\{(\mathbf{S}(\boldsymbol{\psi})^{\mathrm{T}}\mathbf{S}(\boldsymbol{\psi}))^{-1}(\mathbf{S}(\boldsymbol{\psi})^{\mathrm{T}} d\mathbf{S}(\boldsymbol{\psi}))\}. \tag{20}$$

Since the differentiation of each element in the sensitivity matrix is

$$d\left(\frac{\partial \mathbf{y}}{\partial \theta_{i_j}}\right) = \frac{\partial^2 \mathbf{y}}{\partial \boldsymbol{\psi}^{\mathrm{T}} \partial \theta_{i_j}} d\boldsymbol{\psi}, \tag{21}$$

the differential of the sensitivity matrix is

$$d\mathbf{S}(\boldsymbol{\psi}) = \left[\frac{\partial^2 \mathbf{y}}{\partial \boldsymbol{\psi}^{\mathrm{T}} \partial \theta_{i_1}} d\boldsymbol{\psi} \quad \frac{\partial^2 \mathbf{y}}{\partial \boldsymbol{\psi}^{\mathrm{T}} \partial \theta_{i_2}} d\boldsymbol{\psi} \cdots \frac{\partial^2 \mathbf{y}}{\partial \boldsymbol{\psi}^{\mathrm{T}} \partial \theta_{i_{ns}}} d\boldsymbol{\psi}\right]. \tag{22}$$

Table 1. Nominal Value of CSTR Parameters

No.	Parameter	Variable	Value
1	Feed temperature	T^{f}	20°C
2	Feed composition	$c_{\mathrm{A}}^{\mathrm{f}}$	2500 mol/m^3
3	Fluid density	ρ	1025 kg/m^3
4	Heat of reaction	ΔH	160 kJ/mol
5	Activation energy	E/R	255 K
6	Pre-exponential factor	k	2.5 h^{-1}
7	Coolant inlet temperature	$T_{\mathrm{c}}^{\mathrm{f}}$	10°C
8	Coolant density	ρ_{c}	1000 kg/m^3
9	Heat transfer coefficient	h	1000 W m^{-2} °C^{-1}
10	Feed flow rate	F	0.1 m^3/h
11	Coolant flow rate	F_{c}	0.15 m^3/h
12	Initial state of composition	c_{A0}	1000 mol/m^3
13	Initial state of reactor temperature	T_0	20°C
14	Initial state of coolant temperature	T_{c0}	20°C
	Reactor volume	V	0.2 m^3
	Cooling jacket volume	V_{c}	0.055 m^3
	Heat transfer area	A	4.5 m^2
	Coolant heat capacity	$C_{\mathrm{P_c}}$	1.2 kJ kg^{-1} °C^{-1}
	Fluid heat capacity	C_{P}	1.55 kJ kg^{-1} °C^{-1}

Substituting Eq. 22 into the optimality criterion results in

$$d\phi_D(\boldsymbol{\psi}) = 2\mathrm{trace}\{[\mathbf{A}_1 d\boldsymbol{\psi} \quad \mathbf{A}_2 d\boldsymbol{\psi} \ \cdots \ \mathbf{A}_{n_s} d\boldsymbol{\psi}]\}, \tag{23}$$

where

$$\mathbf{A}_{i_1} = (\mathbf{S}^{\mathrm{T}}\mathbf{S})^{-1}\mathbf{S}^{\mathrm{T}}\frac{\partial^2 \mathbf{y}}{\partial \boldsymbol{\psi}^{\mathrm{T}} \partial \theta_{i_1}}. \tag{24}$$

Finally,

$$d\phi_D(\boldsymbol{\psi}) = 2(\mathbf{a}_1^{\mathrm{T}} + a_2^{\mathrm{T}} + \cdots + \mathbf{a}_{n_s}^{\mathrm{T}}) d\boldsymbol{\psi}. \tag{25}$$

where $\mathbf{a}_i^{\mathrm{T}}$ is the ith row of the matrix $\mathbf{A}_i$ and the partial derivative of ϕ_D with respect to $\boldsymbol{\psi}$ is

$$\frac{\partial \phi_D}{\partial \boldsymbol{\psi}} = 2\sum_{j=1}^{n_s} \mathbf{a}_j. \tag{26}$$

The magnitude of the gradient is an indicator of the effect that changes in a parameter have on the criterion function value. The sign of the gradient indicates whether a change of the value of a parameter increases or decreases the optimality criterion. The gradient shown in Eq. 26 is in fact the local sensitivity of the criterion function. However, this is not to be confused with the sensitivity of the output. The sensitivity of the output is used to compute the value of the criterion for parameter selection while the sensitivity of the criterion function is used to study the effects that parameter uncertainty has on the criterion value.

Case Studies

Two examples are used to illustrate the developed techniques. The first case study deals with an exothermic continuously-stirred tank reactor, while the second one analyzes a detailed model describing an IL-6 transduction network in liver cells.

Parameter set selection for a CSTR

This model describes an exothermic CSTR in which a first-order reaction A→B is taking place[30]:

$$\mathrm{A} \rightarrow \mathrm{B}, \ R_{\mathrm{A}} = k\exp(-E/RT)c_{\mathrm{A}}. \tag{27}$$

The reactor is described by the following differential equations

$$\begin{aligned} \dot{c}_{\mathrm{A}} &= \frac{F}{V}(c_{\mathrm{A}}^{\mathrm{f}} - c_{\mathrm{A}}) - R_{\mathrm{A}} \\ \dot{T} &= \frac{F}{V}(T^{\mathrm{f}} - T) + \frac{\Delta H}{\rho C_{\mathrm{p}}}R_{\mathrm{A}} - \frac{hA}{\rho C_{\mathrm{p}} V}(T - T_{\mathrm{c}}) \\ \dot{T}_{\mathrm{c}} &= \frac{F_{\mathrm{c}}}{V_{\mathrm{c}}}(T_{\mathrm{c}}^{\mathrm{f}} - T_{\mathrm{c}}) + \frac{hA}{\rho_{\mathrm{c}} C_{\mathrm{P_c}} V_{\mathrm{c}}} \end{aligned} \tag{28}$$

The three states of the system are the concentration of component A, the temperature of the reactor, and the temperature of the coolant jacket. The reactor temperature is chosen as the only output of the system.

All parameters in Eq. 28 are assumed to be constant. It can be seen that ρ and C_{P} never appear by themselves and only in the form of their product in Eq. 28. Because of this only the product of the two parameters can be estimated. The same situation arises for the product of ρ_{c} and $C_{\mathrm{P_c}}$. To take this observation into account, the parameters C_{P} and $C_{\mathrm{P_c}}$ are set to their nominal value and are not considered for parameter set selection. This leaves nine parameters (Nos. 1–9 in Table 1) as candidates considered for estimation. The feed flow rate and the coolant flow rate are the two input variables. These 11 variables plus the 3 initial conditions of the states make up the augmented parameter vector for sensitivity analysis. The reactor volume, the cooling jacket volume, and the heat transfer area are design parameters whose values are exactly known. Thus there is no need to consider them for parameter estimation.

The sensitivities of the reactor temperature with respect to the parameters are calculated according to Eqs. 7–9 and normalized. In the next step, the orthogonalization method is applied. The results are shown in Table 2, where the overall sensitivity and the rank value are shown for each parameter. It can be seen that while the output may be sensitive to some parameters that these parameters may nevertheless have a small rank value as they are highly correlated to parameters already chosen for the set. The method indicates that the coolant density ρ_{c}, the pre-exponential factor k, and the fluid density ρ form a set of three parameters that have the largest effect on the reactor temperature. The rank value of the 4th parameter is less than 0.7% of sum of the first three, and therefore the size of the parameter set is chosen to be three ($n_{\mathrm{s}} = 3$). The set $\{\rho_{\mathrm{c}}, k, \rho\}$ is a suboptimal selection under

Table 2. Parameters of the CSTR Model Ordered by the Orthogonalization Method

	Parameter								
	ρ_{c}	k	ρ	$c_{\mathrm{A}}^{\mathrm{f}}$	h	ΔH	$T_{\mathrm{c}}^{\mathrm{f}}$	E/R	T^{f}
Rank value	9.29	0.79	0.13	0.07	0.008	0.001	0	0	0
Sensitivity	9.29	1.30	0.56	2.09	7.11	2.66	3.72	1.07	0.29

Table 3. Collection of Suboptimal Subsets for CSTR Model

	No.									
	1	2	3	4	5	6	7	8	9	10
	Parameter Subset									
	ρ	c_A^f	c_A^f	c_A^f	ρ	ρ	ρ	k	c_A^f	c_A^f
	k	k	ΔH	E/R	k	E/R	ΔH	ρ_c	ΔH	k
	ρ_c	ρ_c	ρ_c	ρ_c	h	ρ_c	ρ_c	h	h	h
Criterion value	−0.15	−0.23	−0.27	−0.53	−0.54	−0.61	−0.68	−0.85	−0.94	−0.98
Mean criterion value	−0.13	−1.35	−1.35	−1.38	−0.89	−0.27	−0.34	0.21	−0.77	−0.70
Probability to be the optimum	0.150	0.061	0.029	0.053	0.037	0.129	0.170	0.153	0.083	0.137

the *D*-optimality. In fact the set is the optimal in this case but this is not always true (the next case is an example).

The total number of the possible subsets of parameters is $C^3_9 = 84$ and, it is therefore possible to perform an exhaustive search evaluating each set of parameters. The ten sets with the highest criterion value are shown in Table 3. Two thousand simulations with the augmented parameters varying from 0.5 to 2 (normalized value) have been performed to investigate the change of the criterion value with the uncertainty. The mean value of the criterion for each set for these 2000 simulations is listed in Table 3. It can be seen that there are significant differences between the criterion values at the nominal point and the mean values of the criteria for changes in the nominal value of the parameters. For example, the 8th set of parameters results in a higher mean value of the criterion under the influence of uncertainty in the parameter values than the best set for the nominal point. It can be concluded that determining a set of parameters to be estimated from data at a nominal point may not lead to an optimal conclusion. The last row in Table 3 denotes the probability for each subset to be the optimal set for the simulations that were run for the uncertain parameters. The probability is computed by the number of simulations where a subset has the largest criterion value divided by the total number of the simulations that were performed. The 7th parameter set from Table 3 has the largest probability to be the optimal set.

It can be seen from this analysis that there is not one set of parameters that will be the best one for both criteria if uncertainty is taken into account. Instead it is more useful to provide a collection of parameter sets as well as criteria to evaluate them and to have a user chose certain set based upon experience with a process. For example, even though the 8th subset has a slightly lower probability to be the best set compared to the 7th set, the mean criterion value of the 8th parameter set is larger than the one for the 7th set. Therefore, the 8th parameter set is the best choice for parameter estimation for this example. However, other criteria, e.g., experience that one has with a process, may also play a factor when choosing one parameter set over another one.

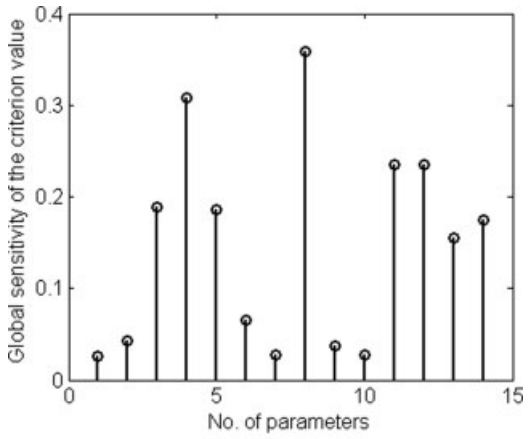
Table 4 lists the smallest variation of the augmented parameters required to change the order of a parameter set with the 1st set. From Table 4, it can be concluded that a small change in the nominal value of the parameters (0.7%) can change the selection of an optimal parameter set. Since the optimal set at the nominal point is extremely sensitive to the nominal values and since these nominal values are by definition imprecise, which is the reason why they need to be estimated, it is questionable if choosing an optimal parameter set simply based upon local sensitivity analysis returns meaningful results.

Another important conclusion that can be drawn from the results shown in Table 4 is that the magnitude of the smallest perturbation required to change the order of two subsets is not proportional to the difference of the criterion value between the two sets. The difference of criterion value between the 6th subset and the 1st subset is less than that between the 10th subset and the 1st subset. However, the variation required to change the order of the 6th subset with the 1st subset is much larger than the one required to make the 10th set more important than the one currently ranked 1st.

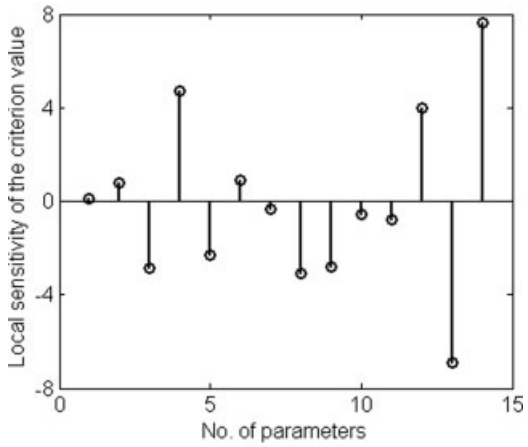
The global sensitivities of the criterion values with respect to the parameters are calculated to identify the influential uncertain sources. The *KS* statistic of the 1st parameter set, as one representative of a global sensitivity measure, is computed from the sampled points (Figure 2a). To study how variations of the parameters affect the criterion values, the gradient of the criterion function are also computed and shown in Figure 2b. The gradient is infact the local sensitivity of the criterion function. It can be concluded that the local and global sensitivity results in different information. The initial value of the coolant temperature T_{c0} (No. 14) has the largest magnitude of the local sensitivity, while the coolant density ρ_c (No. 8) has the largest global sensitivity. To

Table 4. Smallest Variation Required to Change Order of a Subset with the 1st One for the CSTR Model

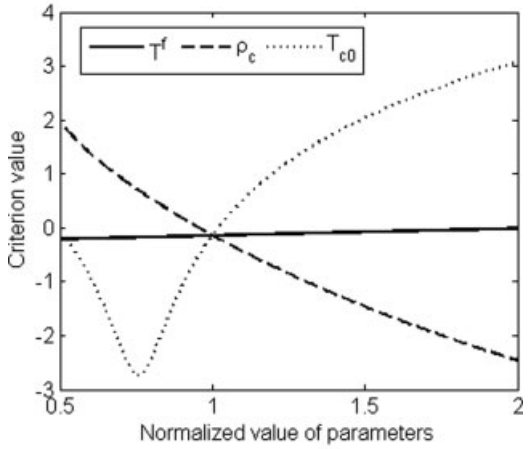
	No. of Subset								
	2	3	4	5	6	7	8	9	10
Variation magnitude	0.007	0.011	0.032	0.200	0.268	0.075	0.163	0.050	0.052



(a)



(b)



(c)

Figure 2. Sensitivity analysis of the criterion value for the subset {ρ, k, ρ_c} for the CSTR model.

(a) Global sensitivity of criterion value; (b) local sensitivity of criterion value; (c) change of criterion value with variation of T^f, ρ_c, and T_0.

investigate the reasons behind these different observations for local and global analysis, the criterion values have been plotted for variations of some specific parameters in Figure 2c. It can be seen that varying T_{c0} strongly changes the criterion value around the nominal value but has a diminishing effect for large values of T_{c0}. Also the criterion value does not decrease monotonically as T_{c0} decreases. The criterion value changes monotonically with changes in ρ_c in the whole range. On the other hand, the feed temperature T^f (No. 1) has only marginal effects by changes in its nominal value and it has small value of both global sensitivity and local sensitivity.

Parameter subset selection of an IL-6 signaling pathway

Modeling and analysis of intracellular signaling networks is an important area in systems biology. Signaling pathways are the cellular information routes by which cells sense their surroundings and adjust to environmental changes or hormonal stimuli. The signaling network includes various components, which detect, amplify, and integrate diverse external signals to generate responses such as changes in enzyme activity or gene expression.

The IL (interleukin)-6-type cytokines are an important family of mediators involved in the regulation of the acute-phase response to injury and infection.[31] Several models of the IL-6 signaling pathway have been proposed and a recently developed model is presented in the paper by Singh et al.,[15] which describes signal transduction in hepatocytes induced by IL-6 (Figure 3). This model contains two signaling mechanisms: Janus-associated kinases (JAK) and signal transducers and transcription factors 3 (STAT3) are activated in one pathway, while the other pathway involves the activation of mitogen-activated protein kinases (MAPK). The model is described by 68 nonlinear ordinary differential equations including 118 parameters. The equations are derived according to the law of mass action or Michaelis–Menten kinetics and the parameters are the kinetic rate constants. The states are the concentrations of the molecules involved in the pathway. The input is the concentration of IL-6 that stimulates the pathway and the output is the concentration of the transcription factor $(STAT3N^*)_2$ (dimer of activated STAT3 in the nucleus). For the detailed model of the differential equations and the nominal values of parameters, one can see Singh et al.[15] and Chu et al.[32]

The investigated model contains a total of 118 parameters. Since the analysis procedure could be computationally prohibitive for such a large number of parameters, only the 50 most important parameters, as identified by the sensitivity value, will be investigated here. Also, 16 of the 68 states have initial conditions different from zero and variations in these initial conditions are also considered in this work.

The order of parameters selected by the orthogonalization method is shown in Table 5. (The parameter k_{fi} is the rate constant of the forward reaction in the ith pathway and k_{bi} is the rate constant of the backward reaction of in the ith pathway.) It can be seen that having more than six parameters does not provide much of a benefit as the additional contribution of the 7th parameter is less than 1% to what can already be achieved by choosing six parameters. Accordingly the size of the subset is determined to be 6 ($n_s = 6$).

The total number of possible sets with six parameters is C^6_{50}, which is roughly 1.6×10^7. It is not possible to perform an exhaustive search in this case due to the large

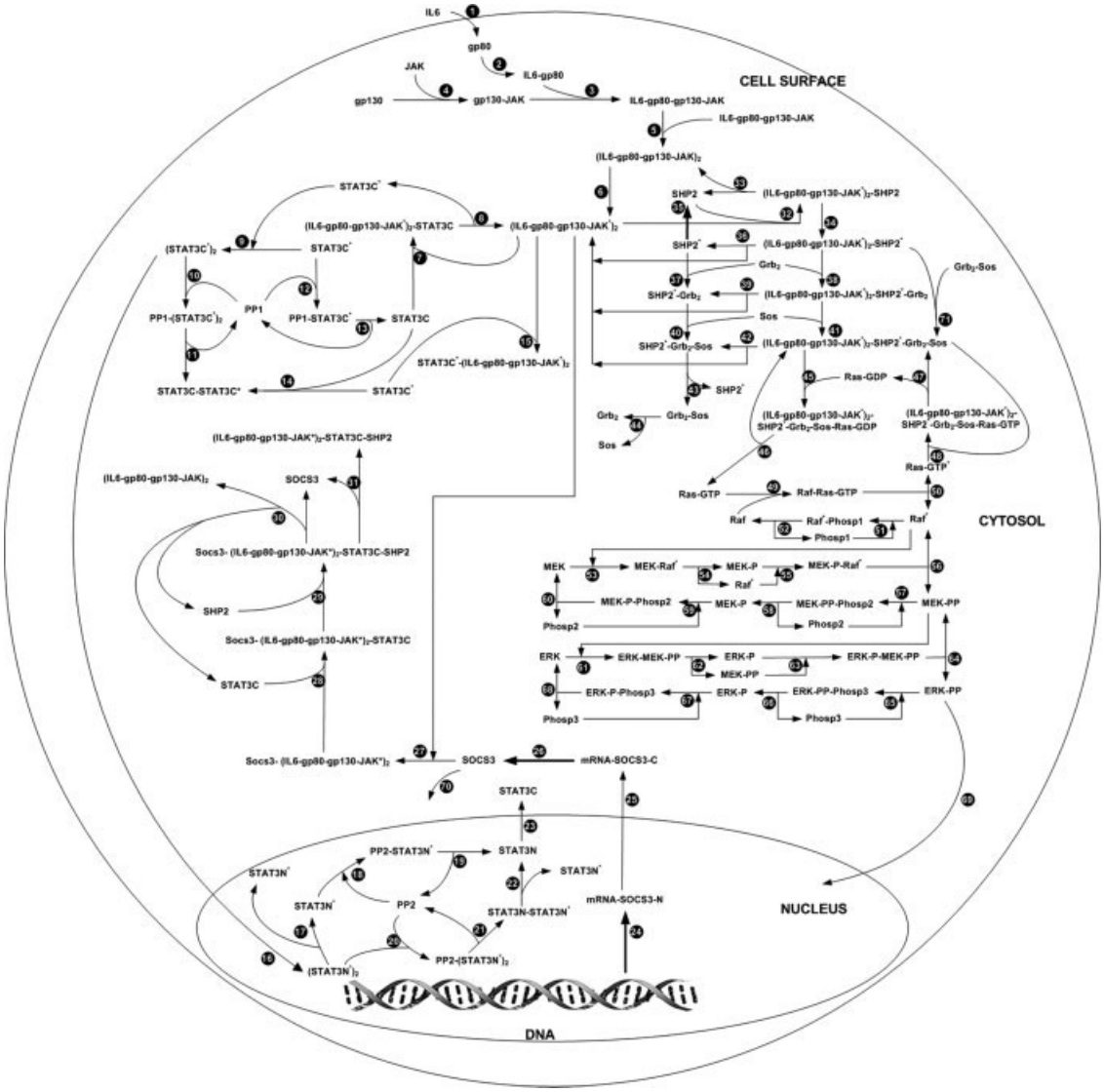


Figure 3. Diagram of IL-6 signaling pathway.

number of possible sets; however, a selection procedure based upon a GA can still be applied. A certain threshold for determining a cutoff of sets to be considered has been found by trial and error. This cutoff has been set to a value of 55 ($\alpha = 55$) resulting in 38 parameter sets, which are considered for further analysis. The ten parameter sets with the largest criterion values are shown in Table 6. It can be seen that the differences in the criterion values between these ten sets is small, which can serve as an indicator that further analysis may be warranted rather than simply using the "optimal" set for parameter estimation.

The mean criterion value is calculated for 2000 simulations where the nominal values of the augmented parameters can vary from 30 to 300% and the results are shown in Figure 4a. The mean criterion value of the 1st subset is about 38.5, while the 10th and the 13th subsets have larger mean value. The probability of a subset to be the optimal one due to uncertainty in the nominal values is shown in Figure 4b. The 1st subset, which is optimal at the nominal point, has small probability to preserve its top position for small changes in the (estimated) nominal value. Its probability of being the top choice is only 8% of the probability of the 28th set, and there are 35 parameter sets that have larger probability of being the optimal choice. It can be concluded that if the uncertainty effects are not considered then the selected parameter set may be far from being the optimal one.

Table 5. First Ten Parameters in the IL-6 Model Ordered by the Orthogonalization Method

	No.									
	1	2	3	4	5	6	7	8	9	10
	k_{f7}	k_{f31}	k_{f21}	k_{f70}	k_{f6}	k_{f48}	k_{f45}	k_{f26}	k_{f18}	k_{f8}
Rank value	637.3	239.6	187.5	63.2	46.3	23.0	11.7	5.3	3.2	1.0

The third column represents the parameters.

Table 6. Ten Sets of Parameters with Largest Performance Indices for the IL-6 Model

	No.									
	1	2	3	4	5	6	7	8	9	10
	Parameter Subset									
	k_{f31}	k_{f31}	k_{f31}	k_{f31}	k_{f7}	k_{f7}	k_{f31}	k_{f31}	k_{f31}	k_{f31}
	k_{f21}	k_{f21}	k_{f21}	k_{f21}	k_{f31}	k_{f31}	k_{f21}	k_{f21}	k_{f21}	k_{f21}
	k_{f70}	k_{f6}	k_{f70}	k_{f6}	k_{f21}	k_{f21}	k_{f70}	k_{f70}	k_{f70}	k_{f6}
	k_{f6}	k_{f48}	k_{f6}	k_{f32}	k_{f6}	k_{f70}	k_{f48}	k_{f16}	k_{f6}	k_{f48}
	k_{f48}	k_{f32}	k_{f32}	k_{f26}	k_{f48}	k_{f6}	k_{f16}	k_{f32}	k_{f32}	k_{f32}
	k_{f32}	k_{f26}	k_{b48}	k_{b48}	k_{f26}	k_{f48}	k_{f32}	k_{b48}	k_{f42}	k_{f27}
Criterion value	56.83	56.82	56.63	56.62	56.58	56.57	56.51	56.32	55.92	55.89

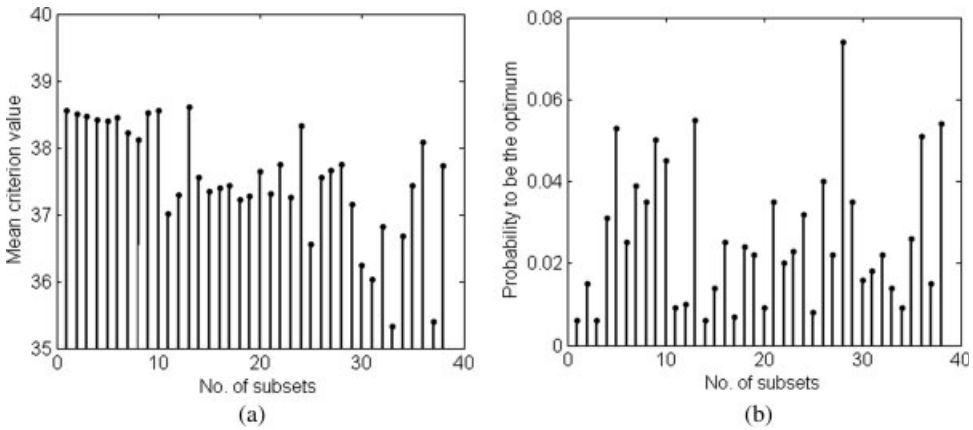


Figure 4. Results calculated by sampling-based method for IL-6 model.

(a) Mean criterion value of a subset; (b) probability for a subset to be the optimal one.

The smallest distance from the nominal point to the point at which another parameter set has larger criterion value than the 1st parameter set is shown in Table 7. It can be seen that a small perturbation of 0.2% of the nominal values in parameter space is able to change the optimal solution. This observation also indicates that the 1st subset, which was determined by local analysis, is likely to lose its top position due to uncertainties in the nominal values of the parameters.

To study how individual parameters change the criterion value, the global sensitivities and the local sensitivities of the criterion with respect to the 1st parameter set are shown in Figure 5. It can be seen that the initial concentration of JAK (No. 54), which has the largest magnitude of the local sensitivity, also has the 3rd largest contribution when global sensitivity analysis is applied. Similarly, the initial concentration of SHP2 (No. 55), which is determined as being most important by global sensitivity, also has the 2nd largest magnitude for local sensitivity.

From the local sensitivities, it can be seen how some biological mechanism affect the estimation accuracy. The initial concentration of STAT3C (No. 55) has the largest local sensitivity and an increase of the initial value raises the optimality criterion. STAT3C is one of the main proteins in the JAK/STAT signaling pathway. The initial concentration of JAK (No. 54) also has large positive sensitivity. JAK is an essential component for forming the receptor complex, which is in turn required to initiate signal transduction. The initial concentration of SHP2 (No. 56) has the largest magnitude among the negative sensitivities. Increase of the initial value of SHP2 will decrease the value of the optimality criterion for this parameter set. SHP2 is an important protein for signaling through the MAPK pathway. The initial conditions of the two inhibitors PP1 (No. 57), which deactivates STAT3C in the cytoplasm, and PP2 (No. 58), which deactivates STAT3N in the nucleus, also have a negative effect on the value of the objective function.

Similar argument can be made for important parameters of the signal transduction pathway model. The parameter k_{f7} (No. 1) has the larger positive sensitivity than any other parameters. k_{f7} is involved in the reaction where STAT3 is activated by the receptor complex, and large values of k_{f7} increase the rate of activation. The parameter k_{f32} (No. 2) has the largest negative sensitivity. k_{f32} is involved in the reaction where SHP2 enables signal transduction through the MAPK pathway, which limits the transduction through the JAK/STAT pathway.

Table 7. Smallest Variation Required to Change the Order of a Subset with the 1st One for the IL-6 Model

	No. of Subset								
	2	3	4	5	6	7	8	9	10
Smallest variation	0.002	0.07	0.05	0.06	0.08	0.08	0.10	0.13	0.15

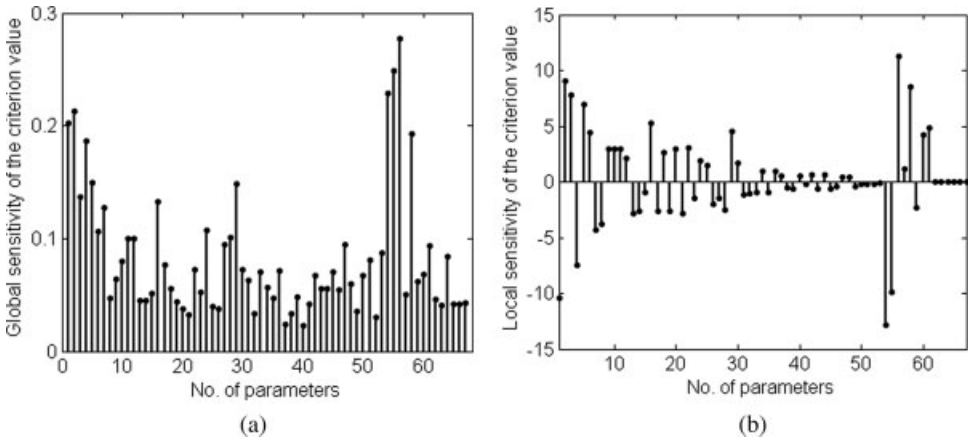


Figure 5. Sensitivity analysis of criterion value of the best subset in the IL-6 model.

(a) Global sensitivities of criterion value; (b) local sensitivities of criterion value.

It is important to point out that the concentration of the cytokine IL-6 (No. 51) has only a mildly positive effect as is determined by local sensitivity analysis. The reason for this is that the nominal value of the input is so large that the cells are saturated with IL-6 and a change in the value of IL-6 will only have a minor effect on the output. However, it should be pointed out that this behavior will be very different if the IL-6 concentration were lower by an order or magnitude or more.

Conclusion

Selection of parameters, which can be estimated accurately from data, is a prerequisite for successful estimation. While it is straightforward to perform parameter sensitivity analysis to determine a set of parameters to be estimated, it may happen that the determined set is not the best one for estimation. The reason for this is that results from local parameter sensitivity analysis depend upon the nominal values of the parameter, which are by definition not precisely known, and on values of the initial conditions and inputs. This article investigated these points as a family of parameter sets can be selected by the *D*-optimality criterion in combination with the orthogonalization method, where the optimization was performed based upon a genetic algorithm.

In a second step, the smallest perturbation required to change the optimal solution is determined to check if the results returned by the local method are acceptable. It has been illustrated in the case studies that the optimal solution can be extremely sensitive to parameter uncertainty and a more detailed analysis may be required. This analysis should start by determining which sources of uncertainty are affecting the value of an optimality criterion. A method based upon global sensitivity analysis and another technique based upon local sensitivity analysis of the criterion value are presented in this work. Furthermore, the mean criterion value and the probability for a subset to be the optimal one for a specified region of the parameter space are used to evaluate the chosen sets of parameters.

The result of the presented technique is a collection of candidate sets of parameters for estimation with detailed information about the effect of uncertainty in the parameter values, initial conditions, and inputs on the optimality criterion. The provided information is also helpful for evaluating data used for parameter estimation or designing future experiments.

Acknowledgments

The authors gratefully acknowledge partial financial support from the ACS Petroleum Research Fund (Grant PRF# 43229-G9) and from the National Science Foundation (Grant CBET# 0706792).

Literature Cited

1. Silvey SD. *Optimal Design: An Introduction to the Theory for Parameter Estimation*. London: Chapman and Hall, 1980.
2. Pazman A. *Foundations of Optimum Experimental Design*. Dordrecht: D. Reidel Publishing Company, 1986.
3. Atkinson AC, Donev AN. *Optimum Experimental Designs*. Oxford: Oxford University Press, 1992.
4. Walter E, Pronzato L. Qualitative and quantitative experiment design for phenomenological models—a survey. *Automatica*. 1990;26:195–213.
5. Faller D, Klingmuller U, Timmer J. Simulation methods for optimal experimental design in systems biology. *Simul Trans Soc Model Simul Int*. 2003;79:717–725.
6. Kiefer J. Optimum experimental designs. *J R Stat Soc Ser B Stat Methodol*. 1959;21:272–319.
7. Weijers SR, Vanrolleghem PA. A procedure for selecting best identifiable parameters in calibrating activated sludge model no. 1 to full-scale plant data. *Water Sci Technol*. 1997;36:69–79.
8. Brun R, Kuhni M, Siegrist H, Gujer W, Reichert P. Practical identifiability of ASM2d parameters—systematic selection and tuning of parameter subsets. *Water Res*. 2002;36:4113–4127.
9. Yao KZ, Shaw BM, Kou B, McAuley KB, Bacon DW. Modeling ethylene/butene copolymerization with multi-site catalysts: parameter estimability and experimental design. *Polym React Eng*. 2003;11:563–588.
10. Li RJ, Henson MA, Kurtz MJ. Selection of model parameters for off-line parameter estimation. *IEEE Trans Control Syst Technol*. 2004;12:402–412.
11. Gadkar KG, Gunawan R. Doyle FJ III. Iterative approach to model identification of biological networks. *BMC Bioinformat*. 2005a;6:155.
12. Gadkar KG, Varner J, Doyle FJ III. Model identification of signal transduction networks from data using a state regulator problem. *Syst Biol*. 2005b;2:17–30.

13. Issanchou S, Cognet P, Cabassud M. Sequential experimental design strategy for rapid kinetic modeling of chemical synthesis. *AIChE J*. 2005;51:1773–1781.
14. Schoeberl B, Eichler-Jonsson C, Gilles ED, Muller G. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat Biotechnol*. 2002;20:370–375.
15. Singh A, Jayaraman A, Hahn J. Modeling regulatory mechanisms in IL-6 signal transduction in hepatocytes. *Biotechnol Bioeng*. 2006;95:850–862.
16. Chaloner K, Verdinelli I. Bayesian experimental design: a review. *Stat Sci*. 1995;10:273–304.
17. Han C, Chaloner K. Bayesian experimental design for nonlinear mixed-effects models with application to HIV dynamics. *Biometrics*. 2004;60:25–33.
18. Muller CH. Maximin efficient designs for estimating nonlinear aspects in linear-models. *J Stat Plan Infer*. 1995;44:117–132.
19. Dette H. Designing experiments with respect to 'standardized' optimality criteria. *J R Stat Soc Ser B Methodol*. 1997;59:97–110.
20. Jaqaman K, Danuser G. Linking data to models: data regression. *Nat Rev Mol Cell Biol*. 2006;7:813–819.
21. Yue H, Brown M, Knowles J, Wang H, Broomhead DS, Kell DB. Insights into the behaviour of systems biology models from dynamic sensitivity and identifiability analysis: a case study of an NF-κ B signalling pathway. *Mol Biosyst*. 2006;2:640–649.
22. Brown PN, Byrne GD, Hindmarsh AC. VODE—a variable-coefficient ode solver. *SIAM J Sci Stat Comput*. 1989;10:1038–1051.
23. Brown PN, Hindmarsh AC, Petzold LD. Using Krylov methods in the solution of large-scale differential-algebraic systems. *SIAM J Sci Stat Comput*. 1994;15:1467–1488.
24. Hindmarsh AC, Brown PN, Grant KE, Lee SL, Serban R, Shumaker DE, Woodward CS. SUNDIALS: suite of nonlinear and differential/algebraic equation solvers. *ACM Trans Math Software*. 2005;31:363–396.
25. Mckay MD, Beckman RJ, Conover WJ. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*. 2000;42:55–61.
26. Ross SM. *Simulation*, 4th edition. Amsterdam: Elsevier Academic Press, 2006.
27. Helton JC, Johnson JD, Sallaberry CJ, Storlie CB. Survey of sampling-based methods for uncertainty and sensitivity analysis. *Reliabil Eng Syst Safety*. 2006;91(10/11):1175–1209.
28. Goldberg DE. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading: Addison-Wesley, 1989.
29. Michalewicz Z. *Genetic Algorithms + Data Structures = Evolution Programs*. Berlin: Springer-Verlag, 1994.
30. Muske KR, Georgakis C. Optimal measurement system design for chemical processes. *AIChE J*. 2003;49:1488–1494.
31. Heinrich PC, Behrmann I, Haan S, Hermanns HM, Muller-Newen G, Schaper F. Principles of interleukin (IL)-6-type cytokine signalling and its regulation. *Biochem J*. 2003;374:1–20.
32. Chu Y, Singh A, Jayaraman A, Hahn J. Parameter sensitivity analysis of IL-6 signaling pathways. *IET Trans Syst Biol*. In press.

Manuscript received May 2, 2007, and revision received Aug. 1, 2007.