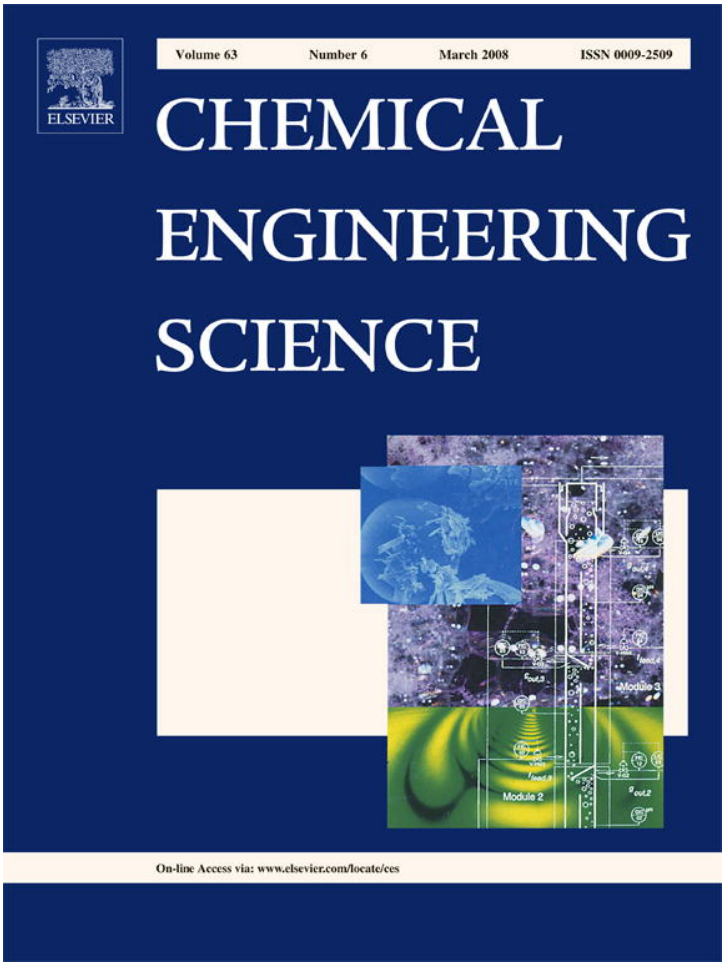

ELSEVIER

Chemical Engineering Science 63 (2008) 1542–1552

Chemical Engineering Science

www.elsevier.com/locate/ces

Nonlinear parameter estimation through particle swarm optimization

Marcio Schwaab, Evaristo Chalbaud Biscaia, Jr., José Luiz Monteiro, José Carlos Pinto*

Programa de Engenharia Química/COPPE, Universidade Federal do Rio de Janeiro, Cidade Universitária-CP: 68502, 21941-972 Rio de Janeiro, RJ, Brazil

Received 16 April 2007; received in revised form 4 November 2007; accepted 15 November 2007
Available online 19 November 2007

Abstract

Parameter estimation procedures are very important in the chemical engineering field for development of mathematical models, since design, optimization and advanced control of chemical processes depend on model parameter values obtained from experimental data. Model nonlinearity makes the estimation of parameter and the statistical analysis of parameter estimates more difficult and more challenging. In this work, it is shown that many of these difficulties can be overcome with the use of heuristic optimization methods, such as the particle swarm optimization (PSO) method. Parameter estimation problems are solved here with PSO and it is shown that the PSO method is efficient for both minimization and construction of the confidence region of parameter estimates. Moreover, it is shown that the elliptical approximation of confidence regions of nonlinear model parameters can be very poor sometimes and that more accurate likelihood confidence regions can be constructed with PSO, allowing for more reliable statistical analysis of the significance of parameter estimates.

Keywords: Parameter estimation; Particle swarm optimization; Confidence regions; Nonlinear models

1. Introduction

Mathematical models are commonly used for interpretation of experimental data, understanding of process behavior, in process design, process optimization and process control. Generally, mathematical models are composed of sets of equations (algebraic and/or differential) that represent the relationships among the independent (or controllable) variables and the dependent (or observable) variables. During model development, some variables that cannot be measured (or whose measurement is very difficult) may be defined in the model equations. In order to evaluate these variables (called model parameters) and to improve the model reliability, model parameter values must be estimated from available experimental data through minimization of some objective function that weighs the distance between model predictions and available experimental results.

For linear models, the minimization procedure allows for development of analytical solutions. Assuming that deviations between predicted and experimental data follow the normal probability distribution, the confidence region of parameter estimates defines a hyper-ellipsoid in the parameters space, with the point estimate of model parameters placed at the center of the hyper-ellipsoid (Draper and Smith, 1998).

In the case of nonlinear models, analytical solutions are unavailable and numerical iterative procedures must be used both for minimization of the objective function and evaluation of confidence regions of parameter estimates. Besides, model nonlinearity makes the minimization of the objective function more difficult and some regard must be taken about: (a) the size of the parameter space, (b) the existence of local minima, (c) the continuity of the objective function and (d) the sensitivity of the objective function to each of the model parameters (Hibbert, 1993). The first two points are related to definition of the initial guesses of parameter values, while the last two points are related to the definition of the derivatives of the objective function. These last two points are closely related to the numerical difficulties experienced by traditional methods of optimization, such as Newton-like methods, during minimization of the objective function.

With regard to the confidence region of parameter estimates in nonlinear models, even when the experimental deviations are normally distributed, the parameter deviations do not necessarily follow the normal distribution. The assumption of

* Corresponding author. Tel.: +55 21 25628337; fax: +55 21 25628300.
E-mail address: pinto@peq.coppe.ufrj.br (J.C. Pinto).

0009-2509/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ces.2007.11.024

the elliptical shape of the confidence region is very often a poor approximation of the confidence region (Donaldson and Schnabel, 1987; Bates and Watts, 1988). In spite of that, elliptical approximations of the confidence regions are widely used due to its simplicity: the parameter estimates are assumed to follow the normal distribution and only the information about the point estimate and the covariance matrix are necessary (Bard, 1974).

In order to obtain more accurate description of confidence regions, different methods have been proposed in the literature, such as the likelihood method (Beale, 1960), the lack-of-fit method (Halperin, 1963; Hartley, 1964; Williams, 1962) and more recently the profiling t-plots (Bates and Watts, 1988; Watts, 1994). Profiling t-plot methods were designed more specifically for computation of the confidence intervals of parameter estimates; confidence regions can then be obtained through interpolation. The likelihood method is exact only for linear models, although the shapes of the confidence regions obtained for nonlinear model parameters are very close to the exact ones. The lack-of-fit method can be used to produce exact confidence regions for all model parameters; however, it is more computationally expensive than likelihood method, as it also requires definition of model derivatives. As the confidence regions obtained with the likelihood method are usually very similar to the ones obtained with the lack-of-fit method (Donaldson and Schnabel, 1987), the use of the likelihood method is normally preferred.

Despite the previous remarks, it should be clear that the determination of the confidence regions of model parameters with the likelihood method does not constitute an easy task. Published works propose that a large number of model simulations be performed in order to obtain a large number of objective function values near the solution. These values are then used to construct the likelihood confidence region (Vanrolleghem and Keesman, 1996). Particularly, Klepper and Hendrix (1994) developed a procedure for obtaining uniform distribution of points inside the confidence region.

The use of heuristic optimization methods, such as the genetic algorithm (GA), simulated annealing (SA) and particle swarm optimization (PSO) for parameter estimation is very promising. (Heuristic optimization methods are general-purpose methods based on empirical evolutionary rules that frequently mimic successful optimization strategies found in nature. They are usually very flexible and can be applied to many types of objective functions and constraints, which explains the widespread interest in these techniques. The word "heuristic" is used to refer to these methods because it is not possible to guarantee that the exact optimum solution will be found, although a useful and good approximation of the searched optimum can frequently be obtained.) These methods can be used in problems that contain many model parameters, are not very sensitive to the initial parameter guesses, do not need derivatives of the objective functions and are able to perform global optimization through extensive calculation of the objective function in the parameter space. Many works have already reported the use of heuristic methods to perform parameter estimation and data reconciliation. Park and Froment (1998), Marseguerra et al. (2003) and Wongrat et al. (2005) used GA procedures; Costa et al. (2000) and Eftaxias et al. (2002) used SA procedures; Parsopoulos et al. (2001) used PSO techniques. Marseguerra et al. (2003) were the only ones to observe that a large number of solution points are generated during minimization and that these solution points can reveal some of the model characteristics, such as the model sensitivity to parameter estimates and how model sensitivity affects the convergence of the search procedure. However, only the final point estimates were considered in the proposed analysis, which is not sufficient to characterize the quality of the parameter estimates.

Some measure of the parameter uncertainties is necessary for proper analysis of the estimation results. This can be achieved by using the solution points generated by any of the heuristic optimization algorithms during function minimization. Detailed description of the confidence regions of parameter estimates can be obtained with heuristic procedures, providing a statistically rigorous analysis of parameter estimates, since the elliptical approximation is often inadequate for nonlinear model parameters. Therefore, the high number of function evaluations, which is traditionally regarded as a disadvantage of the heuristic methods, can be used for detailed statistical analysis and can constitute a major benefit of these algorithms.

According to our previous experience (Schwaab, 2005), PSO methods allow for improved parameter estimation performance with less computational efforts, when compared to GA and SA algorithm. For this reason, only PSO will be used in this work, although the determination of confidence regions of model parameters can be performed similarly with any other heuristic optimization method.

In this work, the PSO method is used for estimation of model parameters in nonlinear models. It is shown that PSO methods are capable of minimizing the objective function and describing the likelihood confidence region of model parameters very successfully.

This paper is organized as follows. In Section 2 the parameter estimation problem is formulated. Special attention is given to the definition of the confidence regions of parameter estimates. The PSO method is described in Section 3 and the obtained results are presented in Section 4. Linear parameter estimation is performed first in order to show that the PSO technique can recover the elliptical confidence region of simpler problems. Then some nonlinear problems are formulated and solved. It is shown that the obtained confidence regions may be very different from the elliptical approximations and that disconnected confidence regions can be found in some simple kinetic problems. Finally, Section 5 presents some concluding remarks.

2. Parameter estimation

As the experimental data are uncertain due to existence of experimental errors, the results obtained after the parameter estimation are also uncertain to some extent. Characterization of this uncertainty is of fundamental importance for proper evaluation of the final results. Definition of the maximum likelihood function may be convenient for interpretation of the parameter

estimation procedure. The experimental data can be regarded as random variables, whose joint probability distribution can be defined as

$$P(\mathbf{z}^e; \mathbf{z}^*, \mathbf{V}), \tag{1}$$

which describes the probability to obtain the experimental values $\mathbf{z}^e$, given the real unknown values $\mathbf{z}^*$ and a measure of the experimental errors $\mathbf{V}$. The maximum likelihood estimation consists in maximizing Eq. (1), given the model constraints

$$\mathbf{g}(\mathbf{z}^*, \boldsymbol{\theta}) = \mathbf{0}, \tag{2}$$

where $\boldsymbol{\theta}$ is a vector of model parameters and $\mathbf{g}$ is a vector of model functions. Assuming that the model is perfect and that the experiments are well done, it is reasonable to admit that the experimental results are the most probable ones. Therefore, attempts should be made to maximize the probability of obtaining these experimental results (Bard, 1974). When the experimental errors follow the normal distribution, maximization of the likelihood function is equivalent to minimization of the function

$$S(\boldsymbol{\theta}) = (\mathbf{z}^* - \mathbf{z}^e)^{\mathrm{T}} \mathbf{V}^{-1} (\mathbf{z}^* - \mathbf{z}^e), \tag{3}$$

where $\mathbf{z}$ is a vector that contains the independent $\mathbf{x}$ and dependent $\mathbf{y}$ variables and $\mathbf{V}$ is the covariance matrix of measurements. When the objective function defined in Eq. (3) is used, the procedure is usually called Data Reconciliation. Assuming that the independent variables $\mathbf{x}$ are known with great precision, the objective function becomes

$$S(\boldsymbol{\theta}) = (\mathbf{y}^* - \mathbf{y}^e)^{\mathrm{T}} \mathbf{V}_{\mathbf{y}}^{-1} (\mathbf{y}^* - \mathbf{y}^e) \tag{4}$$

and the model equations can be rewritten as

$$\mathbf{y}^* = f(\mathbf{x}^*, \boldsymbol{\theta}) \tag{5}$$

that assumes that the dependent variables can be calculated (explicitly or numerically) as function of the independent variables and of the model parameters.

If the experimental measurements of the dependent variables are uncorrelated, the matrix $\mathbf{V}_{\mathbf{y}}$ is diagonal and Eq. (4) takes the form of the known weighted least-squares function

$$S(\boldsymbol{\theta}) = \sum_{i=1}^{\mathrm{NE}} \sum_{j=1}^{\mathrm{NY}} \frac{(y^*_{i,j} - y^e_{i,j})^2}{\sigma^2_{i,j}}, \tag{6}$$

where $\sigma^2_{i,j}$ is the variance of the experimental fluctuations of the dependent variable j in the experiment i. NE and NY are the number of experiments and the number of dependent variables, respectively. When all variances are equal (which is usually assumed for models with only one dependent variable), Eq. (6) can then be simplified to take the form of the well-known least-squares function

$$S(\boldsymbol{\theta}) = \sum_{i=1}^{\mathrm{NE}} (y^*_i - y^e_i)^2. \tag{7}$$

The sum over different dependent variables was removed, since the use of the least-squares function is usually inappropriate for multi-response models. This is due to differences among the magnitudes of the responses and due to confusing different physical units. Matrix $\mathbf{V}$ provides the proper normalization and dimensionalization of each term in the sum.

Care should be taken regarding simplifications of matrix $\mathbf{V}$ and/or $\mathbf{V}_{\mathbf{y}}$, since over simplification of the covariance matrix may remove the statistical significance of the objective function and, consequently, leads to meaningless results. Although the maximum likelihood method may allow for estimation of the elements of the covariance matrix and of the model parameters simultaneously, this should not be encouraged as the covariance matrix may be nearly singular (Bard and Lapidus, 1968) and meaningless parameter estimates and parameter uncertainties may be obtained, as pointed out by Santos and Pinto (1998). Variable correlation should only be considered when it can be evaluated experimentally, independently from the parameter estimation procedure. Therefore, experimental determination of the covariance matrix through replication is encouraged, as it allows for independent definition of the experimental error structure. Additionally, the covariance matrix of experimental deviations may contain significant amount of information about the process (Cerqueira et al., 1999; Larentis et al., 2003; Rawet et al., 2001).

After definition of the objective function, several numerical methods can be used for minimization of the objective function. The methods used most often are the derivative-based ones. According to these methods, the minimization is performed along a direction that combines gradient vector (vector of first derivatives with respect to model parameters) and the Hessian matrix (matrix of second derivatives with respect to model parameters) of the objective function. A group of methods named direct search methods perform the minimization of the objective function based only on evaluation of the objective function, without the calculation of derivatives. Although the idea of minimization without the calculation of derivatives is appealing, Bard (1974) reports that gradient methods outperform direct search methods both in reliability and speed of convergence. Both gradient and direct search methods may be regarded as local search methods, since the search starts from an initial parameter guess and then evolves to a minimum. A good compilation of derivative-based and direct search methods can be found in Bard (1970, 1974) and Edgar and Himmelblau (1988).

Minimization of the objective function in parameter estimation problems, particularly in the field of chemical engineering, may lead to difficult numerical problems. Difficulties are related to the large number of model parameters, high correlation between model parameters and multimodal nature of the objective function. In order to overcome these difficulties, the use of heuristic optimization method, such as GA (Goldberg, 1989), SA (Kirkpatrick et al., 1983) and PSO (Kennedy and Eberhart, 1995), may be considered. These algorithms are characterized by the large number of function evaluations and a random search character, which assures a higher probability to find the global minima, when compared to derivative-based and direct search methods. These algorithms do not require initial guesses for model parameters and do not use derivatives. According to our previous experience, the PSO methods outperform other heuristic algorithms, particularly in parameter

estimation problems (Schwaab, 2005). For this reason, the PSO method will be described in detail in Section 3.

After finding a solution, the results have to be analyzed in statistical terms. Due to existence of experimental errors it is also necessary to evaluate the uncertainty of the model parameters.

The second-order Taylor expansion of the objective function around the minimum point estimate can be written as

$$S(\boldsymbol{\theta}) = S(\hat{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^{\mathrm{T}} \nabla \mathbf{S}_{\hat{\boldsymbol{\theta}}} + \tfrac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^{\mathrm{T}} \mathbf{H}_{\hat{\boldsymbol{\theta}}} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}), \tag{8}$$

where $\hat{}$ denotes the estimate of the point of minimum, $\nabla \mathbf{S}$ represents the gradient vector and $\mathbf{H}$ stands for the Hessian matrix of the objective function, defined as

$$\mathbf{H}_{\hat{\boldsymbol{\theta}}} = 2\left(\frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}}\right)^{\mathrm{T}} \mathbf{V}_{\mathbf{y}}^{-1} \left(\frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}}\right), \tag{9}$$

where Eq. (4) was assumed to be valid. As the Hessian matrix is also related to the covariance matrix of model parameters according to (Bard, 1974)

$$\mathbf{H}_{\hat{\boldsymbol{\theta}}} = 2\mathbf{V}_{\boldsymbol{\theta}}^{-1}, \tag{10}$$

Eq. (8)can be rewritten as

$$S(\boldsymbol{\theta}) - S(\hat{\boldsymbol{\theta}}) = (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^{\mathrm{T}} \mathbf{V}_{\boldsymbol{\theta}}^{-1} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \equiv \chi_p^2. \tag{11}$$

Eq. (11) follows a chi-square distribution with p degrees of freedom χ_p, where p is the number of parameters. As the objective function can be seen as a random variable that follows the chi-square distribution with $n-p$ degrees of freedom, where n is the total number of data points (NY · NE), the following expression can be written as

$$\frac{[S(\boldsymbol{\theta}) - S(\hat{\boldsymbol{\theta}})]/p}{S(\hat{\boldsymbol{\theta}})/(n-p)} = \frac{[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^{\mathrm{T}} \mathbf{V}_{\boldsymbol{\theta}}^{-1} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})]/p}{S(\hat{\boldsymbol{\theta}})/(n-p)} \equiv F_{p,n-p}. \tag{12}$$

The ratios in both sides of Eq. (12) follow an F-distribution with p and $n-p$ degrees of freedom. According to the right side of Eq. (12), a confidence region with confidence level $(1-\alpha)$ can be defined as

$$(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^{\mathrm{T}} \mathbf{V}_{\boldsymbol{\theta}}^{-1} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leqslant \frac{p}{n-p} S(\hat{\boldsymbol{\theta}}) F_{p,n-p}^{1-\alpha}. \tag{13}$$

Eq. (13) defines a hyper-ellipsoid in the parameter space that is the confidence region of the parameter estimates, according to the traditional procedure employed for parameter analysis. The confidence interval of model parameters can be defined as

$$\theta_i \pm t_{n-p}^{1-\alpha/2} (s^2 v_{ii})^{1/2}, \tag{14}$$

where v_{ii} is the ith diagonal element of matrix $\mathbf{V}_\theta$ and s^2 is defined as

$$s^2 = S(\hat{\boldsymbol{\theta}})/(n-p). \tag{15}$$

$t_{n-p}^{1-\alpha/2}$ is the t-Student variable defined for $n-p$ degrees of freedom and confidence level of $(1-\alpha/2)$.

The uncertainty of parameter estimates can be analyzed in terms of the elliptical confidence regions or in terms of the confidence intervals. The confidence regions should be preferred, as more detailed information about model parameters can be obtained, particularly about correlation among the estimated parameters. Interesting discussions about confidence intervals and confidence regions of model parameters can be found in Draper and Guttman (1995) and Draper and Smith (1998).

For nonlinear models, even when experimental deviations are normally distributed, parameter deviations do not necessarily follow the normal distribution. Therefore, the elliptical shape of the confidence region only provides an approximation of the confidence region (Bates and Watts, 1988; Donaldson and Schnabel, 1987). According to the left side of Eq. (12), a confidence region with confidence level $(1-\alpha)$ can be defined as

$$S(\boldsymbol{\theta}) \leqslant S(\hat{\boldsymbol{\theta}}) \left(1 + \frac{p}{n-p} F_{p,n-p}^{1-\alpha}\right). \tag{16}$$

Beale (1960) was the first to propose the use of Eq. (16) for computation of confidence regions. This equation is exact for linear models, when experimental errors follow the normal distribution. In a more general case, when nonlinear models are used and/or when experimental errors fluctuate according to an arbitrary probability distribution, Eq. (16) should be rewritten as: $S(\boldsymbol{\theta}) \leqslant cS(\hat{\boldsymbol{\theta}})$; where c is a constant that depends on the required confidence level and on the defined objective function. Unfortunately, the exact definition of the probability distributions of experimental fluctuations (and, consequently, of the objective function) and of parameter uncertainties (due to the nonlinear character of the model) may constitute a very difficult task, making hard the correct definition of the constant c. However, as Eq. (16) does not require the confidence region to have the elliptical shape, very good approximations of the true confidence region can be obtained with this method (Donaldson and Schnabel, 1987). The constant c is estimated with the help of Eq. (16), as discussed previously.

The confidence regions obtained with this method (Eq. (16)), called likelihood confidence regions, can be disjoint and unbounded, as shown in the following sections. This is because the contours of a complex nonlinear function can be disjoint and unbounded. For two-parameter models, the likelihood region can be determined with standard contouring methods (Bates and Watts, 1988). However, obtaining the contours of confidence regions in problems that involve more than two parameters may be difficult, requiring the evaluation of a very large number of points to produce a satisfactory contour.

The difficulties described in the previous paragraph can be minimized when the PSO method is used for objective function minimization. As the PSO method performs a high number of objective function evaluations (as other heuristic search procedures), the likelihood confidence region can be determined through proper selection of the points that satisfy Eq. (16) along the search path. Plotting of these points produce the confidence regions. Therefore, the only additional computational effort is the selection of the points that satisfy Eq. (16), since multiple evaluations of the objective function is naturally performed by the PSO method (and other related heuristic procedures) during objective function minimization. As a consequence, the use of linear approximations of the confidence regions of parameter

estimates are not necessary when heuristic procedures are used for parameter estimation.

3. Particle swarm optimization

The PSO technique was originally proposed by Kennedy and Eberhart (1995), based on the social behavior of collection of animals. Each individual of the swarm, called particle, remembers the best solution found by itself and by the whole swarm along the search trajectory. The particles move along the search space and exchange information with other particles, in accordance with the following equations:

$$v_{p,d}^{k+1} = w v_{p,d}^{k} + c_1 r_1 (x_{p,d}^{\text{ind}} - x_{p,d}^{k}) + c_2 r_2 (x_{d}^{\text{glo}} - x_{p,d}^{k}), \tag{17}$$

$$x_{p,d}^{k+1} = x_{p,d}^{k} + v_{p,d}^{k+1}. \tag{18}$$

In Eqs. (17)–(18), p denotes the particle, d is the search direction, k represents the iteration number, v is the velocity (or pseudo-velocity) of the particle and x is the position of particle. x^{ind} and x^{glo} represent the regions of the search space where the objective function attains low (optimum) values. x^{ind} is the best position found by the particle itself, while x^{glo} is the best position found by whole swarm. r_1 and r_2 are two random numbers with uniform distribution in the range [0, 1]. w, c_1 and c_2 are search parameters. The parameters c_1 and c_2 are the cognition and the social parameters. The parameter w is called inertial weight and was not present in the original form of the algorithm. This parameter was inserted into the algorithm by Shi and Eberhart (1998) in order to assure the convergence of particles to the best point in the course of the search. Shi and Eberhart (1998) also proposed the use of a linear decreasing value of w, in order to increase the exploration of the parameter space during the initial search iterations and increase the exploitation of the parameter space during the final steps of the search.

The PSO presents interesting characteristics along the iterations. In the initial iterations, the random character of the search is high and the particles conduct a global search over the search region. As the iterations evolve, the particles concentrate around the more promising regions, found during the exploration stage. This local search is called exploitation and leads to improvement of the solution. Proper balancing between exploration and exploitation is of fundamental importance for successful searches, assuring simultaneously global search characteristic and good precision of the final results. When exploration is emphasized, convergence may not be attained. When exploitation is emphasized, premature convergence can occur in some region that is far from the global minimum. The proper selection of the search parameters is the key for proper balancing between exploration and exploitation. Complete theoretical analysis of the particle trajectories, particle convergence and search parameters effects can be found in Clerc and Kennedy (2002), Van den Berg and Engelbrecht (2006) and Trelea (2003). The PSO method has been used in several fields, for optimization of the operations of styrene polymerization process (Costa Jr., et al., 2003), for multi-objective optimizations (Parsopoulos and Vrahatis, 2002), for nonlinear dynamic analysis of chemical processes (Ourique et al., 2002) and for parameter estimation in error-in-variables problems (Parsopoulos et al., 2001).

The greatest disadvantage of the PSO method (and also of other heuristic algorithms) is the high number of objective function evaluations, which requires longer computation times when compared to the traditional methods. However, as already discussed in Section 2, this high number of objective functions can be used for rigorous statistical analysis of the final results, so that this disadvantage becomes a gain over traditional methods for determination of confidence regions of parameter estimates. Besides, PSO methods are not very sensitive to initial guesses of model parameters, which makes its use appealing when large number of unknown parameters are present in the model. The PSO algorithm is presented in detail in Scheme 1.

Scheme 1: The PSO algorithm

1. Initialize the search parameters:
   Niter: number of iterations;
   Npt: number of particles;
   Nd: number of searched dimensions;
   $\mathbf{x}^{\text{MIN}}$ and $\mathbf{x}^{\text{MAX}}$: vectors of length Nd with searching limits;
   c_1, c_2, w_o, w_f: PSO searching parameters;
   set $k = 0$ (iteration counter).
2. Calculate the maximum particle velocities along each direction d:

$$v_d^{\text{MAX}} = (x_d^{\text{MAX}} - x_d^{\text{MIN}})/2.$$

3. Calculate initial particle positions and velocities:

$$x_{p,d}^{k} = x_d^{\text{MIN}} + r(x_d^{\text{MAX}} - x_d^{\text{MIN}}),$$

$$v_{p,d}^{k} = v_d^{\text{MAX}}(2r - 1).$$

4. Evaluate the objective function for each particle.
5. Write the particle positions and particle objective functions in a file to be used for construction of the confidence region.
6. Update $\mathbf{x}^{\text{glo}}$, a vector with dimension Nd that contains the best position found by the whole particle swarm.
7. When the maximum number of iterations is achieved ($k =$ Niter), the search is terminated.
8. Update $\mathbf{x}_p^{\text{ind}}$, Npt vectors with dimension Nd that contain the best position found by each particle of the swarm.
9. Calculate the inertial weight value:

$$w = w_o + (w_f - w_o)\frac{k}{\text{Niter}}.$$

10. Update the particle velocities for p=1...Npt; d=1...Nd:

$$v_{p,d}^{k+1} = w v_{p,d}^{k} + c_1 r_1 (x_{p,d}^{\text{ind}} - x_{p,d}^{k}) + c_2 r_2 (x_{d}^{\text{glo}} - x_{p,d}^{k}).$$

11. If the absolute particle velocity is higher than the maximum permitted value then:

$$v_{p,d}^{k+1} = v_d^{\text{MAX}} \text{sign}(v_{p,d}^{k+1}).$$

12. Update the particle positions:

$$x_{p,d}^{k+1} = x_{p,d}^{k} + v_{p,d}^{k}.$$

13. If the particle position is not inside the searching limits, the particle is placed at the violated searching limit.
14. Add an iteration to the iteration counter ($k = k + 1$) and return to step 4.

Due to the simplicity of the PSO algorithm, the implementation of the algorithm is very easy. It is important to say that the computer time required for PSO computations is very small when compared to the time required for computation of model predictions and evaluation of objective function values for all particles. Consequently, the total time required for optimization is basically the time required for model evaluations.

4. Results

Three parameter estimation problems are used to illustrate the procedure proposed here. These problems show that the PSO is capable of minimizing the objective function and describing the confidence regions of model parameters successfully. The PSO algorithm and the analyzed mathematical models were implemented in FORTRAN, as part of ESTIMA (a software package for parameter estimation and design of experiments). A desktop computer (Pentium 4 3.0 MHz processor and 1024 Mb of RAM memory) was used to perform the computations presented here.

Problem 1. Linear parameter estimation

The first problem consists of a very simple linear parameter estimation problem. Since the confidence region of parameter estimates is elliptical, the shape of the confidence region is known a priori. Thus, one can observe where the PSO is able to describe the confidence region of the parameter estimates as obtained rigorously through theoretical analysis.

The linear model is shown in the following equation and has two parameters, θ_1 and θ_2, which were estimated using the data presented in Table 1:

$$y = \theta_1 x + \theta_2. \tag{19}$$

Table 1
Data for linear parameter estimation

x	y
1.0	9.92
2.0	16.89
3.0	17.12
4.0	26.03
5.0	30.71
6.0	33.28
7.0	39.83
8.0	42.44
9.0	50.44
10.0	53.63

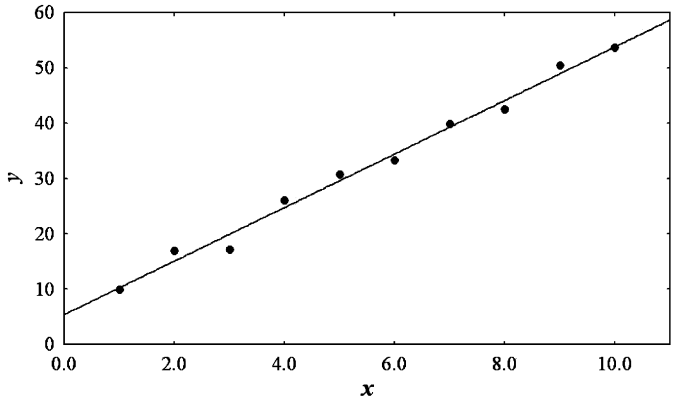

Fig. 1. Linear model adjust to data in Problem 1.

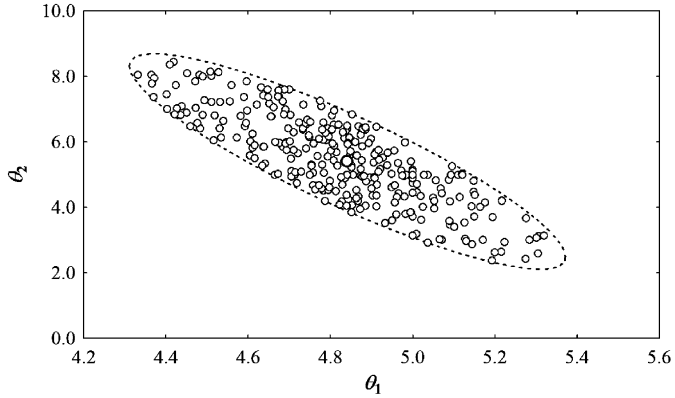

Fig. 2. Elliptical confidence region (dashed line) and likelihood confidence region obtained through particle swarm optimization (points) in Problem 1.

The search parameters c_1 and c_2 were made equal to 2.0 and w was made equal to 0.9. Twenty particles were used and 100 iterations were performed. These parameter values are used quite often in most publications regarding the PSO. The search space was defined as the interval [0, 10] for both parameters. The minimum objective function value was 20.8 with parameter estimates equal to 4.85 and 5.28 for θ_1 and θ_2, respectively. These numbers are the same ones obtained with the usual least-squares analytical solution. The typical computer time required to solve this simple problem was about 1 s. As model calculations are very simple in this case, the number of particles and iterations required to provide good numerical results are very small, as presented previously.

Fig. 1 illustrates the obtained linear fit and Fig. 2 shows the elliptical (Eq. (13)) and likelihood (Eq. (16)) confidence regions. The likelihood confidence region of Fig. 2 was constructed with 305 points that had been used previously by the PSO procedure during the search. The selected points were the ones that led to objective functions smaller than 43.8, as calculated from Eq. (16). As it can be seen in Fig. 2, the likelihood confidence region is similar to the elliptical one, showing that points used by the PSO during the minimization can be used effectively for definition of the parameter confidence region.

Table 2
Biochemical oxygen demand along the time

x (days)	y (mg/L)
1.0	8.3
2.0	10.3
3.0	19.0
4.0	16.0
5.0	15.6
7.0	19.8

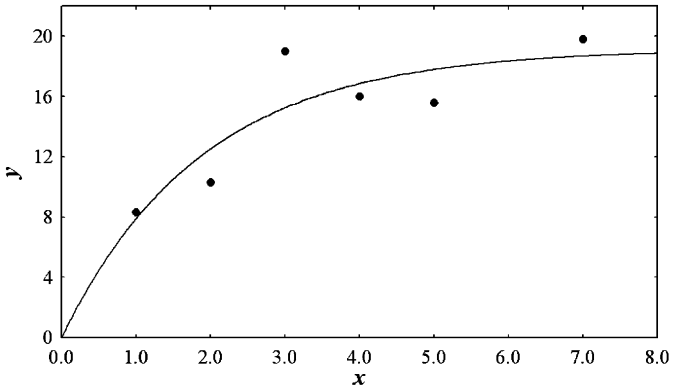

Fig. 3. Model adjusts to data in Problem 2.

Example 2. First-order kinetics

This problem consists of a nonlinear model that describes a first-order irreversible reaction (A → B) in a batch stirred tank. The equation that describes how product B concentration varies with the time can be described as

$$y = \theta_1(1 - e^{-\theta_2 x}). \tag{20}$$

This problem was taken from Bates and Watts (1988). The model was used to describe the variation of the biochemical oxygen demand (defined as y) along the time (defined as x). There are two parameters to be estimated which are related to unknown initial conditions (θ_1) and to the kinetic rate constant (θ_2). The available experimental data are presented in Table 2.

The search parameters c_1 and c_2 were made equal to 2.0, while w was made equal to 0.9. One thousand iterations were performed and 40 particles were used. Compared to the previous example, these numbers were increased to provide better representation of the confidence region of parameter estimates, as analytical solutions are not available in this case. Parameter search space was restricted to the interval [0, 100] to both parameters. Search regions were enlarged on purpose, to allow for larger exploration of the parameter space. The minimum found was equal to 26.0 for parameter values of 19.15 and 0.5273, respectively. The typical computer time required to solve Example 2 was about 1 s, as in the previous example.

Fig. 3 shows the model fit to the experimental data. Fig. 4 shows both the likelihood and the elliptical confidence regions of parameter estimates.

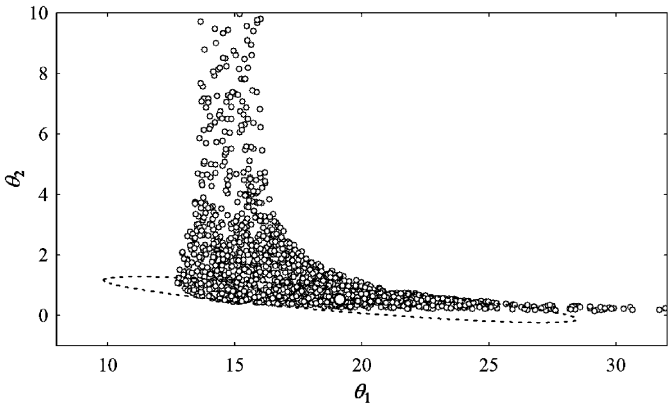

Fig. 4. Elliptical confidence region (dashed line) and likelihood confidence region obtained through particle swarm optimization (points) in Problem 2.

Different from Problem 1, where both confidence regions were equal, Fig. 4 shows that the likelihood confidence region is very different from the elliptical one. As the model is nonlinear, the elliptical confidence region provides only a very poor approximation of the confidence region of the parameter estimates. Fig. 4 shows that the likelihood confidence region is non-convex and is unbounded, since parameter θ_2 does not posses an upper limit. The occurrence of this type of confidence region is due to model structure. As the exponential term approaches zero when θ_2 increases, the model provides a constant output response that is equal to parameter θ_1, as discussed by Bates and Watts (1988). This means that the model is unable to discriminate high values of θ_2, as high values of θ_2 lead to similar model fits.

Despite its simplicity, this problem shows very clearly that the use of elliptical confidence regions in nonlinear parameter estimation can lead to very wrong conclusions. This can be avoided with the use of likelihood confidence regions described by function evaluations performed by the PSO procedure along the search.

Problem 3. Polymerization kinetics

This problem consists in the estimation of kinetic parameters for the ethylene polymerization carried out with nickel complexes. Variation of reaction rates along the reaction course and the characteristics of the final polymer products suggest that the active species change along the time. Details about the reaction procedures and apparatus can be found in Silva (2003) and Crossetti et al. (2004). The model that describes the rates of ethylene consumption along the reaction time takes into account the modification of the active species and can be defined as

$$\mathrm{Rp} = \sum_{i=0}^{m} \left[\left(\sum_{n=i}^{m} \mathrm{Kp}_n A_i^n \right) \exp(-k_i t) \right]. \tag{21}$$

Rp is the rate of ethylene consumption along the time t. Kp_n is the polymerization rate constant for active specie n. k_i is the rate constant for transformation of active species i into active species $(i + 1)$. m is the maximum number of active species

in the reaction medium of catalyst. A_i^n are coefficients defined recursively as

$$A_0^0 = 1, \tag{22}$$

$$A_i^n = A_i^{n-1} \frac{k_{n-1}}{k_n - k_i}, \quad i = 1 \ldots n-1, \quad n > 0, \tag{23}$$

$$A_n^n = -\sum_{i=0}^{n-1} A_i^n. \tag{24}$$

The model contains $2m$ kinetic parameters to be estimated: m transformation rate constants $(k_0, \ldots, k_{m-1})$ and m polymerization rate constants $(\mathrm{Kp}_1, \ldots, \mathrm{Kp}_m)$. k_m is assumed to be equal to zero, as the last species is stable. Kp_0 is assumed to be equal to zero, as the initial catalyst species is not active for polymerization. The detailed description of model development will be reported elsewhere.

In order to overcome the high correlation among model parameters, which is typical of exponential models, and to guarantee the obtainment of precise parameter estimates, 15 000 iterations were used with 25 particles. Both c_1 and c_2 were made equal to 1.5 and w was made equal to 0.8. Additionally, the search was reinitiated whenever convergence of particles was detected, which was assumed to occur when the difference between the average value of the objective function of a particular iteration and the best value found for the objective function was smaller than 10^{-5}. Compared to the previous examples, the use of smaller values for c_1, c_2 and w accelerates the convergence of particles to assure a good precision of the parameter, while reinitiation of the PSO after detection of convergence assures the good exploration of the parameter space, which is necessary to provide good description of the confidence regions.

The experimental data used in this problem were obtained by Silva (2003). Initially, the occurrence of two active species ($m = 2$) was assumed, meaning that four parameters had to be estimated. Some preliminary searches were performed to determine the appropriate search intervals for each parameter: for k_0 [0, 100]; for k_1 [0, 50]; for Kp_1 [0, 1000]; and for Kp_2 [0, 10]. Intervals were obtained by assuming that all search intervals were equal to [0, 1000] and performing some PSO iterations. The objective function decreases very fast in the proposed search intervals. The minimum obtained value for the objective function was equal to 1000.53. The estimated parameters were: $k_0 = 6.3810\,\mathrm{min}^{-1}$; $k_1 = 6.381\,\mathrm{min}^{-1}$; $\mathrm{Kp}_1 = 150.6\,\mathrm{mol/min}$; and $\mathrm{Kp}_2 = 6.278\,\mathrm{mol/min}$. The typical computer time required to solve Example 3 was about 10 min, which was much larger than in the previous cases because of the larger number ofparticles, iterations and experimental data.

Fig. 5 shows the quality of the model fit for a particular experiment. It can be observed that the initial moments of the reaction are well adjusted by the model, although the model is unable to describe the observed long-term activity modifications.

The likelihood confidence region for each pair of model parameters is shown in Fig. 6. Except for parameter Kp_2, all confidence intervals are large, since most of the observed transformation occurs during the first minute of reaction and no precision can be obtained for the model parameters. Kp_2 is the only well-estimated parameter because it represents the constant rate value of Fig. 5. It must be noticed that the confidence regions are non-convex and open. Some model parameters do not present upper limits, as discussed in Problem 2.

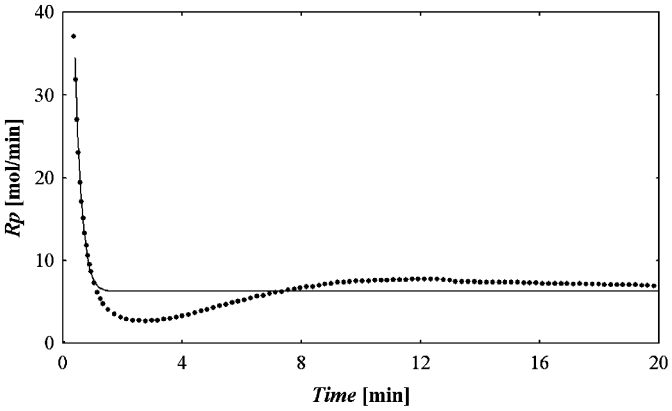

Fig. 5. Experimental data (points) and model predictions when two active species (line) are used in Problem 3.

It becomes clear once more that elliptical approximations of the confidence regions may be completely inadequate and should not be used in many problems, as this may lead to erroneous conclusions about the confidence on model parameters.

One might wonder whether the complex shapes of the confidence regions were related to the poor performance of the model responses. In order to improve the model performance, an additional active species was assumed, leading to three sequential active species transformations and six kinetic parameters. Ten thousand iterations were performed and 50 particles were used. c_1 and c_2 were made equal to 1.5 and w was made equal to 0.8, as in the previous case. The search intervals were defined after some preliminary searches as previously. Search intervals were: for k_0 [0, 100]; for k_1 [0, 50]; for k_2 [0, 1]; for Kp_1 [0, 1000]; for Kp_2 [0, 1]; and for Kp_3 [0, 10]. The minimum value achieved for the objective function was equal 106.39.

Surprisingly, this minimum value was achieved for two distinct sets of parameters, which means that there are two global minima for the proposed objective function. The estimated model parameters are presented in Table 3. The quality of the model fit to experimental data is illustrated in Fig. 7 and is very good in this case.

As shown in Table 3, the parameters k_2, Kp_2 and Kp_3 are the same for two observed minima. Parameters k_0 and k_1, however, interchange their values from one minimum to the other. Parameter Kp_1 assumes very different values in each minimum. According to the model results, the second species is inactive for polymerization. If species 1 decays too fast into species 2, then propagation rates must be high to explain the results. If species 1 is transformed into species 2 slowly, then the rate constants should be smaller. This allows for similar fits when parameters change.

Fig. 8 shows the likelihood confidence regions for some selected pairs of parameters (as the minima for parameters Kp_2 and Kp_3 are the same, the confidence regions for these parameters were omitted for the sake of conciseness). The different

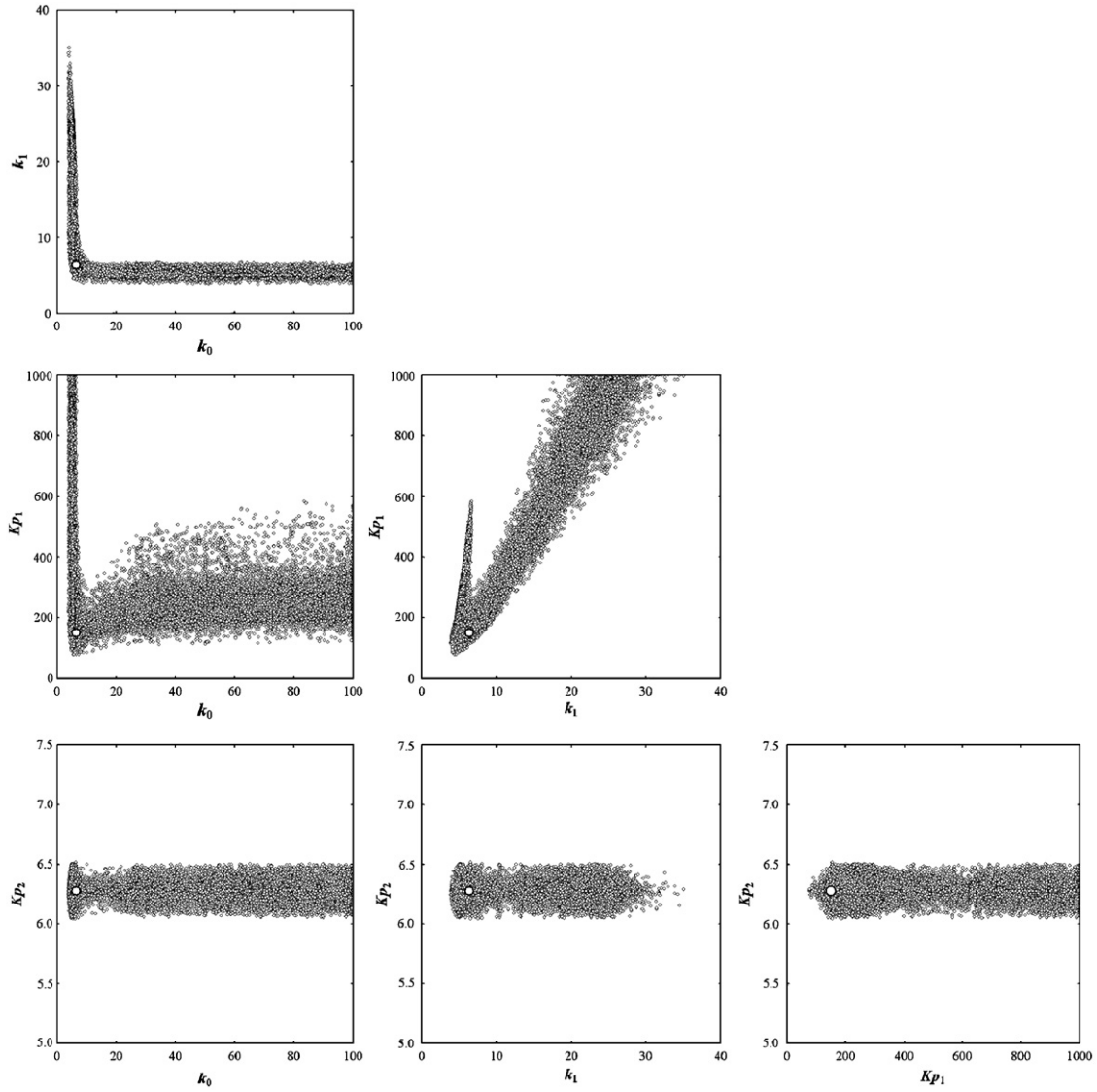

Fig. 6. Likelihood confidence regions for each pair of parameters of model with two active species in Problem 3.

Table 3
Estimated model parameters for Problem 3 with $m = 3$

Parameter	Minimum A	Minimum B
k_0 (min^{-1})	10.31	2.313
k_1 (min^{-1})	2.313	10.30
k_2 (min^{-1})	0.232	0.232
Kp_1 (mol/min)	55.98	249.6
Kp_2 (mol/min)	0.000	0.000
Kp_3 (mol/min)	7.587	7.587

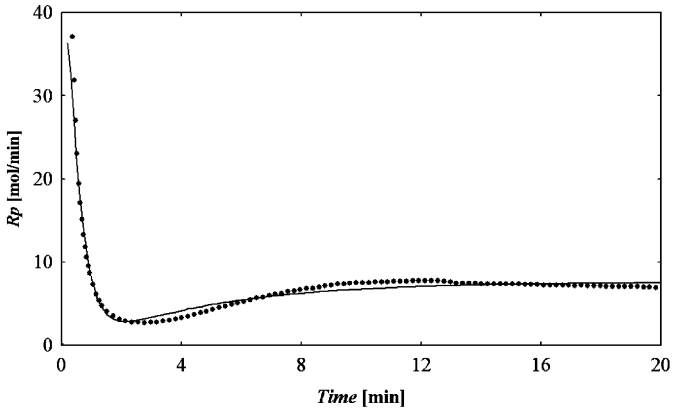

Fig. 7. Experimental data (points) and model predictions with active species (line) in Problem 3 with $m = 3$.

minima lead to disjoint confidence regions. The confidence region presents disconnected parts for the parameters that assume different values at the distinct minima. For minimum A, parameter k_0 does not present an upper bound. For minimum B, the parameter Kp_1 does not present an upper bound. It is also interesting to observe that the correlation between parameters Kp_1 and k_1 for minimum A is very low; however for minimum B this correlation is practically 1. This poses a very difficult numerical problem for minimization, since the point of minimum is located in a very narrow region of the parameter space. In spite of that, the PSO method is able to find the op-

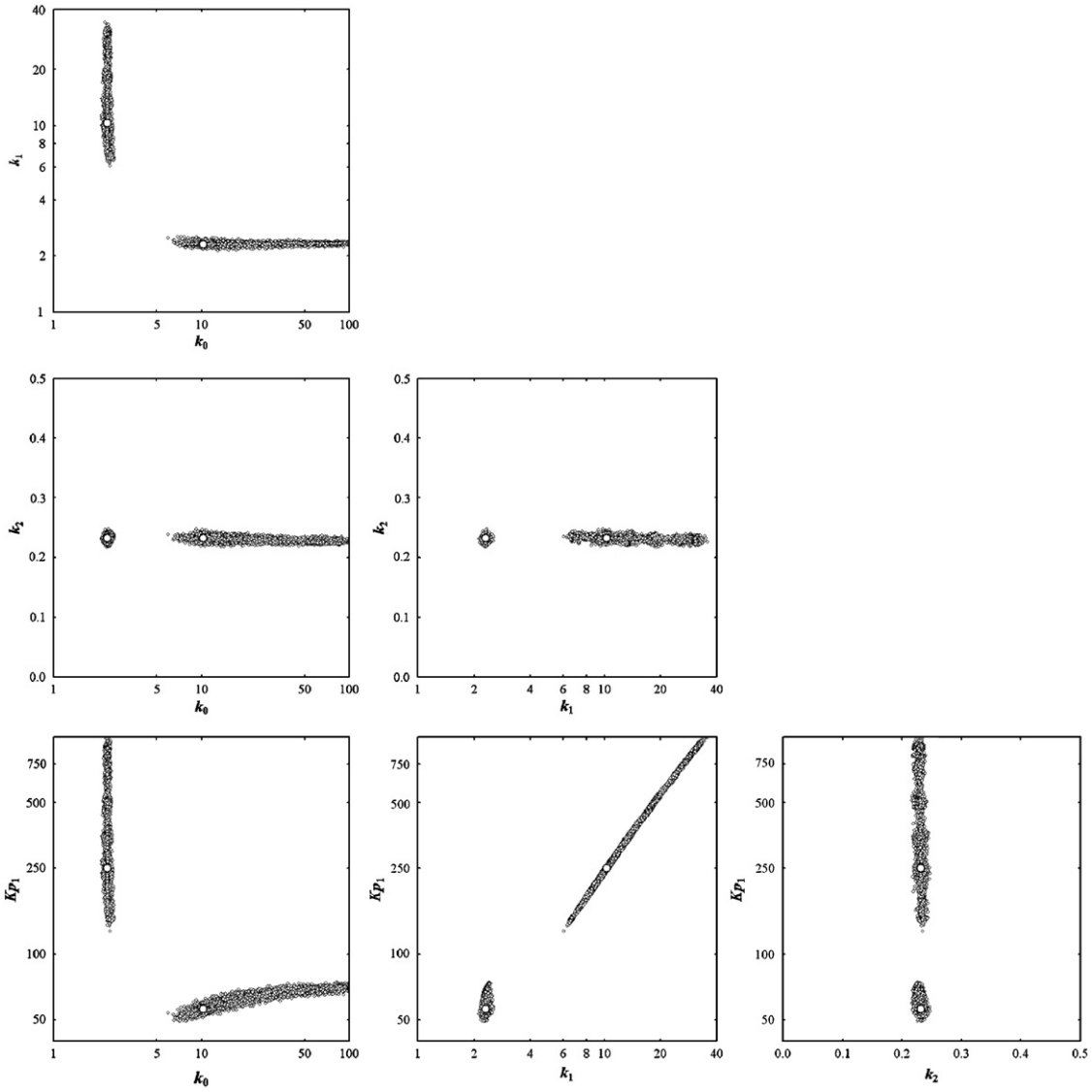

Fig. 8. Likelihood confidence regions for some selected pairs of model parameters in Problem 3 with $m = 3$.

tima. It must be noticed that derivative-based and direct search methods failed repeatedly to find parameter estimates, for the already described reasons.

Based on the obtained results, it is not possible to determine which minimum is the correct one, since both minima lead to similar model performances. This can only be performed through detailed analysis of the parameter values, of their physical meanings and comparison with other experiments. However, this is beyond the scope of this text.

Based on the previous results, it can be said once more that the elliptical approximation of the confidence region of the parameter estimates may be inappropriate and that the PSO is able to characterize both the parameter estimates and their confidence regions appropriately.

5. Conclusions

Combination of the particle swarm optimization (PSO) procedure and the likelihood confidence region method allowed for proper analysis of parameter estimates without introduction of unnecessary simplifications that may lead to erroneous conclusions when nonlinear models are considered. It was shown that the elliptical confidence region method may lead to inappropriate approximations of the confidence region and should be used with care. It was shown that the shape of the confidence regions of parameter estimates can assume shapes that may be very different from the elliptical shape. Besides, confidence regions can be unbounded, non-convex and composed of unconnected parts.

Elliptical approximations of confidence regions of parameter estimates are simpler to use, since only the point estimate and the covariance matrix are necessary to describe all the statistics of the estimated model parameters. However, in general, the confidence regions are not elliptical, the point estimates are not placed at the center of the confidence region and the probability distribution of model parameters is not normal. The likelihood confidence region obtained with the PSO procedure may reveal how the elliptical approximation

is not adequate to describe the confidence region of model parameters.

When the PSO method is used for estimation of model parameters, the approach presented here does not require any additional calculations to be performed. It is only necessary to select the points evaluated by the PSO that satisfy the likelihood confidence region constraint. Also, the computation time required is not a strong limitation, especially nowadays when the computer speed is constantly increasing. It is important to emphasize that this approach can be successfully applied with any other heuristic optimization method, such as the genetic algorithm and the simulated annealing.

Finally, since the PSO procedure requires neither derivatives nor initial parameter guesses, it can solve problems where parameter correlations are high, sensitivity of the objective function to model parameters is low and the objective function is discontinuous. As shown in this work, it can also provide likelihood confidence regions of model parameters very easily. Therefore, the use of PSO method in parameter estimation problems should be encouraged.

References

Bard, Y., 1970. Comparison of gradient methods for the solution of nonlinear parameter estimation problems. SIAM Journal on Numerical Analysis 7 (1), 157–186.

Bard, Y., 1974. Nonlinear Parameter Estimation. Academic Press, New York.

Bard, Y., Lapidus, L., 1968. Kinetic analysis by digital parameter estimation. Catalysis Reviews 2 (1), 67–112.

Bates, D.M., Watts, D.G., 1988. Nonlinear Regression Analysis and its Applications. Wiley, New York.

Beale, E.M.L., 1960. Confidence-regions in non-linear estimation. Journal of the Royal Statistical Society B-Statistical Methodology 22 (1), 41–88.

Cerqueira, H.S., Rawet, R., Pinto, J.C., 1999. The influence of experimental errors during laboratory evaluation of FCC catalysts. Applied Catalysis A-General 181 (1), 209–220.

Clerc, M., Kennedy, J., 2002. The particle swarm—explosion, stability, and convergence in a multidimensional complex space. IEEE Transactions on Evolutionary Computation 6 (1), 58–73.

Costa, A.L.H., Silva, F.P.T., Pessoa, F.L.P., 2000. Parameter estimation of thermodynamic models for high-pressure systems employing a stochastic method of global optimization. Brazilian Journal of Chemical Engineering 17 (3), 349–354.

Costa Jr., E.F., Lage, P.L.C., Biscaia Jr., E.C., 2003. On the numerical solution and optimization of styrene polymerization in tubular reactors. Computers and Chemical Engineering 27 (11), 1591–1604.

Crossetti, G.L., Dias, M.L., Queiroz, B.T., Silva, L.P., Ziglio, C.M., Bomfim, J.A.S., Filgueiras, C.A.L., 2004. Ethylene polymerization with imine and phosphine nickel complexes containing isothiocyanate. Applied Organometallic Chemistry 18 (7), 331–336.

Donaldson, J.R., Schnabel, R.B., 1987. Computational experience with confidence-regions and confidence-intervals for nonlinear least-squares. Technometrics 29 (1), 67–82.

Draper, N.R., Guttman, I., 1995. Confidence intervals versus regions. Statistician 44 (3), 399–403.

Draper, N.R., Smith, H., 1998. Applied Regression Analysis. Wiley, New York.

Edgar, T.F., Himmelblau, D.M., 1988. Optimization of Chemical Processes. McGraw-Hill, New York.

Eftaxias, A., Font, J., Fortuny, A., Fabregat, A., Stüber, F., 2002. Nonlinear kinetic parameter estimation using simulated annealing. Computers and Chemical Engineering 26 (12), 1725–1733.

Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization and Machine Learning. Addison Wesley Longman Inc., Boston.

Halperin, M., 1963. Confidence interval estimation in non-linear regression. Journal of the Royal Statistical Society B-Statistical Methodology 25 (2), 330–333.

Hartley, H.O., 1964. Exact confidence regions for the parameters in non-linear regression laws. Biometrika 51 (3–4), 347–353.

Hibbert, D.B., 1993. Genetic algorithms in chemistry. Chemometrics and Intelligent Laboratory Systems 19 (3), 277–293.

Kennedy, J., Eberhart, R., 1995. Particle swarm optimization. In: Proceedings of the IEEE International Conference on Neural Networks, Perth, Australia, vol. 4, pp. 1942–1948.

Kirkpatrick, S., Gelatt Jr., C.D., Vecchi, M.P., 1983. Optimization by simulated annealing. Science 220 (4598), 671–680.

Klepper, O., Hendrix, E.M.T., 1994. A comparison of algorithms for global characterization of confidence regions for nonlinear models. Environmental Toxicology and Chemistry 13 (12), 1887–1899.

Larentis, A.L., Bentes Jr., A.M.P., Resende, N.S., Salim, V.M.M., Pinto, J.C., 2003. Analysis of experimental errors in catalytic tests for production of synthesis gas. Applied Catalysis A-General 242 (2), 365–379.

Marseguerra, M., Zio, E., Podofillini, L., 2003. Model parameters estimation and sensitivity by genetic algorithms. Annals of Nuclear Energy 30 (14), 1437–1456.

Ourique, C.O., Biscaia Jr., E.C., Pinto, J.C., 2002. The use of particle swarm optimization for dynamical analysis in chemical processes. Computers and Chemical Engineering 26 (12), 1783–1793.

Park, T.Y., Froment, G.F., 1998. A hybrid genetic algorithm for the estimation of parameters in detailed kinetic models. Computers and Chemical Engineering 22 (Suppl.), S103–S110.

Parsopoulos, K.E., Vrahatis, M.N., 2002. Recent approaches to global optimization problems through particle swarm optimization. Natural Computing 1 (2–3), 235–306.

Parsopoulos, K.E., Laskari, E.C., Vrahatis, M.N., 2001. Solving L1 norm errors-in variables problems using particle swarm optimizer. In: Hamza, M.H. (Ed.), Artificial Intelligence and Applications. IASTED/ACTA Press, Anahein, CA, USA, pp. 185–190.

Rawet, R., Cerqueira, H.S., Pinto, J.C., 2001. The influence of covariances during laboratory evaluation of FCC catalysts. Applied Catalysis A-General 207 (1–2), 199–209.

Santos, T.J., Pinto, J.C., 1998. Taking variable correlation into consideration during parameter estimation. Brazilian Journal of Chemical Engineering 15 (1), 1–20.

Schwaab, M., 2005. Evaluation of heuristic optimization algorithms in parameter estimation problems.

Shi, Y., Eberhart, R., 1998. A modified particle swarm optimizer. In: Proceedings of the IEEE International Conference on Evolutionary Computation, Anchorage, Alaska, pp. 69–73.

Silva, L.P., 2003. Synthesis and Characterization of polyolefins obtained with nickel diimine complexes containing pseudo-halides.

Trelea, I.C., 2003. The particle swarm optimization algorithm: convergence analysis and parameter selection. Information Processing Letters 85 (6), 317–325.

Van den Berg, F., Engelbrecht, A.P., 2006. A study of particle swarm optimization particle trajectories. Information Sciences 176 (8), 937–971.

Vanrolleghem, P.A., Keesman, K.J., 1996. Identification of biodegradation models under model and data uncertainty. Water Science and Technology 33 (2), 91–105.

Watts, D.G., 1994. Estimating parameters in nonlinear rate equations. Canadian Journal of Chemical Engineering 72 (4), 701–710.

Williams, E.J., 1962. Exact fiducial limits in non-linear estimation. Journal of the Royal Statistical Society B-Statistical Methodology 24 (1), 125–139.

Wongrat, W., Srinophakun, T., Srinophakun, P., 2005. Modified genetic algorithm for nonlinear data reconciliation. Computers and Chemical Engineering 29 (5), 1059–1067.