

# Parameter Set Selection for Dynamic Systems under Uncertainty via Dynamic Optimization and Hierarchical Clustering

Wei Dai, Loveleena Bansal, and Juergen Hahn

Dept. of Biomedical Engineering and Dept. of Chemical & Biological Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180

Daniel Word

Dept. of Chemical Engineering, Texas A&M University, College Station, TX 78743

DOI 10.1002/aic.14265

Published online November 6, 2013 in Wiley Online Library (wileyonlinelibrary.com)

*It is common that only a subset of the parameters of models can be accurately estimated. One approach for identifying a subset of parameters for estimation is to perform clustering of the parameters into groups based upon their sensitivity vectors. However, this has the drawback that uncertainty cannot be directly incorporated into the procedure as the sensitivity vectors are based upon the nominal values of the parameters. This article addresses this drawback by presenting a parameter set selection technique that can take uncertainty in the parameter space into account. This is achieved by defining sensitivity cones, where a sensitivity cone includes all sensitivity vectors of a parameter for different values, resulting from the uncertainty, in the parameter space. Parameter clustering can then be performed based upon the angles between the sensitivity cones, instead of the angle between sensitivity vectors. The presented technique is applied to two case studies. © 2013 American Institute of Chemical Engineers AICHE J, 60: 181–192, 2014*

**Keywords:** parameter estimation, parameter selection, sensitivity analysis, hierarchical clustering, dynamic optimization

## Introduction

Mathematical models composed of ordinary differential equations (ODEs) or differential algebraic equations (DAEs) are widely used to describe the behavior of dynamic systems, ranging from ecological systems,<sup>1</sup> power systems,<sup>2</sup> chemical processes,<sup>3</sup> biochemical reaction networks,<sup>4</sup> to pharmaceutical processes.<sup>5</sup> The accuracy of these models not only depends on the structure of the model which is determined by the physics, chemistry, and biology of the system, but also relies on adjustable parameters in the model, many of which are either taken from the literature or estimated using experimental data.

A number of studies have investigated various aspects of parameter estimation.<sup>6–8</sup> The parameter estimation problem can be formulated as an optimization problem which minimizes the norm of the error between data and model predictions while treating the equations of the system as constraints. The problem can then be solved by either a sequential approach or a simultaneous approach,<sup>9</sup> with both approaches offering advantages for certain classes of problems. However, before any parameter estimation is performed, it is important to determine if all parameters are numerically identifiable and, if not, then what subset of parameters can be accurately estimated.<sup>10</sup>

A variety of methods for parameter set selection based on sensitivity analysis have been proposed in the literature. These methods include, but are not limited to, genetic algorithms,<sup>11</sup> collinearity index methods,<sup>1</sup> column pivoting methods,<sup>12</sup> Gram–Schmidt orthogonalization methods,<sup>13</sup> and clustering methods.<sup>14</sup> A systematic scheme for parameter set selection is based on optimality criteria computed from the Fisher information matrix which is closely related to the parameter-output sensitivity matrix.<sup>15</sup>

It is important to note that all of the methods mentioned above for parameter set selection utilize local sensitivity analysis. The main drawback of local sensitivity analysis is that the sensitivity vectors are dependent on the parameter values that are not precisely known prior to parameter estimation. This may result in identification of parameter subsets that are suboptimal, which can have a significant impact on the model's prediction accuracy if the parameter uncertainty is large.<sup>16</sup> One alternative to local sensitivity analysis is global sensitivity analysis which simultaneously varies multiple parameters, often over a large range. Additionally, global sensitivity analysis can incorporate the uncertainty description of the parameters into the sensitivity analysis procedure. Unfortunately, the results from global sensitivity analysis, for example, the Morris method,<sup>17</sup> sampling-based method,<sup>18</sup> or a variance-based method,<sup>19</sup> are nontrivial to interpret for experimental design or parameter set selection<sup>16</sup> as they do not rely on the concept of sensitivity vectors.

This article addresses the challenges mentioned above for parameter set selection of dynamic systems under

Correspondence concerning this article should be addressed to J. Hahn at hahnj@rpi.edu.

uncertainty. This is achieved by combining a hierarchical clustering method<sup>14</sup> and dynamic optimization techniques, in order to simultaneously vary all parameters over a large range, to quantify the effect of the uncertainty in the parameter space on the sensitivity vectors. As the uncertainty of the parameter values has an effect on the sensitivity vectors, a sensitivity cone is computed for each parameter. All sensitivity vectors associated with one parameter, which correspond to different values of the parameters according to their uncertainty, are contained inside a sensitivity cone. Unlike the local approach where parameters are clustered according to their sensitivity vectors,<sup>14</sup> the presented approach clusters the parameters based upon the sensitivity cones introduced in this work. Important challenges that are dealt with in this work arise from the computation of the sensitivity cones, not only because of the uncertainty description over all parameters, but also because not all cones have the same angle as the sensitivity vectors of some parameters are significantly more affected by uncertainty than those corresponding to other parameters.

The article is structured as follows: preliminaries concerning sensitivity equations, hierarchical clustering, and dynamic optimization are presented. Then, the problem formulation and the solution approach for quantifying an upper bound on the uncertainty will be discussed. A new scheme for parameter selection is then introduced based upon hierarchical clustering of sensitivity cones. Two models will be presented in the case studies.

## Preliminaries

### Sensitivity equations

One form of dynamic system containing  $n$  states,  $m$  parameters, and  $l$  inputs can be represented as

$$\frac{dx}{dt} = f(x, p, u) \quad (1)$$

where,  $x \in R^{n \times 1}$  is the state vector,  $p \in R^{m \times 1}$  is the parameter vector, and  $u \in R^{l \times 1}$  is the input vector. The sensitivity equation is derived by taking the derivative of Eq. 1 with respect to the parameters and by applying the chain rule. The resulting dynamic sensitivity equation for the state  $x_j$  and the parameter  $p_i$  is given by

$$\frac{d}{dt} \left( \frac{\partial x_j}{\partial p_i} \right) = \sum_{k=1}^n \left( \frac{\partial f_j}{\partial x_k} \cdot \frac{\partial x_k}{\partial p_i} \right) + \frac{\partial f_j}{\partial p_i}$$

which is equivalent to

$$\frac{ds_{ij}}{dt} = \sum_{k=1}^n \left( \frac{\partial f_j}{\partial x_k} \cdot s_{ik} \right) + \frac{\partial f_j}{\partial p_i} \quad (2)$$

Here,  $s_{ij}$  can be defined as the sensitivity of the state  $x_j$  with respect to the parameter  $p_i$ . As,  $i \in \{1, \dots, m\}$  and  $j \in \{1, \dots, n\}$ , the total number of sensitivity equations is  $m \times n$ . The extended model consists of the original model and the sensitivity equations which corresponds to  $(m + 1) \times n$  ODEs. The sensitivities can be calculated by integrating all of the ODEs simultaneously using an ODE solver. However, as the original model (Eq. 1) is independent of the sensitivity equations and the sensitivity equations in Eq. 2 are independent for different  $p_i$ , only  $2n$  equations (Eqs. 1 and 2) for a particular  $p_i$  need to be integrated at a time.

One advantage of utilizing sensitivity equations rather than automatic differentiation to calculate the sensitivity is that the integration of the extended ODE model, given by Eqs. 1 and 2, has a higher guaranteed accuracy than automatic differentiation.<sup>20–22</sup> This is important as a small perturbation of a single-model parameter might not generate a sufficiently large output variation, if this calculation is performed with automatic differentiation, whereas a relatively large perturbation might cause the dynamic system to become unstable.<sup>21</sup> Additionally, due to the unique structure of the extended ODE model, the sensitivities of all the states can be calculated with respect to only one parameter at a time by integrating  $2n$  equations, which makes the scheme computationally inexpensive.

After calculating the sensitivity variables using numerical integration, the sensitivity vector can be obtained by interpolating between the values for uniformly spaced time points  $t = (t_1, t_2, \dots, t_h)^T$ , where  $h$  is the time step. The  $i$ -th column of the sensitivity matrix  $S$  represents the sensitivity vector of the output  $x_j$  with respect to the parameter  $p_i$ .

$$S = \begin{bmatrix} s_{1j}(t_1) & \cdots & s_{mj}(t_1) \\ \vdots & \ddots & \vdots \\ s_{1j}(t_h) & \cdots & s_{mj}(t_h) \end{bmatrix} \quad (3)$$

If there is number of  $r$  outputs are considered at the same time, that is,  $x_{j_1}, x_{j_2}, \dots, x_{j_r}$ , then the matrix from Eq. 3 would result in a tensor

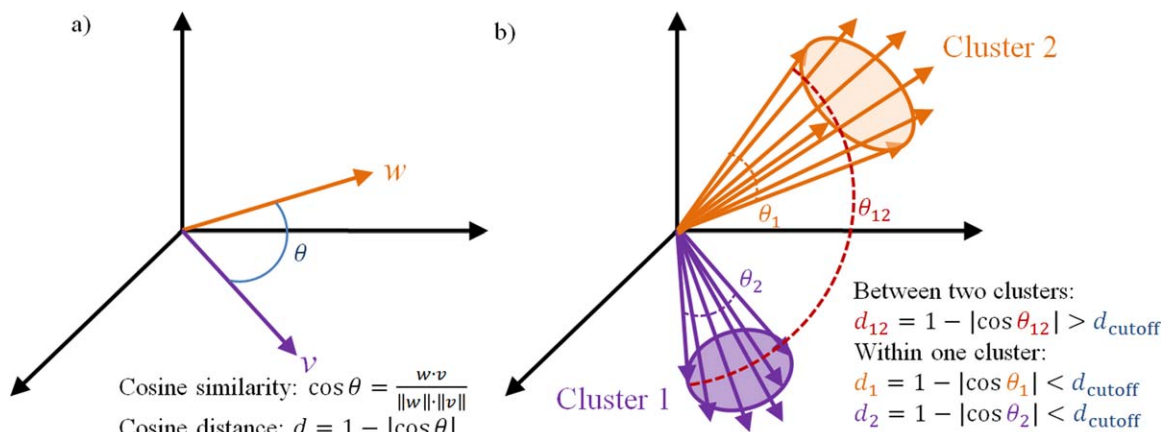
$$S = \begin{bmatrix} s_{1j_1} & \cdots & s_{mj_1} \\ s_{1j_2} & & s_{mj_2} \\ \vdots & \ddots & \vdots \\ s_{1j_r} & \cdots & s_{mj_r} \end{bmatrix} \quad (4)$$

where  $s_{ij}$  represents the sensitivity vector of the  $j$ -th output with respect to the  $i$ -th parameter. The sensitivity matrix shown in Eq. 3 or the tensor from Eq. 4 are normalized by dividing each sensitivity vector by  $(x_j^{ss}/p_i^0)$ , where  $x_j^{ss}$  is the steady-state value of its corresponding output and  $p_i^0$  is the nominal value of its corresponding parameter.

As the local sensitivity matrix is dependent on the initial value of the adjustable parameters, the performance of parameter set selection using this matrix is heavily influenced by the nominal values. To address this challenge, this work develops a formulation to incorporate the uncertainty in the parameter space for sensitivity analysis of a dynamic system with respect to the parameters. The approach for quantifying this uncertainty for the sensitivity vectors will be discussed in the next section.

### Hierarchical clustering

Hierarchical clustering is a technique to group data based upon similarities of statistical properties. Results from hierarchical clustering are commonly presented in a dendrogram. A dendrogram does not only represent a single set of clusters, but rather a multilevel hierarchy, where clusters at one level are grouped into clusters to form the next level and the different levels correspond to degrees of similarity that are assigned to objects within clusters.<sup>23</sup> Strategies for



**Figure 1. Illustration of cosine distance between (a) sensitivity vectors and (b) two clusters of sensitivity vectors.**

[Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

hierarchical clustering generally fall into two types: agglomerative and divisive. Agglomerative clustering is a bottom-up approach where each object starts in its own cluster, and pairs of clusters are merged as the threshold value moves up the hierarchy. Divisive clustering is a top-down approach where all objects start in one cluster, and splits are performed recursively as the threshold value moves down the hierarchy.

For parameter set selection, hierarchical clustering is used to reduce the number of model parameters considered by determining several groups of parameters that are pairwise indistinguishable (i.e., they cannot be uniquely estimated for a reasonable level of noise in the measurements). It is then possible to only consider one parameter per group for estimation. Hierarchical clustering via agglomeration is implemented using the following steps:

1. Calculate the distance between every pair of objects in the dataset. For a dataset containing  $m$  objects, there are  $m(m-1)/2$  pairs for which distances need to be measured.
2. Group the objects into a dendrogram based on the distance.
3. Determine a threshold distance cutoff value for the similarity which directly affects the number of clusters.

In step one, the choice of an appropriate distance metric is very important, as some objects may be close to one another according to one distance metric and farther away according to another, and this will significantly influence the clustering algorithm. For example, the Euclidean distance is commonly applied to data scattered in Euclidean space, however, for high-dimensional vectors, cosine similarity is usually used to measure the angle between the vectors.<sup>24,25</sup> As explained previously, each column of the normalized sensitivity matrix represents the sensitivity vector for a corresponding parameter. A large angle between two sensitivity vectors implies a large distance (less similarity) between these vectors (shown in Figure 1a). As the sign of the direction of the sensitivity vector has no influence on parameter set selection, the cosine distance between two sensitivity vectors is defined using the following equation

$$d = 1 - |\cos \theta| = 1 - \frac{|w \cdot v|}{\|w\| \cdot \|v\|} \quad (5)$$

Here,  $w$  and  $v$  represent two different sensitivity vectors, and  $d$  is the defined cosine distance.

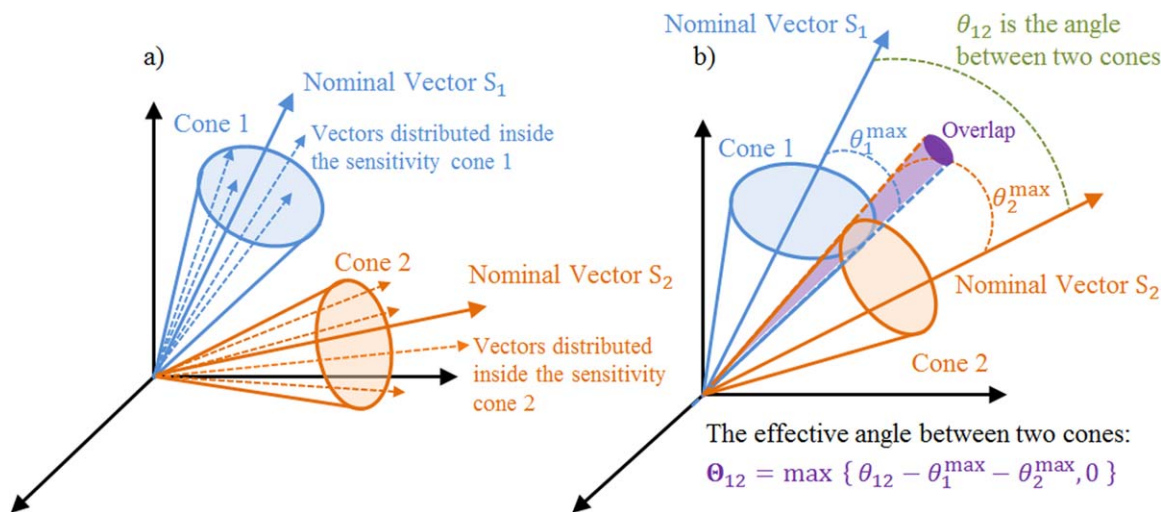
In step two, the data are clustered using a dendrogram based on the distance defined in step one. Commonly used linkage criteria between two clusters include complete linkage clustering, where the longest distance between two clusters is found, single-linkage clustering, where the shortest distance is found, average linkage clustering, and minimum energy clustering.<sup>23</sup> For parameter set selection, parameters need to be clustered into several groups such that the parameters within a group are pairwise indistinguishable, therefore, the complete linkage criterion is preferred.<sup>14</sup>

A threshold cutoff value for the similarity is determined in the last step. This threshold value affects the number of clusters that the sensitivity vectors are partitioned into (shown in Figure 1b). As a rule of thumb, the cutoff value is chosen to be small enough so that the angle between any pair of sensitivity vectors within a cluster is small. However, it should be noted that if the uncertainty in the parameter space is large, then the parameters cannot be distinguished no matter how small the cutoff value is chosen. This point will be discussed in more detail in the next section.

One of the key contributions of this work is to translate the uncertainty of the parameters into the angle of each sensitivity cone and then use the angles between the sensitivity cones for clustering. Hierarchical clustering will be performed based on the corresponding cosine distance of the angle between sensitivity cones.

### Dynamic optimization

Dynamic optimization refers to a category of optimization problems that address time-varying systems. Generally, these problems seek to maximize or minimize an objective function by determining a group of input profiles that may change over time, while the dynamic model, described by initial-value ODEs or DAEs, is considered as constraints. It is worth noting that, for parameter estimation applications, the optimization problem seeks to minimize some measure of the prediction error and the parameters are assumed to be constant over time. As DAE systems can be treated as ODE systems with equality constraints imposed on the state varia-



**Figure 2. Visualization of effect of uncertainty in parameter space on sensitivity vectors.**

[Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

bles, all of the dynamic systems mentioned below are described by ODEs. The mathematical formulation of the dynamic optimization problem is given by

$$\max_{u(t), p} y = y(x, t, p) \quad (6)$$

$$\text{s.t. } \frac{dx}{dt} = f(x, p, u(t)), x(0) = x_0 \quad (7)$$

$$\Psi(x, u(t)) \leq 0, \Omega(x(t_f)) \leq 0 \quad (8)$$

Here,  $y$  is the performance index to be optimized,  $u(t)$  is the input profile,  $p$  are the parameters. Equation 7 represents the ODE system over the time horizon  $(0, t_f)$  with states  $x$ , and known initial values  $x_0$ . Path constraints are denoted as  $\Psi$  and  $\Omega$  are the terminal constraints.

Two numerical approaches can be used to solve dynamic optimization problems. The sequential approach parameterizes the inputs using a finite number of decision variables, integrates the system states iteratively to compute the performance index, and then uses an optimization algorithm to update the values of the decision variables. Although the sequential approach is straightforward to implement, it can be very computationally expensive, especially when dealing with a large number of degrees of freedom and inequality path constraints.<sup>26</sup> The simultaneous approach parameterizes the input variable and also discretizes the dynamic system, typically using some collocation method. The discretized system is then included as algebraic constraints in a large-scale optimization problem. This approach increases the size of the optimization problem considerably, but the performance can be significantly better as the model (represented by the discretized equality constraints) is solved simultaneously with the optimization problem. The discretized problem can be especially large for stiff systems due to the fine discretization required, and powerful Nonlinear Programming (NLP) solvers must be used to deal with these problems.<sup>27</sup> All the dynamic optimization problems presented in this work are solved using the simultaneous approach because of its computational efficiency.

## Parameter Set Selection for Dynamic System Under Uncertainty

### Visualization of the effect of uncertainty in the parameter space on the sensitivity vectors

As discussed in previous section, the distance of each pair of sensitivity vectors is reflected by the angle between the vectors in a high-dimensional Euclidean space. A large angle represents a long distance (low similarity) and a small angle represents a short distance (high similarity). Based on the agglomerative clustering strategy and the complete linkage criterion, sensitivity vectors can be clustered into different groups in a dendrogram. If parameter uncertainty is considered for the dynamic system, the sensitivity vector generated from the sensitivity equation for each parameter will not be a single-fixed vector but a group of vectors distributed around the nominal vector as the sensitivity vector changes for perturbations in the parameter values corresponding to the uncertainty range (shown in Figure 2a). A sensitivity cone can be defined where all the sensitivity vectors corresponding to one parameter for different values of all parameters lie within the cone. Any vectors inside the sensitivity cone cannot be distinguished from each other due to uncertainty in the parameter values. Furthermore, vectors associated with different parameters may have sensitivity cones that overlap due to uncertainty (shown in Figure 2b). One result of this is that parameters that have sensitivity cones which overlap, or have only small angles between them, need to be grouped in the same cluster.

In order to capture the largest uncertainty for each parameter  $p_i$ ,  $\theta_i^{\max}$  is defined as the largest angle between the vectors inside the sensitivity cone and their corresponding nominal vector

$$\theta_i^{\max} = \max \cos^{-1} \left( \frac{|s_i(p) \cdot s_i(p_0)|}{\|s_i(p)\| \cdot \|s_i(p_0)\|} \right) \quad \forall i \in \{1, \dots, m\} \quad (9)$$

Here,  $s_i(p_0)$  and  $s_i(p)$  are the sensitivity vectors of the  $i$ -th parameter at the nominal values  $p_0$ , and at other values,  $p$ , chosen from the uncertainty range of all parameters, respectively. As the sign (or direction) of the sensitivity vector has no impact on the parameter set selection, only the absolute



value of  $\mathbf{s}_i(p_0) \cdot \mathbf{s}_i(p)$  is considered, and  $\theta_i^{\max}$  is an acute or right angle.

It can easily be seen that two sensitivity cones will not overlap when  $\theta_{12} > \theta_1^{\max} + \theta_2^{\max}$  (illustrated in Figure 2b), thus the corresponding parameters  $p_1$  and  $p_2$  can be distinguished as long as the cutoff value is small enough. In other words, if  $\theta_{12} \leq \theta_1^{\max} + \theta_2^{\max}$ , then the two sensitivity vectors may be colinear, and the two parameters cannot be distinguished no matter how small the cutoff value is. It should be noted that the condition  $\theta_{12} \leq \theta_1^{\max} + \theta_2^{\max}$  is conservative as it will account for the worst possible case that can happen for a given level of uncertainty in the parameter values. As the sum of  $\theta_i^{\max}$  and  $\theta_j^{\max}$  will always be used to compare with  $\theta_{ij}$ , for different pair of parameters  $p_i$  and  $p_j$ , an effective angle  $\Theta_{ij}$  is defined here using Eq. 10. The corresponding cosine distance between two sensitivity cones becomes  $1 - \cos \Theta_{ij}$ . Hierarchical clustering using complete linkage criterion is then performed based on the cosine distance of the effective angle between each pair of sensitivity cones. It is important to note that calculating the pairwise cosine distance of the effective angle between two sensitivity cones and then clustering them are two independent steps. Both steps are needed to group all sensitivity cones, where any pairwise distance is smaller than a specified cutoff value, into one cluster.

$$\Theta_{ij} = \max \left\{ \theta_{ij} - \theta_i^{\max} - \theta_j^{\max}, 0 \right\} \quad (10)$$

This approach incorporates the uncertainty of the parameter values into the parameter set selection procedure while retaining existing methods for generating a local sensitivity matrix and hierarchical clustering. The next subsection will describe the numerical implementation of this approach.

### Problem formulation

Each sensitivity cone can be described by the sensitivity vector corresponding to the nominal values of the parameters and the angle between the cone surface and the nominal sensitivity vector. For each specific parameter  $p_{i^*}$  ( $i^*$  is a fixed index here), this angle  $\theta_{i^*}^{\max}$  can be calculated by solving the following dynamic optimization problem

$$\min_p y = \cos^2(\theta_{i^*}) = \frac{[\mathbf{s}_{i^*j}(p) \cdot \mathbf{s}_{i^*j}(p_0)]^2}{\|\mathbf{s}_{i^*j}(p)\|^2 \cdot \|\mathbf{s}_{i^*j}(p_0)\|^2} \quad (11)$$

s.t.

$$\frac{dx_j}{dt} = f(x, p, u), x_j(t_0) = x_j^0, j = \{1, \dots, n\} \quad (12)$$

$$\frac{ds_{i^*j}}{dt} = \sum_{k=1}^n \left( \frac{\partial f_j}{\partial x_k} \cdot s_{i^*k} \right) + \frac{\partial f_j}{\partial p_{i^*}}, s_{i^*j}(t_0) = 0, j = \{1, \dots, n\} \quad (13)$$

$$\mathbf{s}_{i^*j}(p) = [s_{i^*j}(t_1) \cdots s_{i^*j}(t_h)]^T \quad (14)$$

Here,  $y$  is the square of the cosine of the angle between the sensitivity vector and the nominal vector. The square is used rather than the absolute value so that the objective function will be continuously differentiable (a common requirement for NLP solvers). Equation 12 is the original

dynamic system and Eq. 13 represents the sensitivity equations of the state  $x_j$  for the parameter  $p_{i^*}$ . In this problem, the input  $u$  is fixed and treated as a known parameter, whereas all the parameters  $p$  are perturbed to determine the maximum angle  $\theta_{i^*}^{\max}$ . The perturbation range of  $p$  can be specified to reasonable values according to its physical meaning or prior knowledge. It is worth noting that  $p$  can be either simultaneously or partially perturbed (e.g., fix some parameters at their nominal values) based on different conditions. For  $m$  different parameters, there will be  $m$  different maximum angles  $\theta_{i^*}^{\max}$  that describe the largest possible angle between any vector inside the sensitivity cone and its corresponding nominal vector.

### Solution method

The dynamic optimization problem formulated in Eqs. 11–14 is solved using the simultaneous approach, which discretizes the extended ODE model using a collocation method. The discretization method used in this study is a Legendre–Gauss–Radau collocation on finite elements. Using this discretization technique the differential equation

$$\frac{dx}{dt} = f(x, t) \quad (15)$$

is transformed to a set of algebraic equations given by

$$x_{i,j} = x_{i-1, \text{last}}(CP) + h \sum_{k \in CP} a_{i,jf}(x_{i,k}, t_{i,k}) \quad (16)$$

Here,  $x$  denotes the state vector consisting of entries of the states at different points in time  $t$ . In the algebraic equations,  $i$  is the finite element number,  $j$  and  $k$  are the collocation point numbers, and  $CP$  is the set of collocation points. The finite element step length is  $h$ , which is specified based on the tradeoff between the accuracy needed for the discretization, which increases with smaller step sizes, and an acceptable size of the optimization problem, which becomes computationally more expensive for smaller step sizes. To obtain a more accurate discretization, the finite element length can be reduced, or a higher order collocation strategy can be used. Although variant-step collocation can be used if the system stiffness varies dramatically across the entire time horizon, this work uses a fixed-step three-point Radau collocation. The Radau collocation coefficient matrix for this method (shown in Eq. 17) is found from computing the roots of the Lagrange interpolating polynomials.

$$a = \begin{pmatrix} 0.1968154772 & 0.3944243147 & 0.3764030627 \\ -0.0655354258 & 0.2920734116 & 0.5124858261 \\ 0.0237709743 & -0.0415487521 & 0.1111111111 \end{pmatrix} \quad (17)$$

Discretizing the dynamic optimization problem creates a nonlinear programming problem that can be solved using standard NLP solvers. Here, the optimization problem is formulated in A Mathematical Programming Language (AMPL),<sup>28</sup> which is an algebraic modeling language that provides first- and second-order derivatives through automatic differentiation. The problem is solved using the interior-point nonlinear solver IPOPT<sup>29</sup> which is a software package for

**Table 1. Algorithm for Parameter Set Selection under Uncertainty**

Step 1	Calculate the normalized sensitivity matrix for the nominal parameter values.
Step 2 (optional)	Fix parameters whose nominal sensitivity vectors have small lengths (e.g., less than 5% of the largest one) at their nominal values.
Step 3	Calculate the maximum angle $\theta_i^{\max}$ of each sensitivity cone associated with parameter $p_i$ that is not fixed at its nominal value in Step 2.
Step 4	Cluster the parameters into a dendrogram by hierarchical clustering on the basis of the pairwise cosine distance between two cones $1 - \cos \Theta_{ij}$ , where $\Theta_{ij} = \max \{ \theta_{ij} - \theta_i^{\max} - \theta_j^{\max}, 0 \}$ .
Step 5	Choose a cutoff value (usually smaller 0.05) to partition the parameters into $n$ different clusters.
Step 6	Select the parameters with the largest nominal sensitivity vectors of each of the $n$ clusters as representatives from these clusters. These $n$ parameters form the subset that needs to be estimated.

large-scale nonlinear optimization which has excellent convergence properties ( $q$ -quadratic). It is important to note that, when using interior-point methods, the computation time required for solving the linear system to calculate the Newton steps can dominate the total solution time. When using IPOPT, the choice of the linear solver can dramatically affect the solution speed. For the results in this article, MA86<sup>30</sup> was used as the linear solver in IPOPT. MA86 is designed to solve large sparse symmetric linear systems in parallel. Other solvers can also be used and their convergence may depend upon a variety of factors, including, but not limited to, the conditioning of the problem and the solver itself.

Initialization is another important consideration when addressing nonlinear optimization problems using a simultaneous approach. If a problem is initialized far from its optimum or in some infeasible region, the optimizer may require a very large number of iterations to converge, whereas a good initialization might require only a few iterations. A two-step approach is used in this work to provide a reasonable initialization for this problem: the model is first simulated using the nominal parameter values, and then the values of the states computed using these nominal parameter values are used to initialize the optimization problem.

#### **Schematic for parameter set selection under uncertainty**

Now that all the preliminaries have been introduced, and the technical details for quantifying the effect of uncertainty on the sensitivity vectors have been discussed, the steps of the parameter set selection algorithm under uncertainty are summarized in Table 1. Step 2 represents an optional preliminary screening procedure for reducing the parameter set, as parameters with small sensitivity vector lengths are unlikely to be chosen for estimation. Step 3 solves an optimization problem to compute the sensitivity cones of all parameters still under consideration for parameter set selection. Step 4 performs clustering on the basis of the cosine distance of the effective angle between two cones which is defined in Eq. 10. An appropriately small cutoff value is chosen for partitioning the clusters in Step 5. In Step 6, the parameters with the largest nominal sensitivity vectors of all groups are chosen as the representatives for these groups of parameters.

These selected parameters form the parameter subset that should be estimated.

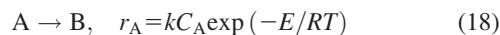
The limitation of the proposed approach is that, as the approach considers the worst situation, the dendrogram generated from clustering the sensitivity cones will be quickly pushed toward zero when dealing with large dynamic systems, even for relatively small uncertainty ranges. Modifications dealing with this limitation will be investigated in future work.

## **Case Studies**

This section presents two examples that illustrate the details of the technique developed previously. The first example is a dynamic model of a continuous stirred-tank reactor (CSTR) involving an exothermic reaction. Two states are measured and the D-optimality criterion is used to evaluate the performance of different combinations of parameter sets. The second example is a signal transduction pathway model representing a biochemical reaction network in liver cells exposed to the cytokine IL-6. A single state is measured and the performance of parameter set selection is evaluated using cross-validation involving a norm of the prediction error.

### **CSTR model**

This model describes an exothermic CSTR in which a first-order reaction,  $A \rightarrow B$ , is taking place



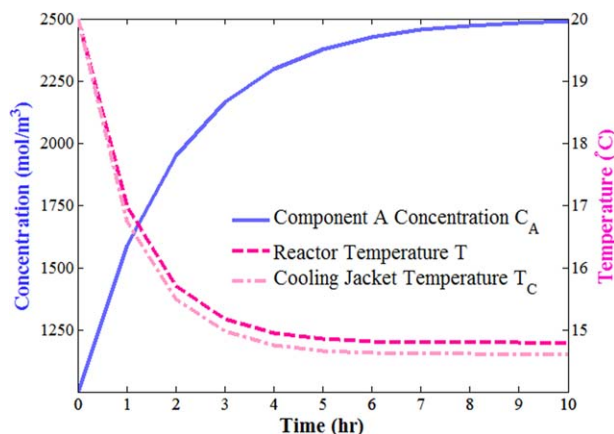
The model is described by the following differential equations

$$\begin{aligned} \dot{C}_A &= \frac{F}{V} (C_A^f - C_A) - R_A \\ \dot{T} &= \frac{F}{V} (T^f - T) + \frac{\Delta H}{\rho C_P} R_A - \frac{hA}{\rho C_P V} (T - T_c) \\ \dot{T}_c &= \frac{F_c}{V_c} (T_c^f - T_c) + \frac{hA}{\rho_c C_{Pc} V_c} (T - T_c) \end{aligned} \quad (19)$$

The three states of the system are the concentration of component A, the temperature inside the reactor, and the temperature of the cooling jacket. The temperature of the reactor and the cooling jacket are measured as the outputs. There are 16 parameters in the model and their nominal values are taken from the reference<sup>10</sup> and are listed in a table in the Appendix. Figure 3 shows a simulation of the system using nominal parameter values and the initial conditions listed in the Appendix.

The values of the parameters  $\rho$ ,  $C_p$ ,  $k$ ,  $E$ ,  $R$ ,  $\Delta H$ , and  $h$  are not known as precisely as the parameters fixed by the design of the process. Therefore, some of these parameters need to be estimated to improve the prediction capability of this dynamic model.

Considering that  $\rho$  and  $C_p$  never appear independently, but only in the form of their product,  $\rho C_p$ , and  $E$  and  $R$  only ever appear as  $(-E/R)$ , these two expressions are treated as two new parameters rather than as four parameters. With this substitution, there are a total of five parameters that can be considered for estimation. They are renamed and listed in Table 2. It is worth noting that if there are only five parameters in a system then the correlations among these



**Figure 3. Simulation of CSTR model at nominal parameter values.**

[Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

parameters are unlikely to cause significant over-parameterization. However, this simple model is useful for illustrating the utility of the presented technique for parameter set selection under uncertainty. The application of the developed technique to a larger model will be presented in the second case study.

First, an extended ODE model is formulated by combining the original model and the sensitivity equations. Equation 20 is the original CSTR model and Eq. 21 shows the sensitivity equations based on the output for each parameter  $p_i$  ( $i = 1, \dots, 5$ ). For different parameters, the differences between each group of sensitivity equations will be reflected in the

last terms shown in Eq. 21, that is,  $\partial f_1/\partial p_i$ ,  $\partial f_2/\partial p_i$ , and  $\partial f_3/\partial p_i$ , while the remaining portions of Eq. 21 remain the same for all values of  $i = 1, \dots, 5$ . According to simulations (Figure 3), the system reaches steady state after approximately 10 h, so the sensitivity vector is generated by interpolating the sensitivity variables from zero to 10 h with a step size of 0.1 h. Step 2 in Table 1 is skipped here as there are only five parameters and no further reduction of the parameter space is needed. Assuming that there is no uncertainty, the parameters are grouped in a dendrogram, as shown in Figure 4a, using the complete linkage clustering. The parameters can be grouped into different numbers of clusters, based upon the cutoff value of the cosine distance. For instance, if a cutoff of 0.05 is chosen, the parameters can be grouped into three distinguishable clusters containing  $\{p_1, p_3, p_4\}$ ,  $\{p_2\}$ , and  $\{p_5\}$ . The corresponding parameter subset that should be estimated is  $\{p_2, p_3, p_5\}$  as these are the parameters with the largest norms of the nominal sensitivity vectors in each of the three clusters. If the cutoff value is increased to 0.35, the parameters can now only be grouped into two clusters containing  $\{p_1, p_2, p_3, p_4\}$  and  $\{p_5\}$ , where  $p_2$  and  $p_5$  should be estimated.

$$\begin{aligned}\dot{x}_1 = f_1 &= 0.5(2500 - x_1) - p_1 x_1 \exp\left(\frac{p_3}{x_2}\right) \\ \dot{x}_2 = f_2 &= 0.5(20 - x_2) + \frac{p_4}{k_2} p_1 x_1 \exp\left(\frac{p_3}{x_2}\right) - \frac{22.5 p_5}{p_2} (x_2 - x_3) \\ \dot{x}_3 = f_3 &= 2.7272(10 - x_3) + 0.06818 p_5 (x_2 - x_3)\end{aligned}\quad (20)$$

For  $i = 1, \dots, 5$

$$\begin{aligned}\dot{s}_{i1} &= \left(-0.5 - p_1 \exp\left(\frac{p_3}{x_2}\right)\right) s_{i1} + \left(-p_1 x_1 \exp\left(\frac{p_3}{x_2}\right) \left(-\frac{p_3}{x_2^2}\right)\right) s_{i2} + \frac{\partial f_1}{\partial p_i} \\ \dot{s}_{i2} &= \left(\frac{p_4}{p_2} p_1 \exp\left(\frac{p_3}{x_2}\right)\right) s_{i1} + \left(-0.5 + \frac{p_4}{k_2} p_1 x_1 \exp\left(\frac{p_3}{x_2}\right) \left(-\frac{p_3}{x_2^2}\right) - \frac{22.5 p_5}{p_2}\right) s_{i2} + \frac{\partial f_2}{\partial p_i} \\ \dot{s}_{i3} &= (0.06818 p_5) s_{i2} + (-2.7272 - 0.06818 p_5) s_{i3} + \frac{\partial f_3}{\partial p_i}\end{aligned}\quad (21)$$

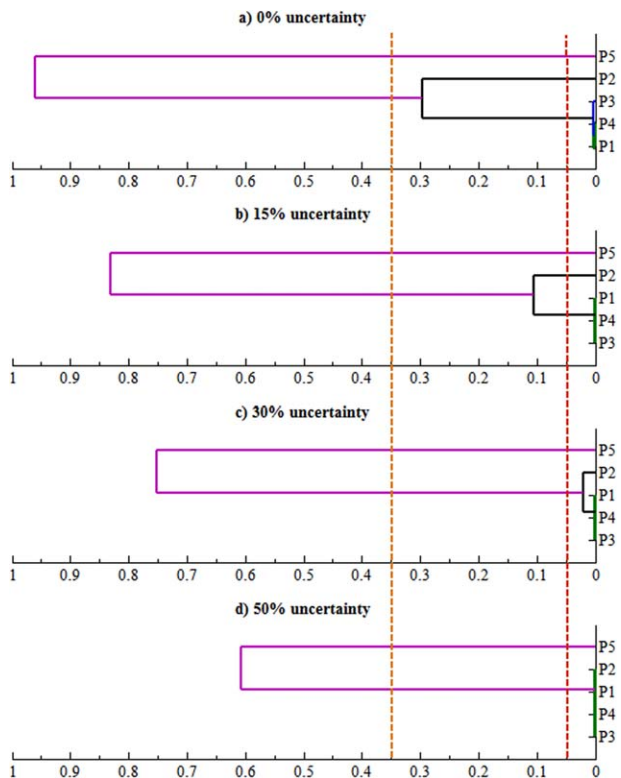
The cluster number is also influenced by the uncertainty range in the parameter space. The larger the uncertainty range, the shorter the cosine distance between each pair of sensitivity cones will be, thus fewer clusters can be partitioned using the same cutoff value. Furthermore, once two sensitivity cones overlap, the corresponding two parameters cannot be distinguished no matter how small the cutoff value is chosen. In this example, the largest angle of each sensitivity cone for different uncertainty ranges is listed in Table 3, the cosine distance between each pair of sensitivity cones is listed in Table 4, and the dendrograms for different uncertainty ranges are shown in Figure 4. It can be seen from Figure 4 that an increase of the uncertainty range results in the content of the dendrogram being pushed toward zero and the sensitivity cones start to overlap, resulting in clusters of indistinguishable parameters. For example, if 15% uncertainty is considered, the cosine distances between the sensi-

tivity cones of  $p_1$ ,  $p_3$ , and  $p_4$  are zero. Thus, no matter how small the cutoff value is, the parameters can at most be partitioned into three clusters (i.e.,  $\{p_1, p_3, p_4\}$ ,  $\{p_2\}$ , and  $\{p_5\}$ ). If the uncertainty is increased to 50%, at most two clusters can be partitioned. Therefore, both the cutoff value and the uncertainty range determine the cluster number, and the uncertainty range determines the largest number of clusters that can be partitioned.

The D-optimality criterion,<sup>31</sup> which is the most popular experimental design criterion, is used in this example to

**Table 2. Parameters that Need to be Estimated in CSTR Model**

Original Parameter	$k$	$\rho C_p$	$(-E/R)$	$\Delta H$	$h$
New parameter	$p_1$	$p_2$	$p_3$	$p_4$	$p_5$
Nominal value	2.5	1590	-255	160	1000



**Figure 4. Hierarchical clustering of sensitivity cones under different uncertainty in CSTR model.**

[Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

verify the performance of the parameter set selection. This criterion minimizes the volume of the confidence ellipsoid with an arbitrary fixed confidence level for a least-squares estimator. If the uncertainty of parameters is close to zero, then the criterion value is obtained by calculating the nominal sensitivity matrix. However, if a larger uncertainty is considered, for example, 15% or 50%, a Monte Carlo simulation is performed where 10,000 parameter sets are chosen randomly within the uncertainty range and the average criterion value is obtained from calculating the sensitivity matrix for these different parameter values. Therefore, the magnitude of the criterion value reflects the performance of the parameter selection for estimation. As there are only five parameters in this model, it is possible to compare all of the  $2^5 - 1$  different combinations of parameter subsets that are listed in Table 5. There are several trends that can clearly be seen when analyzing the data from Table 5. One trend is that parameter sets that only consist of one or two parameters are hardly affected by the uncertainty. This is in stark contrast to parameter sets with more parameters, where the criterion value reduces significantly for larger uncertainty, for example, estimating all five parameters results in a criterion value that is three orders of magnitude lower for 50% uncertainty than for no uncertainty (the values shown are on a logarithmic scale). What is particularly striking is that this general trend holds for every set that can potentially be estimated from the five parameters under consideration. A second point worth noting is that parameter sets considered good by the procedure introduced in this work are significantly less affected by increases of the uncertainty than other sets with the same number of parameters. Lastly, parameter

sets that are found to be optimal by the presented procedure, that is,  $\{p_2\}$  for estimating one parameter,  $\{p_2, p_5\}$  for estimating two parameters, and  $\{p_2, p_3, p_5\}$  for estimating three parameters, are indeed found to have by far the largest criterion values for all sets with the same number of parameters. It is important to note that there is likely no perfect parameter set for estimation of uncertain systems, especially as the choice of an appropriate cutoff value will be part of future work; however, it is important to determine one or more potential sets of parameters which are good candidates for estimation and differentiate those from other sets which would clearly result in worse prediction accuracy.

### IL-6 signaling pathway model

Modeling and analysis of intracellular signaling networks is an important area in systems biology. Signaling pathways initiate essential processes for regulating cell growth, division, apoptosis, or responses to environmental stimuli. These pathways include a large number of components, which detect, amplify, and integrate diverse external signals to generate responses, such as changes in enzyme activity or gene expression. It is infeasible to measure all the components in these pathways which limit the number of parameters that can be estimated. Therefore, the values of most of the kinetic parameters are taken from the literature and contain a significant level of uncertainty. An IL-6 signaling pathway model is used in this case study to illustrate the selection of a subset of uncertain parameters for estimation using limited experimental data. This model consists of 13 state variables and 19 parameters. The IL-6 concentration is the input and the concentration of the transcription factor STAT3 in the nucleus, STAT3N\*-STAT3N\*, is the only measured output. The mathematical model is described by Eq. 22<sup>32</sup> and the initial values of all states, the value of the input, and the descriptions and nominal values of all the parameters are listed in the Appendix.

The sensitivity matrix is generated by interpolating the sensitivity variables from 0 to 10 h using a 15-minute time step. After Step 2 (shown in Table 1),  $p_1$ ,  $p_3$ ,  $p_5$ ,  $p_6$ ,  $p_{14}$ , and  $p_{15}$  are left as the candidates for parameter clustering. The largest angle of each sensitivity cone is presented in Table 5, and dendrograms for different uncertainty ranges obtained using hierarchical clustering for these six candidate parameters are shown in Figure 5.

It can be seen from Table 6 that with an increase in the uncertainty range, the largest angles of the sensitivity cones of  $p_1$ ,  $p_5$ ,  $p_6$ ,  $p_{14}$ , and  $p_{15}$  increase monotonically. For example, the largest angles of the sensitivity cones of  $p_6$  and  $p_{14}$  are very close to zero when the uncertainty range is within 5%. This means that perturbation of the entire set of

**Table 3. Maximum Angle (°) of the Sensitivity Cones for Different Uncertainty Ranges**

Uncertainty Range (%)	$p_1$	$p_2$	$p_3$	$p_4$	$p_5$
5	7.476	1.404	7.476	7.432	3.032
10	11.96	1.621	12.23	11.96	4.211
15	16.72	1.985	17.14	16.72	5.438
20	21.52	2.292	21.94	21.50	6.686
25	26.14	2.688	26.47	26.14	7.904
30	30.54	3.032	30.67	30.54	9.069
50	48.91	7.343	47.49	48.94	13.49



**Table 4. Cosine Distance Between Each Pair of Sensitivity Cones (i.e.,  $1 - \cos \Theta_{ij}$ ) for Different Uncertainty Ranges**

Uncertainty Range (%)	$p_1-p_2$	$p_1-p_3$	$p_1-p_4$	$p_1-p_5$	$p_2-p_3$	$p_2-p_4$	$p_2-p_5$	$p_3-p_4$	$p_3-p_5$	$p_4-p_5$
0	0.2972	0.0020	0.0000	0.9336	0.2554	0.2972	0.9596	0.0020	0.9429	0.9336
5	0.1958	0.0000	0.0000	0.7527	0.1613	0.1962	0.8824	0.0000	0.7618	0.7535
10	0.1498	0.0000	0.0000	0.6583	0.1172	0.1498	0.8582	0.0000	0.6627	0.6583
15	0.1061	0.0000	0.0000	0.5621	0.0778	0.1061	0.8308	0.0000	0.5640	0.5621
20	0.0698	0.0000	0.0000	0.4699	0.0471	0.0699	0.8041	0.0000	0.4717	0.4701
25	0.0413	0.0000	0.0000	0.3863	0.0245	0.0413	0.7766	0.0000	0.3893	0.3863
30	0.0210	0.0000	0.0000	0.3127	0.0102	0.0210	0.7510	0.0000	0.3178	0.3127
50	0.0000	0.0000	0.0000	0.0849	0.0000	0.0000	0.6068	0.0000	0.0993	0.0848

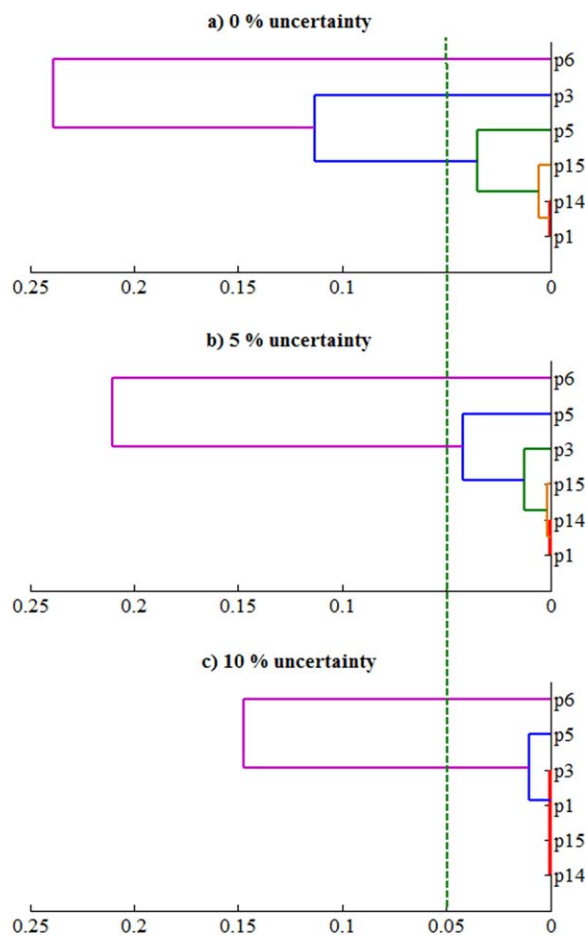
parameters would not cause  $p_6$  and  $p_{14}$  to noticeably deviate from their nominal sensitivity vectors. However, the changes of  $p_3$  due to the uncertainty are significantly larger than those of the other parameters. When the uncertainty range reaches 30%, the corresponding maximum angle for  $p_3$  is  $90^\circ$ , that is, all the parameters have to be partitioned into one cluster that is pairwise indistinguishable. When this happens, a similar dynamic optimization problem is solved while fixing  $p_3$  and perturbing all other parameters. However, the solution of this problem still returns a very large angle of the sensitivity cone for  $p_3$ . This result can be explained if  $p_3$  is strongly correlated with other parameters when uncertainty is taken into account.

It can be seen from Figure 5, that at most three clusters can be formed if the uncertainty range is 10%. If the cutoff

**Table 5. Different Combinations of Parameter Subsets and the Corresponding Criterion Values<sup>a</sup> (the Top Five Choices for Each Uncertainty Level are Highlighted)**

Parameter Subset	0%	15%	30%	50%
$\{p_1\}$	-6.025	-6.025	-6.025	-6.020
$\{p_2\}$	3.068	3.067	3.061	3.049
$\{p_3\}$	-3.726	-3.728	-3.733	-3.747
$\{p_4\}$	-6.037	-6.035	-6.032	-6.024
$\{p_5\}$	0.3060	0.3075	0.3119	0.3240
$\{p_1, p_2\}$	-3.253	-3.267	-3.309	-3.414
$\{p_1, p_3\}$	-12.09	-12.11	-12.18	-12.29
$\{p_1, p_4\}$	-15.93	-16.02	-16.61	-18.18
$\{p_1, p_5\}$	-5.721	-5.720	-5.716	-5.702
$\{p_2, p_3\}$	-1.012	-1.029	-1.085	-1.234
$\{p_2, p_4\}$	-3.270	-3.282	-3.321	-3.420
$\{p_2, p_5\}$	3.374	3.373	3.369	3.362
$\{p_3, p_4\}$	-12.22	-12.23	-12.27	-12.34
$\{p_3, p_5\}$	-3.421	-3.422	-3.423	-3.429
$\{p_4, p_5\}$	-5.732	-5.730	-5.723	-5.706
$\{p_1, p_2, p_3\}$	-10.29	-10.38	-10.56	-10.87
$\{p_1, p_2, p_4\}$	-13.32	-13.41	-14.09	-16.59
$\{p_1, p_2, p_5\}$	-2.955	-2.968	-3.009	-3.108
$\{p_1, p_3, p_4\}$	-22.39	-22.47	-23.14	-25.62
$\{p_1, p_3, p_5\}$	-11.79	-11.82	-11.88	-11.99
$\{p_1, p_4, p_5\}$	-15.64	-15.72	-16.36	-18.71
$\{p_2, p_3, p_4\}$	-10.70	-10.77	-10.87	-11.06
$\{p_2, p_3, p_5\}$	-0.7140	-0.7308	-0.7839	-0.9277
$\{p_2, p_4, p_5\}$	-2.972	-2.983	-3.020	-3.114
$\{p_3, p_4, p_5\}$	-11.92	-11.93	-11.97	-12.04
$\{p_1, p_2, p_3, p_4\}$	-21.24	-21.29	-21.98	-24.38
$\{p_1, p_2, p_3, p_5\}$	-9.995	-10.08	-10.25	-10.57
$\{p_1, p_2, p_4, p_5\}$	-13.02	-13.11	-13.79	-16.25
$\{p_1, p_3, p_4, p_5\}$	-22.10	-22.18	-22.84	-25.22
$\{p_2, p_3, p_4, p_5\}$	-10.41	-10.47	-10.58	-10.76
$\{p_1, p_2, p_3, p_4, p_5\}$	-20.95	-20.99	-21.68	-24.16

<sup>a</sup>D-optimality criterion:  $\max \log_{10}(\det(S^T S))$ ,  $S$  is the normalized sensitivity matrix of the subset.



**Figure 5. Hierarchical clustering of IL-6 signaling pathway model under uncertainty.**

[Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

**Table 6. Largest Angle ( $^\circ$ ) of the Sensitivity Cones for Different Uncertainty Ranges**

Uncertainty Range (%)	$p_1$	$p_3$	$p_5$	$p_6$	$p_{14}$	$p_{15}$
5	2.688	8.271	2.563	0.000	0.000	1.621
10	5.062	17.73	3.970	4.997	4.930	3.032
15	6.486	31.00	5.377	7.607	5.438	4.365
20	7.904	61.14	6.832	9.970	5.789	6.975
30	10.55	90.00	9.769	13.88	10.95	9.530
40	12.99	90.00	13.17	17.82	12.42	12.37
50	19.30	90.00	17.86	22.00	15.53	15.80

**Table 7. Prediction Error using 10-Fold Cross-Validation (CV) When Different Combinations of Parameter Subsets are being Estimated from Simulated Data**

CV Data Set #	$\{p_1\}$	$\{p_6\}$	$\{p_1, p_6\}$	$\{p_3, p_6\}$	$\{p_5, p_6\}$	$\{p_{14}, p_{15}\}$	$\{p_1, p_5, p_6\}$
1	14.84	22.76	21.75	11.92	16.82	158.0	20.26
2	16.55	23.58	23.91	64.73	13.83	40.78	52.85
3	22.73	30.03	20.25	52.70	18.36	174.0	19.14
4	21.06	21.04	20.76	22.79	31.10	26.04	21.96
5	18.40	16.44	9.85	26.44	21.42	15.66	29.31
6	19.58	18.52	22.93	14.43	20.66	65.67	20.56
7	21.38	14.50	17.62	137.4	278.9	21.61	76.03
8	44.31	15.64	20.77	17.07	21.44	19.57	17.98
9	19.53	39.82	19.69	70.34	112.8	307.7	22.80
10	15.45	26.03	22.40	13.14	33.97	25.33	150.2
Mean	21.38	22.83	19.99	43.09	56.93	85.43	43.11

value is chosen to be 0.05, two clusters are partitioned and the corresponding parameters considered for estimation are  $p_1$  and  $p_6$ . In this example, the performance of the parameter set selection is compared by using cross validation with measured output datasets that are randomly partitioned into 10 subsets, nine of which are used for training the parameter for each run, whereas the remaining data set is used to measure the prediction error. The uncertainty range or the bounds on the parameters for estimation is  $\pm 10\%$ . There are seven different parameter subsets listed in Table 7. These are chosen for comparison because the subset  $\{p_1, p_6\}$  is obtained from the presented approach when the uncertainty is 10%.  $\{p_1, p_5, p_6\}$  is the subset for estimation when there is no uncertainty and  $\{p_1\}$  is the subset when the uncertainty increases to an extent that all the sensitivity cones overlap and merge into one cluster.  $\{p_{14}, p_{15}\}$  is chosen as a counter example when two parameters are from the same cluster.  $\{p_3, p_6\}$  is also selected for comparison purposes because  $p_3$  has a strong correlation with other parameters as discussed previously. It can be seen from Table 7 that the average prediction errors associated with set  $\{p_1, p_6\}$ ,  $\{p_1\}$ , and  $\{p_6\}$  are significantly smaller than that of the set  $\{p_1, p_5, p_6\}$  or any of the other sets. The reason for this is that estimating fewer parameters is generally preferable if large uncertainty is considered. The prediction error associated with set  $\{p_{14}, p_{15}\}$  is the largest because the effects of the two parameters are strongly correlated and neither of them can describe the dynamics of the model when there is noise in the experimental data. The reason that the set  $\{p_3, p_6\}$  has an acceptable performance is that the effect of changes in  $p_3$  are strongly correlated with those for  $p_1$ ,  $p_{14}$ ,  $p_{15}$ , and  $p_3$  can partly compensate the function of  $p_1$ .

This example illustrates the utility of the presented approach by comparing the performance of estimating different sets of parameters using cross-validation. The results returned by the presented method are reasonable, while choosing other parameter sets can result in varying performance. It should be noted that the cross validation would not be performed in practice; it was simply used here to illustrate that the sets of parameters determined for estimation indeed perform better than other sets of parameters. Furthermore, both of the case studies illustrate that the number of parameters that can be estimated is affected by the magnitude of the uncertainty. Although a uniform distribution of the parameters over their uncertainty intervals is assumed in these examples, other uncertainty distribu-

tions can also be implemented within the presented framework.

## Conclusions

This work presented an approach for parameter set selection for dynamic systems under uncertainty. The technique extends existing methods that are based upon clustering of parameters according to their sensitivity vectors. The extension is made by realizing that the sensitivity vectors vary due to the uncertainty and that a sensitivity cone can be defined, where all sensitivity vectors of a parameter computed for different parameter values lie within the sensitivity cone. Computation of the sensitivity cones is nontrivial, and this article presents a dynamic optimization technique for doing so. The angle between two sensitivity cones, each one corresponding to the sensitivity of a different parameter, can be used as an indicator of the similarity of the effect that changes in the parameter have on the outputs under uncertainty. The cosine distance corresponding to this angle is used for clustering the sensitivity cones. The main advantage of this approach is that existing techniques for parameter set selection can be used, for example, hierarchical clustering, while at the same time the effect of uncertainty on the sensitivity vectors is incorporated into the procedure by determining sensitivity cones.

## Acknowledgments

The authors gratefully acknowledge partial financial support by the National Science Foundation (Grant CBET#0941313) and the American Chemical Society (ACS-PRF#50978-ND9).

## Literature Cited

- Brun R, Reichert P, Künsch HR. Practical identifiability analysis of large environmental simulation models. *Water Resour Res.* 2001; 37(4):1015–1030.
- Hiskens IA. Nonlinear dynamic model evaluation from disturbance measurements. *IEEE Trans Power Syst.* 2001;16(4):702–710.
- Dai W, Word DP, Hahn J. Modeling and dynamic optimization of fuel-grade ethanol fermentation using fed-batch process. *Control Eng Pract.* In Press.
- Gadkar KG, Varner J, Doyle FJ III. Model identification of signal transduction networks from data using a state regulator problem. *IET Proceedings on Systems Biology*, 2005.
- Gernaey KV, Gani R. A model-based systems approach to pharmaceutical product-process design and analysis. *Chem Eng Sci.* 2010; 65(21):5757–5769.

6. Poyton A, Varziri MS, McAuley KB, McLellan P, Ramsay JO. Parameter estimation in continuous-time dynamic models using principal differential analysis. *Comput Chem Eng.* 2006;30(4):698–708.
7. Ramsay JO, Hooker G, Campbell D, Cao J. Parameter estimation for differential equations: a generalized smoothing approach. *J R Stat Soc Ser B (Stat Methodol).* 2007;69(5):741–796.
8. Lin Y, Stadtherr MA. Deterministic global optimization for parameter estimation of dynamic systems. *Ind Eng Chem Res.* 2006;45(25):8438–8448.
9. Tjoa IB, Biegler LT. Simultaneous solution and optimization strategies for parameter estimation of differential-algebraic equation systems. *Ind Eng Chem Res.* 1991;30(2):376–385.
10. Chu Y, Hahn J. Parameter set selection for estimation of nonlinear dynamic systems. *AIChE J.* 2007;53(11):2858–2870.
11. Chu Y, Hahn J. Integrating parameter selection with experimental design under uncertainty for nonlinear dynamic systems. *AIChE J.* 2008;54(9):2310–2320.
12. Velez-Reyes M, Verghese G. Subset selection in identification, and application to speed and parameter estimation for induction machines. Proceedings of the 4th IEEE Conference on IEEE Control Applications, 1995, Albany, NY, USA, 1995.
13. Lund BF, Foss BA. Parameter ranking by orthogonalization—applied to nonlinear mechanistic models. *Automatica.* 2008;44(1):278–281.
14. Chu Y, Hahn J. Parameter set selection via clustering of parameters into pairwise indistinguishable groups of parameters. *Ind Eng Chem Res.* 2008;48(13):6000–6009.
15. Walter E, Pronzato L. Qualitative and quantitative experiment design for phenomenological models—a survey. *Automatica.* 1990;26(2):195–213.
16. Chu Y, Hahn J. Quantitative optimal experimental design using global sensitivity analysis via quasi-linearization. *Ind Eng Chem Res.* 2010;49(17):7782–7794.
17. Morris MD. Factorial sampling plans for preliminary computational experiments. *Technometrics.* 1991;33(2):161–174.
18. Hornberger GM, Spear R. Approach to the preliminary analysis of environmental systems. *J Environ Manag.* 1981;12(1):7–18.
19. Atherton R, Schainker R, Ducot E. On the statistical sensitivity analysis of models for chemical kinetics. *AIChE J.* 1975;21(3):441–448.
20. Dickinson RP, Gelinas RJ. Sensitivity analysis of ordinary differential equation systems—a direct method. *J Comput Phys.* 1976;21(2):123–143.
21. Griewank A. A mathematical view of automatic differentiation. *Acta Numer.* 2003;12(1):321–398.
22. Feehery WF, Tolsma JE, Barton PI. Efficient sensitivity analysis of large-scale differential-algebraic systems. *Appl Numer Math.* 1997;25(1):41–54.
23. Jain AK, Murty MN, Flynn PJ. Data clustering: a review. *ACM Comput Surv (CSUR).* 1999;31(3):264–323.
24. van der Laan MJ, Pollard KS. A new algorithm for hybrid hierarchical clustering with visualization and the bootstrap. *J Stat Plan Inference.* 2003;117(2):275–303.
25. Zhao Y, Karypis G. Evaluation of hierarchical clustering algorithms for document datasets. Proceedings of Eleventh International Conference on Information and Knowledge Management, McLean, VA, USA, 2002.
26. Vassiliadis V, Sargent R, Pantelides C. Solution of a class of multi-stage dynamic optimization problems. 1. Problems without path constraints. *Ind Eng Chem Res.* 1994;33(9):2111–2122.
27. Kameswaran S, Biegler LT. Simultaneous dynamic optimization strategies: recent advances and challenges. *Comput Chem Eng.* 2006;30(10):1560–1575.
28. Fourer R, Gay D, Kernighan B. AMPL: a modeling language for mathematical programming. *Duxbury Press*, 2002.
29. Wächter A, Biegler LT. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math Program.* 2006;106(1):25–57.
30. HSL2013. Collection of Fortran codes for large-scale scientific computation. 2013. Available at: <http://www.hsl.rl.ac.uk>. Last accessed on April 1, 2013.
31. John RS, Draper NR. D-optimality for regression designs: a review. *Technometrics.* 1975;17(1):15–23.
32. Huang Z, Chu Y, Hahn J. Model simplification procedure for signal transduction pathway models: an application to IL-6 signaling. *Chem Eng Sci.* 2010;65(6):1964–1975.

## APPENDIX

**Table A1. Nominal Parameter Values and Initial Values of CSTR Model**

Nominal Parameter Values		
$T^f$	Feed temperature	20°C
$C_{Af}$	Feed composition	2500 mol/m <sup>3</sup>
$\rho$	Fluid density	1025 kg/m <sup>3</sup>
$\Delta H$	Heat of reaction	160 kJ/mol
$E/R$	Activation energy	255 K
$K$	Pre-exponential factor	2.5h <sup>-1</sup>
$T_{cf}$	Coolant inlet temperature	10°C
$\rho_c$	Coolant density	1000 kg/m <sup>3</sup>
$H$	Heat transfer coefficient	1000 Wm <sup>-2</sup> /C
$F$	Feed flow rate	0.1 m <sup>3</sup> /h
$F_c$	Coolant flow rate	0.15 m <sup>3</sup> /h
$V$	Reactor volume	0.2 m <sup>3</sup>
$V_c$	Cooling jacket volume	0.055 m <sup>3</sup>
$A$	Heat transfer area	4.5 m <sup>2</sup>
$C_{pc}$	Coolant heat capacity	1.20 kJ/kg/C
$C_p$	Fluid heat capacity	1.55 kJ/kg/C
Initial Values		
$C_{A0}$	Initial composition	1000 mol/m <sup>3</sup>
$T_0$	Initial reactor temperature	20°C
$T_{c0}$	Initial coolant temperature	20°C

**Table A2. IL-6 Signaling Pathway Model**

$$\begin{aligned}
 \dot{x}_1 &= p_1 u^2 R^2 - p_2 x_1 - p_3 x_1 x_2 + p_4 x_3 + p_5 x_3 \\
 &\quad - p_{11} x_1 x_5 + p_{12} x_6 - p_{14} x_1 x_7 + p_{15} x_8 \\
 \dot{x}_2 &= -p_3 x_1 x_2 + p_4 x_3 + 2p_6 x_4 \\
 \dot{x}_3 &= p_3 x_1 x_2 - p_4 x_3 - p_5 x_3 \\
 \dot{x}_4 &= 0.5 p_5 x_3 - p_6 x_4 \\
 \dot{x}_5 &= p_7 x_4 / (p_9 + x_4) \cdot \text{step}(t - p_3) - p_{10} x_5 \\
 &\quad - p_{11} x_1 x_5 + p_{12} x_6 + p_{13} x_6 \\
 \dot{x}_6 &= p_{11} x_1 x_5 - p_{12} x_6 - p_{13} x_6 \\
 \dot{x}_7 &= -p_{14} x_1 x_7 + p_{15} x_8 \\
 \dot{x}_8 &= p_{14} x_1 x_7 - p_{15} x_8 \\
 \dot{x}_9 &= p_{16} x_8 x_{10} / (p_{17} + x_{10}) - p_{18} x_9 \\
 \dot{x}_{10} &= -p_{16} x_8 x_{10} / (p_{17} + x_{10}) + p_{18} x_9 \\
 \dot{x}_{11} &= p_{13} x_6 \\
 \dot{x}_{12} &= -2p_{19} x_9 x_{12}^2 \\
 \dot{x}_{13} &= p_{19} x_9 x_{12}^2
 \end{aligned} \tag{22}$$

**Table A3. Initial Values of IL-6 Signaling Pathway Model**

$x^1$	(IL6-gp80-gp130-JAK) <sub>2</sub> *	0 (nM)
$x_2$	STAT3C	1000
$x_3$	(IL6-gp80-gp130-JAK) <sub>2</sub> *-STAT3C	0
$x_4$	STAT3N*-STAT3N*	0
$x_5$	SOCS3	0
$x_6$	(IL6-gp80-gp130-JAK) <sub>2</sub> *-SOCS3	0
$x_7$	SHP2	100
$x_8$	(IL6-gp80-gp130-JAK) <sub>2</sub> *-SHP-SUM	0
$x_9$	Erk-pp	0
$x_{10}$	Erk	16468
$x_{11}$	(IL6-gp80-gp130-JAK) <sub>2</sub>	0
$x_{12}$	C/EBPβi	40.493
$x_{13}$	C/EBPβn	0
$u$	IL-6 (input)	3.83
$R$	Receptor	4

**Table A4. Nominal Parameter Values of IL-6 Signaling  
Pathway Model**

$p_1$	Forward Rate Constant for Reaction #1	$2.336e - 005$
$p_2$	Backward rate constant for Reaction #1	0.002
$p_3$	Forward rate constant for Reaction #2	0.0138
$p_4$	Backward rate constant for Reaction #2	1.502
$p_5$	Forward rate constant for Reaction #3	0.273
$p_6$	Forward rate constant for Reaction #4	$3.282e - 004$
$p_7$	Maximum rate for Reaction #5	0.023
$p_8$	Time delay for Reaction #5	1290
$p_9$	Michaelis–Menten constant for Reaction #5	50.6
$p_{10}$	Forward rate constant for Reaction #6	$2.067e - 004$
$p_{11}$	Forward rate constant for Reaction #7	16.52
$p_{12}$	Backward rate constant for Reaction #7	0.04
$p_{13}$	Forward rate constant for Reaction #8	0.0023
$p_{14}$	Forward rate constant for Reaction #9	$4.059e - 004$
$p_{15}$	Backward rate constant for Reaction #9	$5.086e - 004$
$p_{16}$	Maximum rate constant for Reaction #10	16.00
$p_{17}$	Michaelis–Menten constant for Reaction #10	$5.115e + 003$
$p_{18}$	Forward rate constant for Reaction #11	$1.198e - 005$
$p_{19}$	Forward rate constant for Reaction #12	$1.0e - 006$

*Manuscript received May 30, 2013, and revision received Oct. 3, 2013.*