# Analysis procedure for signal transduction pathways by clustering parameters according to their sensitivity profiles

Yunfei Chu, Zuyi Huang and Juergen Hahn
Artie McFerrin Department of Chemical Engineering
Texas A&M University
College Station, TX 77843-3122

## 1. Introduction

Mathematical modeling and analysis of signal transduction networks plays an important role in systems biology. Models of signal transduction networks generally consist of nonlinear differential equations (Schoeberl et al., 2002; Singh et al., 2006) with a large number of parameters whose values are not precisely known and only a portion of which can be estimated from experimental data.

Sensitivity analysis techniques are commonly used to determine the key parameters of a model. Sensitivity values are used to rank the importance of the effect of changes in parameter values on the output. However, for a dynamic system the sensitivity is a function of time and lumping the effect over time into a scalar value only provides partial information about the importance of a parameter. One reason for this is that parameters which have similar cumulative effects may still cause very different dynamic changes of the outputs. It has been recognized that distinct temporal activation profiles of the same signaling proteins result in diverse physiological responses (Marshall, 1995; Hoffmann et al., 2002; Kholodenko 2006) and, therefore, categorizing the dynamic effects of parameters is of great importance.

The aim of this work is to use the entire time-dependent sensitivity profile of parameters for ranking the importance of parameters of signal transduction networks but also to determine which parameters have similar dynamic sensitivity profiles that cannot be distinguished from one another. A technique based on parameter clustering of the sensitivity profiles is developed to rank the parameters. A similarity measure is defined to quantify correlations among the effects that changes in parameters have on the measurements. If the similarity measure has a value of unity then the effects of two parameters cannot be distinguished from one another, i.e., the effect caused by changes in one of the parameters can be compensated by changing the other one. If the similarity is zero then the two parameters have distinct effects. The parameters can be grouped by a clustering algorithm based upon their sensitivity profile. Since parameters in a group have correlated effects, the magnitude of the sensitivity vector can be used to rank the parameters in a group. The parameter with the longest sensitivity vector in each group can be selected as the representative parameter of that group. Thus, the technique identifies the important parameters of the signal transduction network as the representative parameters for each of the groups.

The advantage of this technique over conventional methods is that the parameters are ranked by their dynamic effects rather than the cumulative effects only. This allows to not only determine a set of parameters that are important for the signal transduction network but also to characterize the effect that changes in parameters have on the output. As a result it is possible to view each cluster of parameters as a set of parameters where uncertainty in the value of any of the parameters can be compensated by the values of other parameters. For illustration purposes this technique is applied to the Jak/STAT and MAPK/NF-IL-6 signal transduction network stimulated by interleukin 6.

## 2. Background

Successful parameter estimation depends, among other things, on parameter identifiability. Parameter identifiability can be determined either analytically, also called structural identifiability, or numerically, also called practical identifiability (Walter 1987; Walter and Pronzato, 1997; Ljung, 1999). Analytical identifiabilty investigates uniqueness of the solution derived from parameter estimation while the numerical identifiability focuses on the stability of the solution. Additionally, analytical identifiabilty can be either global or local. While global identifiability includes local identifiability as a special case, it is significantly more difficult to determine global identifiability (Ljung and Glad, 1994) as approaches based upon Taylor series approximations and similarity transformations (Chappell *et al*., 1990) are restricted to small systems. Local identifiability on the other hand is relatively straightforward to test as the rank of the parameter output sensitivity matrix determines local identifiability. As the techniques introduced in this work are based upon these concepts, the definitions of identifiability as presented in Rothenbe (1971) are briefly reviewed next.

**Definition 1**: A parameter point $\boldsymbol{\theta}_0$ is said to be locally identifiable if there exists an open neighborhood of $\boldsymbol{\theta}_0$ containing no other $\boldsymbol{\theta}$ which produces the identical observations $\mathbf{y}$.

**Condition 1**: Let $\boldsymbol{\theta}_0$ be a parameter point and the sensitivity matrix $\mathbf{S}(\boldsymbol{\theta}) = \partial \mathbf{y} / \partial \boldsymbol{\theta}^{\mathrm{T}}$ has constant rank in a neighborhood of $\boldsymbol{\theta}_0$. Then $\boldsymbol{\theta}_0$ is locally identifiable if and only if $\mathbf{S}(\boldsymbol{\theta}_0)$ is nonsingular.

It should be noted that the sensitivity matrix having constant rank is a necessary condition. If this condition is removed then the nonsingularity of the sensitivity matrix is just a sufficient condition for local identifiability, or, in other words, a rank-deficient sensitivity matrix does not imply that the parameters are not locally identifiable. The condition of constant rank has to be checked analytically and evaluating this condition for one nominal value of the parameters is insufficient.

Analytical identifiability guarantees the existence of a unique solution in at least a small neighborhood of the nominal point. However, analytically identifiable of parameters does not guarantee accurate estimation in practice. If sensitivity matrix is not singular but is ill-conditioned, then noise in the data will result in large variations of the estimated parameter values. To obtain an accurate estimate it is also required that the sensitivity matrix should not be ill-conditioned.

## 3 Investigation of pairwise indistinguishable parameter sets

As the number of parameters in many fundamental models far exceeds the number of parameters that can be accurately estimated from available data, it is necessary to determine a subset of parameters which can be estimated. Parameter selection can be viewed as a special case of model reduction as only the values for some parameters are determined from parameter estimation while all other parameters are assumed to remain at their constant value.

This section presents two techniques for determining subsets of parameters to be estimated. The first technique is an analytical approach which derives the condition where the output of a system with fewer parameters is identical to the one for the entire parameter set. The second method is a numerical approach which does not focus on the outputs being identical but instead investigates the error bound that results from including fewer parameters whose values can change in a model.

### 3.1 Analytical approach

Determining which parameters can be lumped together in a model is a problem that is related to parameter identifiability. Each time a parameter is not locally identifiable, it is possible to

reduce the parameter by setting it to a constant value. If the effects that two parameters have on the output are identical then each parameter may be individually identifiable, however, only one of the two parameters needs to be considered and the other can be set to a constant value. The following definitions and propositions define this situation.

**Definition 2**: A parameter set is said to be a pairwise indistinguishable set when any two parameters in the set are not locally identifiable.

From Condition 1 it can be seen that if the two parameters are not locally identifiable then their sensitivity matrix is rank deficient. This implies that sensitivity vectors of two parameters are parallel and that the two parameters can be replaced by a linear combination of them, e.g., by setting one parameter to its nominal value and using only the other parameter.

**Proposition 1.**
Assume the output $\mathbf{y} \in R^m$ is an analytical function of the parameter $\boldsymbol{\theta} \in R^n$

$$\mathbf{y} = \mathbf{f}(\boldsymbol{\theta}), \tag{1}$$

and $\bar{\boldsymbol{\theta}}$ is a nominal value of the parameter. If the sensitivity value of the output with respect to $\theta_i$ ($i$=1,…,$r$) is nonzero

$$\frac{\partial f_s}{\partial \theta_i} \neq 0, \quad i = 1, \cdots, r \quad \text{and} \quad s = 1, \cdots, m. \tag{2}$$

then for any $\boldsymbol{\theta}$ in a neighborhood of $\bar{\boldsymbol{\theta}}$, there exist a function $\psi(\boldsymbol{\theta})$ such that

$$\mathbf{f}(\theta_1, \cdots, \theta_{r-1}, \theta_r, \theta_{r+1}, \cdots, \theta_n) = \mathbf{f}(\bar{\theta}_1, \cdots, \bar{\theta}_{r-1}, \psi(\theta_1, \cdots, \theta_n), \theta_{r+1}, \cdots, \theta_n) \tag{3}$$

if and only if the sensitivity vectors of the output with respect to $\theta_i$ ($i$=1,…,$r$) are parallel to each other

$$\frac{\partial \mathbf{f}}{\partial \theta_i} + \alpha_i \frac{\partial \mathbf{f}}{\partial \theta_r} = 0, \qquad i = 1, \cdots, r-1 \tag{4}$$

where $\alpha_i$ is a function of $\boldsymbol{\theta}$.

Proposition 1 states that the effects that changes of the values of parameters in a pairwise indistinguishable set have on the outputs can be lumped together and the same changes can be expressed by any one parameter of the set. As a result of this, all other parameters can be fixed at their nominal values. Accordingly, the output function can be re-parameterized to have fewer parameters, where one possible re-parameterization is shown in Proposition 1.

**Illustrative example**
A simple nonlinear regression model is used to illustrate the presented analytical procedure.
Let

$$\mathbf{f}(\theta_1, \theta_2, \theta_3) = \begin{bmatrix} \theta_1\theta_2 + \theta_3 \\ \theta_1\theta_2\theta_3 \end{bmatrix} = \mathbf{g}(\psi, \theta_3), \tag{5}$$

where it can easily be seen that a substitution $\psi = \theta_1\theta_2$ can be made. However, this result is derived here using the procedure presented above.
The sensitivity vectors for $\theta_1$ and $\theta_2$ are computed to be

$$\frac{\partial \mathbf{f}}{\partial \theta_1} = \begin{bmatrix} \theta_2 \\ \theta_2\theta_3 \end{bmatrix} \text{ and } \frac{\partial \mathbf{f}}{\partial \theta_2} = \begin{bmatrix} \theta_1 \\ \theta_1\theta_3 \end{bmatrix}. \tag{6}$$

These two sensitivity vectors are parallel and are related by the following differential equation:

$$\frac{\partial \mathbf{f}}{\partial \theta_1} - \frac{\theta_2}{\theta_1}\frac{\partial \mathbf{f}}{\partial \theta_2} = 0\,. \tag{7}$$

Equation (7) can be used to compute the re-parameterization of $\psi$. The characteristic ordinary differential equation is given by

$$\frac{d\theta_1}{1} = \frac{d\theta_2}{-\theta_2/\theta_1}\,. \tag{8}$$

which can be solved by separation of variables and the solution is

$$\theta_1\theta_2 = C\,, \tag{9}$$

where $C$ is a constant and the first integral is $\psi = \theta_1\theta_2$ which is the new variable to re-parameterize the model.

### 3.2 Numerical approach

The procedure presented in the last section results in a set of characteristic ordinary differential equations which needs to be solved. As it is rarely the case that this expression can be solved analytically, a numerical approach is presented here. This numerical approach does not require the sensitivity vectors to be parallel, however, the angle between the sensitivity vectors should be small. In this case the parameters can be viewed as being pairwise indistinguishable with a certain numerical precision.

A similarity measure of the effect of two parameters on the output can be defined by

$$\cos\phi_{ik} = \frac{\left|\mathbf{s}_i{}^{\mathrm{T}}\mathbf{s}_k\right|}{\left\|\mathbf{s}_i\right\|_2\left\|\mathbf{s}_k\right\|_2}\,, \tag{10}$$

where $\phi_{ik} \in [0, \pi/2]$ is the angle between sensitivity vector $\mathbf{s}_i$ and $\mathbf{s}_k$. The value of the similarity measure ranges from 0 to 1 where a value of one indicates that the two vectors are parallel to one another and that the two parameters cannot be distinguished. A value of zero, on the other hand, refers to the sensitivity vectors of the parameters being orthogonal and that the parameters have distinct effects on the output. It is should be noted that the absolute value is used for the similarity measure as it is of little important which direction the sensitivity vectors have.

Based on the similarity measure, the parameters can be grouped by a clustering algorithms. Agglomerative hierarchical clustering is used here since from the hierarchical tree it is easy to determine how many groups the parameters should be clustered into and how the least similarity value is changed as the number of groups changed. However, other clustering algorithms (Duda et al., 2001; Theodoridis and Koutroumbas, 2006) could be used with little significant change in the outcome.

This method forms groups by repeatedly merging different groups of parameters. Initially, each parameter is in a group by itself. In a second step, the two groups with the largest similarity measure are merged into a new group. The similarity within a group can be controlled by the number of groups that one chooses to have.

Since the parameters in a numerical pairwise distinguishable set have similar effects on the output, a parameter in the set can be selected as the representative for the group. Parameter estimation is simplified due to this procedure as the number of parameters can be reduced to the number of groups. However, since the sensitivity vectors for the parameter in a group are not perfectly parallel, it has to be taken into account that there will be a discrepancy between the parameter-output effect of the original system and the one with a reduced number of parameters.

This discrepancy can be measured by the prediction gap between the two functions

$$d(\boldsymbol{\theta}) = \min_{\boldsymbol{\psi}} \left\|\mathbf{f}(\boldsymbol{\theta}) - \mathbf{g}(\boldsymbol{\psi})\right\|_2\,, \tag{11}$$

where **f** is the original output function and **g** is derived from **f** when only one parameter per group is considered and all other ones are fixed at their nominal values. The individual parameter of $\boldsymbol{\psi}$ are the representative parameters for each group and are a subset of $\boldsymbol{\theta}$. The prediction gap indicates how well the model with a reduced parameter set can approximate the behavior of the original model.

It is non-trivial to compute this discrepancy for general nonlinear functions. Due to this an approximation of d based upon linearization is used in this work. The truncated Taylor series approximation of original function, **f**, with respect to the parameters is given by

$$\mathbf{f}(\boldsymbol{\theta}) \approx \mathbf{f}(\bar{\boldsymbol{\theta}}) + \frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}^{\mathrm{T}}}(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}), \tag{12}$$

and the approximated Taylor expression of the function **g** is

$$\begin{aligned} \mathbf{g}(\boldsymbol{\psi}) &\approx \mathbf{g}(\bar{\boldsymbol{\psi}}) + \frac{\partial \mathbf{g}}{\partial \boldsymbol{\psi}^{\mathrm{T}}}(\boldsymbol{\psi} - \bar{\boldsymbol{\psi}}) \\ &= \mathbf{f}(\bar{\boldsymbol{\theta}}) + \sum_i \frac{\partial \mathbf{f}}{\partial \theta_{s_i}}\left(\psi_i - \bar{\theta}_{s_i}\right) \end{aligned}, \tag{13}$$

where $s_i$ is the index for the unfixed parameters. Then the discrepancy becomes

$$\mathbf{f}(\boldsymbol{\theta}) - \mathbf{g}(\boldsymbol{\psi}) = \mathbf{S}\mathbf{x} - \mathbf{T}\mathbf{y}, \tag{14}$$

where $\mathbf{S} = \left.\frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right|_{\bar{\boldsymbol{\theta}}}$, $\mathbf{T} = \left.\frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}_{\mathbf{s}}{}^{\mathrm{T}}}\right|_{\bar{\boldsymbol{\theta}}}$, $\mathbf{x} = \boldsymbol{\theta} - \bar{\boldsymbol{\theta}}$ and $\mathbf{y} = \boldsymbol{\psi} - \bar{\boldsymbol{\theta}}_{\mathbf{s}}$. The discrepancy is dependent on the value of parameters and the worst case can be considered

$$d = \max_{\|\mathbf{x}\|_2 = 1} \min_{\mathbf{y}} \|\mathbf{S}\mathbf{x} - \mathbf{T}\mathbf{y}\|_2 . \tag{15}$$

Since the discrepancy may increase unbounded with increased of length of **x,** a constraint is placed on the length of **x**.

The similarity of parameters within a group can be controlled by determining the number of groups for the parameter set. In the extreme case where each group only contains one parameter, the discrepancy is between the original function and the one with a reduced parameter set is zero as the two parameter sets are equal. However, the discrepancy will increase as fewer groups are used and the similarity within groups decreases. It will be shown in the following that the discrepancy can be bounded by a decreasing function of the least similarity value in a set.

**Proposition 2**. Let $\mathbf{s}_k$ be the $k$-th column vector of the matrix **S**. Then the discrepancy

$$d = \max_{\|x\|_2 = 1} \min_{y} \|\mathbf{S}\mathbf{x} - \mathbf{s}_k y\|_2 \tag{16}$$

can be bounded by

$$d \le \sqrt{1 - \cos^2 \phi_m} \sqrt{\sum_{i \ne k} \|\mathbf{s}_i\|_2^2}, \tag{17}$$

where $\cos\phi_m$ is the smallest similarity value in a group.

**Proposition 3**. Let $\mathbf{s}_{l_i}$ be the $l_i$-th column of the matrix **S** where $l$ is the index of groups and $i$ is the index of the sensitivity vector in a group. Then the discrepancy (Eq. 15) can be bounded by

$$d \le \sqrt{\sum_l d_l^2}, \tag{18}$$

where $d_l$ is the discrepancy of the $l$-th group.

### 3.3 Application of determining pairwise indistinguishable parameter set for parameter estimation

One important step for parameter estimation is to select the set of parameters to be estimated. It is possible to formulate the parameter set selection procedure as an optimization problem, such as

$$\begin{aligned} &\mathbf{z}^* = \arg\max_{\mathbf{z}} \ \log\det\left(\mathbf{F}(\mathbf{z})\right) \\ \text{s.t. } &\mathbf{F}(\mathbf{z}) = \mathbf{FIM}_{(i_1,\cdots i_{n_s})}^{(i_1,\cdots i_{n_s})} \text{ with } i_j \text{ that } z_{i_j} = 1,\ j = 1\cdots n_s \\ &z_1 + z_2 + \cdots + z_{n_\theta} = n_s \\ &z_i \in \{0,1\},\ i = 1\cdots n_\theta \end{aligned} \quad . \tag{19}$$

where 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 $\theta_i$ belongs to the selected subset with the size of $n_s$. The matrix **FIM** is the Fisher information matrix of all parameters. **F**(**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 non-zero decision variables (the entries of column $i_j$ and row $i_k$, $j$, $k$ = 1…$n_s$).

The optimization problem given by equation (19) is nontrivial to solve as the number of possible combinations of parameters grows exponentially with the number of parameters in the problem. Reducing the number of parameters to be considered can significantly reduce the computational burden. The parameter clustering algorithm can be used as described in the previous subsection. Since only one parameter per can be reliable estimated from data, it is unnecessary to consider all parameters for the parameter set selection problem.

**Algorithm of parameter selection base on parameter clustering**

Step 1. Calculate the sensitivity vectors of the output with respect to the parameters.

Step 2. Determine $n_s$, the number of parameters per set, by singular value decomposition of the sensitivity matrix or the methods of forward selection.

Step 3. Set parameters whose sensitivity vectors have small length (e.g., less than 5% of the largest one) to their nominal values.

Step 4. Cluster the parameters into $n_g$ ($n_g \geq n_s$) groups with the similarity measure (Eq. 10) by hierarchical clustering.

Step 5. Select the parameter which has the largest sensitivity vector in a group as the representative of the group.

Step 6. Select $n_s$ parameters from $n_g$ representatives to optimize the criterion function by solving the optimization problem given by equation (19).

Different approaches exist for calculating the parameter sensitivity in Step 1. One commonly used technique is to calculate the sensitivity value $\partial\mathbf{y}(t)/\partial\boldsymbol{\theta}^{\mathrm{T}}$ by solving the system equations

$$\begin{aligned} \frac{d\mathbf{x}}{dt} &= \mathbf{f}\left(\mathbf{x},\mathbf{u},\boldsymbol{\theta}\right) \\ \mathbf{y} &= \mathbf{h}\left(\mathbf{x},\mathbf{u},\boldsymbol{\theta}\right) \end{aligned} , \tag{20}$$

and the sensitivity equations simultaneously

$$\begin{aligned} \frac{d}{dt}\frac{\partial\mathbf{x}}{\partial\boldsymbol{\theta}^{\mathrm{T}}} &= \frac{\partial\mathbf{f}}{\partial\mathbf{x}^{\mathrm{T}}}\frac{\partial\mathbf{x}}{\partial\boldsymbol{\theta}^{\mathrm{T}}} + \frac{\partial\mathbf{f}}{\partial\boldsymbol{\theta}^{\mathrm{T}}} \\ \frac{\partial\mathbf{y}}{\partial\boldsymbol{\theta}^{\mathrm{T}}} &= \frac{\partial\mathbf{h}}{\partial\mathbf{x}^{\mathrm{T}}}\frac{\partial\mathbf{x}}{\partial\boldsymbol{\theta}^{\mathrm{T}}} + \frac{\partial\mathbf{h}}{\partial\boldsymbol{\theta}^{\mathrm{T}}} \end{aligned} , \tag{21}$$

where **x**, **u**, **y**, **θ** are states, inputs, outputs, and parameters, respectively. The sensitivity vector consists of the sensitivity values at different time points

$$\mathbf{s}_i = \left[ \partial \mathbf{y}^{\mathrm{T}}(t_1)/\partial\theta_i \quad \partial \mathbf{y}^{\mathrm{T}}(t_2)/\partial\theta_i \quad \cdots \quad \partial \mathbf{y}^{\mathrm{T}}(t_n)/\partial\theta_i \right]^{\mathrm{T}}, \tag{22}$$

where $\mathbf{s}_i$ is the sensitivity vector with respect to the parameter $\theta_i$.

The number of parameters per set from Step 2 can be determined by the rank of the sensitivity matrix. Each column of the sensitivity matrix is a sensitivity vector of a parameter. The number of columns is equal to the number of parameters. However, due to correlation between parameters, the sensitivity matrix may be ill-conditioned. The rank of the sensitivity matrix can be determined by the number of singular values greater than a certain threshold. Step 3 represents a simple methodology for reducing the parameter set as no parameter with a small length of the sensitivity vector needs to be considered. Step 4 performs clustering of the remaining parameters into groups and the parameter with the largest sensitivity vector for each group is chosen as the representative of this group in Step 5. Step 6 selects the parameters to be considered for solution of the optimization problem from equation (19) by taking one parameter per cluster as described in Step 5.

The presented technique can significantly reduce the computational burden for solving the optimization problem given by equation (19) as the computational effort for solution of this problem grows exponentially with the number of parameters to be considered.

## 4 Case study

Modeling signal transduction networks is one important component of systems biology. Signal transduction networks are biochemical reaction networks that can contain a large number of proteins and the number of parameters in a model usually exceeds the number of proteins. Since it is not possible to determine the majority of parameters from experimental data, some form of parameter set selection has to take place before parameter estimation is performed (Gadkar et al., 2005; Yue et al., 2006; Jaqaman and Danuser 2006).

To illustrate the technique presented in this work, a model of a signal transduction pathway for hepatocytes stimulated by Interleukin-6 is used (Singh et al., 2006). The model, shown in Figure 1, contains two pathways: Janus-associated kinases & signal transducers and transcription factors are activated in one pathway while the other pathway involves the activation of mitogen-activated protein kinases. This model consists of 66 nonlinear ordinary differential equations and includes 115 parameters. The state variables are the concentrations of the proteins in the pathway and the input variable is the concentration of Interleukin-6 outside of the cell that initiates signal transduction. The output variable is the concentration of $(STAT3N^*)_2$ (dimer of activated STAT3 in the nucleus) as this transcription factor can be indirectly measured using a green fluorescent protein (GFP) reporter system. A detailed description of the original version of the model and the nominal values of the parameters can be found in the literature (Singh et al., 2006; Chu et al., 2007), however, the model has been updated to describe the mechanism that SOCS3 and SHP2 compete for the same binding site on the receptor (Huang et al., 2007).

The Fisher information matrix is computed in a first step. The sensitivity value is sampled every minute during the time interval from 0 to 12 hr to form the sensitivity vector. Singular value decomposition of the sensitivity matrix determines that the $9^{th}$ through $115^{th}$ singular values are close to zero. Accordingly, a parameter set consisting of 8 parameters will be selected. In a second step, the lengths of the sensitivity vectors are analyzed. 70 of the parameters have sensitivity vectors with a length that is less than 5% of the length of the largest sensitivity vector. These 70 parameters will be set to their nominal values and not considered further. The problem to be solved turns into a problem where a combination of 8 parameters needs to be chosen from a set of 45 parameters such that the *D*-optimality criterion is

maximized. If an exhaustive search were to be performed then the number of possible parameter sets that would have to be evaluated would be $\sim 2\times 10^{8}$. For the purpose of comparison, the forward selection (the orthogonalization method), a solution of the optimization problem via genetic algorithm, and the clustering method introduced in this paper are applied and discussed in the following.

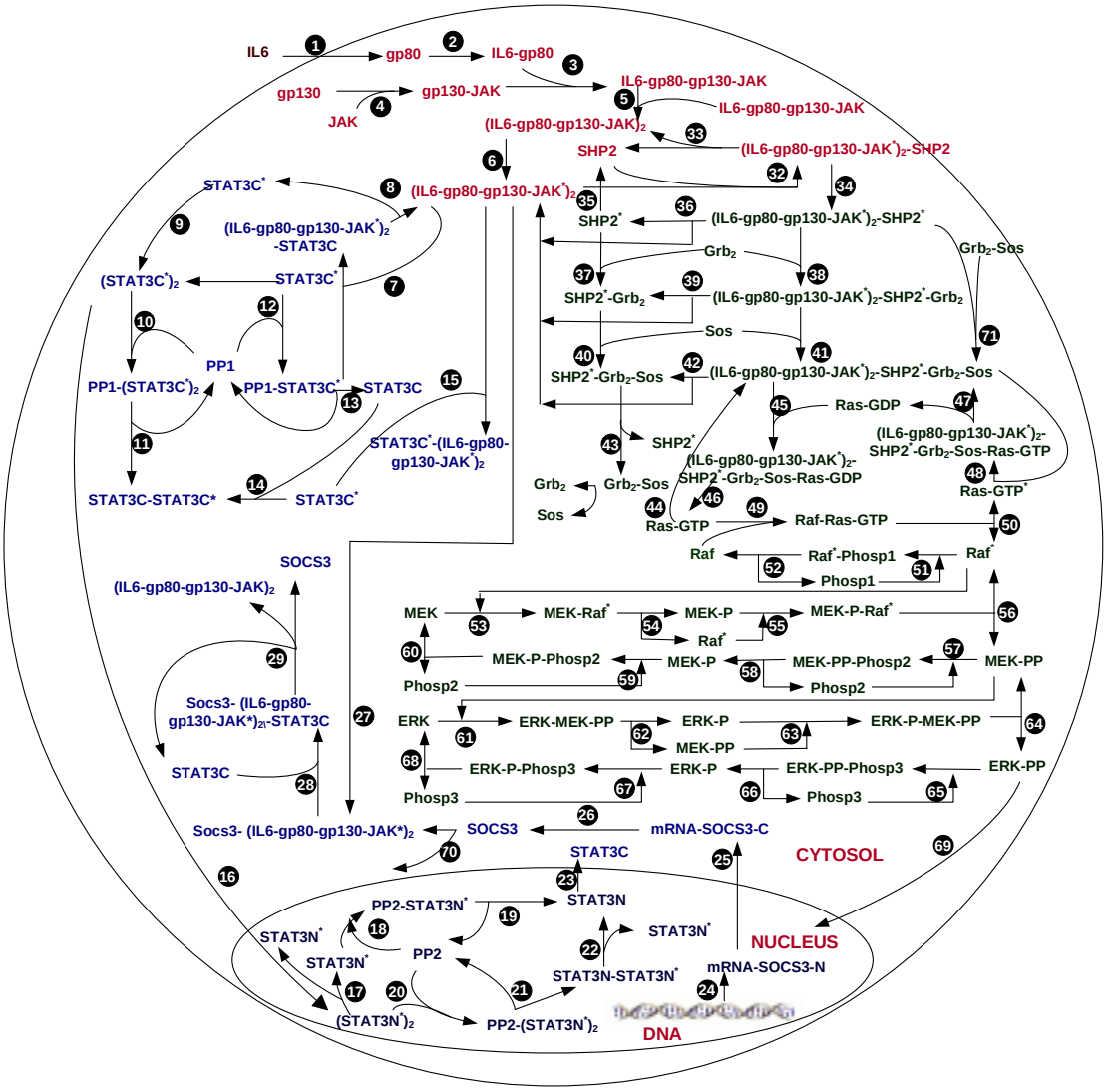


Figure 1. Model of the Interleukin-6 signaling pathway

Figure 2 shows the dendrogram of hierarchical clustering of parameters. It can be concluded that the similarity values between some of the parameters is very high as their sensitivity vectors are almost parallel. The diagram also illustrates how the selection of the similarity value influences the number of group. For example, if 11 groups are used then the smallest similarity value is equal to 0.941 which is illustrated by the dashed line. An increase in the number of groups leads to an increase of the lowest similarity value of the system.

Reducing the parameter space via parameter clustering can be viewed as one type of model reduction. The discrepancy between the original model and the reduced model is important as it indicated how many groups need to be selected to appropriately represent the original model. As discussed in Proposition 2 and Proposition 3, the model discrepancy can be bounded by the least similarity measure. At the same time, the smallest similarity measure can be determined by

choosing the number of groups from the dendrogram in Figure 2. Therefore the number of groups can be determined by assuming the discrepancy to be less than a certain threshold value. Table 1 lists the least similarity measure and the discrepancy value for different number of groups. If the parameters are clustered into more groups, then the least similarity measure is increased and the discrepancy value is decreased. In this case the number of groups is determined to be equal to 11 as the discrepancy value drops below 0.05.

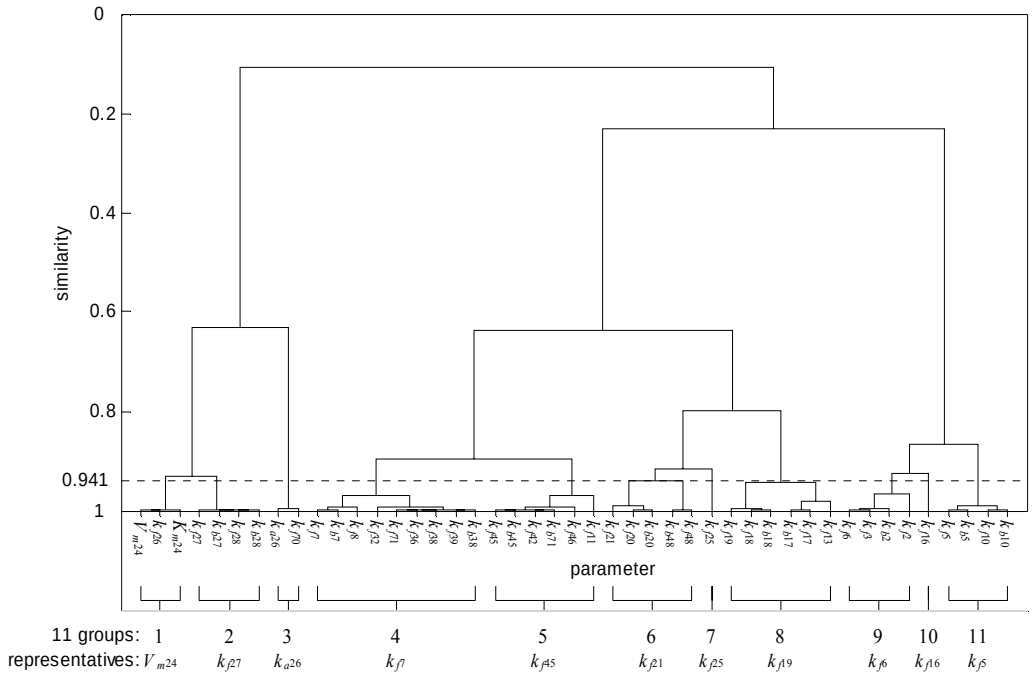


Figure 2. Dendrogram of hierarchical clustering of parameters

The parameter with the largest sensitivity vector in each group is chosen as the representative parameters for the group. The parameters selected for estimation are now chosen from the set of 11 representative parameters instead of the original 45 parameters. The optimization problem has reduced to determining a set of 8 parameters out of 11 possible parameters to maximize the *D*-optimality criterion of the Fisher information matrix, as compared to the original problem that involved choosing a set of 8 parameters out of 45 parameters. The computational effort decreases significantly, from $\sim 2\times 10^8$ possible combinations to 165, due to this reduction in the number of parameters that need to be considered.

**Discussions**

The phenomenon of highly correlated parameters is an important feature of the complex biochemical network. It has been demonstrated that the phenomenon is universe in the existing models in systems biology via analysis of the eigenvalue of the Hessian matrix (Gutenkunst et al., 2007). Similar inferences were also made by global inversion of a metabolic model (Piazza et al., 2008). The parameter clustering method developed and applied to the signal transduction network in this work supports the inference from another perspective. Some parameters are highly correlated parameters and have very high value of the similarity measure.

The effect of the correlated parameters on estimation is two folds. On one side the correlated parameters result in an ill-conditioned problem for finding the optimal parameter value. Over-fitting occurs and the estimation which is sensitivity to the noise deteriorates the predictability of the model. On the other side, correlated parameters have the overlapping effects on the output

and it is unnecessary to estimate all parameters accurately to generate an accurate prediction. Parameter selection is both a procedure to overcome the over-fitting by controlling the model complexity and an approach to simplify the parameter optimization by reducing the dimension of the decision variables.

Analysis of the highly correlated parameters is essential to not only estimation of the dynamic model but also investigation of the underlying mechanism in a biochemical network. The hierarchy tree clustered provides a clear map of the parameter relationship. Sensitivity analysis is a widely used technique to identify the important parameters. For dynamic system the parametric sensitivity is a function of time and is recorded as a sensitivity vector by sampling at specified time points. The length of the sensitivity vector which represents the accumulative effect of a parameter is often used to rank a parameter. However, it is the whole sensitivity profile that records all the information of dynamic effects of a parameter rather than the accumulative sensitivity value. The parameters which have the distinct sensitivity profiles may have the similar accumulative sensitivity value (Chu and Hahn, 2007). It has been recognized that distinct temporal activation profiles of the same signaling proteins result in diverse physiological responses (Marshall, 1995; Hoffmann et al., 2002; Kholodenko 2006) and, therefore, discriminating the dynamic effects of the parameters is of great importance.

Clustering of parameters provide a new approach to investigate the importance of a parameter taking the dynamic information into account. Since the importance of a parameter will change with time not all parameters are comparable. However, for the correlated parameters since their sensitivity profiles are similar the parameter which has large accumulative sensitivity value is more important than the parameter which has small sensitivity value. So in a group the representative parameter selected for estimation which has the largest sensitivity value is also the most important parameter in the group. Parameters from different groups have to be investigated separately since their sensitivity profiles are different.

It should be noted that some groups only contain one parameter, such as $k_{f16}$, $k_{f25}$ in the shown example. These parameters have distinct sensitivity profiles and their effects cannot be compensated for by other parameters. These parameters reveal some key parts in the signal transduction pathway. The parameter $k_{f16}$ is in the reaction where the transcription factor enters the nucleus. The parameter $k_{f25}$ controls the reaction where the mRNA leaves the nucleus and the mRNA initiates the down regulating via activating the suppressor SOCS3.

For groups containing multiple parameters it is interesting to know if the parameters in a group are close to each other or distributed around the pathway. From the results in the case study, the answer is that both situations exist. The parameters represented by $k_{f21}$ are all in the down regulating reactions associated with suppressor PP2 in the nucleus while the parameters represented by $k_{f27}$ are all in the down regulating reactions associated with suppressor SOCS3. However, the parameters in the group represented by $k_{f7}$ are distributed more widely. The group contains parameters from both JAK/STAT pathways and MAPK pathway. This provides the insight into the coupling of the two pathways and presents the clue for further study.

Investigation of correlated parameters is also very helpful to analyze the robustness with variations of parameter value. It is well known that robustness is a common property of biological networks, e.g., resilience to perturbations in current conditions (Stelling et al., 2004). The kinetic parameters in a biochemical network can change due to alteration of enzyme activity caused by a mutation or a disease. However, biological networks have a certain tolerance regarding variations in kinetic parameters while still being able to maintain their functions. One important reason for robustness is redundancy built into a network (Chen et al., 2005). As some components of a network may fail, there are other components that have a similar effect which allows the network to function properly. Parameter clustering can reveal some of these redundancies inherent to a system. The variations of some parameters by the change of the

environment can be compensated by change of other parameters in the group via self regularization of the cell.

**Conclusion**

The great number of correlated free parameters contrasted with the limited number of noisy experimental data is the main difficulty in estimation of a complex network. The model is too complex to fit the data and over-fitting will occur. A method of parameter selection via parameter clustering to improve model predictability is developed. The hierarchical clustering method is used to group parameters according to their effects on the output. The parameters in a group have similar effects on the output and the effect of a parameter can be compensated by another. The representative parameters from every group are the parameters selected for estimation. The parameter selection circumvents the over-fitting. It results in slightly larger value in the loss function, however, it produces a smaller mean squared error and also a much more accurate prediction outside the experimental conditions for estimation. Parameter clustering also sheds insights into the relationship among parameters and provides a very useful tool for analysis of the biochemical network.

**References**


1. Chappell MJ, Godfrey KR, Vajda S. Global identifiability of the Parameters of Nonlinear-systems with specified inputs-A comparision of methods. Mathematical Biosciences 102 (1): 41-73 1990.
2. Chen BS, Wang YC, Wu WS, Li WH. A new measure of the robustness of biochemical networks. Bioinformatics 21 (11): 2698-2705 2005.
3. Chu Y, Jayaraman A, Hahn J. Parameter Sensitivity Analysis of IL-6 Signaling Pathways. IET Systems Biology 1 (6): 342-352 2007.
4. Chu Y, Hahn J. Parameter Set Selection for Estimation for Nonlinear Dynamic Systems. AIChE Journal 53 (11): 2858-2870 2007.
5. Duda RO, Hart PE, Stork DG. Pattern classification (2nd edition). New York: Wiley, 2001.
6. Gadkar KG, Varner J, Doyle FJ. Model identification of signal transduction networks from data using a state regulator problem. IEE Proceedings Systems Biology 2 (1): 17-30 2005.
7. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP. Universally sloppy parameter sensitivities in systems biology models. PLOS Computational Biology 3 (10) 1871-1878 2007.
8. Hoffmann A, Levchenko A, Scott ML, Baltimore D. The IκB-NF-κB signaling module: temporal control and selective gene activation. Science 298 (5596): 1241-1245, 2002.
9. Huang Z, Chu Y, Senocak F, Jayaraman A, Hahn J. Model Update of Signal Transduction Pathways in Hepatocytes based upon Sensitivity Analysis. Proceedings Foundations of Systems Biology. 45-50 Stuttgart, Germany 2007
10. Kholodenko BN. Cell-signalling dynamics in time and space. Nature Reviews Molecular Cell Biology 7 (3): 165-176 2006.
11. Ljung L, Glad T. On global identifiability for arbitrary model parametrizations. Automatica 30 (2): 265-276 1994.
12. Ljung L. System identification: theory for the user (2nd edition). Upper Saddle River: Prentice Hall PTR, 1999.
13. Marshall CJ. Specificity of receptor tyrosine kinase signaling: Transient versus sustained extracellular signal-regulated activation. Cell 80 (2): 179-185 1995.
14. Piazza M (Piazza, Matthew), Feng XJ (Feng, Xiao-Jiang), Rabinowitz JD (Rabinowitz, Joshua D.), Rabitz H. Diverse metabolic model parameters generate similar methionine cycle dynamics. Journal of Theoretical Biology 251 (4) 628-639 2008.
15. Rothenbe TJ. Identification in parametric models. Econometrica 39 (3): 577-591 1971.

16. Singh A, Jayaraman A, Hahn J. Modeling regulatory mechanisms in IL-6 signal transduction in hepatocytes. Biotechnology and Bioengineering 95 (5): 850-862, 2006.
17. 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. Nature Biotechnology 20 (4): 370-375, 2002.
18. Stelling J, Sauer U, Szallasi Z, Doyle FJ, Doyle J. Robustness of cellular functions. Cell 118 (6): 675-685 2004.
19. Theodoridis S, Koutroumbas K. Pattern recognition (3rd edition). Boston: Elsevier/Academic Press, 2006.
20. Walter E. Identifiability of parametric models. New York: Pergamon Press, 1987.
21. Walter E, Pronzato L. Identification of parametric models from experimental data. Berlin: Springer, 1997.
22. Yue H, Brown M, Knowles J, Wang H, Broomhead DS, Kell DB. Insights into the behaviour of systems biology models from dynamic sensitivty and identifiability analysis: a case study of an NF-kappa B signalling pathway. Molecular Biosystems. 2 (12): 640-649, 2006.