

Published in final edited form as:

Comput Methods Programs Biomed. 2014 February ; 113(2): 413–432. doi:10.1016/j.cmpb.2013.10.003.

Statistical Identifiability and Convergence Evaluation for Nonlinear Pharmacokinetic Models with Particle Swarm Optimization

Seongho Kim^{1,*} and Lang Li^{2,*}

¹Biostatistics Core, Karmanos Cancer Institute, Wayne State University, Detroit, MI, 48201

²Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, 46032

Abstract

The statistical identifiability of nonlinear pharmacokinetic (PK) models with the Michaelis-Menten (MM) kinetic equation is considered using a global optimization approach, which is particle swarm optimization (PSO). If a model is statistically non-identifiable, the conventional derivative-based estimation approach is often terminated earlier without converging, due to the singularity. To circumvent this difficulty, we develop a derivative-free global optimization algorithm by combining PSO with a derivative-free local optimization algorithm to improve the rate of convergence of PSO. We further propose an efficient approach to not only checking the convergence of estimation but also detecting the identifiability of nonlinear PK models. PK simulation studies demonstrate that the convergence and identifiability of the PK model can be detected efficiently through the proposed approach. The proposed approach is then applied to clinical PK data along with a two-compartmental model.

Keywords

Michaelis-Menten kinetic equation; nonlinear models; particle swarm optimization; pharmacokinetics; statistical identifiability

1. Introduction

The nonlinear modeling is a routine but absolutely necessary statistical method in analyzing drug concentration data measured over time in pharmacokinetics (PK). In PK studies, Michaelis-Menten (MM) equation is often employed to describe the intrinsic clearance

$$CL_{int} = \frac{V_{max}}{K_m + C(t)}$$

© 2013 Elsevier Ireland Ltd. All rights reserved.

*Corresponding authors' kimse@karmanos.org and lali@iupui.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflict of Interest:
None declared

where V_{max} is the maximum enzyme activity; K_m is an inverse function of the affinity between drug and enzyme; $C(t)$ is an unbound drug concentration. K_m is also called the MM constant having the units of $C(t)$. The deterministic and statistical identifiabilities of parameters in the MM equation have been examined (Tong and Metzler, 1980; Metzler and Tong, 1981; Godfrey and Fitch, 1984). The deterministic identifiability is concerned with whether the model parameters can be identified with noise-free data, while the statistical identifiability is the possibility of identifying the model parameters with noise data.

Although numerous methods have been presented to detect the non-identifiable parameters deterministically, such as the Laplace transform (Godfrey and DiStefano, 1987), the similarity transformation approach (Vajda et al., 1989), the Voterra and generating power series approaches (Lecourtier et al., 1987), the differential algebra approach (Saccomani et al., 2003), and the alternating conditional expectation algorithm (Hengl et al., 2007), there has been much less development in statistical identifiability analysis of PK models. One of the empirical approaches to assessing the statistical identifiability is the local sensitivity analysis. The local sensitivity analysis in the statistical identification uses the first partial derivatives of the differential equations with respect to the parameters, and depends on the non-singularity of the Fisher information matrix, which is equivalent to the Taylor series method and differential algebra method (Hidalgo and Ayesa, 2001; Wynn and Parkin, 2001).

However, the local sensitivity analysis is likely to make a wrong decision if the estimate is far from the true value or the model has very complicated dynamics. Yue et al. (2008) thus proposed the global sensitivity analysis for robust experimental design based on the modified Morris method (Morris, 1991), but it still requires an initial guess or prior knowledge concerning the underlying relation of the parameters. Therefore, we propose an approach not only to accessing the identifiability globally but also to requiring no preprocessing to obtain an initial guess or prior knowledge.

A number of estimation approaches were developed for population PK analysis (Beal and Sheiner 1982; Lindstrom and Bates 1990; Vonesh and Carter 1992; Wolfinger, 1993; Kim and Li, 2011). Most approaches are a derivative-based local optimization method, however. A well-known challenge of the local optimization, such as the Newton and alike methods, is stuck at the saddle points or a local optimum so that the initial values are required to lie within a relatively small neighborhood of the true optimum to find a global optimum, and the derivative-based method is often terminated earlier due to the singularity. The singularity problem can become more prominent when the model is statistically non-identifiable. These issues urge us to use a derivative-free global optimization algorithm since it can avoid the singularity problem as well as seek the best parameter estimates of nonlinear models regardless of the presence of multiple local optima.

One interesting evolution based global optimization approach, particle swarm optimization (PSO), was developed by Kennedy and Eberhart (Eberhart and Kennedy, 1995; Kennedy and Eberhart, 1995). PSO algorithm is a derivative-free approach and becoming very popular due to its simplicity of implementation and robust convergence capability. Using PSO algorithm, Kim and Li (2011) developed a global search algorithm, P-NONMEM, for nonlinear mixed-effects models to meet the challenges of the local optimization in NONMEM, which is one of the most popular approaches in PK studies. However, NONMEM uses a Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton algorithm, which is a derivative-based approach, so that it is not free from the singularity problem. For this reason, we develop a modified version of PSO algorithm, which is the PSO coupled with a derivative-free local optimization algorithm (LPSO), in order to estimate the parameters regardless of the identifiability.

One challenge of PSO algorithm is the lack of convergence criteria. The number of function evaluations is often used as a stopping criterion along with incorporating the choice of a problem-dependent parameter, which relies on the gradient or difference between the previous and the current estimates. However, this approach doesn't take the random or stochastic behavior of PSO into account so that it will make the estimation stopped before reaching a global optimum. It also focuses only on the identifiable situations. Therefore, it is desirable to have a reliable convergence criterion for detecting when the optimization process has found the global optimum even for non-identifiable conditions. We thus propose several approaches to not only diagnosing the convergence of PSO but also detecting the statistical identifiability.

In Section 2, a brief description of a two-compartment model with Michaelis-Menten kinetic equation is given. The nonlinear PK models with PSO are introduced in Section 3. In Section 4, the proposed PSO algorithm and its convergence criteria are described in details. Simulation studies are performed to evaluate the proposed approaches and real clinical PK data then are applied in Section 5. In Section 6, conclusions are reached.

2. Michaelis-Menten Kinetic Equation and Two Compartmental Pharmacokinetics Model

Statistical Identifiability with the Michaelis-Menten kinetic equation

It is well known that the drug metabolism rate follows the Michaelis-Menten (MM) kinetics equation:

$$V(t) = \frac{dC(t)}{dt} = \frac{V_{max} \cdot C(t)}{K_m + C(t)},$$

where $V(t)$ is the velocity of the reaction, V_{max} is the maximum velocity, K_m is the MM constant, and $C(t)$ is the drug concentration. Monod (1949) first applied the MM equation to microbiology for the growth rate of microorganisms.

The MM equation generally describes the relationship between the rates of substrate conversion by an enzyme to the concentration of the substrate. In this relationship, $V(t)$ is the rate of conversion, V_{max} is the maximum rate of conversion, and $C(t)$ is the substrate concentration. The MM constant K_m is equivalent to the substrate concentration at which the rate of conversion is half of V_{max} . K_m approximates the affinity of enzyme for the substrate. A small K_m indicates high affinity, and a substrate with a smaller K_m will approach V_{max} more quickly. Very high $C(t)$ values are required to approach V_{max} , which is reached only when $C(t)$ is high enough to saturate the enzyme (Hein and Niemann, 1962).

In pharmacology research, the statistical identifiability often occurs with the MM equation. Suppose the observed data $y(t)$ follows a normal distribution with the MM equation at a time point t given the parameter $\theta = (V_{max}, K_m)$:

$$y(t) \sim ND \left\{ \log f(\theta, t), \sigma^2 \right\},$$

where $f(\theta, t) = V(t)$ and ND stands for a normal distribution. However, when K_m is much higher than the concentration $C(t)$ (i.e., $K_m \gg C(t)$), the function $f(\theta, t)$ is close to

$\frac{V_{max}}{K_m} \cdot C(t)$ in the equation below:

$$f(\theta, t) = \frac{Vmax \cdot C(t)}{Km + C(t)} \approx \frac{Vmax}{Km} \cdot C(t) \quad \text{if } C(t) \ll Km.$$

In addition, when Km is much smaller than the concentration $C(t)$ (i.e., $Km \ll C(t)$), $f(\theta, t)$ is close to $Vmax$ in the equation below:

$$f(\theta, t) = \frac{Vmax \cdot C(t)}{Km + C(t)} \approx Vmax \quad \text{if } C(t) \gg Km.$$

In other words, if the concentration $C(t)$ is much either less or greater than Km , one will not be able to estimate both Km and $Vmax$ separately due to identifiability.

Two Compartmental Intravenous Pharmacokinetic Models with the Michaelis-Menten kinetic equation

Compartmental PK analysis uses kinetic models to describe and predict the concentration-time curve for both oral (PO) and intravenous (IV) administration. PK compartmental models are often similar to kinetic models used in other scientific disciplines such as chemical kinetics and thermodynamics. The simplest PK compartmental model is the one-compartmental PK model with oral dose administration and first-order elimination (Chang, 2010). A two-compartmental IV model with the MM equation is considered for this study. In this case, its PK is described by the system of the ordinary differential equations (ODEs):

$$\begin{aligned} \frac{dA_1(t)}{dt} &= -CL \cdot \frac{A_1(t)}{V_1} + CL_{12} \cdot \left(\frac{A_2(t)}{V_2} - \frac{A_1(t)}{V_1} \right); \\ \frac{dA_2(t)}{dt} &= -CL_{12} \cdot \left(\frac{A_2(t)}{V_2} - \frac{A_1(t)}{V_1} \right); \\ CL &= \frac{Qh \cdot CL_{int}}{Qh + CL_{int}}; \\ CL_{int} &= \frac{Vmax}{Km + \frac{A_1(t)}{V_1}}; \\ (A_1(t), A_2(t))|_{t=0} &= (Dose, 0), \end{aligned}$$

where $(A_1(t), A_2(t))$ are amounts of drug in systemic and peripheral compartments at time t , respectively, (V_1, V_2) are volumes of distribution in systemic and peripheral compartments, respectively, CL_{12} is the inter-compartment rate constant, CL is the systemic clearance, CL_{int} is the intrinsic hepatic clearance, $Vmax$ is the maximum of velocity, Km is MM constant, and Qh is the hepatic blood flow known as 80 l/h.

Because the ODEs are nonlinear, there exists no closed-form solution and a numerical approach should be used to solve the differential equations. We use the R package *odesolve* to deal with the ODEs. Due to the nature of the clinical study, only the systemic concentrations are observable from PK study and its predicted concentration at time t is given by

$$\log f(\theta, t) = \log \frac{A_1(t)}{V_1},$$

where $\theta = (\log V_1, \log V_2, \log CL_{12}, \log Vmax, \log Km)$.

3. Pharmacokinetic Nonlinear Models and Particle Swarm Optimization

We illustrate the nonlinear PK model with the MM equation in this section. The observed drug concentration is described by a non-linear model,

$$\log y_i \sim ND \left(\log f(\theta, t_i), \sigma^2 \right), \quad i=1, \dots, N,$$

where N is the number of time points, y_i the drug concentration at time t_i , and $f(\cdot)$ a nonlinear function of population PK parameter vector θ , and its log-transformed value

$\log f(\theta, t) = \log \frac{A_1(t)}{V_1}$. The PK model is often assumed to follow a log-normal error model since the observed systemic concentration y_i is greater than zero. Then the log-likelihood function for (θ, σ^2) is

$$-\frac{N}{2} \log(2\pi) - N \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N (\log y_i - \log f(\theta, t_i))^2.$$

Since the function $f(\cdot)$ has no closed form solution, the parameters are estimated using numerical methods, such as Newton-type approximation, Laplace asymptotic approximation, and Markov chain Monte Carlo simulation. However, these approaches are derivative-based methods so that it might not be free from singularity, especially when the model is not identifiable. Furthermore, most derivative-based algorithms are local optimization approaches. For these reasons, we adapt a global derivative-free optimization algorithm, particle swarm optimization (PSO), to deal with both identifiable and non-identifiable models.

PSO was originally developed by Kennedy and Eberhart (1995) as a population-based global optimization method. Its evolutionary algorithm stochastically evolves a group of particles. PSO allows each particle to maintain a memory of its best fitting. Each particle's trace in the search space is then determined by its own memory of best fittings. Individual particle moves towards a stochastically weighted average of these positions, until they converge to the global best. It is used to solve a wide array of different optimization problems because of its attractive advantages, such as the ease of implementation and its gradient free stochastic algorithm. It has been proved to be an efficient method for many global optimization problems, and not suffering from the difficulties encountered by other evolutionary computation techniques. For instance, PSO does not suffer from some of genetic algorithm (GA)'s difficulties, such as interaction in the group enhances rather than detracts from progress toward the solution. In addition, PSO has memory, which GA does not have. Change in genetic populations results in destruction of previous knowledge of the problem. In PSO, individuals who fly past optima are tugged to return toward them, meaning that knowledge of good solutions is retained by all particles (Kennedy and Eberhart, 1995; Eberhart and Kennedy, 1995). For an overview of PSO and its variants, see Engbrecht (2007).

Let particle s be an element of the population. Its position vector is $x^s = (x_k^s)_{k=1, \dots, K}'$, and updating velocity vector is $v^s = (v_k^s)_{k=1, \dots, K}'$, where K is the total number of iterations of PSO and S is the population size ($s = 1, \dots, S$). Its best previous positions of itself (i.e. local best) and the population (i.e. global best) are represented as x_{lbest}^s and x_{gbest} , respectively. The velocity v_{k+1}^s and the position x_{k+1}^s at the $(k+1)$ th iteration of particle s are calculated according to the following equations:

$$v_{k+1}^s = w_k v_k^s + c_1 r_1 (x_{lbest}^s - x_k^s) + c_2 r_2 (x_{gbest}^s - x_k^s); \quad (4)$$

$$x_{k+1}^s = x_k^s + v_{k+1}^s; \quad (5)$$

where w_k is called inertia weight ($0 < w_k < 1$), c_1 and c_2 are the two positive constants called cognitive and social coefficient, respectively, r_1 and r_2 are the two random sequences in the range $[0,1]$, and k is the iteration number. The low values of constants c_1 and c_2 allow each particle to roam far from the target regions before being tugged back, while the high values result in abrupt movement towards target regions. For this reason, these constants are conventionally set as 2.0, as we did in our simulation studies. The inertia weight w_k is

$$w_k = w_{max} - \frac{k}{K} (w_{max} - w_{min}), \quad (6)$$

where w_{min} and w_{max} are user-defined constants in the range $[0,1]$ and $w_K = w_{min}$, $w_k = w_{max}$ at $k=0$. The inertia weight is adapted to control the impact of the previous history of velocities on the current velocity and to influence the trade-off between global (wide-ranging) and local (nearby) exploration abilities of the “flying points”. A larger inertia weight facilitates global exploration (searching new areas) while a smaller inertia weight tends to facilitate local exploration to fine-tune the current search area. Suitable selection of the inertia weight can provide a balance between global and local exploration abilities and thus require less iteration on average to find the global optimum (Shi and Eberhart, 1998). In order to use these properties of the inertia weight, we use a dynamic inertia weight by linearly decreasing it as described in (6) so that PSO can escape from premature convergence when it gets stagnated (Zhang and Cai, 2009).

4. LPSO: Particle Swarm Optimization Coupled with a Local Optimization Algorithm

We propose a modified PSO coupled with a local optimization algorithm to improve the rate of convergence and call the proposed algorithm LPSO. Kim and Li (2011) proposed P-NONMEM that is a combined approach between PSO and NONMEM for mixed-effects models. However, NONMEM, one of the most popular algorithms for PK study, is a derivative-based algorithm so that it will not avoid the singularity problem of non-identifiable models. For this reason, we incorporate a derivative-free local optimization algorithm, Nelder-Mead method, into PSO to deal with the local best. It is a derivative-free direct search method based on evaluating an interesting function at the vertices of a simplex iteratively by shrinking the simplex to find better points until some desired bound (Nelder and Mead, 1965). It is also called a simplex search algorithm developed by Nelder and Mead (1965). Note that the term *simplex* is a generalized triangle in a certain dimension. Nelder-Mead method requires no derivative information, making it suitable for problems with non-smooth functions or/and discontinuous functions. Its general algorithm is composed of the following two steps: construct the initial working simplex and repeat the transformation of the working simplex until it converges. There are four transformations to compute the new working simplex for the current one: reflect, expand, outside, contract, and shrink. Our second improvement over PSO is to establish a novel approach to diagnosing the convergence of the estimation. To do this, we propose three types of diagnostic measures: the local best-quartile method, the global best-variance method, and the local best-quartile-variance method.

The local best-quartile method uses the first and third quartiles and the correlation structure of the population. Suppose θ^k is the $S \times p$ matrix of the population (local best) of size S and the p parameters at k th iteration, i.e.,

$$\theta^k = \begin{bmatrix} \theta_{11}^k & \cdots & \theta_{1p}^k \\ \vdots & \ddots & \vdots \\ \theta_{S1}^k & \cdots & \theta_{Sp}^k \end{bmatrix} \text{ with } X,$$

where θ_{ij}^k is the local best of i th particle of j th parameter at k th iteration, $1 \leq i \leq S, 1 \leq j \leq p$, X is the set of indices of each particle from 1 to S and $|X| = S$.

We first calculate the first (lower) and third (upper) quartiles, Q_1^{kj} and Q_3^{kj} , for each parameter j at k th iteration, where $j = 1, 2, \dots, p$, and then obtain the reduced matrix $\hat{\theta}^k$ using the first and third quartiles as follows:

$$\hat{\theta}^k = \begin{bmatrix} \hat{\theta}_{11}^k & \cdots & \hat{\theta}_{1p}^k \\ \vdots & \ddots & \vdots \\ \hat{\theta}_{M1}^k & \cdots & \hat{\theta}_{Mp}^k \end{bmatrix} \text{ with } X^k,$$

where $M = |X^k|$ and $X^k \subset X$. In particular, $X^k = \bigcap_{j=1}^p X_j^k$ and $X_j^k = \{i | Q_1^{kj} \leq \theta_{ij}^k \leq Q_3^{kj}, i \in X\}, j = 1, 2, \dots, p$. Then the difference between the first and third quartiles for each parameter is calculated based on $\hat{\theta}^k$, i.e.,

$$d_j^k = |Q_1^{kj} - Q_3^{kj}|, \text{ where } j = 1, \dots, p,$$

obtaining the maximum difference of all the parameters as the following

$$d^k = \max_{j=1,2,\dots,p} d_j^k.$$

In addition, the $p \times p$ correlation matrix of θ^k , i.e.,

$$\Omega^k = \begin{bmatrix} 1 & \cdots & \omega_{1p}^k \\ \vdots & \ddots & \vdots \\ \omega_{p1} & \cdots & 1 \end{bmatrix}$$

is computed and then its maximum and minimum eigenvalues, λ_{max}^k and λ_{min}^k , are estimated to calculate the ratio of two eigenvalues, $\rho^k = \left| \frac{\lambda_{min}^k}{\lambda_{max}^k} \right|$. If at least one parameter has $d_j^k = 0$, then the eigenvalues cannot be obtained, so we will assign zero to ρ^k in this case.

The global best-variance method considers the standard deviation of each estimate of the parameters according to the different window size. Suppose ψ^k is the $k \times p$ matrix consisting of the global best for each parameter up to k th iteration,

$$\psi^k = \begin{bmatrix} \psi_1^1 & \cdots & \psi_p^1 \\ \vdots & \ddots & \vdots \\ \psi_1^k & \cdots & \psi_p^k \end{bmatrix},$$

where ψ_j^i is the global best of j th parameter at i the iteration and l^k is the vector of the loglikelihood of each global best of size such as $l^k = (l^1, l^2, \dots, l^k)$. Then the reduced matrix is obtained based on the user-defined window size, w . That is, the reduced matrix with w is

$$\psi_w^k = \begin{bmatrix} \psi_1^{k-w+1} & \cdots & \psi_p^{k-w+1} \\ \vdots & \ddots & \vdots \\ \psi_1^k & \cdots & \psi_p^k \end{bmatrix},$$

where $k - w < 0$, and the reduced loglikelihood vector is $l_w^k = (l^{k-w+1}, l^{k-w+2}, \dots, l^k)$. Then we compute the standard deviations for each parameter and loglikelihood:

$$SD(\psi_w^k) = \max_{j=1, \dots, p} SD_w(\psi_j^k);$$

$$SD(l_w^k) = \sqrt{Var(l^{k-w+1}, l^{k-w+2}, \dots, l^k)},$$

where $SD_w(\psi_j^k) = \sqrt{Var(\psi_j^{k-w+1}, \psi_j^{k-w+2}, \dots, \psi_j^k)}$.

Likewise, the local best-quartile-variance method is based on the standard deviation of the measures, d^k and ρ^k , of the local best-quartile method. That is,

$$SD(d_w^k) = \sqrt{Var(d^{k-w+1}, d^{k-w+2}, \dots, d^k)};$$

$$SD(\rho_w^k) = \sqrt{Var(\rho^{k-w+1}, \rho^{k-w+2}, \dots, \rho^k)}.$$

Until now, we introduce six different measures to diagnose the convergence of PSO at k th iteration: d^k , ρ^k , $SD(\psi_w^k)$, $SD(l_w^k)$, $SD(d_w^k)$, and $SD(\rho_w^k)$. We compare the performances of each approach in terms of the number of iterations to converge and the estimates at that iteration through the simulation in the next section.

The detailed procedure of the proposed LPSO is described as follows:

Step 1. Initialization

For the parameter θ and the measurement error σ^2 , their populations are initialized randomly by their uniform distribution:

$$x_0 = (x_0^1, x_0^2, \dots, x_0^s, \dots, x_0^S) \sim Unif(R(\theta) \times R(\sigma^2)),$$

where $s = 1, \dots, S$; S is the size of population; $x_0^s = (\theta_0^s, (\sigma^2)_0^s)$ is the s th particle for θ and σ^2 ; $R(z)$ is the range of a random variable (vector) z ; $Unif$ stands for the multivariate uniform distribution.

Step 2. A derivative-free local optimization-based estimation

For particle s at iteration k , the local optimization-based estimation, Nelder-Mead method, is rendered by an objective (loglikelihood) function. Let the current position of particle s at iteration k , $x_k^s = (\theta_k^s, (\sigma^2)_k^s)$ be an initial value for the estimation. Their estimates \hat{x}_k^s are obtained and then the current position x_k^s is updated with the estimate \hat{x}_k^s which is converged and estimated by the local optimization algorithm, i.e.,

$$x_k^s \leftarrow \hat{x}_k^s.$$

In case of failing to converge, there is no update on that particle.

Step 3. Finding local and global best positions

The loglikelihood for particle s at iteration k (l_k^s) is calculated given the update current position x_k^s ,

$$l_k^s = \log L(y|x_k^s).$$

l_k^s is then compared to the best previous local and global best goodness of fits (i.e., l_{lbest}^s and l_{gbest}), and the current local and global bests are updated as follows:

a. Updating the local best position

$$x_{lbest}^s \leftarrow x_k^s \quad \text{and} \quad l_{lbest}^s \leftarrow l_k^s \quad \text{if} \quad l_k^s > l_{lbest}^s,$$

and there is no update otherwise;

b. Updating the global best position

$$x_{gbest} \leftarrow x_k^s \quad \text{and} \quad l_{gbest} \leftarrow l_k^s \quad \text{if} \quad l_k^s > l_{gbest},$$

and there is no update otherwise.

Step 4. Convergence

If iteration k reaches the pre-specified maximum K or all the particles are converged by the convergence diagnosis, the proposed algorithm LPSO will stop. Otherwise it will go to the next step.

Step 5. Updating the current positions

If the current iteration is not satisfied with one of the convergence criteria described, the current positions $\{x_k^s\}_{s=1}^S$ will be updated to $\{x_{k+1}^s\}_{s=1}^S$ by (4) and (5). That is,

$$x_{k+1}^s = x_k^s + v_{k+1}^s,$$

where v_{k+1}^s is defined in (4).

5. Simulation and Application

Simulation Experiment

The two-compartment IV model described in Section 2 was employed for the simulation experiments. This model has five PK parameters $\theta = (\log V1, \log V2, \log CL12, \log Vmax, \log Km)$ and variance σ^2 for the measurement error. For the sake of simplicity, we fixed *Dose* and $\log V1$ as one and zero, respectively, in order for the systemic concentrations to range zero to one. The identifiable and non-identifiable cases were used for comparison analysis. As for the identifiable case, the true values for $\log Vmax$ and $\log Km$ were zero and -2.3 , respectively, and, as for the non-identifiable case, their values were zero and 15, respectively. We generated the simulated data without measurement error and its generated data are depicted in Figure 1. As a result, the parameters to estimate are $\theta = (\log Vmax, \log Km)$. In Figure 1, we compare the nonlinear ODE based PK model with the linear ODE based PK model. We call them linear and nonlinear in the sense that the intrinsic clearance is dependent to time. The linear ODE-based PK model uses the following intrinsic clearance instead:

$$CL_{int} = \frac{Vmax}{Km}.$$

That is, the intrinsic clearance is the ratio of $Vmax$ to Km so that there is no influence from the concentration. In fact, the trace plots are much different from each other when the model is identifiable (black solid and dotted lines with circle), while both trace plots are very similar to each other in case of non-identifiable models (read solid and dotted lines with triangle) can be seen in Figure 1.

The constants of PSO were taken as $(c1, c2, w_{max}, w_{min}, K) = (2, 2, 0.9, 0.3, 5000)$, and the number of particles of each parameter was ten (for PSO) or five (for LPSO). The parameter boundaries are $(-20, 20)$. The true values are $\theta^{true} = (0, -2.3)$ for the identifiable case and $(0, 15)$ for non-identifiable case. For both PSO and LPSO, the same seed number was used to generate the initial population. We also use the five different cutoff values to consider PSO and LPSO as converging to a global optimum such as 0, 10^{-5} , 10^{-3} , 10^{-1} , and 1. After 5000 iterations, we can observe that both PSO and LPSO converge to the true value for the identifiable case, but the non-identifiable case doesn't converge to the true value for both PSO and LPSO as shown in Table 1. In addition, we can see that, in case of non-identifiable model, LPSO will not be stopped if d^k is used with a cutoff value less than or equal to one.

Figure 2 (a) and (c) display the trace plots for the global best of each parameter and their loglikelihood over iteration, while the trace plots for d^k and ρ^k over iteration are depicted in Figure 2 (b) and (d), for identifiable and non-identifiable cases, respectively. We can see that LPSO reaches steady state faster than PSO does in this figure. In Figure 1(c), the global best estimates of each parameter have bigger variation than its loglikelihood for both PSO and LPSO due to the non-identifiability. This is because several estimates share the exactly same loglikelihood so that loglikelihood reaches steady state earlier for non-identifiable cases. In terms of the number of iterations to converge, ρ^k converges earlier than d^k , and LPSO converges faster than PSO, as shown in Table 1. As well, the estimates of each parameter

are closer to the true values in case of LPSO. However, in case of non-identifiable, although PSO and LPSO converge to certain estimates when ρ^k or d^k is less than a cutoff value, its estimate is far from the true value. Especially, $\log Vmax$ has much biased estimates for both PSO and LPSO in Table 1. Interestingly, LPSO converges to the estimates close to the true value already just after 1st iteration in case of the identifiable model.

Since PSO is a stochastic method and the direction of each particle is selected randomly, the loglikelihood is not increased monotonically different from a gradient based approach. In other words, although it is not converged yet, PSO will stay in the same estimate or loglikelihood value, causing the users to consider PSO as converging to a global optimum, especially, if the first derivative (or the difference) of the previous and current estimates is employed. To overcome this issue, we instead consider the standard deviation according to the five different window sizes, $w = 10, 20, 30, 40,$ and 50 , as described in the previous section. Figure 3 shows the trace plots for the standard deviation (SD) of the four diagnostic measures, ψ_w^k , l_w^k , d_w^k , and ρ_w^k . In the identifiable case, LPSO behaviors more stable than PSO in terms of the trace plots of each standard deviation, while PSO reaches steady state earlier than LPSO when the model is non-identifiable as shown in Figure 3. This is because the direction of each particle in LPSO is the same as that of a local optimum by its combined local optimization. If the model is non-identifiable, there are many estimates having the exactly same loglikelihood so that the local optimization will finish to the different estimates according to the different initial values. For this reason, the trace plot of LPSO as depicted in the right column of Figure 3(d) is not stable. However, PSO relies on the previous local and global bests so that it will fluctuate less than LPSO. Tables 2 and 3 show the number of iterations to converge and the estimates according to the different window sizes (w) and cutoff values (CV). In case of the identifiable models, if the window size is large, it seems that all the methods can stop both PSO and LPSO when these algorithms converge to the true values for all the cutoff values. However, if the window size becomes smaller, then PSO is often stopped before it converges to the true values, while LPSO always stops when it converge to the true values regardless of the diagnostic measures. This demonstrates that LPSO converges much faster than PSO. Overall, LPSO has a better performance than PSO

in terms of the number of iterations to converge and the estimates, and $SD(\rho_w^k)$ with the cutoff values less than 10^{-3} and the window size of 10 seems to detect the correct iteration to stop both PSO and LPSO as can be seen in Tables 2 and 3.

Figure 4 shows the scatter-box plots for PSO and LPSO at 1st, 500th, and 5000th iterations. As for the identifiable model, LPSO reaches closely the true value after 1st iteration as depicted in the upper row of Figure 4 (a). Both PSO and LPSO display the evidence of being non-identifiable after 500th iteration in the non-identifiable case as shown in the bottom row of Figure 4 (b). After 5000th iteration, although PSO reaches the true value, it seems that $\log Km$ may still need more iterations in the sense that there are a lot of outliers in its box plot in case of the identifiable model as can be seen in the upper row of Figure 4(c). It is noteworthy that even though the pattern of the identifiable case of LPSO in Figure 4(c) is similar to the non-identifiable, the ranges of x- and y-axis of each parameter are much narrower than those of the non-identifiable model. Upon investigating of these properties, the following convergence diagnostics is proposed for LPSO.

Convergence diagnostics

If $SD(\rho_w^k)$ is less than equal to the user-defined cutoff value (α) with the window size of w , LPSO will be considered as converged to a global optimum. Furthermore, if d^k is greater than the user-defined cutoff value (β), the model is considered as non-identifiable, where k is

the number of iterations to converge which is identified by $SD(\rho_w^k)$. The general guideline for α and β is 0.001 and one, respectively.

Midazolam pharmacokinetic data analysis

We describe the analysis of a Midazolam (MDZ) pharmacokinetic data using the proposed approaches of convergence diagnostics with LPSO. MDZ is a benzodiazepine used to cause relaxation or sleep before surgery and to block the memory of the procedure. It can be administrated in both oral and intravenous formulations. The MDZ PK study was conducted in the General Clinical Research Center (GCRC) at Indiana University. Twenty-two subjects were recruited into this study. Blood samples for MDZ assays were collected in non-heparinized evacuated blood collection tubes at 0.5, 0.75, 1, 1.5, 2, 4, 6, and 9 hours after intravenously dosing MDZ (2.98 mg ~ 4.8 mg). We investigate four of 24 subjects' clinical trial data for this study. Here the MDZ PK is assumed to follow a two-compartmental IV model with the MM equation as described in Section 2. Likewise, this model has six log-transformed parameters composed of five PK parameters and variance, resulting in $\theta = (\log V1, \log V2, \log CL12, \log Vmax, \log Km, \log \sigma^2)$. The constants of LPSO were set to $(c_1, c_2, w_{max}, w_{min}, K) = (2, 2, 0.9, 0.3, 500)$ as the simulation studies did, and the number of particles of each parameter was set to three. The parameter boundary for $\log \sigma^2$ is $(-20, 0)$ and others are set to $(-15, 15)$. The $SD(\rho_w^k)$ is used as a stopping rule with the cutoff value of $\alpha = 0.001$ and the window size of $w = 10$, meaning that LPSO will stop if $SD(\rho_{10}^k) \leq 0.001$, where k is the current iteration.

The converged global optima for MDZ data are shown in Table 4. The estimates of Km for Subject 1 and 2 are larger, while those for Subject 3 and 4 are close to zero, indicating that the parameter Km of these four MDZ data might be non-identifiable. Namely, in case of Subject 1 and 2, the estimate of Km is much larger than their concentrations, while the estimate of Km is much smaller than their concentrations. As a result, the estimates of $Vmax$ for Subject 1 and 2 become larger since only the ratio of $Vmax$ to Km ($Vmax/Km$) is identifiable. The number of iterations to converge ranges from 67 to 388. Although the MDZ data set of Subject 3 has the largest number of iterations to converge, it has the worst MSE (MSE = 3.8851), while Subject 1 has the smallest MSE (MSE = 0.0480) with the largest variance ($\hat{\sigma}^2 = 0.15$). Note that MSE stands for the mean squared error between observed and predicted concentrations after log-transformed.

Figure 5 displays several trace plots of the estimation results of LPSO for four MDZ individual data from Subject 1 to 4. For each subject, the trace plots of global estimates of each parameter, the trace plots of loglikelihood and $SD(\rho_{10}^k)$, the trace plots of d^k , and the prediction plot are depicted in (a)-(d), respectively.

If we consider the trace plots of loglikelihood in Figure 5(a), all subjects reach the steady state after ~10 iterations of LPSO, while $SD(\rho_{10}^k)$ still fluctuates until it converges. On the other hand, compared to the trace plots of loglikelihood, the global optima for each individual parameter still fluctuate as shown in Figure 5(b). In Figure 5(b), we can further observe several steady-state-like periods before reaching the convergence. For example, in case of Subject 1, all traces of the parameters have the steady-state period between 50 and 100 iterations before stopped. Therefore, if the stopping rule relies on the gradient or difference between two consecutive estimates, it is possible that the estimation will stop between 50 and 100 iterations and then give us a local optimum instead of a global

optimum. This demonstrates an advantage of the proposed convergence criteria over the gradient or difference based approaches.

The trace plots of d^k for $\log Km$ are larger than the cutoff of $\beta = 1$ across all the subjects, which indicate the non-identifiability as shown in Figure 5(c). In other words, according to the proposed *Convergence Diagnostics*, we can decide that PK models of these four subjects are non-identifiable since there are one or more parameters having d^k greater than the cutoff of $\beta = 1$ until it converges. We can see an interesting fact in Figure 5(d). The MSEs of Subject 1 and 2 are smaller than those of Subject 3 and 4. As shown in Table 4, the difference between Subject 1 and 2 and Subject 3 and 4 is the size of estimates of Km. As for the first two subjects, their estimates of Km are large values, while the estimates of Km are almost zero for the last two subjects.

Overall, the clinical PK data analysis is consistent with the simulation studies and shows that the proposed LPSO and Convergence Diagnostics are able to not only diagnose the convergence of LPSO but also detect the identifiability.

6. Conclusion

A novel version of PSO is proposed with enhancing the convergence of the local best using a derivative-free local optimization algorithm, which is called LPSO. In fact, the simulation studies and MDZ PK data analysis show that LPSO converges to a global optimum much faster than PSO does. Since PSO is a derivative-free algorithm and a derivative-free local optimization is combined, the proposed LPSO becomes a derivative-free global optimization algorithm so that LPSO can be applied to the parameter estimation regardless of the identifiability. Furthermore, several convergence diagnostic measures are proposed and evaluated through both the simulation studies and clinical PK data analysis. Of these measures, using the maximum of the difference between the first and third quartiles and the standard deviation of the ratio of the minimum and maximum of eigenvalues can detect when to stop LPSO as well as indicate whether the model is identifiable or not.

Acknowledgments

Dr. Seongho Kim's research is partially supported by NSF Grant DMS-1312603 (S.K.). Dr. Lang Li's research is partially sponsored by NIH Grants R01 GM574217 and U54 CA113001 (L.L.). The Biostatistics Core is supported, in part, by NIH Center Grant P30 CA022453 to the Karmanos Cancer Institute at Wayne State University.

References

- Beal SL, Sheiner LB. Estimating population kinetics. *CRC Crit. Rev. Bioeng.* 1982; 8:195–222.
- Chang, M. Monte Carlo simulation for the pharmaceutical industry: concepts, algorithms, and case studies. CRC Press; Boca Raton, FL: 2010.
- Eberhart, R.; Kennedy, J. A New Optimizer Using Particle Swarm Theory.. Proceedings of the Sixth International Symposium on Micro Machine and Human Science; Nagoya, Japan. 1995; p. 39-43.
- Engelbrecht, AP. Computational Intelligence: An Introduction. John Wiley & Sons; 2007.
- Godfrey, KR.; DiStefano, JJ. Identifiability of model parameters in identifiability of parametric models. Pergamon Press; Oxford: 1987.
- Godfrey KR, Fitch WR. The deterministic identifiability of nonlinear pharmacokinetic models. *J. Pharmacokin. Biopharm.* 1984; 12:177–191.
- Hein GE, Niemann C. Steric course and specificity of a chymotrypsin catalyzed reactions, I. *Journal of the American Chemical Society.* 1962; 84:4487–4494.
- Hengl S, Kreutz C, Timmer J, Maiwald T. Data-based identifiability analysis of non-linear dynamical models. *Bioinformatics.* 2007; 23:2612–2618. [PubMed: 17660526]

- Hidalgo ME, Ayesa E. A numerical identifiability test for state-space models – application to optimal experimental design. *Water Science & Technology*. 2001; 43:339–346. [PubMed: 11385866]
- Kennedy, J.; Eberhart, R. Particle Swarm Optimization.. Proceedings of IEEE International Conference on Neural Networks; Perth, Australia. Piscataway, NJ: IEEE Service Center; 1995. p. 1942-1948.
- Kim S, Li L. A novel global search algorithm on non-linear mixed-effect models using particle swarm optimization. *Journal of Pharmacokinetics and Pharmacodynamics*. 2011; 38:471–495. [PubMed: 21717235]
- Lecourtier, Y.; Lamnabhi-Lagarrigue, F.; Walter, E. Volterra and generating power series approaches to identifiability testing, in identifiability of parametric models. Walter, E., editor. Pergamon Press; Oxford: 1987.
- Lindstrom M, Bates D. Nonlinear mixed effects models for repeated measures data. *Biometrics*. 1990; 46:673–687. [PubMed: 2242409]
- Metzler CM, Tong DDM. Computational problems of compartment models with Michaelis-Menten type elimination. *J. Pharm. Sci*. 1981; 70:733–737. [PubMed: 7264916]
- Monod J. The growth of bacterial cultures. *Ann Rev. Microb*. 1949; 3:371–394.
- Nelder JA, Mead R. A simplex method for function minimization. *Comput. J*. 1965; 7:308–313.
- Morris MD. Factorial sampling plans for preliminary computational experiments. *Technometrics*. 1991; 33:161–174.
- Saccomani MP, Audoly S, D'Angio L. Parameter identifiability of nonlinear systems: the role of initial conditions. *Automatica*. 2003; 39:619–632.
- Shi, YH.; Eberhart, RC. A Modified Particle Swarm Optimizer. IEEE International Conference on Evolutionary Computation; Anchorage, Alaska. May 4-9, 1998;
- Tong DDM, Metzler CM. Mathematical properties of compartment models with Michaelis-Menten type elimination. *Mathematical Biosciences*. 1980; 48:293–306.
- Vajda S, Godfrey KR, Rabitz H. Similarity transformation approach to identifiability analysis of nonlinear compartmental-models. *Mathematical Biosciences*. 1989; 93:217–248. [PubMed: 2520030]
- Vonesh EF, Carter RL. Mixed-effects nonlinear regression for unbalanced repeated measures. *Biometrics*. 1992; 44:673–683.
- Wolfinger RD. Laplace's approximation for nonlinear mixed models. *Biometrika*. 1993; 80:791–795.
- Wynn, HP.; Parkin, N. Sensitivity analysis and identifiability for differential equation models.. Proceedings of the 40th IEEE Conference on Decision and Control; Orlando, FL. 2001; p. 3116-3121.
- Yue H, Brown M, He F, Jia J, Kell DB. Sensitivity analysis and robust experimental design of a signal transduction pathway system. *International Journal of Chemical Kinetics*. 2008; 40:730–741.
- Zhang T, Cai JD. A new chaotic PSO with dynamic inertia weight for economic dispatch problem. Sustainable Power Generation and Supply, 2009. SUPERGEN '09. International Conference on. : 1–6. 6–7.

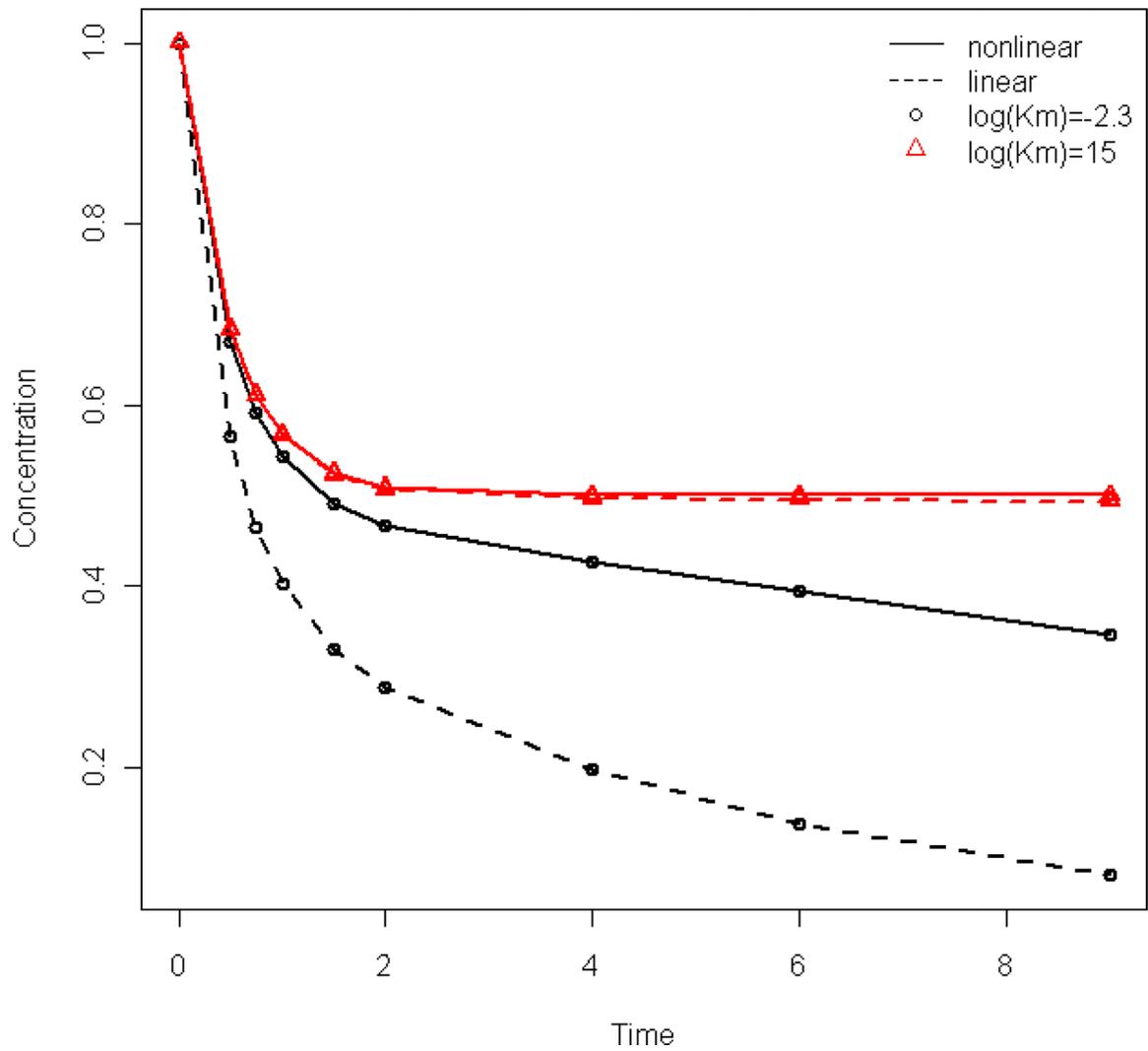
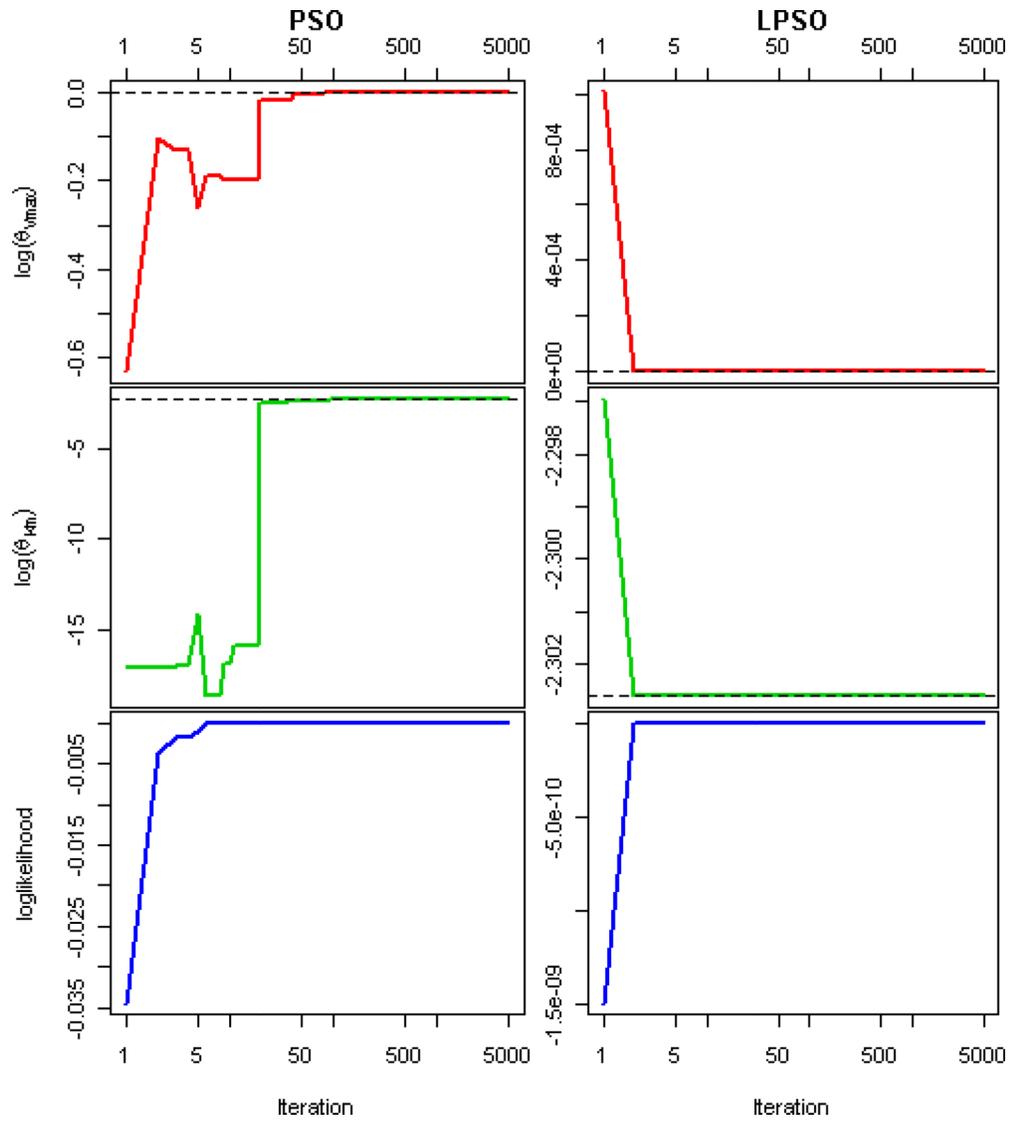
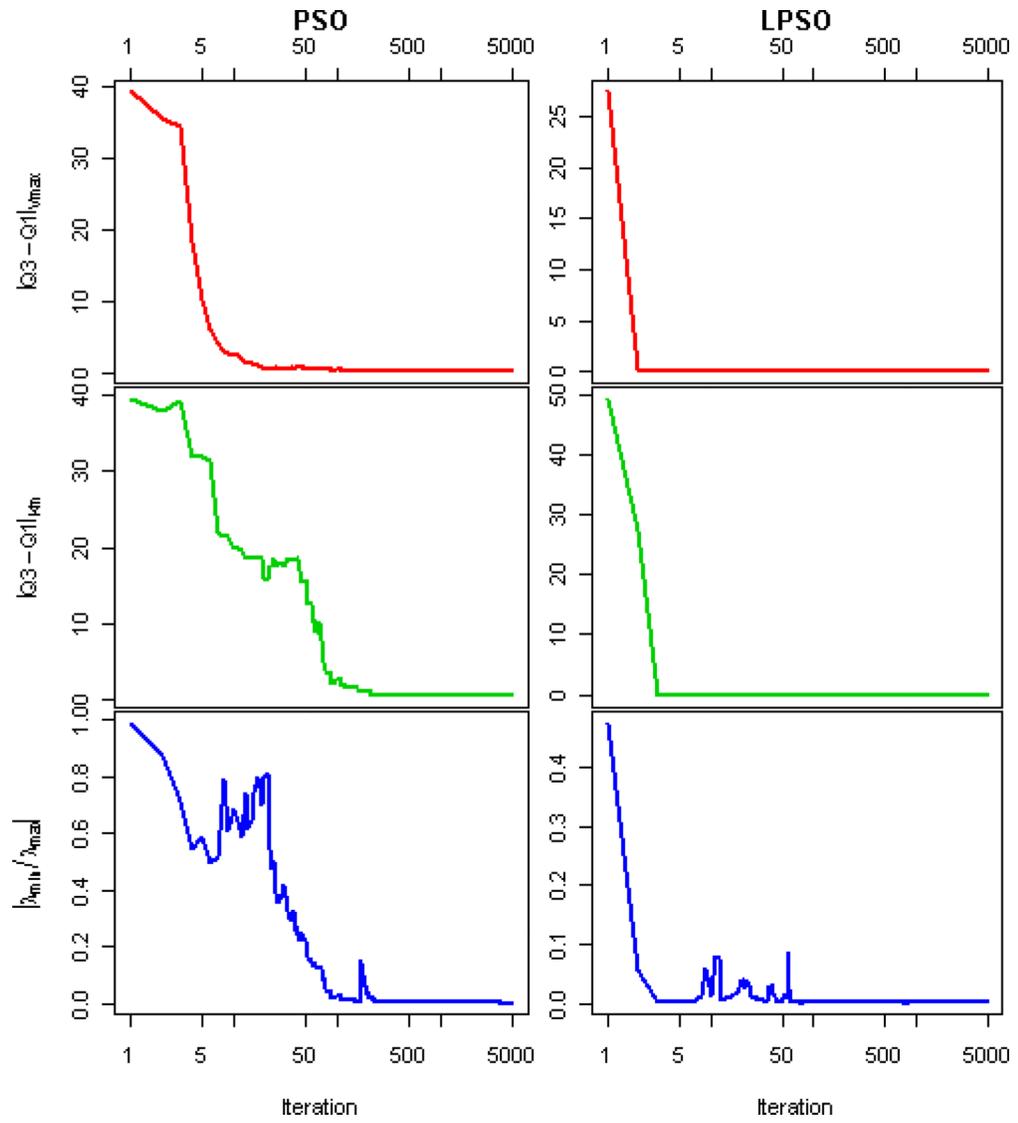
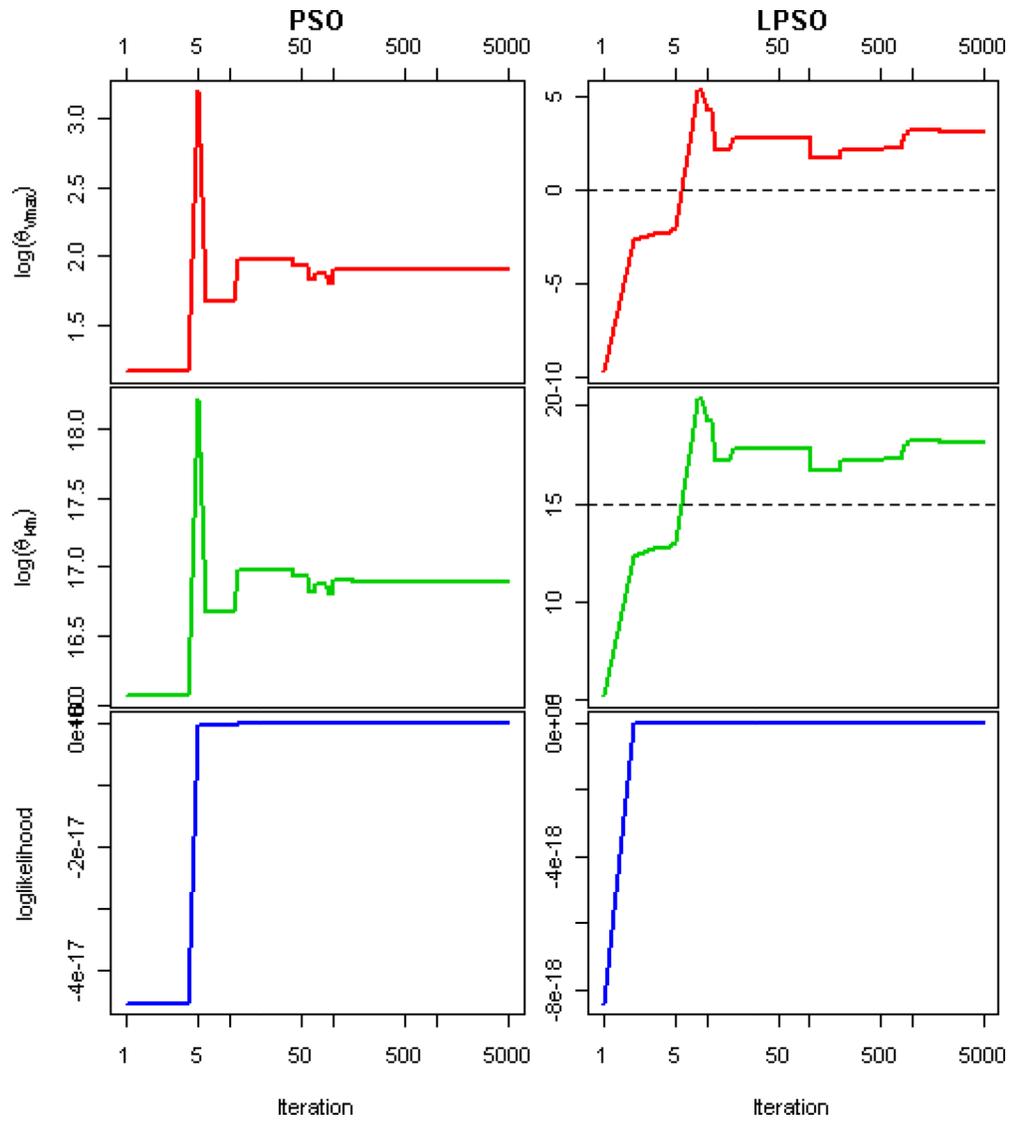


Figure 1. The trace plots between time and concentration for simulation experiments
The solid line is for the nonlinear ODEs and the dotted line is for linear ODEs. When $\log(Km) = -2.3$, the open black circle is used and the red triangle is used when $\log(Km) = 15$.







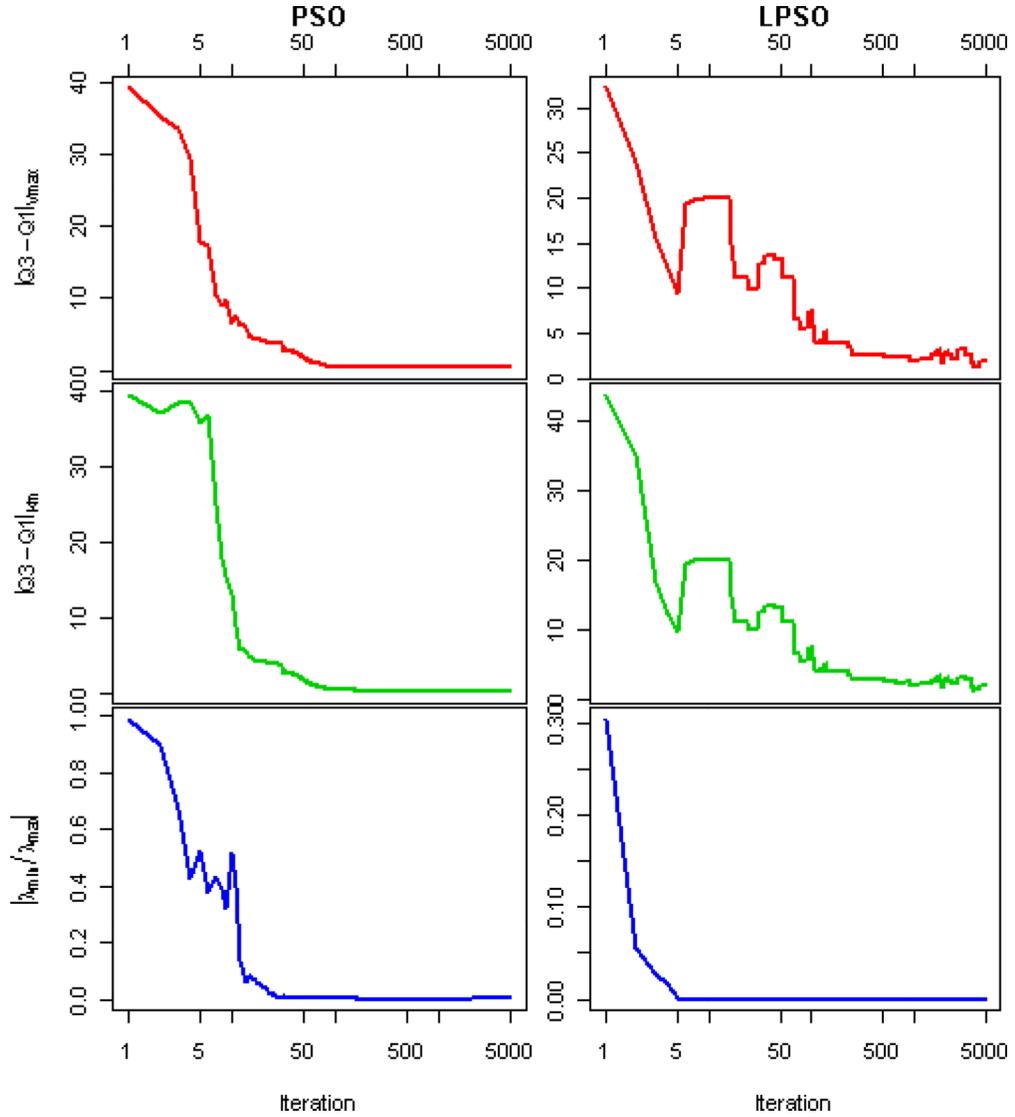
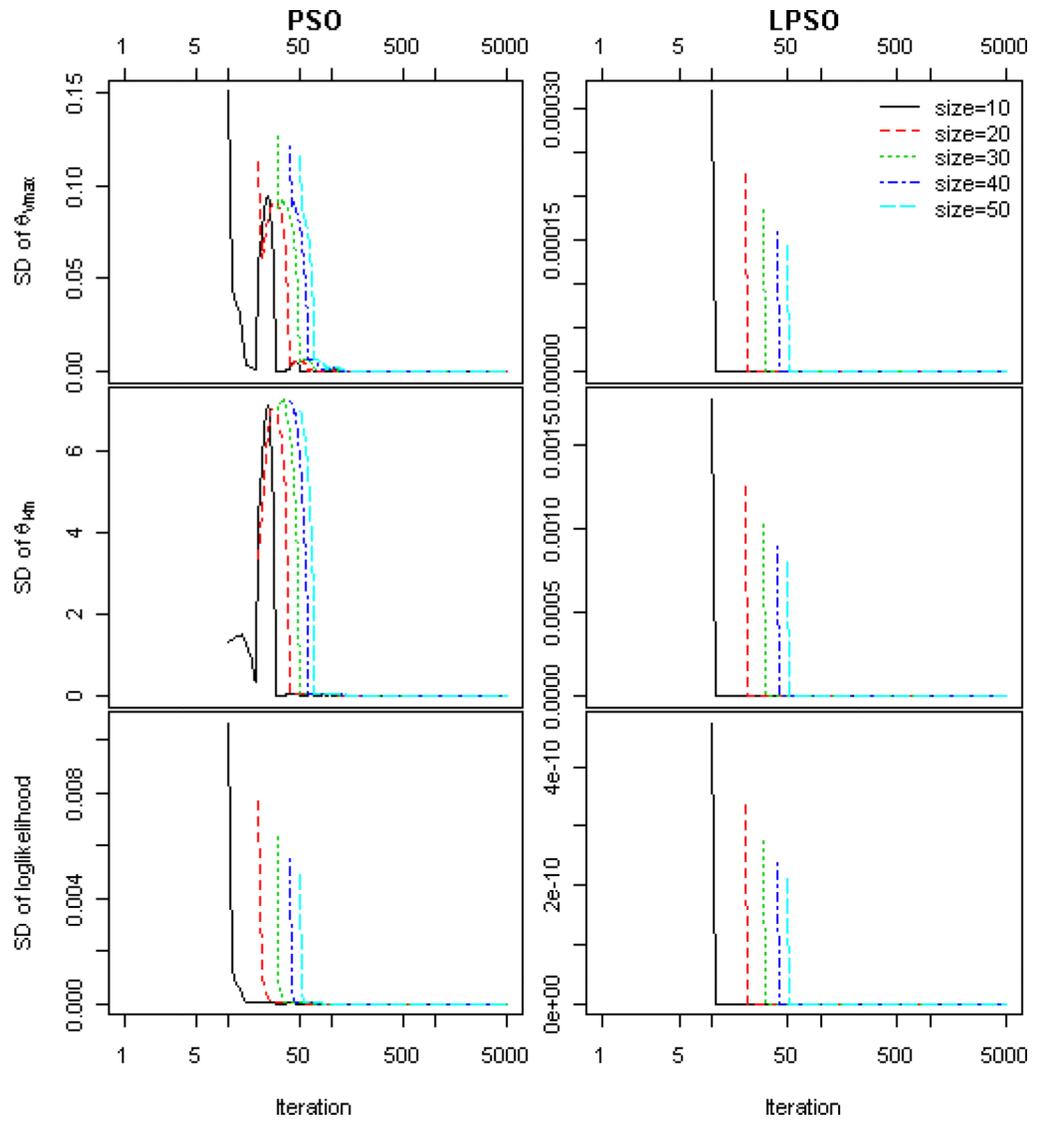
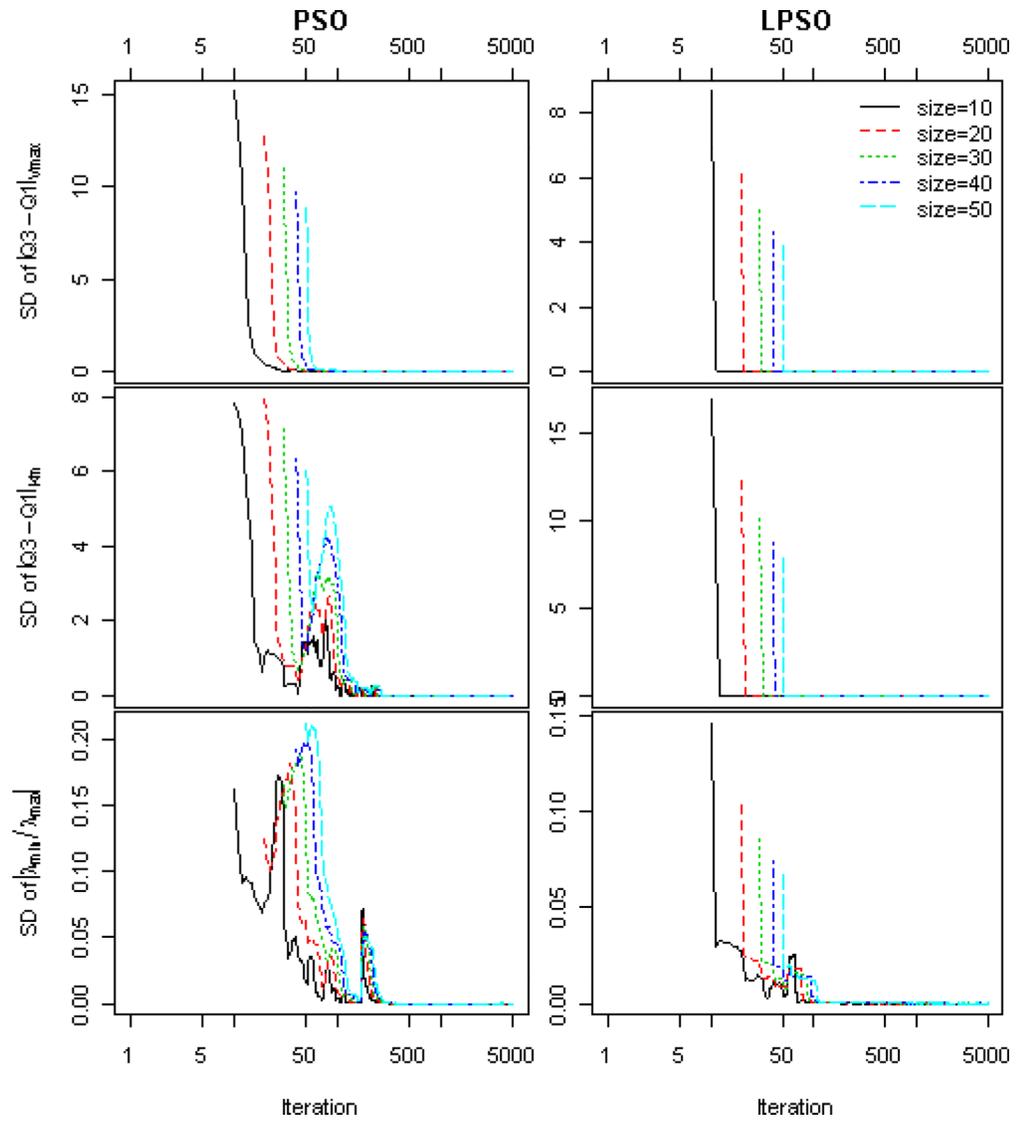
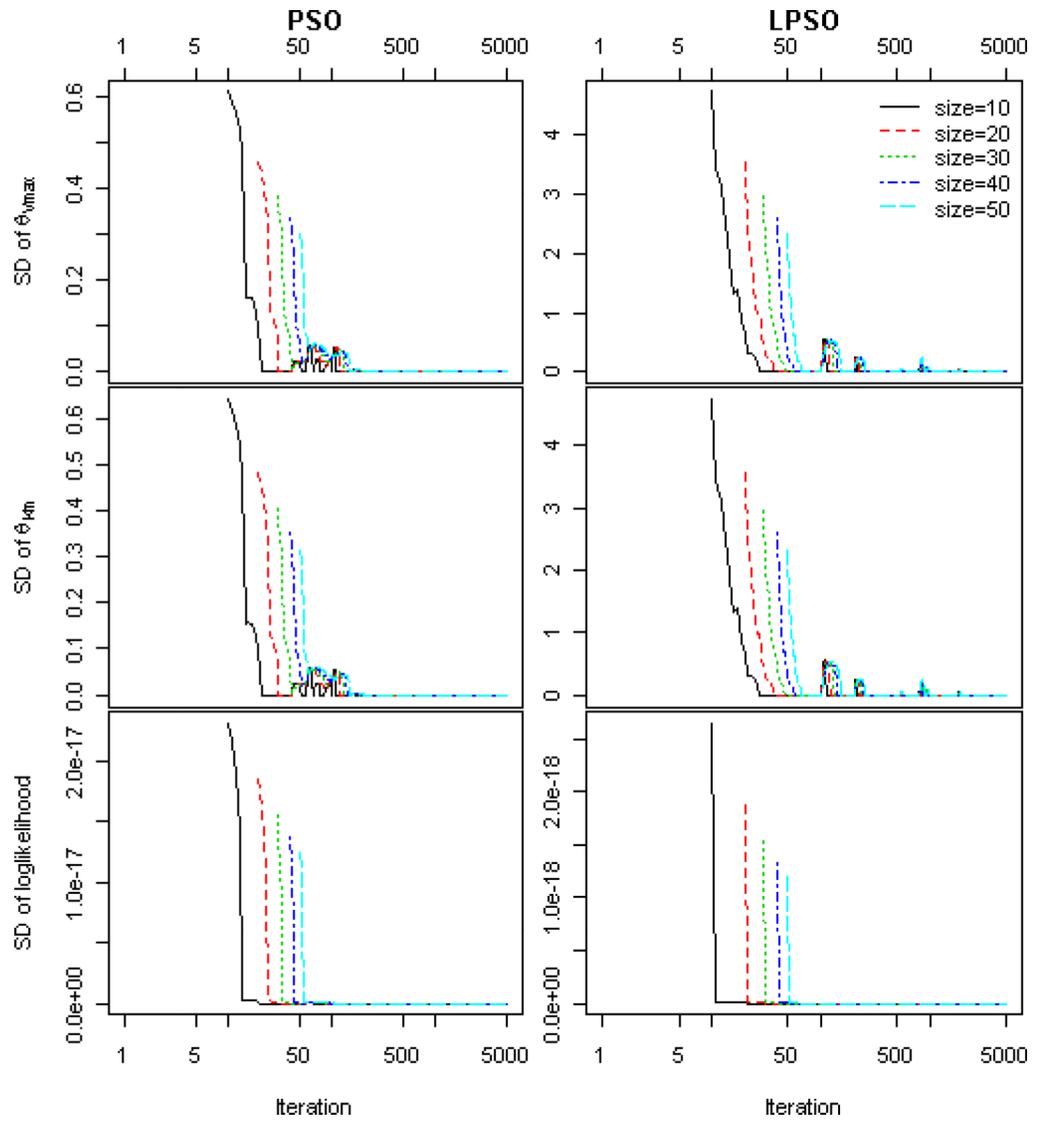


Figure 2. The trace plots between iteration and the global best and its log-likelihood
 In each plot, the first column is by PSO and the result of LPSO is the second column. The trace plots for the global best and its loglikelihood of $\log V_{max}$ and $\log K_m$ at each iteration are depicted when (a) $\log K_m = -2.3$ and (c) $\log K_m = 15$ where the dotted line indicates the true value. The trace plots for the difference of the first and third quartiles ($Q_1 - Q_3$) for $\log V_{max}$ and $\log K_m$ and the ratio between the minimum and maximum eigenvalues ($|\frac{\lambda_{min}^k}{\lambda_{max}^k}|$) of correlation matrix at each iteration are depicted when (b) $\log K_m = -2.3$ and (d) $\log K_m = 15$.







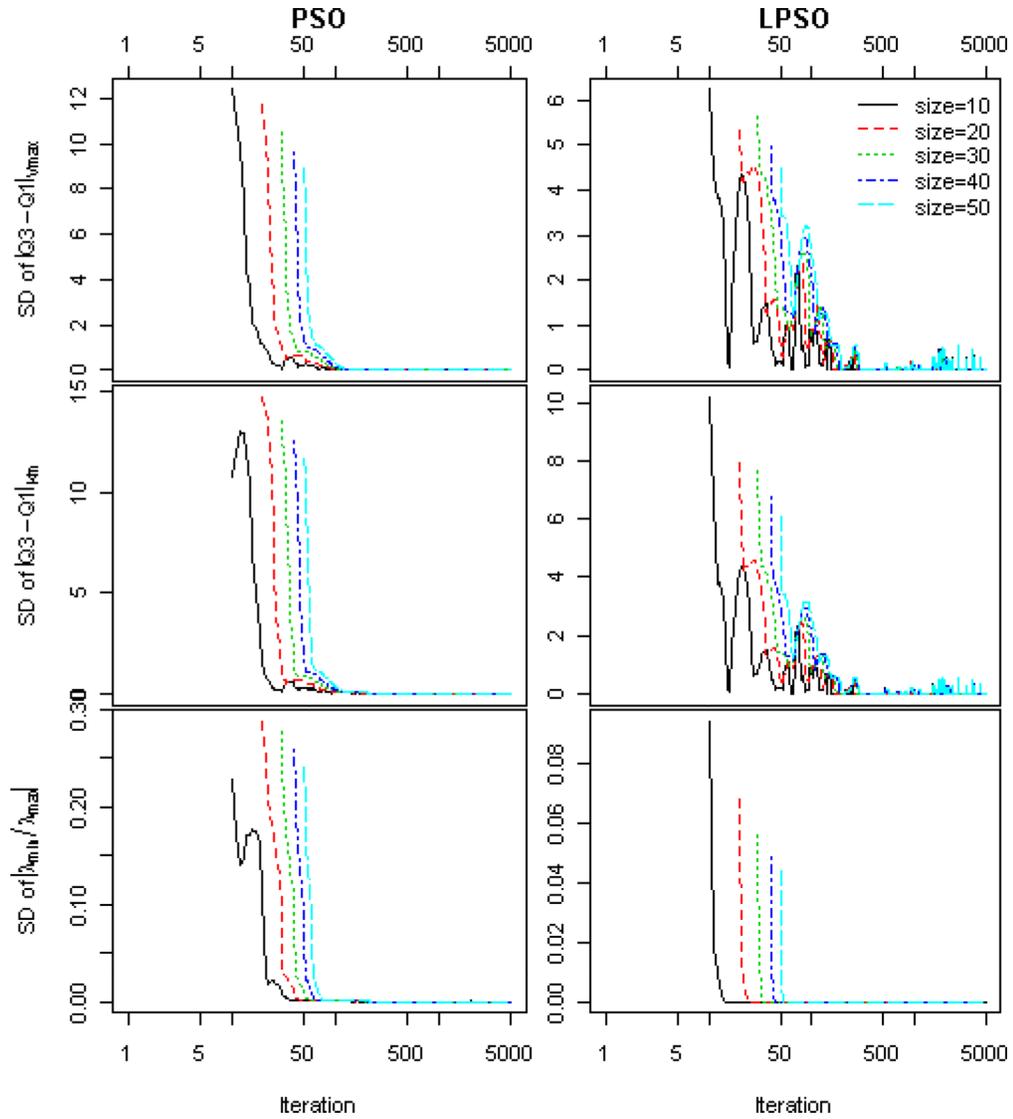
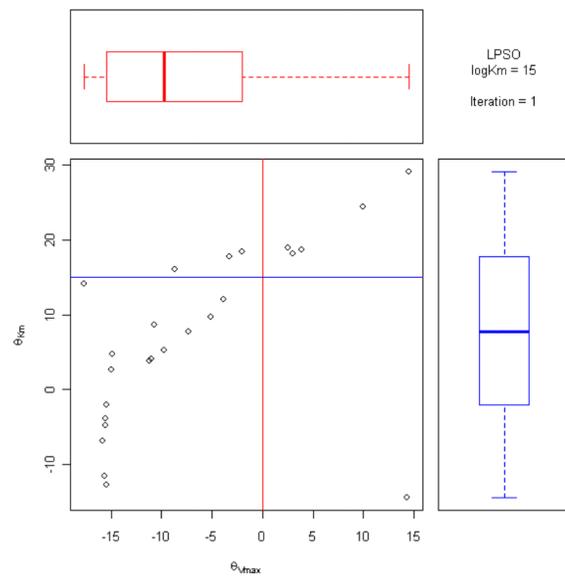
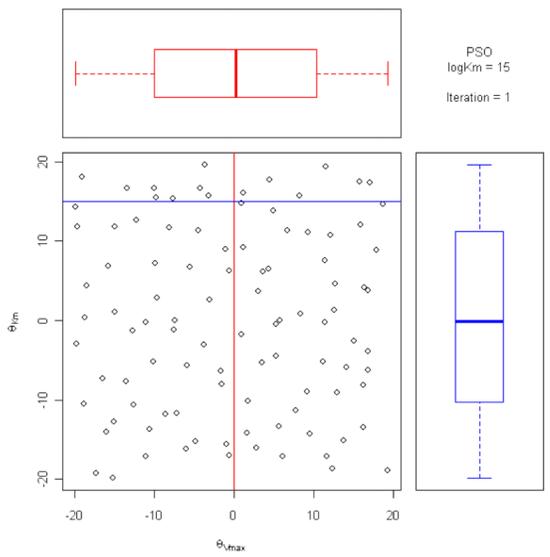
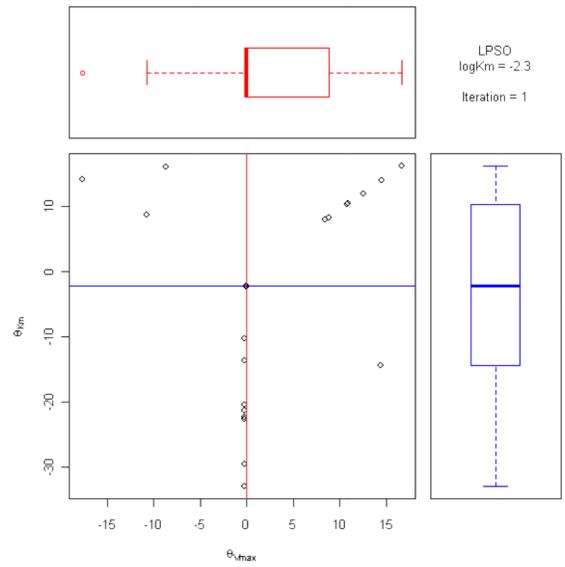
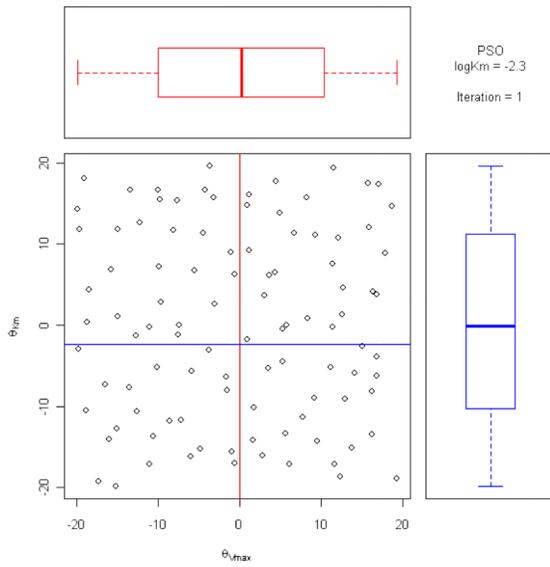
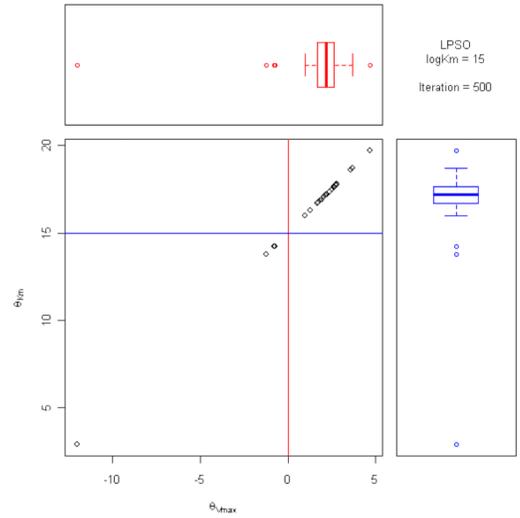
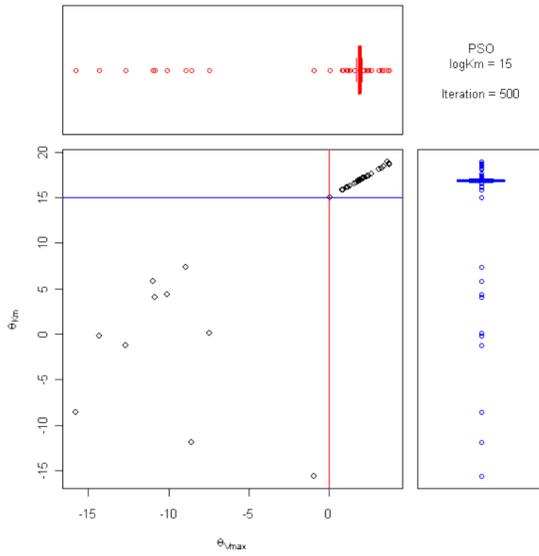
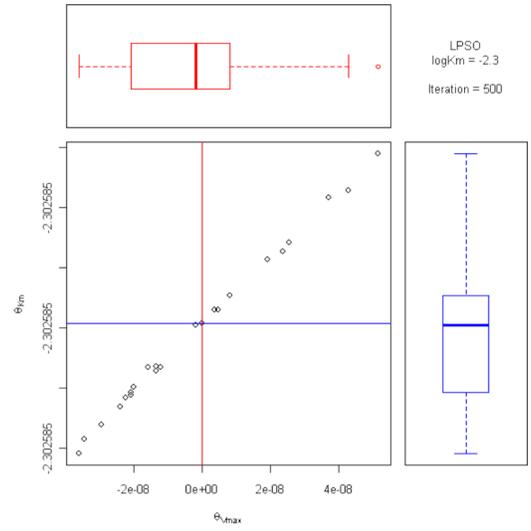
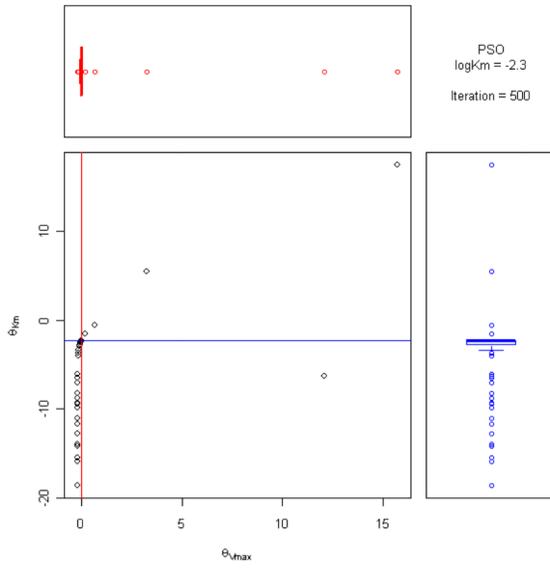


Figure 3. The standard deviation plots for convergence

The standard deviation plots of the global best and its loglikelihood at each iteration are depicted when (a) $\log Km = -2.3$ and (c) $\log Km = 15$ according to the five different window sizes (10, 20, 30, 40, 50). The standard deviation plots for the difference of the first and third quartiles ($|Q_1 - Q_3|$) and for the ratio between the minimum and maximum eigenvalues of correlation matrix at each iteration are depicted when (b) $\log Km = -2.3$ and (d) $\log Km = 15$. In each plot, the first column is by PSO and the result of LPSO is the second column.





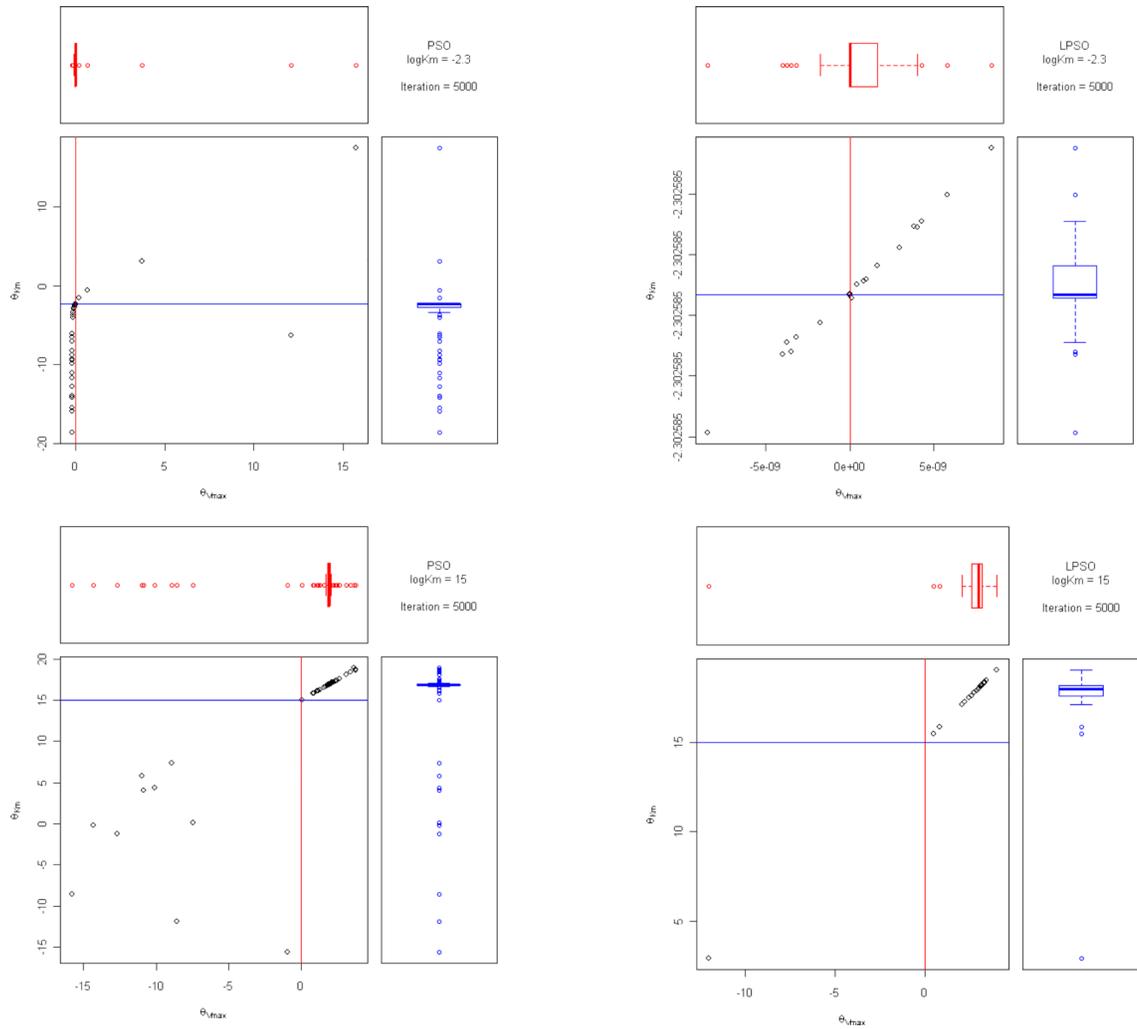
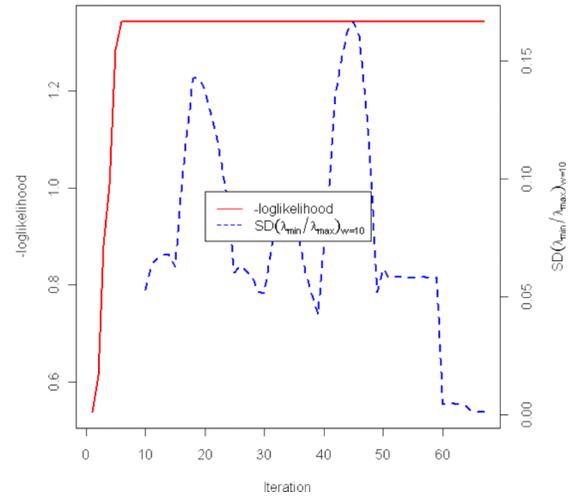
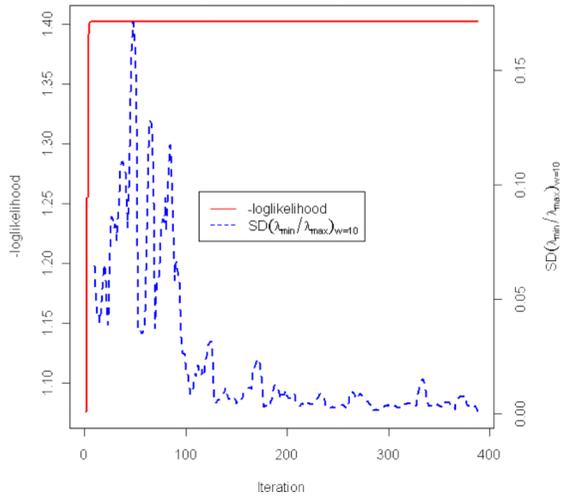
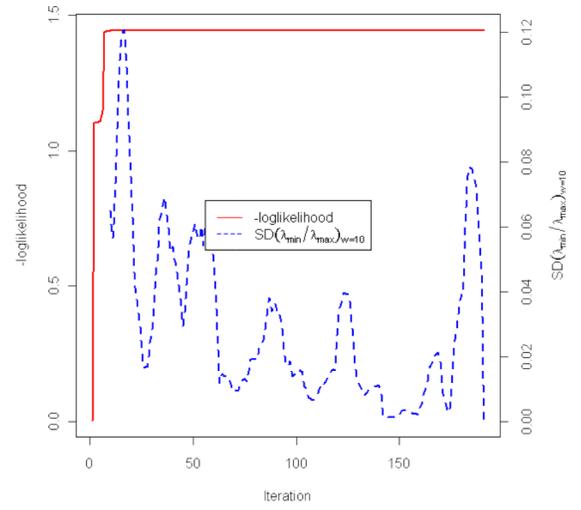
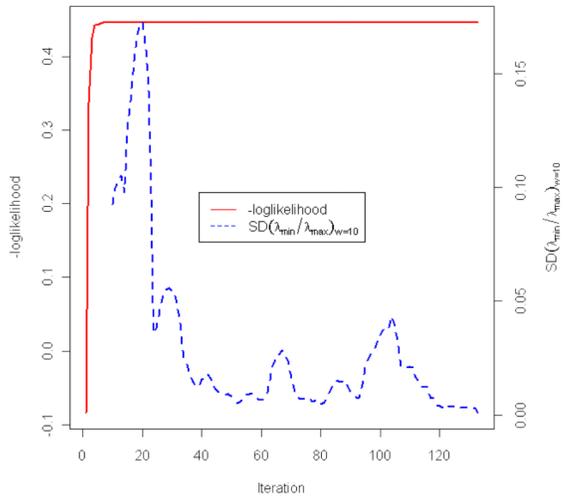
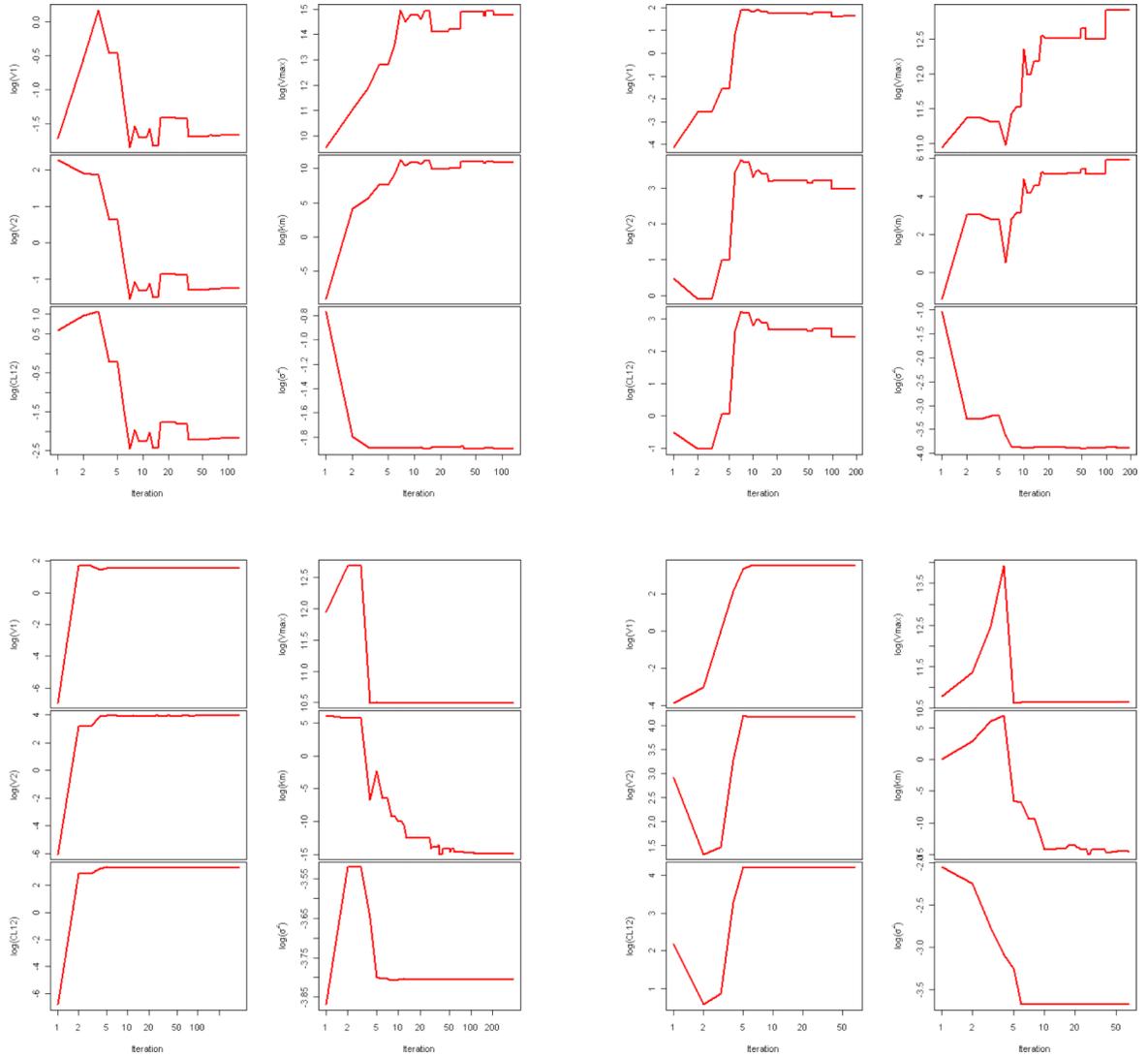
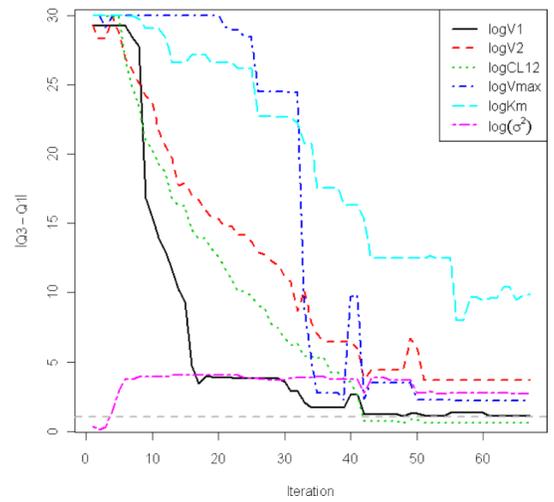
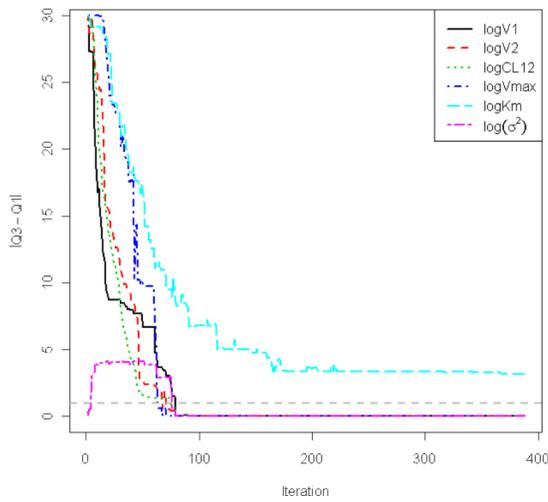
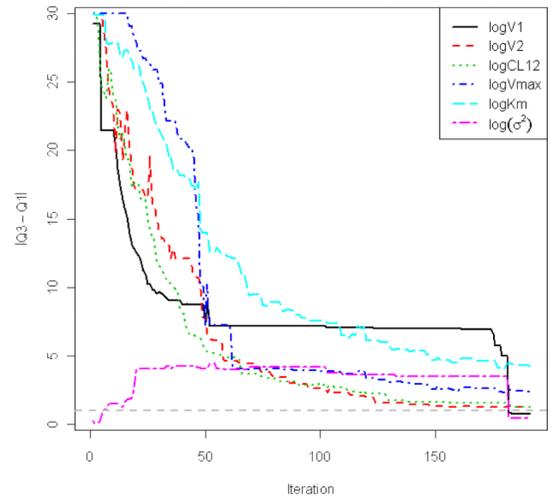
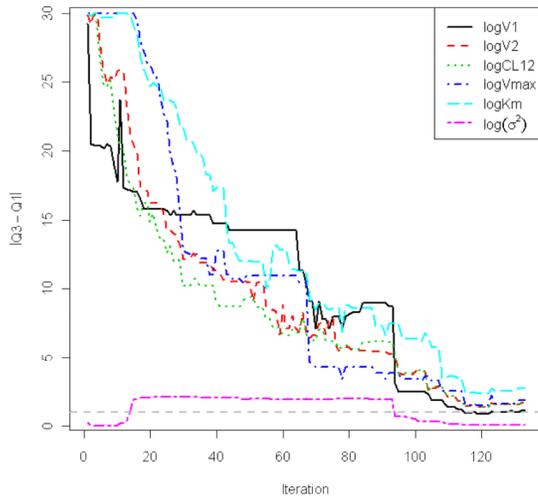


Figure 4. The scatter-box plots between $\log V_{umx}$ and $\log Km$ for PSO and LPSO. The plots in the left and right columns are for PSO and LPSO, respectively, and the first and second rows are for identifiable and non-identifiable cases. The solid lines in the plot indicate the true values for each parameter.







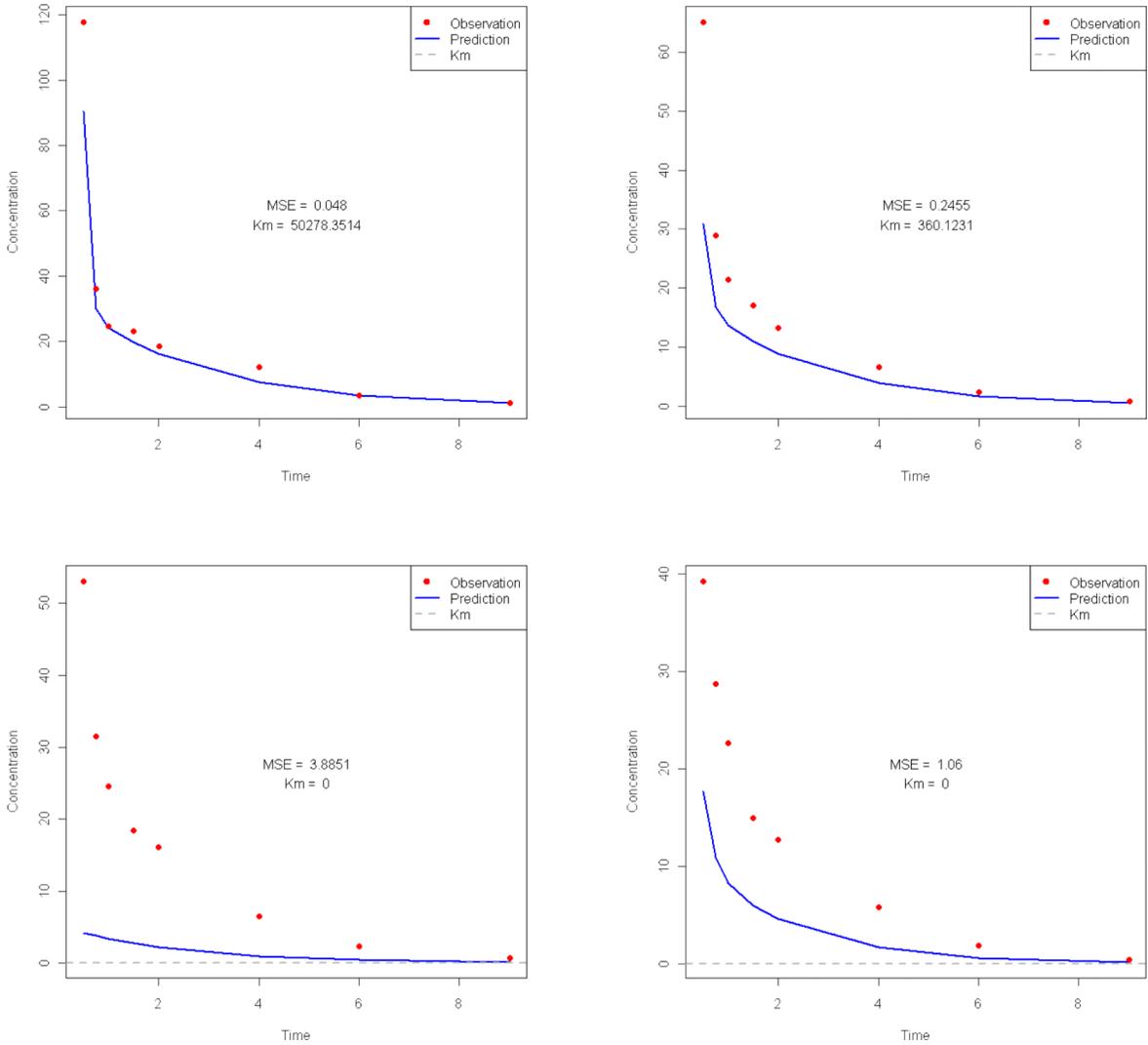


Figure 5. The results of estimation of LPSO for MDZ data

For each subject of four MDZ individual data, the trace plots of global estimates, the trace

plots of the loglikelihood and the standard deviation of the ratio of eigenvalues ($SD(\rho_{10}^k)$), the trace plots of the difference between Q1 and Q3 (d^k), and the prediction plot are displayed in (a)-(d), respectively. The grey dotted horizontal line indicates the threshold of 1 to see whether a parameter is identifiable in (c). In (d), the MSE stands for the mean squared error between the log-transformed observation and the predicted concentration. The estimated Km value is indicated by the dotted grey horizontal line. The plots of Subject 1, 2, 3, and 4 are in the left-top, the right-top, the left-bottom, and the right-bottom, respectively.

Table 1

The number of iterations to converge and its estimate for PSO and LPSO using the difference of the first and third quartiles (d^k) and the ratio of the eigenvalues (ρ^k) at k th iteration.

	PSO			LPSO			True		
	CV	Iter	$\hat{\theta}_{Vmax}$	$\hat{\theta}_{Krm}$	Iter	$\hat{\theta}_{Vmax}$	$\hat{\theta}_{Krm}$	θ_{Vmax}	θ_{Krm}
0									
10^{-5}					57	0.0000	-2.3026		
d^k	10^{-3}				9	0.0000	-2.3026		
	10^{-1}				3	0.0000	-2.3026		
	1	216	0.0000	-2.3026	3	0.0000	-2.3026		
Identifiable	0							0	-2.3
10^{-5}									
ρ^k	10^{-3}	293	0.0000	-2.3026	5	0.0000	-2.3026		
	10^{-5}	75	-0.0054	-2.3350	2	0.0000	-2.3026		
	1	1	-0.6308	-16.9981	1	0.0010	-2.2970		
	5000	0.0000	-2.3026	5000	0.0000	-2.3026			
0									
10^{-5}									
d^k	10^{-3}								
	10^{-1}								
	1	59	1.8228	16.8226					
Nonidentifiable	0								
10^{-5}					6	0.6985	15.6990	0	15
ρ^k	10^{-3}	98	1.8005	16.8005	5	-2.0000	13.0005		
	10^{-1}	13	1.9787	16.9777	2	-2.6834	12.3145		
	1	1	1.1539	16.0780	1	-9.7187	5.2450		
	5000	1.8948	16.8948	5000	3.1095	18.1095			

Table 2

The number of iterations to converge (Iter) and estimates for PSO and LPSO using the standard deviation of the global best ($SD(\psi_w^k)$) and of its loglikelihood ($SD(l_w^k)$) at k th iteration with different cutoff values (CV).

		Window size (w)																
		10		20		30		40		50								
	CV	Iter	$\hat{\theta}_{Vmax}$	$\hat{\theta}_{Km}$	Iter	$\hat{\theta}_{Vmax}$	$\hat{\theta}_{Km}$	Iter	$\hat{\theta}_{Vmax}$	$\hat{\theta}_{Km}$	Iter	$\hat{\theta}_{Vmax}$	$\hat{\theta}_{Km}$					
	(a) PSO	0	29	-0.021	-2.434	78	-0.005	-2.335	981	0.000	-2.303	991	0.000	-2.303	1001	0.000	-2.303	
10^{-5}		29	-0.021	-2.434	78	-0.005	-2.335	221	0.000	-2.303	230	0.000	-2.303	240	0.000	-2.303		
$SD(\psi_w^k)$		10^{-3}	29	-0.021	-2.434	78	-0.005	-2.335	144	0.000	-2.302	154	0.000	-2.303	163	0.000	-2.303	
10^{-1}		29	-0.021	-2.434	39	-0.017	-2.404	49	-0.007	-2.348	59	-0.005	-2.335	69	-0.005	-2.335		
1		17	-0.200	-15.909	39	-0.017	-2.404	49	-0.007	-2.348	59	-0.005	-2.335	69	-0.005	-2.335		
Identifiable	0	29	-0.021	-2.434	78	-0.005	-2.335	580	0.000	-2.303	665	0.000	-2.303	675	0.000	-2.303		
	10^{-5}	18	-0.200	-15.909	39	-0.017	-2.404	49	-0.007	-2.348	59	-0.005	-2.335	68	-0.005	-2.335		
	$SD(l_w^k)$	10^{-3}	12	-0.200	-15.909	21	-0.021	-2.434	31	-0.021	-2.434	41	-0.007	-2.348	51	-0.007	-2.348	
	10^{-1}	10	-0.197	-16.903	20	-0.021	-2.434	30	-0.021	-2.434	40	-0.017	-2.404	50	-0.007	-2.348		
	1	10	-0.197	-16.903	20	-0.021	-2.434	30	-0.021	-2.434	40	-0.017	-2.404	50	-0.007	-2.348		
	Nonidentifiable	0	21	1.979	16.978	31	1.979	16.978	41	1.979	16.978	142	1.902	16.902	152	1.902	16.902	
		10^{-5}	21	1.979	16.978	31	1.979	16.978	41	1.979	16.978	142	1.902	16.902	152	1.902	16.902	
		$SD(\psi_w^k)$	10^{-3}	21	1.979	16.978	31	1.979	16.978	41	1.979	16.978	142	1.902	16.902	152	1.902	16.902
		10^{-1}	20	1.979	16.978	29	1.979	16.978	38	1.979	16.978	47	1.931	16.930	55	1.931	16.930	
		1	10	1.668	16.675	20	1.979	16.978	30	1.979	16.978	40	1.979	16.978	50	1.931	16.930	
Nonidentifiable	0	21	1.979	16.978	31	1.979	16.978	41	1.979	16.978	142	1.902	16.902	152	1.902	16.902		
	10^{-5}	10	1.668	16.675	20	1.979	16.978	30	1.979	16.978	40	1.979	16.978	50	1.931	16.930		
	$SD(l_w^k)$	10^{-3}	10	1.668	16.675	20	1.979	16.978	30	1.979	16.978	40	1.979	16.978	50	1.931	16.930	
	10^{-1}	10	1.668	16.675	20	1.979	16.978	30	1.979	16.978	40	1.979	16.978	50	1.931	16.930		
	1	10	1.668	16.675	20	1.979	16.978	30	1.979	16.978	40	1.979	16.978	50	1.931	16.930		

(b) LPSO

Window size (w)		10		20		30		40		50		
CV	Iter	$\hat{\theta}_{Vmax}$	$\hat{\theta}_{Km}$	Iter	$\hat{\theta}_{Vmax}$	$\hat{\theta}_{Km}$	Iter	$\hat{\theta}_{Vmax}$	$\hat{\theta}_{Km}$	Iter	$\hat{\theta}_{Vmax}$	$\hat{\theta}_{Km}$
Identifiable												
0	31	0.000	-2.303	41	0.000	-2.303	51	0.000	-2.303	106	0.000	-2.303
10^{-5}	11	0.000	-2.303	21	0.000	-2.303	31	0.000	-2.303	41	0.000	-2.303
$SD(\psi_w^k)$	10^{-3}	11	0.000	-2.303	21	0.000	-2.303	31	0.000	-2.303	40	0.000
	10^{-1}	10	0.000	-2.303	20	0.000	-2.303	30	0.000	-2.303	40	0.000
1	10	0.000	-2.303	20	0.000	-2.303	30	0.000	-2.303	40	0.000	-2.303
0	31	0.000	-2.303	41	0.000	-2.303	51	0.000	-2.303	106	0.000	-2.303
10^{-5}	10	0.000	-2.303	20	0.000	-2.303	30	0.000	-2.303	40	0.000	-2.303
$SD(l_w^k)$	10^{-3}	10	0.000	-2.303	20	0.000	-2.303	30	0.000	-2.303	40	0.000
	10^{-1}	10	0.000	-2.303	20	0.000	-2.303	30	0.000	-2.303	40	0.000
1	10	0.000	-2.303	20	0.000	-2.303	30	0.000	-2.303	40	0.000	-2.303
Nonidentifiable												
0	27	2.778	17.778	37	2.778	17.778	47	2.778	17.778	57	2.778	17.778
10^{-5}	27	2.778	17.778	37	2.778	17.778	47	2.778	17.778	57	2.778	17.778
$SD(\psi_w^k)$	10^{-3}	27	2.778	17.778	37	2.778	47	2.778	17.778	57	2.778	17.778
	10^{-1}	27	2.778	17.778	37	2.778	47	2.778	17.778	56	2.778	17.778
1	19	2.778	17.778	26	2.778	17.778	35	2.778	17.778	45	2.778	17.778
0	27	2.778	17.778	37	2.778	17.778	47	2.778	17.778	57	2.778	17.778
10^{-5}	10	4.271	19.271	20	2.778	17.778	30	2.778	17.778	40	2.778	17.778
$SD(l_w^k)$	10^{-3}	10	4.271	19.271	20	2.778	17.778	30	2.778	40	2.778	17.778
	10^{-1}	10	4.271	19.271	20	2.778	17.778	30	2.778	40	2.778	17.778
1	10	4.271	19.271	20	2.778	17.778	30	2.778	17.778	40	2.778	17.778

Table 3

The number of iterations to converge (Iter) and the estimates for PSO and LPSO using the standard deviation of the difference of the first and third quartiles ($SD(d_w^k)$) and of the ratio of the eigenvalues ($SD(\rho_w^k)$) at kth iteration with different cutoff values (CV)

		Window size (w)													
		10		20		30		40		50					
CV	Iterk	θ_{Vmax}	θ_{Km}	Iter	θ_{Vmax}	θ_{Km}	Iter	θ_{Vmax}	θ_{Km}	Iter	θ_{Vmax}	θ_{Km}	Iter	θ_{Vmax}	θ_{Km}
		0	106	-0.001	-2.307	207	0.000	-2.303	245	0.000	-2.303	255	0.000	-2.303	265
10^{-5}	106	-0.001	-2.307	207	0.000	-2.303	245	0.000	-2.303	255	0.000	-2.303	265	0.000	-2.303
$SD(d_w^k)$	10^{-3}	106	-0.001	-2.307	150	0.000	-2.302	214	0.000	-2.303	255	0.000	-2.303	265	0.000
	10^{-1}	42	-0.007	-2.348	134	0.000	-2.303	143	0.000	-2.302	151	0.000	-2.302	160	0.000
	1	19	-0.200	-15.909	30	-0.021	-2.434	39	-0.017	-2.404	114	0.001	-2.299	124	0.000
Identifiable															
	0	1411	0.000	-2.303	1437	0.000	-2.303	1765	0.000	-2.303	1775	0.000	-2.303	1785	0.000
	10^{-5}	318	0.000	-2.303	334	0.000	-2.303	347	0.000	-2.303	359	0.000	-2.303	369	0.000
$SD(\rho_w^k)$	10^{-3}	106	-0.001	-2.307	146	0.000	-2.302	156	0.000	-2.303	282	0.000	-2.303	292	0.000
	10^{-1}	12	-0.200	-15.909	41	-0.007	-2.348	51	-0.007	-2.348	64	-0.005	-2.335	75	-0.005
	1	10	-0.197	-16.903	20	-0.021	-2.434	30	-0.021	-2.434	40	-0.017	-2.404	50	-0.007
Nonidentifiable															
	0	478	1.895	16.895	602	1.895	16.895	1154	1.895	16.895	1164	1.895	16.895	1174	1.895
	10^{-5}	145	1.902	16.902	155	1.902	16.902	196	1.895	16.895	206	1.895	16.895	216	1.895
$SD(d_w^k)$	10^{-3}	37	1.979	16.978	47	1.931	16.930	62	1.823	16.823	72	1.872	16.872	82	1.872
	10^{-1}	21	1.979	16.978	30	1.979	16.978	40	1.979	16.978	50	1.931	16.930	59	1.823
	1	10	1.668	16.675	20	1.979	16.978	30	1.979	16.978	40	1.979	16.978	50	1.931

(b) LPSO

Window size (w)		10		20		30		40		50		
CV	Iter	$\hat{\theta}_{Vmax}$	$\hat{\theta}_{Km}$	Iter	$\hat{\theta}_{Vmax}$	$\hat{\theta}_{Km}$	Iter	$\hat{\theta}_{Vmax}$	$\hat{\theta}_{Km}$	Iter	$\hat{\theta}_{Vmax}$	$\hat{\theta}_{Km}$
Identifiable												
0	143	0.000	-2.303	193	0.000	-2.303	203	0.000	-2.303	213	0.000	-2.303
10^{-5}	62	0.000	-2.303	71	0.000	-2.303	81	0.000	-2.303	91	0.000	-2.303
$SD(d_w^k)$	10^{-3}	12	0.000	-2.303	22	0.000	-2.303	32	0.000	-2.303	42	0.000
	10^{-1}	12	0.000	-2.303	22	0.000	-2.303	32	0.000	-2.303	42	0.000
1	12	0.000	-2.303	22	0.000	-2.303	32	0.000	-2.303	42	0.000	-2.303
0	187	0.000	-2.303	755	0.000	-2.303	812	0.000	-2.303	822	0.000	-2.303
10^{-5}	187	0.000	-2.303	206	0.000	-2.303	209	0.000	-2.303	530	0.000	-2.303
$SD(\rho_w^k)$	10^{-3}	70	0.000	-2.303	94	0.000	-2.303	103	0.000	-2.303	112	0.000
	10^{-1}	11	0.000	-2.303	21	0.000	-2.303	30	0.000	-2.303	40	0.000
1	10	0.000	-2.303	20	0.000	-2.303	30	0.000	-2.303	40	0.000	-2.303
Nonidentifiable												
0	63	2.778	17.778	164	1.729	16.729	174	1.729	16.729	184	1.729	16.729
10^{-5}	63	2.778	17.778	164	1.729	16.729	174	1.729	16.729	184	1.729	16.729
$SD(d_w^k)$	10^{-3}	63	2.778	17.778	164	1.729	16.729	174	1.729	16.729	184	1.729
	10^{-1}	16	2.202	17.202	164	1.729	16.729	174	1.729	16.729	184	1.729
1	15	2.202	17.202	49	2.778	17.778	60	2.778	17.778	108	1.729	16.729
0	144	1.729	16.729	173	1.729	16.729	183	1.729	16.729	326	2.211	17.211
10^{-5}	14	2.202	17.202	24	2.778	17.778	34	2.778	17.778	44	2.778	17.778
$SD(\rho_w^k)$	10^{-3}	14	2.202	17.202	24	2.778	17.778	34	2.778	44	2.778	17.778
	10^{-1}	10	4.271	19.271	20	2.778	17.778	30	2.778	40	2.778	17.778
1	10	4.271	19.271	20	2.778	17.778	30	2.778	17.778	40	2.778	17.778

Table 4

The global optima for the parameters of MDZ data

The MSE stands for the mean squared error between observed and predicted concentrations after log-transformed.

Subject	<i>V1</i>	<i>V2</i>	<i>CL12</i>	<i>Vmax</i>	<i>Km</i>	σ^2	MSE	The number of iterations
1	0.19	0.28	0.11	2593897	50278.35	0.15	0.0480	133
2	5.01	19.55	11.45	406634.8	360.12	0.02	0.2455	191
3	4.77	51.07	27.87	36187.57	3.21E-7	0.02	3.8851	388
4	33.39	65.02	67.37	41628.6	4.54E-7	0.03	1.0600	67