On the capabilities and limitations of harmony search for parameter estimation in vapor-liquid equilibrium modeling

Adrian Bonilla-Petriciolet
On the capabilities and limitations of harmony search for parameter estimation in vapor–liquid equilibrium modeling

A. Bonilla-Petriciolet

Instituto Tecnológico de Aguascalientes, Aguascalientes 20256, Mexico

ABSTRACT

The modeling of vapor–liquid equilibrium data using local composition models is an interesting and challenging global optimization problem in the context of chemical engineering and applied thermodynamics. Until now, several deterministic and stochastic global optimization strategies have been used for modeling vapor–liquid equilibrium (VLE) data. Stochastic optimization methods may offer several advantages for solving global optimization problems and, until now, some meta-heuristics have been tested for modeling phase equilibrium data. However, these optimization strategies usually show a robust performance but, in some challenging problems, they may fail to locate the global optimum. In particular, Harmony Search (HS) is a direct-search method with attractive characteristics for its use in phase equilibrium modeling and calculations. However, to the best of our knowledge, this stochastic optimization strategy has not been used to perform this type of thermodynamic calculations. This study introduces the HS method for solving the non-linear parameter estimation problem involved in the modeling of VLE data. Specifically, the performance of this meta-heuristic has been tested and analyzed using several sets of binary VLE data with local composition models and both the classical approach of the least squares regression and the error-in-variable formulation. Results of this study are used to identify the capabilities and limitations of HS for VLE data modeling. In summary, HS is a promising meta-heuristic for processing these phase equilibrium data using the classical least square formulation and may offer a better performance than those obtained using current stochastic methods such as Genetic Algorithm or Particle Swarm Optimization. However, the reliability of the traditional HS is poor for VLE parameter estimation using the error-in-variable formulation. Finally, this paper discusses and analyzes alternatives to improve the performance of HS in VLE data modeling especially for the error-in-variable approach. Results indicate that the HS variants called the Improved Harmony Search and the Global-Best Harmony Search offer a better performance for solving EIV parameter estimation problems.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Non-linear parameter estimation is an important numerical procedure in fields of applied sciences and engineering [1–3]. In particular, the thermodynamic modeling of phase equilibrium data plays a major role in the design, development, operation, optimization, and control of processes involved in chemical engineering [4]. A common problem for modeling phase equilibrium is to determine the parameters of a non-linear thermodynamic model used to represent a specific set of experimental information such as vapor–liquid or liquid-liquid equilibrium data [5]. Specifically, the task is to establish the values of model parameters that provide the best fit to measured data using a proper objective function. This objective function can be formulated using either the least squares or the maximum likelihood criterion [4,5]. Usually, the least squares formulation [6] is preferred for data correlation in engineering applications and also is widely applied in phase equilibrium modeling. On the other hand, the error-in-variable approach is useful to perform the simultaneous data reconciliation and parameter estimation for phase equilibrium data processing [4,5]. To solve non-linear parameter estimation problems using both formulations, we have to face the following global optimization problem: for a given objective function \( f: \Omega \rightarrow \mathbb{R}, \Omega \subseteq \mathbb{R}^d \), the point \( \mathbf{u}^* \) is to be found such that \( f(\mathbf{u}^*) = \min_{\mathbf{u} \in \Omega} f(\mathbf{u}) \). Note that this optimization problem may involve constraints depending on the thermodynamic model used.

Formally, the non-linear parameter estimation in vapor–liquid equilibrium (VLE) modeling can be performed either by direct optimization of the objective function or by solving an equivalent system of non-linear equations that is obtained from the stationary conditions of the optimization problem [7,8]. The optimization techniques commonly used for non-linear parameter
estimation in thermodynamic models include the Simplex, Brent, Levenberg–Marquardt and quasi-Newton methods [5]. However, this type of optimization problems is complex and difficult to solve using traditional local optimization methods due to [7–9]: (a) the presence of several local minima for the objective function used as optimization criterion, (b) the objective function may be flat or with discontinuities in some regions of solution domain, and (c) the model parameters may vary over a wide range of values. In fact, results reported by several studies [4,7–16] indicate that the parameter estimation problem has complex non-linear objective functions even using thermodynamic models with a relatively small number of adjustable parameters. In addition, we have to face both large-scale and non-convex optimization problems if the error-in-variable formulation is used. Under these conditions, traditional optimization methods are easy to be trapped in local optima and do not guarantee convergence to the global optimum. Therefore, the development of reliable methods for solving parameter estimation problems in phase equilibrium modeling is still a challenge and a current research topic [14,15]. In this context, it is also convenient to remark that there is a need of better understanding of the performance of available numerical strategies for solving global optimization problems involved in phase equilibrium modeling.

Recently, deterministic and stochastic global optimization methods have been suggested for reliable parameter estimation in VLE modeling using both least squares and error-in-variable formulations [4]. Studies on deterministic optimization have been mainly focused on the application of a branch and bound procedure with convex underestimating functions [7] and interval analysis using an interval-Newton/generic bisection algorithm [8,10,16]. Although these methods are promising, they are model dependent and, may require problem reformulations, or a significant computational time when several decision variables are involved [13].

On the other hand, several studies have shown the potential of stochastic (or meta-heuristic) optimization methods to solve non-linear parameter estimation problems for modeling phase equilibrium data [11–15]. A meta-heuristic method is a high-level algorithm that combines rules (or heuristics) and randomness to imitate natural phenomena and to solve optimization problems. These optimization methods contain global and local search phases for exploration and exploitation of search space for finding the global optimum. The global phase is used to explore and to identify the potentially promising areas where the global optimum could be located, while the objective of the local phase search is speed up the algorithm convergence for finding that optimal solution on the identified promising areas. These methods offer several advantages for solving global optimization problems such as generality, robust performance, few information requirements of the optimization problem to be solved, easy implementation, and a reasonable computational time. To date, different stochastic methods have been studied and tested for parameter estimation using VLE data, and they include: Simulated Annealing, Genetic Algorithm, Differential Evolution, and Particle Swarm Optimization, among others [4,11–15]. These strategies usually show a robust performance but, in some challenging problems, they may fail to locate the global optimum [4,13]. Note that the failure to find the globally optimal parameters of a thermodynamic model may cause significant errors and uncertainties in process design [8]. Thus, based on the fact that the correct processing of experimental data using reliable numerical strategies is crucial for process system engineering, alternative optimization strategies should be studied to identify a better approach for solving parameter estimation problems in the context of applied Thermodynamics.

This manuscript discusses the application of Harmony Search (HS) for solving the parameter estimation using VLE data. HS is a meta-heuristic that can be easily adapted for several engineering optimization problems [17,18]. In particular, HS is a stochastic method based on the analogy with music improvisation process where music players improve the pitches of their instruments to obtain a better harmony. Literature indicates that HS is simpler both in formulation and in computer implementation than other meta-heuristics [18]. To date, HS has been successfully applied to solve global optimization problems in several engineering fields including few applications in chemical engineering [19]. With respect to the phase equilibrium modeling, previous studies suggest that this stochastic method is reliable for solving challenging global optimization problems [20–22]. In particular, the goal of this study is to gain insights on the numerical performance of HS for VLE data modeling. Therefore, the performance, capabilities and limitations of HS for non-linear parameter estimation using VLE data are discussed in this paper. Results of this study are useful to understand the strong and weak points of HS-based methods for parameter estimation in phase equilibrium modeling. Finally, alternatives to improve the performance of HS method in VLE parameter estimation have been analyzed in this manuscript given a special emphasis to solve the error-in-variable formulation.

2. Formulation of the non-linear parameter estimation problem for vapor–liquid equilibrium modeling

Consider the following model \( y = g(u(k), \theta) \) where \( y \in \mathbb{S}^n \) is the system output, \( u \in \mathbb{S}^m \) is the system input and \( \theta = [\theta_1, \ldots, \theta_m] \in \mathbb{S}^m \) are the parameters to be estimated, and suppose the form of model \( g \) is known (e.g., local composition model or EoS). Based on the fact that the structure of the model is known in advance, the parameter estimation problem implies to compare the output of the system and of the estimated model under the same inputs, and to adjust the parameters \( \theta = [\theta_1, \ldots, \theta_m] \) for minimizing a predefined error function (i.e., objective function). The objective functions used for non-linear parameter estimation in VLE data modeling can be derived from either the least squares principle or the maximum likelihood criterion [45,8]. Literature indicates that the least squares principle is the most popular approach for obtaining parameters of thermodynamic models for fitting VLE and other thermodynamic data [23]. On the other hand, the error-in-variable approach is the standard procedure for simultaneous parameter estimation and data reconciliation [4]. So, in this study both approaches have been used for illustrating the performance of HS as global optimization strategy in the modeling of VLE data.

In the least squares (LS) formulation, the objective function is defined as the difference between the experimental data and the calculated values using a specific model. For the case of VLE data (i.e., vapor–liquid equilibrium compositions and pressure at constant temperature, or vapor–liquid equilibrium compositions and temperature at constant pressure), excess Gibbs energy equations are widely employed for phase equilibrium modeling. The objective function commonly used for data fitting of VLE data is based on activity coefficients and is usually defined as:

\[
F_{LS} = \sum_{j=1}^{n_{exp}} \sum_{i=1}^{c} \left( \frac{y_{exp} - y_{calc}}{y_{exp}} \right)^2
\]

(1)

where \( y_{exp} \) and \( y_{calc} \) are respectively the experimental and calculated values for the activity coefficient of component \( i \), \( n_{exp} \) is the number of VLE experimental data, and \( c \) is the number of components in the mixture, respectively. The activity coefficients \( y_{i} \) are calculated by the solution models for the excess Gibbs free energy \( G^E \) and using the relationship:

\[
\ln y_{i} = \left[ \frac{\partial nC^E_i}{\partial n_i} \right]_{T,P,n_{j \neq i}}
\]

(2)
where $n_i$ is the mole number of component $i$, $R$ is the universal gas constant, $T$ is the temperature and $P$ is the pressure, respectively.

Classical thermodynamics establishes that the equilibrium between vapor and liquid phases in a multi-component system implies that $T$, $P$ and the fugacities of each component must be the same in both phases. Herein, it is necessary to remark that the equality of fugacities is a necessary condition for phase equilibrium and a phase stability analysis [24] must be performed to identify if the solution found corresponds to the global optimum of the Gibbs free energy. At low pressure, the fugacity coefficient of pure components nearly cancels each other and Poynting corrections are usually very close to unity. With these assumptions, $\gamma_i^\text{v}^\text{exp}$ can be calculated from VLE data using the following expression:

$$\gamma_i^\text{v}^\text{calc} = \gamma_i^\text{v}^\text{exp} \frac{P_i^0}{P_i^0}$$

where $\gamma_i^\text{v}^\text{exp}$ and $\gamma_i^\text{v}^\text{calc}$ are respectively the experimental mole fractions of component $i$ in liquid and vapor phases at equilibrium, and $P_i^0$ is the vapor pressure of pure component $i$ at the system temperature $T$. In this study, the Wilson [25], NRTL [26] and UNIQUAC [27] thermodynamic models are used to calculate the activity coefficients $\gamma_i^\text{calc}$, and in parameter estimation we seek the minimization of the energy parameters of these models that will minimize Eq. (1). Formally, the following global optimization problem has to be solved:

$$\min_\theta F_{\text{obj}}$$

subject to $g(z_{ij}^\theta, \theta) = 0, \quad i = 1, \ldots, n_{\text{var}}, \quad j = 1, \ldots, n_{\text{exp}}$

The energy parameters of these thermodynamic models are defined as in the DECHEMA [28] and they are given in Table 1. It is important to note that the global minimization of least squares objective function can be done as an unconstrained optimization problem using these thermodynamic models. However, several studies have shown that, even for relatively simple thermodynamic equations such as those given in Table 1, multiple local optima can occur in non-linear parameter estimation for VLE data modeling [8,11–13]. This is because the highly non-linear form of the thermodynamic models makes $F_{\text{obj}}$ potentially non-convex.

On the other hand, if we assume that there are measurement errors in all the variables $z_{ij}$ (which include both independent and response variables) for the experiments of the system to be modeled, the minimization problem to be solved is the error-in-variable (EIV) formulation of the form:

$$F_{\text{EIV}} = \frac{\sum_{j=1}^{n_{\text{exp}}} \left( z_{ij}^\theta - z_{ij} \right)^2}{\sigma_i^2}$$

(5)

Here, $n_{\text{exp}}$ is the number of state variables, $z_{ij}^\theta$ is the unknown “true” value of $i$th state variable in $j$th measurement, and $\sigma_i$ is the standard deviation associated with the measurement of $i$th state variable. Note that the experimental measurements of phase equilibrium data usually include random and systematic errors caused by the variations in the operation and calibration of measurement instruments. Therefore, the error-in-variable approach is considered as a statistical technique to perform the adjustment of measured experimental values (i.e., the vector of state variables $z_{ij}$) with the aim of that the corrected measurements (i.e., the reconciled values, which are represented by the vector of unknown true values $z_{ij}^\theta$) are consistent with the corresponding equations that describe the phase equilibrium condition. So, the decision variables of EIV problem are the set of $z_{ij}^\theta$ and the model parameters $\theta$. Formally, a constrained global optimization problem has to be solved, which is given by [4]:

$$\min_\theta F_{\text{EIV}}$$

subject to $g(z_{ij}^\theta, \theta) = 0, \quad i = 1, \ldots, n_{\text{var}}, \quad j = 1, \ldots, n_{\text{exp}}$

(6)

where $g$ is a vector of $np$ model functions. In the EIV formulation, there is a substantial increase in the dimensionality of the optimization problem, which depends on the number of experiments. If the model functions $g$ are explicit in $z_{ij}$, then they can be eliminated by direct substitution of $z_{ij}$ in the objective function, and the above
optimization can be solved as an unconstrained problem. For the case of VLE data, the state variables $y_j$ are $x, y, P$ and $T$ with standard deviations ($\sigma_x, \sigma_y, \sigma_P$ and $\sigma_T$). Therefore, the common objective function for VLE data modeling using the EIV approach is defined as [4,16]:

$$
F_{EIV} = \sum_{j=1}^{ne} \sum_{i=1}^{c} \left[ \frac{(x_j^i - x_j) y_j^i - y_j)^2}{\sigma_x^2} + \frac{(y_j^i - y_j)^2}{\sigma_y^2} \right]
$$

which is optimized with respect to $npar+c \times ne$ decision variables. At low pressure, VLE can be described by the next equations:

$$
P = \sum_{i=1}^{c} y_i x_i P_i^0
$$

$$
y_i = \sum_{j=1}^{ne} y_j x_j P_j^0
$$

where $y_i$ is the calculated activity coefficient by the chosen thermodynamic model and $c$ is the number of components involved in the VLE system. Based on this fact, the EIV problem can be formulated as an unconstrained optimization problem using Eqs. (8) and (9) to eliminate $P$ and $y_j$ in the EIV objective function (Eq. 7). For this unconstrained problem, the independent variables are the set of $\bar{z} = (\bar{x}_j, \bar{T})$ for all measurements, while the decision variables are

$$
\theta = (\bar{t}_1, \ldots, \bar{t}_{npar})
$$

and the set of $\bar{z} = (\bar{x}_j, \bar{T})$.

Previous studies [4,7,8,12–16] have shown that the objective functions used for both LS and EIV formulations in VLE modeling are strongly non-linear, potentially non-convex with several local minima within the specified bounds. Therefore, the performance of HS for the global minimization of Eqs. (1) and (7) has been studied.

3. Harmony search for solving global optimization problems

Harmony Search is a meta-heuristic algorithm, which has been conceptualized using the musical process of searching for a perfect state of harmony, such as during jazz improvisation [17,18]. This meta-heuristic is based on the analogy with music improvisation process where music players improvise the pitches of their instruments to obtain a better harmony. In the optimization context, each musician is replaced with a decision variable, and the possible notes in the musical instruments correspond to the possible values for the decision variables. The harmony in music is analogous to the optimization solution vector, and the musician’s improvisations are analogous to local and global search schemes in optimization techniques. Note that the musical performances seek to find pleasing harmony (i.e., a perfect state) as determined by an esthetic standard, just as the optimization process seeks to find a global solution as determined by an objective function [18]. Thus, this stochastic optimization method follows the underlying principles of the musician improvisation of the harmony. Specifically, when musicians improvise they may perform the following steps: playing an existing score from memory, performing variations on an existing piece, or creating an entirely new composition. In the optimization context, HS combines heuristic rules and randomness to imitate this process. A comprehensive explanation of HS is provided by Geem et al. [17] and a flow chart describing its principal stages is given in Fig. 1.

In summary, the traditional HS involves the following parameters: the harmony memory size (HMS), the harmony memory considering rate (HMCR), the pitch adjusting rate (PAR), the bandwidth or step size for variable perturbation during pitch adjustment (bw), and the number of improvisations (NI). The harmony memory is a memory location where a set of solution vectors for decision variables is stored. The parameters HMCR and PAR are used to improve the solution vectors stored in the harmony memory, which contains promising solutions of the global optimum, and to increase the diversity of the search process [17,18]. In HS, a new harmony (i.e., a new solution vector) is generated using these parameters and the following procedures: (a) memory consideration, (b) pitch adjustment, and (c) random selection. To illustrate the concepts of HS, consider the following unconstrained global optimization problem: minimize $f(u)$ such that $lb_j \leq u_j \leq ub_j$, where $u$ is a solution vector of $n_{opt}$ continuous decision variables with lower ($lb_j$) and upper ($ub_j$) bounds for each decision variable (i.e., $u_j$). Recall that the non-linear parameter estimation for modeling VLE data, using both LS and EIV approaches, can be defined using this formulation. So, HS performs the following steps [17,29]:

1. Initialize a harmony memory. First, the parameters of HS (e.g., HMS, HMCR, PAR, bw) are defined by the user and the harmony memory is initialized. This harmony memory (i.e., $u_j$ for $i = 1, \ldots, n_{opt}$ and $j = 1, \ldots, n_{HMS}$) preserves the history of optimization sequence and is useful to identify promising areas for global optimization because good harmonies can be considered as elements of new solution vectors. In fact, this harmony memory stores important information of the optimization problem and is used to guide the search for the global optimum. Usually, the initial values of harmony memory are generated from a uniform distribution in the bounds of decision variables: $u_j = lb_j + rand(ub_j - lb_j)$ where rand $\in (0, 1)$ is a random number.

2. Improve a new harmony. A new harmony vector ($s_i$) is obtained using the following stages: memory consideration, pitch adjustment and random selection. These stages can be summarized using the following pseudo-code [29]:

Fig. 1. Flowchart of Harmony Search for global optimization.
for \( i = 1 \) to \( n_{\text{iter}} \) do
  if \( \text{rand} \in (0, 1) \geq \text{HMCR} \) then perform memory consideration
    begin
      \( v_i = u_i \) where \( j = (1, \ldots, \text{HMS}) \)
    end
  if \( \text{rand} \in (0, 1) \leq \text{PAR} \) then perform pitch adjustment
    begin
      \( v_i = n_i + (v_i - l_i) \)\( \times \)\( \text{rand} \), where \( b_i \) is the bandwidth (i.e., step size)
    end
  else perform random selection
    \( v_i = u_i + (u_i - l_i) \)
  end
end

These operators are used to perform both diversification and intensification stages in HS. The diversification is controlled by the pitch adjustment and random selection operators, while memory consideration is generally associated with the intensification. In particular, HMCR is used to determine the degree of contribution of harmony memory (i.e., promising solutions) during random search. On the other hand, PAR and bw are used to control the additional random perturbation of decision variables when memory consideration is applied. Note that the pitch adjustment is used to generate a slightly different solution during optimization sequence and prevents stagnation of HS. This operator improves the diversity of solutions stored in harmony memory and favors the exploration capabilities of HS method. The pitch adjustment plays a similar role to the mutation operator in Genetic Algorithm. On the other hand, the random selection is useful to explore different regions of objective function ensuring that the new solution vectors are close from existing good solutions and also contributes to increase the diversity of solution vectors for finding the global optimum. It is convenient to point out that the proper combination of these operators is important to favor the performance of HS in global optimization. The generation of a new harmony (i.e., new solution vector) is called improvisation. A boundary violation check must be implemented, principally during pitch adjustment, to verify the feasibility of \( v_i \): if \( v_i \) is infeasible, a new harmony is randomly generated inside lower and upper bounds of decision variables, where \( v_i = l_i + \text{rand} \times (u_i - l_i) \).

3. Update harmony memory. In this stage, a new harmony (\( v \)) replaces the worst harmony in harmony memory only if its objective function value is lower than that of the worst harmony. The decision vectors stored in harmony memory are useful to exploit the history and experience of the search process, being the intensification mechanism of HS method.

4. Check the stopping condition. This iterative procedure (i.e., steps 2 and 3 of this pseudo-code) is repeated until satisfying a proper convergence criterion. Similar to other stochastic methods, the choice of stopping condition can affect significantly the performance of HS. In the literature, the stopping criterions commonly used in HS are based on the number of function evaluations (NFE) or improvisations (NI). Alternatively, a maximum number of successive improvisations (\( S_{\text{NIMAX}} \)) without improvement in the best function value can be used as convergence condition for HS [20]. The best solution found by HS, which is stored in harmony memory, is expected to be a near global optimum solution.

In summary, HS is able to generate a diverse range of solutions using the stages of memory consideration, pitch adjustment and random selection and can explore the search space effectively avoiding to be trapped in a local optimum and can intensify the search around the promising areas for finding the global optimum solution. The search for finding the global optimum in HS is guided by both the harmony memory considering rate and the pitch adjusting rate. In particular, the information stored in harmony memory is used to generate a new solution based on both HMCR and PAR. Therefore, these operators are the driving force of HS for finding the global optimum in non-convex optimization problems. HS incorporates several characteristics of current meta-heuristics used for global optimization. In particular, HS uses the harmony memory for preserving the history of vector of decision variables similar to the Tabu Search, while several vectors of decision variables (i.e., possible solutions of the optimization problem) are handled by HS in a manner similar to Genetic Algorithm or Differential Evolution [30]. On the other hand, the major difference of HS with respect to Genetic Algorithm or Differential Evolution is that HS generates a new vector of decision variables using all harmonies stored in harmony memory, while these methods use two vectors from population to generate the new solution vector.

Usually, a local optimization technique can be used at the end of global search for efficiently improving the accuracy of the best solution obtained by HS. It is convenient to remark that stochastic optimization methods may require a significant computational effort to improve the accuracy of global solution because they explore the search space of decision variables by creating movements using heuristics and randomness instead of determining an optimization trajectory defined by the gradient of the objective function. Similar to other meta-heuristic methods, HS does not guarantee an exact solution for the optimization problem due to its stochastic nature. Thus, this additional intensification step is required for rapid convergence in the final stage of HS and for improving the solution quality obtained by this method. This local optimization aims at locating a refined solution of the optimization problem as fast as possible. This approach for improving solution quality has been used in other studies related to global optimization using stochastic methods, including the modeling of phase equilibrium, e.g., [13,15]. For local optimization, a quasi-Newton method has been used in all calculations performed in this study.

As indicated, the parameters HMS, HMCR, PAR and bw are important to determine the performance (i.e., reliability and efficiency) of HS in global optimization. Therefore, the fine-tuning of these parameters is very crucial for solving global optimization problems. Traditionally, fixed values for HS parameters, which cannot be changed during new improvisations, are used in global optimization. This approach has proven to be competitive with respect to other optimization methods for solving engineering optimization problems [17–19]. So, this standard version of HS algorithm has been used in this study and a Fortran code has been developed. This code is available to interested readers upon request to the corresponding author.

Until now, HS has been successfully applied to perform water network design, vehicle routing, soil stability analysis, heat exchanger design, and transportation energy modeling [18,19]. In the context of applied thermodynamics, there are few studies concerning the application of this stochastic method [20–22] and, to the best of our knowledge, the performance of HS in non-linear parameter estimation for VLE modeling has not yet been reported. Thus, this study introduces the first application of HS for solving this challenging global optimization problem.

4. Results and discussion

4.1. Description of VLE parameter estimation problems

The performance of HS has been tested using several sets of binary vapor–liquid equilibrium data. Tables 2 and 3 provide the details (i.e., conditions of experimental data, thermodynamic models, and global optimum) of all VLE examples used in this study. All the experimental data are taken from DEHEMA [28] collection and the vapor pressure is calculated by Antoine equation using the parameters reported in Gau et al. [8]. These VLE problems have been studied using both deterministic and stochastic optimization
Table 2
Details of vapor–liquid equilibrium data modeling problems selected to test the performance of Harmony Search method using least squares formulation.

<table>
<thead>
<tr>
<th>No.</th>
<th>System</th>
<th>Data</th>
<th>Models</th>
<th>Decision variables</th>
<th>Global optimum</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tert butanol + 1 butanol</td>
<td>P = 100 mm Hg</td>
<td>Wilson and Ideal gas model</td>
<td>( n_{\text{opt}} = 2 \quad \theta_1, \theta_2 \in (\pm 5000, 320,000) )</td>
<td>( F_{\text{opt}} = 0.01026 \quad \theta_1 = 567.59 \quad \theta_2 = 745.33 )</td>
<td>2 minima</td>
</tr>
<tr>
<td>2</td>
<td>No.</td>
<td>n_{\text{opt}} = 9</td>
<td></td>
<td></td>
<td></td>
<td>Gau et al. [8], Alvarez et al. [12], Bonilla-Petriciolet et al. [13]</td>
</tr>
<tr>
<td>3</td>
<td>P = 700 mm Hg</td>
<td>n_{\text{opt}} = 9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>P = 500 mm Hg</td>
<td>n_{\text{opt}} = 9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>No.</td>
<td>n_{\text{opt}} = 18</td>
<td>Wilson and Ideal gas model</td>
<td>( n_{\text{opt}} = 2 \quad \theta_1, \theta_2 \in (\pm 5000, 20,000) )</td>
<td>( F_{\text{opt}} = 1.092134 \quad \theta_1 = 5072.361 \quad \theta_2 = 1921.62 )</td>
<td>2 minima</td>
</tr>
<tr>
<td>6</td>
<td>Water + 1.2-ethanediol</td>
<td>P = 430 mm Hg</td>
<td>UNIQUAC and Ideal gas model</td>
<td>( n_{\text{opt}} = 2 \quad \theta_1, \theta_2 \in (\pm 5000, 20,000) )</td>
<td>( F_{\text{opt}} = 1.048547 \quad \theta_1 = 1111.84 \quad \theta_2 = 3617.65 )</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Benzene + hexafluorobenzene</td>
<td>T = 50 °C</td>
<td>Wilson and Ideal gas model</td>
<td>( n_{\text{opt}} = 2 \quad \theta_1, \theta_2 \in (\pm 5000, 320,000) )</td>
<td>( F_{\text{opt}} = 0.008935 \quad \theta_1 = -424.08 \quad \theta_2 = 983.06 )</td>
<td>2 minima</td>
</tr>
<tr>
<td>8</td>
<td>P = 300 mm Hg</td>
<td>n_{\text{opt}} = 17</td>
<td></td>
<td></td>
<td></td>
<td>Gau et al. [8], Bonilla-Petriciolet et al. [13]</td>
</tr>
<tr>
<td>9</td>
<td>P = 760 mm Hg</td>
<td>n_{\text{opt}} = 29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>T = 30 °C</td>
<td>n_{\text{opt}} = 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

methods, e.g. [7,8,12,13,15,16]. According to these studies, there are at least two local minima (including the global minimum) in the specified interval for decision variables of all VLE problems. The dimension of selected VLE problems ranged from 2 to 36 decision variables. Note that selected VLE examples are particularly challenging for stochastic global optimization algorithms [13]. Therefore, these problems are considered benchmarks and their features are suitable to test and compare the performance and capabilities of HS method. Finally, a phase stability analysis has been performed in all VLE systems via the global minimization of the tangent plane distance function [24] to verify the solutions obtained from parameter estimation. Results showed that all calculated VLE compositions are stable.

In all calculations, the LS objective function has been optimized with respect to the parameters of thermodynamic models inside the interval: \( \theta_1, \theta_2 \in (\pm 5000, 320,000) \) for Wilson model, \( \theta_1, \theta_2 \in (\pm 5000, 20,000) \) for UNIQUAC model, and \( \theta_1, \theta_2 \in (\pm 2000, 5000) \) and \( \alpha_{12} \in (0.01, 10.0) \) for NRTL model where \( \theta_1 \) and \( \theta_2 \) are the corresponding energy parameters. Bounds of thermodynamic model parameters have been defined according to results reported in the literature and taking into account physical considerations [8,11,13,15]. In general, the typical values of adjustable parameters of local composition models (i.e., Wilson, UNIQUAC, NRTL) for VLE modeling fall inside the intervals used in this manuscript. For EIV formulation, the initial intervals on the independent state variables \( z = (x_{ij}, T) \) are set using plus and minus three standard deviations.

Table 3
Details of vapor–liquid equilibrium data modeling problems selected to test the performance of Harmony Search method using error-in-variable formulation.

<table>
<thead>
<tr>
<th>No.</th>
<th>System</th>
<th>Data</th>
<th>Models</th>
<th>Decision variables</th>
<th>Global optimum</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Benzene + hexafluorobenzene</td>
<td>P = 500 mm Hg</td>
<td>Wilson and Ideal gas model</td>
<td>( n_{\text{opt}} = 34 \quad \theta_1, \theta_2 \in (\pm 10,000, 200,000) )</td>
<td>( F_{\text{opt}} = 19.998720 \quad \theta_1 = -429.85 \quad \theta_2 = 1023.32 )</td>
<td>2 minima</td>
</tr>
<tr>
<td>12</td>
<td>T = 30 °C</td>
<td>n_{\text{opt}} = 16</td>
<td></td>
<td></td>
<td></td>
<td>Bonilla-Petriciolet et al. [13], Zhang et al. [15], Gau and Stadtherr [16]</td>
</tr>
<tr>
<td>13</td>
<td>T = 40 °C</td>
<td>n_{\text{opt}} = 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>P = 300 mm Hg</td>
<td>n_{\text{opt}} = 17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>T = 50 °C</td>
<td>n_{\text{opt}} = 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>T = 60 °C</td>
<td>n_{\text{opt}} = 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ory is optimum tuning convergence. As formed. Adjusted HMCR, success is the able to the global optimum with precision ε within the given numerical effort:

$$SR = 100 \frac{N_{\text{Success}}}{N_{\text{Trials}}}$$

where $N_{\text{Success}}$ is the number of successful runs and $N_{\text{Trials}}$ (=100) is the total number of test runs, respectively. This performance metric is used in test problems with a known global optimum and the value ε states the maximum allowed difference in fitness value compared to the known globally optimal value. In this study, a numerical trial is considered successful if the global optimum is obtained with an absolute error (ε) of $10^{-5}$ or less in the objective function value: $|F_{\text{glob}, \text{true}} - F_{\text{glob}, \text{test}}| \leq 10^{-5}$ where $F_{\text{glob}, \text{true}}$ is the known global optimum and $F_{\text{glob}, \text{test}}$ is the solution found by HS, respectively. So, SR is useful to calculate a probability for finding the global optimum using HS as specified numerical conditions (i.e., values of HMs, PAR, HMCR, and NFE). Note that random initial values of harmony memory and different random seeds (which are used for generating the sequence of random numbers needed during execution of HS) have been employed in all calculations for obtaining a general and reliable perspective of the algorithm performance. Additionally, the computational efficiency of HS has been measured in terms of average number of function evaluations NFE. According to the literature [13,15], it is more convenient to use the number of function evaluations as the performance metric of algorithmic efficiency because this metric is independent of the algorithm implementation as well as the computer system where the numerical experiments are performed. Therefore, the average NFE is calculated using successful trials only to avoid biased results.

### 4.2. Parameter tuning of HS method

In this study, the parameter tuning of HS has been performed using examples Nos. 1 and 11 and the algorithm parameters are tuned one at a time while keeping others fixed at predefined nominal values without using the local optimization. Overall, the reliability of HS decreases as HMCR increases, while the performance of HS is better if PAR also increases. Results of parameter tuning suggest that PAR has a more impact on the performance of HS for finding the global optimum. In summary, the suggested parameters values of HS are: HMS = 10$n_{\text{opt}}$ where $n_{\text{opt}}$ is the number of decision variables (i.e., the model parameters to be adjusted and the “true” values of state variables if EIV formulation is used), HMCR = 0.25, PAR = 0.75 and $b_{\text{wi}} = ub_{\text{i}} - lb_{\text{i}}$, respectively. Calculations indicate that these parameter values are a reasonable compromise between numerical effort and reliability of HS method for VLE data modeling. It is convenient to remark that the stopping conditions NI and SNImax also contribute to the trade-off between efficiency and reliability of HS. Therefore, the reliability of HS has been tested by examining different levels of algorithm efficiency, which are obtained by changing the values of both stopping conditions.

### 4.3. Performance of traditional HS method

In particular, the reliability of the optimization strategy for finding the global optimum is a fundamental attribute to perform a proper VLE data modeling. As stated in introduction, recent studies have shown that using the locally optimal parameters (i.e., local optimum of the objective function) for VLE modeling may result in incorrect predictions of the azotrope states with local composition models and in qualitative discrepancies of the phase behavior such as prediction of spurious phase split and modeling of homogenous azotropes as heterogeneous [8,31]. Therefore, it is important to characterize the capability of HS method for finding the global optimum using both LS and EIV formulations. Based on this fact, the reliability of HS is summarized through the global success rate (GSR). Specifically, GSR is defined as the percent of successes (i.e., successful trials in finding the global optimum) out of all calculations performed on the collection of VLE problems tested:

$$\text{GSR} = \frac{1}{N_{\text{Prob}}} \sum_{i=1}^{N_{\text{Prob}}} \frac{N_{\text{Success}}}{SR_i}$$

where $SR_i$ is the success rate of HS in each problem tested. Note that GSR has been determined using the stochastic method with the local optimization strategy.

With illustrative purposes, the results of solving the parameter estimation problems for different values of NI (i.e., NFE) and SNImax and using LS formulation are shown in Fig. 2, while Fig. 3 shows the predicted values of $\gamma_i$ obtained with the Wilson model and the global optimal parameters for selected examples. Calculations indicate that increasing the value of both stopping conditions (i.e., NI or SNImax) improves the performance of HS for finding the global optimum in VLE data modeling. However, the reliability of HS is better using NI as convergence condition and it is clear that SNImax is not proper for determining the convergence of HS in this thermodynamic application. In fact, HS shows a good performance and a high reliability (GSR >90%) for the global minimization of the objective functions involved in VLE problems using LS formulation, if a proper number of improvisations is used (see Fig. 2a and Table 4). However, HS showed a poor performance in parameter estimation using EIV formulation (see Fig. 4). Overall, GSR of HS in EIV problems is <5% independent of the used stopping condition: NI or SNImax. It is well recognized that large-scale problems (i.e., problems with several decision variables) are challenging for optimization algorithms and the performance of optimization methods may degrade under these conditions. For example, Fig. 5 shows the convergence profile of HS for selected EIV problems with and without the local optimization. As expected, HS often find a lot of improving function values at early
iterations but it is trapped at local optima after certain number of function evaluations. In these examples, the diversification mechanisms of traditional HS fail during search process and consequently a premature convergence problem appears in EIV examples. In particular, the local minima of VLE problems are stronger attractors for the stochastic method and cause its poor performance especially in error-in-variable problems. Note that the application of local optimization method improves significantly the performance of HS in both LS and EIV problems. Unfortunately, its use has less impact for increasing the reliability of HS in EIV formulation. For EIV parameter estimation problems, it is clear that HS failed frequently to identify the global optimum with and without the local optimization method. Herein, it is convenient to remark that some calculations have been performed using relatively narrower bounds for decision variables and results indicate that the performance of HS does not significantly improve, especially in EIV problems.

As stated in previous studies [13,15], stochastic methods may present a slow convergence and require several function evaluations to escape from the local minimum region reached in the initial iterations. In particular, the maximum number of function evaluations to find the global optimum cannot be judged a priori for an arbitrary function due to the stochastic nature of this type of optimization methods. In this study, HS can be trapped by the local minima of the objective functions of selected VLE problems especially at early NFE and in EIV problems. This convergence failure is because the minima found for VLE parameter estimation problems is located in a relatively narrow valley in the search space of the decision variables, and also they are comparable (i.e., they have a very similar value of the objective function) in some problems. In addition, there is an increase of algorithm’s complexity when the problem’s dimensionality increases in EIV formulation. In summary, the parameter estimation problems tested in this study imply challenging objective functions [13,15] and they are useful to illustrate the difficulty of locating the globally optimal parameters for VLE data modeling even using thermodynamic models with few adjustable parameters such as Wilson, NRTL and UNIQUAC models.

Fig. 2. Global success rate (GSR) versus (a) NFE/ntot and (b) SNImax/ntot of Harmony Search with local optimization for VLE data modeling using least squares approach.

Fig. 3. Experimental and predicted activity coefficients using Wilson model and the global optimal parameters obtained from LS approach for selected VLE examples.
The solution quality is another metric useful to characterize the performance of stochastic methods for solving global optimization problems. Fig. 6 shows the percent of solutions found by HS without local optimization, from all calculations performed in LS formulation, with absolute errors less than or equal to $e$, where this error is defined as: $e = |F_{obj,ern} - F_{obj,sto}|$. This metric provides a general perspective of the solution quality obtained by HS. Results indicate that the solution quality obtained by HS in the parameter optimization problems with local minima that are stronger attractors, which cause the premature convergence of the stochastic method.

On the other hand, the harmony memory size (HMS) is considered as a key parameter of HS, which can affect the algorithm performance especially the reliability to find the global optimum. For illustration, the effect of HMS on non-linear parameter estimation for modeling VLE data using LS formulation is also reported in Fig. 2. In general, the reliability of HS is almost independent of harmony memory size in VLE data modeling. These results are in agreement with those reported by Omran and Mahdavi [29] using benchmark global optimization problems and a variant of HS method. As stated, the harmony memory is a memory location where a set of potential solution vectors for decision variables is stored. This harmony memory is used in the exploration stage of HS, preserves the history of optimization sequence and is useful to identify promising areas for global optimization because good harmonies can be considered as elements of new solution vectors. Note that the use of harmony memory is similar to the selection of the best-fit individual in Genetic Algorithm. In particular, it is expected that small values of HMS may lead to HS to be trapped in local solutions because the diversity of solutions stored in harmony memory is very limited and, as a consequence, the search space is not effectively explored. Therefore, increasing HMS usually provides better solution vectors but at the expense of more function evaluations. However, HMS has less impact for increasing the reliability of HS in the set of VLE problems tested in this study. This pattern of HS performance also occurred for the EIV formulation (results not reported in this manuscript). These results confirmed again that the selected VLE problems are challenging global

![Fig. 4. Global success rate (GSR) versus (a) NFE/nopt and (b) SNImax of Harmony Search with local optimization for VLE data modeling using error-in-variable formulation.](image)

![Fig. 5. Convergence profiles of HS with and without local optimization in VLE data modeling using the error-in-variable approach for selected examples: (a) No. 14 and (b) No. 13.](image)

![Fig. 6. Percent of solutions based on relative errors of HS in the non-linear parameter estimation for VLE modeling. Convergence condition: NI.](image)
estimation for VLE modeling is dependent on the stopping condition and, as a consequence, is related to the numerical effort. As expected, there is a large variation in the solution quality obtained from HS at early NFE and the objective function value of the best solution obtained by HS is nearer to the global minimum as N1 increases. It is important to remark that plots for EIV formulation are not reported because of HS showed the worst performance for finding the global optimum in these examples. In conclusion, these results highlight the importance of using a local optimization method for improving the quality of solutions obtained by HS method in VLE data modeling.

Results indicate that traditional HS is a suitable method for solving global optimization problems in the modeling of VLE data using LS formulation and offers a better performance than those reported for other meta-heuristics if a proper numerical effort is allowed. With illustrative purposes, Fig. 7 shows the comparison of the performance of Simulated Annealing (SA), Particle Swarm Optimization (PSO), Genetic Algorithm (GA) and Differential Evolution (DE) [13] and HS method for parameter estimation using LS formulation. It is convenient to note that previous studies have reported a robust performance of DE [13], PSO [14,32–34], SA [11,13,32] and GA [12,33] for solving global optimization problems in the modeling of phase equilibrium including parameter estimation problems. To directly examine and compare the performance of HS with those obtained for other stochastic methods using LS approach, we keep their numerical efforts the same via NFE and analyze the results obtained in terms of GSR. These results correspond to the implementation of all stochastic methods with the same local optimization strategy. By examining performance at different levels of computational effort, results indicate that the reliability of all stochastic methods depends on and increases with numerical effort (see Fig. 7). In some parameter estimation problems, the stochastic methods failed in the global minimization of LS formulation. Overall, the reliability of the methods is poor when NFE is limited to low values especially for challenging parameter estimation problems. However, the reliability of HS is better than that of DE, PSO, GA and SA for tested VLE examples using NFE >10,000 and the GSR of HS ranged from 10 to 97%. If a low numerical effort is considered, SA outperformed other stochastic methods for solving VLE problems with LS formulation in terms of global success rate. In fact, GSR of SA ranged from 34 to 81% in VLE problems. On the other hand, DE can achieve a maximum GSR of 59%, while the maximum GSR obtained for PSO and GA was 35 and 23% for this set of VLE problems, respectively. These results indicate that the GSR of SA is better than those obtained for DE, PSO, GA and HS in early iterations. However, as NFE increases, the performance of HS improves and its reliability is higher than those reported for SA. So, it is clear that SA may be prefered for parameter estimation in the VLE modeling if a small number of function evaluations are allowed; otherwise, HS is more reliable.

For illustration, Fig. 8 provides the convergence histories of $F_{obj}$ for HS, PSO and GA in the global minimization of examples Nos. 1, 5 and 9 without local optimization. These plots are based on the average (over 100 runs) of the best objective function recorded by these methods at different NFE. In general, the objective function value of best harmony (i.e., solution vector) obtained by HS is usually lower than that achieved by both PSO and GA in selected examples. In fact, convergence profiles indicate that HS converges faster and the
objective function values also descend rapidly. Therefore, HS can find solution vectors closer to the global minimum solution and it outperformed both PSO and GA. It is convenient to recall that HS, PSO and GA are classed as population-based algorithms and they deal in every iteration of the algorithm with a set (i.e., a population) of solutions rather than with a single solution [36]. The performance of these methods depends strongly on the way of the population is manipulated for the exploration of the search space. In particular, HS has some characteristics that distinguish it from other stochastic methods and they include [36]: (1) the generation of a new vector after considering all the existing vectors, rather than considering only two vectors as in GA, (2) the independent consideration of each decision variable in a vector, and (3) it does not require decimal-binary conversions as in GA. So, these characteristics contribute to that HS can achieve a better tradeoff between the capabilities to explore the search space and to exploit the experience accumulated during the optimization sequence and, as a consequence, it is more robust for non-linear parameter estimation in VLE modeling with LS formulation.

On the other hand, the numerical effort (i.e., NFE and CPU time) of stochastic methods increases directly proportional to problem dimensionality (i.e., number of decision variables), see Table 5. However, compared to other stochastic methods, HS imposes fewer mathematical requirements and its implementation is straightforward. So, HS can be considered an alternative method for non-linear parameter estimation in thermodynamic models with few adjustable parameters and using the LS approach. Unfortunately, the traditional HS is not suitable for solving VLE parameter estimation problems using the EIV approach. In the following section, alternatives for improving the performance of HS in EIV problems are discussed.

### 4.4. Alternatives to improve the performance of HS for modeling VLE data using EIV formulation

Results of previous section have illustrated that finding the global optimum in parameter estimation for VLE data modeling is very challenging especially for EIV formulation. This fact may encourage the development of alternative optimization strategies to identify a better approach for solving this thermodynamic problem. Recently, new strategies have been proposed in order to improve the drawbacks of classical stochastic methods for global optimization. For the case of HS method, these improvements primarily cover two aspects [36]: (1) improvements in terms of parameters setting and (2) improvements in terms of hybridizing HS components with other meta-heuristics. This section briefly discusses the performance of some HS variants in VLE parameter estimation with the EIV formulation.

Specifically, two promising variants of HS called: the Improved Harmony Search [37] and the Global-Best Harmony Search [29] have been used in this study. According to the literature, the dynamic adaptation of HS parameters is the most common approach to overcome the drawbacks of classical HS. Results reported by Mahdavi et al. [37] showed that small PAR values with large bw values affect the performance of HS and increase the calculations needed to find the global optimum. Although, small bw values in final improvisations increase the fine-tuning of solution vectors but, in early iterations, bw should take a bigger value to diversify the solution vectors. Furthermore, large PAR values with small bw values cause the improvement of best solutions in final improvisations. Based on this fact, Mahdavi et al. [37] introduced the Improved Harmony Search (IHS), which uses dynamic values of both parameters PAR and bw. Specifically, the dynamic value of PAR changes with improvisation number as follows:

$$\text{PAR}_{k+1} = \text{PAR}_{\min} + \frac{\text{PAR}_{\max} - \text{PAR}_{\min}}{N_{\text{I}}} \times k$$  \hspace{1cm} (12)

where PAR\( \min \) and PAR\( \max \) are the minimum and maximum pitch adjusting rates, and \( k \) is an improvisation counter. On the other hand, the bandwidth for each improvisation is defined as:

$$bw_{k+1} = bw_{\min} \exp \left( \left( \frac{k}{N_{\text{I}}} \right) \ln \left( \frac{bw_{\min}}{bw_{\max}} \right) \right)$$  \hspace{1cm} (13)

where bw\( \min \) and bw\( \max \) are the minimum and maximum values for bandwidth, respectively. Herein, the parameters of IHS are defined as: HMCR = 0.25, PAR\( \min \) = 0.75, PAR\( \max \) = 0.95, bw\( \min \) = 0.001 and bw\( \max \) = \( l_b_i - l_b_i \), respectively. Fig. 9a provides the pseudo-code of this HS algorithm.

On the other hand, Omran and Mahdavi [29] proposed an alternative version of HS called Global-Best Harmony Search (GHS), which is inspired by the concept of swarm intelligence used in Particle Swarm Optimization. This method modifies the pitch-adjustment step of HS to encourage that a new harmony can mimic the best harmony stored in the harmony memory. In general, GHS has the same structure as HS with the exception of pitch adjustment step used in the improvisation of a new harmony. The pseudo-code used to improvise a new harmony in GHS is given in Fig. 9b. Omran and Mahdavi [29] have suggested that using a constant value of PAR improves the performance of GHS and this scheme is even better than GHS using a dynamical value of PAR. So, this approach has been used in the present study for GHS and its parameters are equal to those used for the traditional HS.

Finally, a hybrid method (HHS) based on the application of operators from Simulated Annealing and Differential Evolution has been introduced to improve the search abilities of HS method. In this algorithm, the new harmony is updated in pitch adjustment using the mutation operator of DE:

$$v_i = v_i + F \cdot (u_{N1} - u_{N2})$$  \hspace{1cm} (14)

where \( F \in (0,1) \) is a random number, \( u_{N1} \) and \( u_{N2} \) are randomly selected from the harmony memory. In addition, the harmony memory is updated using the Metropolis criterion of SA, which is given by:

$$M(T_{SA}) = \min \left\{ 1, \exp \left( \frac{-\Delta F}{T_{SA}} \right) \right\}$$  \hspace{1cm} (15)
where \( \Delta f \) is the change in objective function value from the new harmony and the worst harmony stored in the harmony memory. The cooling schedule used for decreasing \( T_{SA} \) is defined as:

\[
T_{SA,k} = 0.5(T_{SA,0} - T_{SA,F}) \left(1 - \tanh \left(\frac{17k}{N_{\text{max}} - 5}\right)\right) + T_{SA,F}
\]

where \( N_{\text{max}} \) is the maximum number of improvisations for HS, \( T_{SA,k} \) is the annealing temperature at improvisation \( k \), \( T_{SA,0} = 1000 \) and \( T_{SA,F} = 0.01 \) are the initial and final values for the annealing temperature, respectively. Fig. 9c provides the pseudo-code of this hybrid algorithm.

The convergence plots of these HS algorithms over two representative EIV problems (i.e., Nos. 12 and 13) are reported in Fig. 10 and Table 6 shows their success rates using the local optimization method for different improvisations (i.e., NI). These results were obtained from 100 independent runs of stochastic methods with different random seeds and initial values. Overall, the performance of all HS variants is superior to that obtained for the traditional HS. However, results indicate that the hybrid approach HHS moderately increases the performance of the traditional HS method for parameter estimation in VLE modeling using EIV formulation. In fact, this approach slightly improved the capability of HS to deal with the premature convergence in large-scale optimization problems. On the other hand, GHS and IHS are more effective for finding the global optimum in the EIV examples selected for this comparison. Fig. 10 shows that the mean objective function value of best harmony (i.e., solution vector) obtained by GHS and IHS is usually lower than that achieved by HS and HHS in selected examples. Therefore, both variants of HS can be more effective for finding the promising area of the global optimum. These results are in agreement with the observations reported by Mahdavi et al. [37] and Omran and Mahdavi [29]. In particular, GHS and IHS offer a better performance for solving large-scale global optimization problems as those involved in VLE parameter estimation using EIV formulation. Further studies should be focus on the improvement of performance of these HS-based methods via

### Table 6

<table>
<thead>
<tr>
<th>Example</th>
<th>Method</th>
<th>SR for NI/HMS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>50 100 250 500 750 1000 1500</td>
</tr>
<tr>
<td>12</td>
<td>HS</td>
<td>1 1 8 8 10 13 16</td>
</tr>
<tr>
<td></td>
<td>HHS</td>
<td>6 3 10 15 19 23 37</td>
</tr>
<tr>
<td></td>
<td>GHS</td>
<td>26 39 50 53 52 55 60</td>
</tr>
<tr>
<td></td>
<td>HHS</td>
<td>1 7 17 42 56 61 76</td>
</tr>
<tr>
<td>13</td>
<td>HS</td>
<td>0 0 0 0 1 1 2</td>
</tr>
<tr>
<td></td>
<td>HHS</td>
<td>1 1 0 0 0 1 2</td>
</tr>
<tr>
<td></td>
<td>GHS</td>
<td>7 11 10 16 14 18 25</td>
</tr>
<tr>
<td></td>
<td>IHS</td>
<td>0 3 6 8 15 24 49</td>
</tr>
</tbody>
</table>
hybrid approaches using other meta-heuristic components and operators.

Finally, it is convenient to highlight the necessity of studying the performance of HS and other stochastic methods for parameter estimation problems of systems that show a multiphase behavior. These systems may involve more challenging global optimization problems due to the presence of several phases at different conditions of temperature, pressure and composition; and as a consequence, an increment in the complexity of parameter estimation problems is expected because of the presence of several local optimums. This type of parameter estimation problems should be also considered for further studies of stochastic global optimization methods.

5. Conclusions

In this study, the performance of HS has been tested and compared for VLE modeling using experimental data of binary systems and the least squares and error-in-variable formulations. Overall, the traditional HS was found to perform better than other stochastic methods in terms of success rate for parameter estimation in VLE data modeling if a proper numerical effort is used in LS formulation. However, the traditional HS is not suitable for modeling VLE data using the EIV formulation. For these challenging problems, results indicate that modified methods GHS and IHS offer a better performance and are promising for solving this and other thermodynamic calculations involved in phase equilibrium modeling. Finally, this study provides a useful reference and guidelines for the application of HS method in the modeling of VLE data. Further studies should be focused on the development of diversification strategies to improve the reliability of GHS and IHS especially at early improvisations and using the EIV approach. In addition, other convergence criterions should be studied and tested for reliably determining the global convergence of HS-based optimization methods.

List of symbols

<table>
<thead>
<tr>
<th>Notation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>bw</td>
<td>bandwidth</td>
</tr>
<tr>
<td>c</td>
<td>number of components in the mixture</td>
</tr>
<tr>
<td>EIV</td>
<td>error-in-variable formulation</td>
</tr>
<tr>
<td>$F_{obj}$</td>
<td>objective function</td>
</tr>
<tr>
<td>G</td>
<td>Gibbs free energy</td>
</tr>
<tr>
<td>GSR</td>
<td>global success rate</td>
</tr>
<tr>
<td>HMCR</td>
<td>harmony memory considering rate</td>
</tr>
<tr>
<td>HS</td>
<td>harmony memory size</td>
</tr>
<tr>
<td>LS</td>
<td>least square formulation</td>
</tr>
<tr>
<td>lb</td>
<td>lower bound of decision variable</td>
</tr>
<tr>
<td>$n_{sat}$</td>
<td>number of VLE experimental data</td>
</tr>
<tr>
<td>$n_{opt}$</td>
<td>number of decision variables</td>
</tr>
<tr>
<td>$n_i$</td>
<td>mole number of component i</td>
</tr>
<tr>
<td>$n_{par}$</td>
<td>number of adjustable parameters</td>
</tr>
<tr>
<td>$N_{prob}$</td>
<td>number of tested problems</td>
</tr>
<tr>
<td>NFE</td>
<td>number of function evaluations</td>
</tr>
<tr>
<td>Ni</td>
<td>number of improvisations</td>
</tr>
<tr>
<td>P</td>
<td>pressure</td>
</tr>
<tr>
<td>$p_0$</td>
<td>vapor pressure of pure component i</td>
</tr>
<tr>
<td>PAR</td>
<td>pitch adjusting rate</td>
</tr>
<tr>
<td>R</td>
<td>gas universal constant</td>
</tr>
<tr>
<td>rand</td>
<td>random number</td>
</tr>
<tr>
<td>SR</td>
<td>success rate</td>
</tr>
<tr>
<td>T</td>
<td>temperature</td>
</tr>
<tr>
<td>$T_{SA}$</td>
<td>annealing temperature used in Metropolis criterion</td>
</tr>
<tr>
<td>x</td>
<td>liquid mole fraction</td>
</tr>
<tr>
<td>u</td>
<td>vector of decision variables</td>
</tr>
<tr>
<td>$ub$</td>
<td>upper bound for decision variable</td>
</tr>
<tr>
<td>$v$</td>
<td>new harmony vector</td>
</tr>
<tr>
<td>$y$</td>
<td>vapor mole fraction</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>activity coefficient</td>
</tr>
<tr>
<td>$\theta$</td>
<td>model parameter</td>
</tr>
</tbody>
</table>

Acknowledgements

Author acknowledges the financial support provided by CONACyT, DGEST and Instituto Tecnológico de Aguascalientes.

References